Essays on Identification of Structural VARMA models

by

Miguel Ángel Cabello Pérez

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Economics

Universidad Carlos III de Madrid

Advisor(s):

Carlos Velasco Gómez

Tutor:

Carlos Velasco Gómez

June

To my mom, Epifania, and my wife, Sara, whose support has been invaluable.

ACKNOWLEDGEMENTS

I am deeply indebted to my professor and supervisor, Carlos Velasco Gómez, for his priceless comments, suggestions, and patience. I would also like to express my gratitude to my defense committee, who profusely provides knowledge and expertise. Furthermore, this research would not have been feasible without the financial support from the Spanish government.

I am also grateful to professors Jesús Gonzalo, Juan José Dolado, Juan Carlos Escanciano, and Miguel Ángel Delgado, classmates, and cohort members, especially the participants in the Econometrics Reading Group, for their helpful feedback. Lastly, I would be negligent in not mentioning my family, especially my mother, spouse, and child. Their belief in me has kept my spirits and motivation high during this arduous process.

OTHER RESEARCH MERITS

This research has been possible thanks to the financial support from the scholarship MCIN/BES-2017-081997 from the Spanish Agency of Science and Innovation (Ministerio de Ciencia e Innovación).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In macroeconometrics, structural vector autoregressive, moving-average models are widely employed since they provide a simple, practical representation of structural theoretical models. For decades, the most preferred framework has been the structural vector autoregressive models, under the belief that structural economics models are fundamental, which intuitively means that information generated from the history (current and past values) of observable, endogenous variables is equivalent to the information constructed from the history of unobservable economic shocks.

Under such an assumption, most empirical studies only focus on the identification challenges of the contemporaneous effects matrix or the static component of the model. Nevertheless, the fundamentalness assumption may need to be revised, especially when the econometrician's information set is smaller than the information set of the theoretical model. Moreover, it has been documented that this case may not be quite an unrealistic situation. Therefore, it might be necessary to develop methods for estimating structural vector autoregressive, moving-average models without imposing the location of roots but determining them from observable data. However, this leads to abandoning the Gaussian behavior assumption regarding the structural shocks. Otherwise, it is unfeasible to distinguish fundamental from non-fundamental models.

In recent years, several methodological works have proposed alternatives for identifying and estimating structural vector autoregressive moving-average models without imposing the location of roots by exploiting the non-Gaussian behavior of structural shocks. Most of these investigations assumed that shocks are serial independent, characterize the non-Gaussianity by its non-zero third and (or) fourth-order cumulants, and impose the finiteness of a large amount of higher-order moments. The present research attempts to fill this gap in the literature by proposing an estimation method that does not require the finiteness of a large number of higher-order moments or the knowledge of a specific joint density distribution of structural shocks. Also, the proposed method may be adapted to structural shocks exhibiting higher-order serial dependence.

Furthermore, one of the critical identification assumptions of the statistical strategy for identifying structural vector autoregressive moving-average models is the non-Gaussian behavior of all shocks in the system. However, in the literature, few efforts have been posed to determine the non-Gaussian dimension in a structural system. Furthermore, these investigations assume that the structural model is fundamental, which makes it unfeasible to apply such proposals to more general contexts. Consequently, this research also proposes a method for determining the non-Gaussian dimensión. Our proposal is robust to the location of the roots of the dynamic polynomials in the model, consequently, can be applied to fundamental or non-fundamental contexts. Finally, this work explores under which conditions it is feasible to employ economically motivated economic sign restrictions for selecting a permutation under the context of fundamental structural vector autoregressive models.

2. CHAPTER I: IDENTIFICATION AND ESTIMATION OF SVARMA MODELS USING PAIRWISE DEPENDENCE MEASURES

Causal structural vector autoregressions (SVAR) models may deliver biased estimates of causal effects of economic shocks when data comes from a non-fundamental structural model. A way to address non-fundamentalness in SVAR or SVARMA models is to allow roots of lag polynomials to lie inside the unit circle. However, relaxing the location of the lag polynomial roots adds an extra identification problem, consisting of the observational equivalence between fundamental and non-fundamental representations of a stationary process when only second-order information is exploited. Instead of imposing identification restrictions, we employ a *statistical identification approach*, whose primary identification condition is that each structural shock in the system is non-Gaussian distributed. In this line, the contributions of this paper are twofold. On the one hand, we extend a well-known identification result for SVARMA models with independent shocks to a situation in which disturbances can exhibit a more general dependence structure; on the other hand, we propose an estimation procedure that minimizes the distance of the *generalized cumulative spectral distribution* of unrestricted and restricted model errors. This procedure does not require the imposing of a particular joint non-Gaussian distribution of structural errors, as well as it does not need the existence of a large number of higher-order moments.

2.1. Introduction

Structural vector autoregressive (hereafter SVAR) models have an extended usage in empirical macroeconomics because they provide a simple framework for identifying economic shocks and estimating their causal effects over outcomes of interest, such as GDP growth, unemployment, and inflation, among others. Identifying the structural shocks using SVAR models entails a well-documented identification problem. According to Sims (1980), this identification issue consists of the observational equivalence of any orthogonal rotation of structural shocks, i.e., any orthogonal rotation of structural disturbances reproduces the second-order information of observed data.

In mainstream macroeconometrics, the strategies for overcoming this identification problem resort to using external information. Early proposals impose zero or linear restrictions over the matrix of contemporaneous effects (see Bernanke and Mihov (1998); Blanchard and Perotti (2002); Christiano et al. (1996); Sims (1980)). Another type of restriction is imposed on the long-run effects of economic shocks over some outcomes of interests, also known as long-run identification restrictions, such as the one employed in Blanchard and Quah (1989); Gali (1992). An alternative strategy is based on sign restrictions, which only restrain the direction (sign) of the response of some endogenous variable to changes in a single or a subset of economic shocks (Arias et al. (2018); Mountford and Uhlig (2009); Rubio-Ramirez et al. (2010); Uhlig (2005)). For instance, it is a consensus to assume that a positive demand shock increases output and inflation in the short-run. On the other hand, a more economic-based identification strategy is the *narrative approach*. This scheme constructs a measure for an unobserved economic shock by analyzing extensive policy change information. For instance, C. D. Romer and D. H. Romer (1989) construct a measure of monetary policy shock by analyzing historical records of Federal Reserve meetings (see Ramey (2011); Ramey and M. D. Shapiro (1998); C. D. Romer and D. H. Romer (2010) for applications to identify government spending shocks). Recently, an identification strategy that uses external measures as a proxy or instrumental variables for identifying a single shock or a subset of structural shocks has been introduced (Mertens and Ravn (2013); Olea et al. (2021); Stock and Watson (2018)).

Despite the usefulness of these strategies, they possess essential drawbacks. In the case of zero, linear or long-run identification restrictions, most of these conditions are not subject to empirical evaluation unless they are overidentification restrictions. Their support must be done by appealing to economic intuitions, some of which may be arguable. Besides, some of the identification restrictions may be challenging to implement. For instance, zero long-run restrictions are commonly approximated by restricting the cumulative response of a shock over a finite time horizon. Regarding identification based on sign restrictions, these can only identify a set of models. Consequently, this strategy entails two sources of variability: sample and model uncertainty. Isolate each of these sources of variance is challenging. Finally, regarding the strategy using proxy or instrumental variables, its identification power depends on the exogeneity and relevance conditions. Besides, as it is documented in Angelini et al. (2022), when proxy SVARs feature multiple target shocks, it requires additional point or sign restrictions. Last but not least, most of these strategies have in common that they were developed for identifying shocks in a causal (or fundamental) structural VAR model.

The fundamental representation of macroeconomic models has been questioned lately, as SVAR models' suitability for properly representing macroeconomic data. Ravenna (2007) shows that the solution of a dynamic stochastic general equilibrium (DSGE) model is described much better by an SVARMA model rather than by a causal SVAR model. Moreover, depending on the chosen parameterization, the structural VARMA model may have a non-invertible MA component. This characteristic implies that linear combinations of current and future values of endogenous, observable variables can recover unobserved structural shocks. Under such a situation, a finite lag, stable VAR model would produce inconsistent estimates of the causal effects of economic shocks, even though the researcher knew the true structure of the contemporaneous effect matrix.

An approach to address non-fundamental behavior in SVARMA models is to permit the roots of AR and MA lag polynomials to lie inside the unit circle. However, once the fundamentalness assumption of the structural model is abandoned and only unit roots are ruled out, a new identification problem appears. This identification issue consists of the observational equivalence between fundamental and non-fundamental representations of a stationary process. This problem means that fundamental and non-fundamental representations can reproduce a stationary sequence's first and second-order information. Velasco (2022a) and Velasco (2022b) call this issue as the *dynamic identification problem*. Hence, employing only second-order information for identification requires that model errors are serially uncorrelated; thus, it is impossible to distinguish which representation is correct because both types of representations generate white-noise model errors. In this sense, imposing that the roots of AR and MA polynomials lie outside the unit circle is an identification restriction that prevents the consideration of non-fundamental representations.

In this paper, the root location of lag polynomials is not imposed. Instead, we follow a *data-driven approach* for identifying the location of roots and the parameters of an SVARMA model. To achieve this, exploiting information beyond the second order is crucial. Therefore, it is necessary to assume that structural shocks are non-Gaussian distributed. Otherwise, only second-order moments are informative. In Time Series literature, in a univariate context, Rosenblatt (2000) and Velasco (2022a) show that the parameters in lag polynomials of a univariate stationary ARMA process are identified if and only if the random disturbance is an independent identically distributed (hereafter i.i.d.) process with non-zero third order cumulant and finite fourth order moment. In the multivariate scenario, Chan et al. (2006) states that any stationary linear process is identified up to shift-permutation if and only if the structural shocks in the process are i.i.d. and each of them is non-Gaussian distributed with non-zero third or fourth order cumulants. Velasco (2022b) reach a similar conclusion but characterize non-Gaussianiaty by third and fourth-order cumulants.

Two standard features of these identification results are that they rely on the independence of structural shocks and the sufficient condition of having non-zero third-order cumulants and finite fourth-order moments. Therefore, the first aim of this work is to provide a necessary condition for the identification of structural VARMA models without imposing the existence of a large number of cumulants of non-Gaussian shocks, as well as to provide an identification result for structural models with dependent shocks. Regarding this latter scenario, Lanne and Luoto (2021); Lanne, Meitz, et al. (2017) can prove the identification of structural parameters by only assuming that errors are serially uncorrelated but restricted to fundamental SVAR models.

On the other hand, once identification is assured, the following step is to provide estimates of structural parameters. In the literature, we can find several estimation approaches. For the context of fundamental SVAR models, we can point out the works of Gouriéroux et al. (2017); Guay (2021); Lanne and Luoto (2021); Lanne, Meitz, et al. (2017). The first one employs a conditional likelihood approach for estimating the parameters of the causal SVAR, for which it is necessary to assume a joint non-Gaussian density function for structural errors. The common choice is to assume that each error follows a t-student density function with a few degrees of freedom. In contrast, Guay (2021); Lanne and Luoto (2021) use second and higher-order moment conditions; thus, their estimators of structural parameters are based on the GMM approach. The work of Gouriéroux et al. (2017) combines moment- and likelihood-based approaches. In the context of non-fundamental models, Gouriéroux et al. (2019); Lanne and Saikkonen (2013) estimates the structural parameters by maximum likelihood estimation. It is worth mentioning that the structural model in Lanne and Saikkonen (2013) is a non-causal SVAR model. He constructs the likelihood of data by assuming that the AR polynomial of the model is the product of two polynomials, one capturing the causal part and the other non-causal component. On the contrary, Gouriéroux et al. (2019) estimate a structural VARMA model with a non-invertible MA component. Unlike these works, Velasco (2022b) employs a GMM approach using third and fourth-order spectrum cumulant densities for an SVARMA model with possibly non-fundamental roots.

The second aim of this paper is to develop an estimation method that does not rely on assuming or imposing a particular joint density function for structural errors, which may expose the estimates to misspecification bias. Additionally, the proposed method should only assume the existence of a small number of higher-order moments. When third or fourth-order moments or cumulants are employed for estimation, then it is necessary to assume the finiteness of the sixth or eighth-order moments for calculating the standard errors of estimates. This assumption is very problematic because of two reasons. First, it focuses attention only on particular types of non-Gaussian behavior. For instance, in a world where shocks are distributed as t-student, the existence of at least eighth-order moments narrows the attention to distributions with at least nine degrees of freedom, implying that the fat-tail property is less pronounced and closer to the tail behaviour of a Normal distribution. Second, estimating accurately higher order moments by their sample counterparts requires a considerable number of observations[1].

For identification and estimation, our approach follows the idea applied by Velasco (2022a) to a univariate context and employs a generalized dependence measure of pair-

[1]For instance, a t-distributed random variable with five degrees of freedom has a population excess of kurtosis coefficient of 6, but its estimated values are 4.0 and 5.6 with sample sizes of 500 and 5000 observations, respectively.

wise dynamics. With independent shocks, our dependence measure is the centered joint characteristic function of model errors at two different periods. This measure can be interpreted as the autocovariance of the complex exponential transformation of model errors. This measure summarizes all the information in the pairwise joint density function of model errors. Thus, imposing a particular joint density function for structural errors is unnecessary. Besides, this measure is well-defined, disregarding the existence of moments of structural errors. For the context of dependent shocks, we assume that structural errors are a *martingale difference sequence* (hereafter m.d.s.). In this case, we modify the dependence measure and use the autocovariance between model errors and the complex exponential transformation of shifted model errors.

Once the pairwise dependence measure for model errors is defined, it is necessary to summarize the information over all lags and (or) leads. In the case of independent shocks, we have two alternatives. One is to employ the *generalized spectral density* (GSD) introduced by Hong (1999). Based on the GSD, the second option is to compute the *cumulative spectral distribution* (CSD). Both functions are well-defined for any joint density function of structural errors and do not require the existence of moments of structural shocks. We use the CSD because its sample counterpart does not require a smoothing kernel. With the CSD, we construct our population loss function to identify the structural model, while the sample loss function is employed for estimation purposes. Our population loss function consists of the distance between the CSD for unrestricted model residuals and the CSD imposing the corresponding dependence structure and the mutual independence components assumption. Using this loss function, we show that identification (up to sign permutation) of structural parameters requires non-Gaussian behavior in a broad sense and only the existence of standard second-order moments in the context of independent shocks. Furthermore, in the case of dependent shocks, identification (up to sign-permutation) requires the existence of second-order moments of structural shocks and peculiar non-Gaussian behavior, detailed below, consisting of non-linearity of expectations of structural errors conditional to non-constant linear filters. Regarding the asymptotic properties of our estimators, in the case of i.i.d. shocks, the finiteness of third-order moments is the only requisite to find the asymptotic distribution of the structural parameter estimator; in the case of m.d.s. structural errors, it required the finiteness of at least fifth-order moments for finding the asymptotic distribution of the estimator.

The simulation evidence shows that our method satisfactorily identifies the location of the roots with a relatively small sample size ($T = 250$). In the case of the bivariate structural model, the rate of correct root location is above 90% when structural shocks are asymmetric. In contrast, this rate decays when the errors follow a symmetric non-Gaussian distribution, especially for VARMA models with non-zero AR and MA components. To analyze our method's robustness, we study how well the proposal identifies the location of roots when structural shocks behave like a multivariate Gaussian random element. The rate of correct root location obtained with the Montecarlo experiment is quite close to the theoretical rate when only errors are required to be white noise.

Finally, we applied our proposal to the dataset analyzed in Blanchard and Quah (1989). This dataset contains only two endogenous variables, GDP growth, and unemployment rate. We select two different specifications. The first case assumes that the roots of the AR polynomial lie outside the unit circle, but the location of the roots of the MA polynomial is not determined. The second specification is an unrestricted structural model, which does not restrict the root location of both the AR and MA polynomials. In both scenarios, the overall degree of the polynomials is set to $p = 1, q = 1$. The estimation results are the following: for the first restricted specification, our proposed method identifies that the roots of the MA polynomial lie outside the unit circle, i.e., our method identifies that the MA polynomial is invertible. Based on the estimation of the structural model, we compute the estimated causal effects of both identified shocks (i.e., impulse response functions of each shock). The analysis of these causal effects shows that we identify a shock with a non-zero long-run effect and another with zero long-run effect over GDP growth. It is worth mentioning that the identification of these shocks did not require imposing any external identification restriction. In the scenario of not restricting the SVARMA model's root location, our method identifies MA roots outside the unit circle but AR roots inside the unit circle. In other words, our proposal identifies a non-causal but invertible SVARMA model.

The remainder of this chapter is organized as follows: Section 2.2 details the structural model and the technical assumptions and explains the methodology we propose. Section 2.3 deals with estimating the parameters of the structural VARMA model and states the asymptotic properties. Section 2.4 presents the simulation study results and some robustness analysis of the methodology for independent shocks. Section 2.5 shows the simple empirical application of our proposed procedure. Finally, Section 2.6 concludes.

2.2. Model, Assumptions and Identification

2.2.1. Structural VARMA Model

Let $\boldsymbol{Y}_t$ be a d-dimensional vector of observable, zero-mean, endogenous variables. $\boldsymbol{Y}_t$ is the stationary solution of the following structural model,

$$\boldsymbol{A}(L)\boldsymbol{Y}_t = \boldsymbol{M}(L)\boldsymbol{\varepsilon}_t, \tag{2.1}$$

where $\boldsymbol{A}(L) = \boldsymbol{A}_0 - \sum_{k=1}^{p} \boldsymbol{A}_k L^k$ and $\boldsymbol{M}(L) = \sum_{k=0}^{q} \boldsymbol{M}_k L^k$ are matrix lag polynomials of degree p and q, respectively. $\{\boldsymbol{A}_{k=0}^{p}, \boldsymbol{M}_{k=1}^{q}\}$ are time-invariant $d \times d$ matrices; $\det(\boldsymbol{A}_0)\det(\boldsymbol{A}_p)\det(\boldsymbol{M}_0)\det(\boldsymbol{M}_q) \neq 0$. L represents the lag (back-shift) operator ($L^k \boldsymbol{Y}_t = \boldsymbol{Y}_{t-k}$). $\boldsymbol{\varepsilon}_t$ represents the d-dimensional vector of unobservable structural shocks, which receive an economic interpretation. A general behavior assumed for structural errors is that they are an ergodic, white-noise, or serially uncorrelated process. Later, we state explicit additional structure on these unobserved structural shocks.

The structural model in (4.1) can be expressed alternatively as

$$\boldsymbol{\Phi}(L)\boldsymbol{Y}_t = \boldsymbol{\Theta}(L)\boldsymbol{B}\boldsymbol{\varepsilon}_t, \tag{2.2}$$

with $\boldsymbol{\Phi}(L) = \boldsymbol{I}_d - \sum_{k=1}^{p} \boldsymbol{\Phi}_k L^k$, $\boldsymbol{\Theta}(L) = \boldsymbol{I}_d + \sum_{k=1}^{q} \boldsymbol{\Theta}_k L^k$. The relationship between coefficients in equations (4.1) and (4.2) is the following: $\boldsymbol{\Phi}_k = \boldsymbol{A}_0^{-1}\boldsymbol{A}_k$ for $k = 1, \dots, p$; $\boldsymbol{\Theta}_k = \boldsymbol{A}_0^{-1}\boldsymbol{M}_k\boldsymbol{M}_0^{-1}\boldsymbol{A}_0$ for $k = 1, \dots, q$; and $\boldsymbol{B} = \boldsymbol{A}_0^{-1}\boldsymbol{M}_0$. Although both equations (4.1) and (4.2) are equivalent, we use the expression in (4.2) for referring to an SVARMA model throughout this work.

Notice that the structural model in (4.2) resembles the standard notation for the SVAR model in macroeconometrics. In fact, if $q = 0$, then MA polynomial $\boldsymbol{\Theta}(L) = \boldsymbol{I}_d$ and the SVARMA model collapses to the SVAR model. When $p = 0$, our model is an SVMA. Besides, polynomials $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ drive the dynamics of endogenous variables, while $\boldsymbol{B}$ accounts for the static, contemporaneous relation between endogenous variables and unobserved shocks. The parameters of interest, known as structural parameters, are $\left(\{\boldsymbol{\Phi}_k\}_{k=1}^{p}, \{\boldsymbol{\Theta}_k\}_{k=1}^{q}, \boldsymbol{B}\right)$.

Unlike SVAR models, SVARMA models have received less attention in macroeconometrics literature. This lack of attention can be explained by the various identification problems a structural VARMA model faces. These issues can be grouped into three categories: (1) the location of roots of $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ polynomials (*dynamic identification problem*) ; (2) identification of static component, $\boldsymbol{B}$ (*static identification problem*); and (3) the issue of unique representation. This latter has been explored by different authors (Deistler and Hannan (1981); Deistler and Hannan (1988); Hannan (1976); Lütkepohl (2005); Reinsel (2003)), and consists on the problem of having different AR and MA polynomials that can lead to the same structural model as in (4.2). This problem is ruled out in a univariate context by assuming that AR and MA polynomials are coprime, meaning that AR and MA polynomials do not share any root.

In a multivariate environment, assuming that $\mathbf{\Phi}(L)$ and $\mathbf{\Theta}(L)$ are left coprime, it is not enough to avoid the unique representation problem. According to Lütkepohl (2005), there exist cases in which it is possible to multiply both polynomials by a uni-modular factor, generating an equivalent representation to the structural model in (4.2). This identification problem is out of the scope of this paper. To ensure a unique representation of the SVARMA model, we imposed that dynamic polynomials in equation (4.2) satisfy the conditions stated in Hannan (1976).

2.2.2. Dynamic Identification Problem

Dynamic identification problem consists of the observational equivalence between fundamental and non-fundamental representations of any stationary process. In our context, where the vector of observable endogenous variables is the stationary solution of the model in (4.2), the dynamic identification problem can be interpreted in the following way: there exist a pair of polynomials $\tilde{\mathbf{\Phi}}(L)$ and $\tilde{\mathbf{\Theta}}(L)$ which are different from the true ones, with no roots over the unit circle, such that the model errors $\boldsymbol{u}_t = \tilde{\mathbf{\Theta}}^{-1}(L)\tilde{\mathbf{\Phi}}(L)\boldsymbol{Y}_t$ are serially uncorrelated.

Let us see this problem with a simple example. Assume $\boldsymbol{Y}_t$ is generated by a non-invertible SVMA(1) model, i.e. $\boldsymbol{Y}_t = \boldsymbol{\varepsilon}_t + \mathbf{\Theta}_1^0 \boldsymbol{\varepsilon}_{t-1}$ with $\boldsymbol{\varepsilon}_t \sim WN(0, \boldsymbol{I}_2)$ and $\mathbf{\Theta}_1^0 = \operatorname{diag}\left(\theta_{11,0}, \theta_{22,0}\right)$ with $|\theta_{ii,0}| > 1$ for $i = 1,2$. The second order information of $\boldsymbol{Y}_t$ is

$$\boldsymbol{\gamma}_h^Y = \mathbb{C}\mathrm{ov}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-k}) = \begin{cases} \boldsymbol{I}_2 + \mathbf{\Theta}_1^0 \mathbf{\Theta}_1^{0\prime}, & h = 0 \\ \mathbf{\Theta}_1^0, & h = 1 \\ \mathbf{\Theta}_1^{0\prime}, & h = -1 \\ \mathbf{0}, & |h| \geq 2 \end{cases}.$$

Now, if the researcher represents $\boldsymbol{Y}_t$ as an invertible SVMA(1), i.e. $\boldsymbol{Y}_t = \left(\boldsymbol{I}_2 + \tilde{\mathbf{\Theta}}_1 L\right)\boldsymbol{u}_t$ with $\tilde{\mathbf{\Theta}}_1 = \operatorname{diag}\left(\frac{1}{\theta_{11,0}}, \frac{1}{\theta_{22,0}}\right)$. Then, the model errors are $\boldsymbol{u}_t = \left(\boldsymbol{I}_2 + \tilde{\mathbf{\Theta}}_1 L\right)^{-1} \boldsymbol{Y}_t$. Since the roots of $P(z) \equiv \det(\boldsymbol{I}_2 + \tilde{\mathbf{\Theta}}_1 z)$ are outside the unit circle, hence $P^{-1}(z) = \boldsymbol{I}_2 - \tilde{\mathbf{\Theta}}_1 z + \tilde{\mathbf{\Theta}}_1^2 z^2 -$

$\tilde{\boldsymbol{\Theta}}_1^3 z^3 + \ldots$; therefore

$$\boldsymbol{u}_t = \left(\boldsymbol{I}_2 - \tilde{\boldsymbol{\Theta}}_1 L + \tilde{\boldsymbol{\Theta}}_1^2 L^2 - \tilde{\boldsymbol{\Theta}}_1^3 L^3 + \ldots\right)\left(\boldsymbol{I}_2 + \boldsymbol{\Theta}_1^0 L\right)\boldsymbol{\varepsilon}_t = \sum_{l=0}^{\infty} \boldsymbol{\delta}_l \boldsymbol{\varepsilon}_{t-l},$$

with $\boldsymbol{\delta}_0 = \boldsymbol{I}_2$ and $\boldsymbol{\delta}_l = \left[-\tilde{\boldsymbol{\Theta}}_1\right]^l + \left[-\tilde{\boldsymbol{\Theta}}_1\right]^{l-1}\boldsymbol{\Theta}_1^0 = \begin{bmatrix} (-1)^l \theta_{11,0}^{-l}\{1-\theta_{11,0}^2\} & 0 \\ 0 & (-1)^l \theta_{22,0}^{-l}\{1-\theta_{22,0}^2\} \end{bmatrix}$ for $l \geq 1$.

After some algebraic manipulations, we have

$$\mathbb{E}\left[\boldsymbol{u}_t \boldsymbol{u}_{t-h}'\right] = \boldsymbol{\gamma}_h^u = \begin{cases} \sum_{l=0}^{\infty} \boldsymbol{\delta}_l \boldsymbol{\delta}_l' = \boldsymbol{\Theta}_1^0 \boldsymbol{\Theta}_1^{0\prime}, & h = 0 \\ \sum_{l=h}^{\infty} \boldsymbol{\delta}_l \boldsymbol{\delta}_{l-h}' = \boldsymbol{0}, & |h| \geq 1 \end{cases},$$

which means that $\{\boldsymbol{u}_t\}_{t\in\mathbb{Z}}$ is a white-noise sequence.

The autocovariance function based on the invertible model is

$$\boldsymbol{\gamma}_h^{Y(u)} = \begin{cases} \boldsymbol{\gamma}_0^u + \tilde{\boldsymbol{\Theta}}_1 \boldsymbol{\gamma}_0^u \tilde{\boldsymbol{\Theta}}_1' = \boldsymbol{\Theta}_1^0 \boldsymbol{\Theta}_1^{0\prime} + \boldsymbol{I}_2, & h = 0 \\ \boldsymbol{\Theta}_1^{(1)} \boldsymbol{\gamma}_0^u = \boldsymbol{\Theta}_1^0, & h = 1 \\ \boldsymbol{\gamma}_0^{\prime u} \boldsymbol{\Theta}_1^{\prime(1)} = \boldsymbol{\Theta}_1^{0\prime}, & h = -1 \\ \boldsymbol{0}, & |h| \geq 2 \end{cases},$$

which is the same second-order information as the one obtained for the data, $\boldsymbol{\gamma}_h^Y$.

Therefore, we say that both representations $\boldsymbol{Y}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1^0 \boldsymbol{\varepsilon}_{t-1}$ and $\boldsymbol{Y}_t = \boldsymbol{u}_t + \tilde{\boldsymbol{\Theta}}_1 \boldsymbol{u}_{t-1}$ are observationally equivalent, which implicates that requiring model errors to be serially uncorrelated, i.e., a white noise process, does not provide enough information for distinguishing these two representations. This observational equivalence between non-invertible and invertible models has a tremendous impact on identifying the causal effect of a shock. The impulse response function (the causal effect measure of structural shocks) obtained from both representations are the following:

$$\frac{\partial y_{t+h,1}}{\partial \varepsilon_{t,1}} = \begin{cases} 1, & h = 0 \\ \theta_{11,0}, & h = 1 \\ 0, & h \geq 2 \end{cases} \qquad \text{and} \qquad \frac{\partial y_{t+h,1}}{\partial u_{t,1}^*} = \begin{cases} \theta_{11,0}, & h = 0 \\ 1, & h = 1 \\ 0, & h \geq 2 \end{cases},$$

where $\frac{\partial y_{t+h,1}}{\partial \varepsilon_{t,1}}$ represents the response of the first endogenous variable to first structural shock, while $\frac{\partial y_{t+h,1}}{\partial u_{t,1}^*}$ is the causal effect to a standardized movement of the first error in the fundamental representation. Without loss of generality, we can assume that $\theta_{11,0} > 1$, then under the correct specification, the largest causal effect of structural shock ε_1 happens one period after the change in the shock. On the contrary, using the fundamental approximation, the largest causal effect is contemporaneous. Although the sign of both responses is the same, the economic implications are quite different.

Risks of imposing causal and invertible representations

In our previous example, we have seen that if the researcher imposes a fundamental representation of data generated by a non-fundamental model, this will produce biased impulse response functions, which lead to erroneous conclusions. Nonetheless, is it still possible to employ alternative methods that do not require estimating an SVAR or SVARMA model for computing the causal effects of shocks? Local projections (LPs) have become quite popular in empirical macroeconometrics because they allow the computing of impulse response functions using simple linear regression techniques (Jordá (2005)). Plagborg-Møller and Wolf (2021) proves that LPs and SVAR models provide the same impulse responses. However, one of the assumptions in their work is that all roots of the AR polynomial are outside the unit circle, i.e., the correct dynamic representation is a fundamental one. When the actual model is non-invertible, which implies structural shocks cannot be recovered using present and past information of endogenous variables, Stock and Watson (2018) propose to employ instrumental variables approach for achieving identification. Such a procedure can handle non-invertibility. Nonetheless, this approach depends on the existence of valid instruments for each shock if we are interested in estimating all the causal effects. Additionally, inference may be problematic in the case of weak instruments.

If the SVAR model is non-fundamental, LPs may not be a suitable alternative. Let us see this with a simple example. The correct structural model is given by $(\boldsymbol{I}_2+\boldsymbol{\Phi}_1^0 L)\boldsymbol{Y}_t = \boldsymbol{\varepsilon}_t$ with $\boldsymbol{\varepsilon}_t \sim WN(0,\boldsymbol{I}_2)$ and $\det(\boldsymbol{I}_2+\boldsymbol{\Phi}_1^0 z) \neq 0$ for all $|z| \geq 1$, i.e. $\boldsymbol{Y}_t$ is non-causal SVAR(1). The researcher approximates data using a fundamental model, $\boldsymbol{Y}_t = \tilde{\boldsymbol{\Phi}}(L)\boldsymbol{u}_t$ with $\tilde{\boldsymbol{\Phi}}(L) = \boldsymbol{I}_2 + \tilde{\boldsymbol{\Phi}}_1 L$, whose roots outside the unit circle, and $\boldsymbol{u}_t = \tilde{\boldsymbol{\Phi}}^{-1}(L)\boldsymbol{Y}_t$ is a white noise process.

Local projections estimate the following linear regression, assuming that $\boldsymbol{u}_t$ is observed:

$$\boldsymbol{Y}_{t+s} = \boldsymbol{\beta}_s \boldsymbol{X}_t + \boldsymbol{\alpha}_s \boldsymbol{u}_t + v_{t+s},$$

where $\boldsymbol{X}_t$ is a vector containing lagged values of $\boldsymbol{Y}_t$.

Therefore, if the correct specification is a fundamental model, then $\boldsymbol{u}_t$ is $\boldsymbol{\varepsilon}_t$. Moreover, according to Jordá (2005), $\boldsymbol{\alpha}_s$ represents the causal effect of shock $\boldsymbol{u}_t$ and can be consistently estimated using the least squares method. The explanation behind this result is that under fundamentalness, endogenous variables can be represented as an infinite, causal moving average of structural shocks, implying that variables in $\boldsymbol{X}_t$ are not correlated with the error term v_{t+s}.

Nonetheless, when the correct specification is a pure non-causal SVAR(1), then the endogenous variables are a linear combination of future values of structural shocks, $\boldsymbol{Y}_t = \sum_{k=1}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t+k}$. Furthermore, under this context, the errors from a fundamental representation, $\boldsymbol{u}_t$, are a dynamic combination of structural shocks' past, current and future values. Therefore, local projections fail to deliver consistent estimates of the causal effects of structural error changes for two reasons. First, the variables in $\boldsymbol{X}_t$ may be correlated with the error term. Second, according to Hamilton (1994), the impulse response (or the causal effect) of a structural shock over a variable of interest can be defined as $IR(t,s,\boldsymbol{d}_i) = \mathbb{E}(\boldsymbol{Y}_{t+s}|\varepsilon_t = \boldsymbol{d}_i, \boldsymbol{X}_t) - \mathbb{E}(\boldsymbol{Y}_{t+s}|\varepsilon_t = \boldsymbol{0}, \boldsymbol{X}_t)$. And, under non-Gaussianity, in the non-fundamental case these conditional expectations are unknown non-linear functions. Therefore, LPs only estimate the best linear approximation, which is not equal to the actual causal effect in this case.

2.2.3. Static Identification Problem

Like in the SVAR context, the structural model described by (4.2) is subject to *static identification problem*. Assuming that AR and MA polynomials are identified, then the model errors are $\boldsymbol{u}_t = \boldsymbol{B}\boldsymbol{\varepsilon}_t$. Thus, the static identification problem implies that model errors based on orthonormal rotations to structural shocks are observationally equivalent. Formally, assuming $\mathbb{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \boldsymbol{I}_d$, then the variance of model errors is $\Sigma_u = \mathbb{E}\boldsymbol{u}_t\boldsymbol{u}_t' = \boldsymbol{B}\boldsymbol{B}'$. Now, defining $\tilde{\boldsymbol{u}}_t = \tilde{\boldsymbol{B}}\tilde{\boldsymbol{\varepsilon}}_t$ with $\tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{Q}\boldsymbol{\varepsilon}_t'$ and $\tilde{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{Q}'$. Then, $\mathbb{E}\tilde{\boldsymbol{u}}_t\tilde{\boldsymbol{u}}_t' = \boldsymbol{B}\boldsymbol{B}' = \Sigma_u$ for any matrix $\boldsymbol{Q}$ that holds $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}_d$. Consequently, infinite orthonormal rotations of structural innovations exist that reproduce the same variance structure of model errors.

2.2.4. Parameterization

In this section, we show a convenient parameterization of the structural model, which makes it easier to discuss the identification and study of the statistical properties of estimators. Besides, our parameterization makes it simple to compute model errors in the time domain. Let $\boldsymbol{\vartheta}$ be a K-dimensional vector containing the parameters of the structural model in (4.2). We assume $\boldsymbol{\vartheta} \in \mathcal{V} \subset \mathbb{R}^K$, with $\mathcal{V}$ compact parameter set. Additionally, the vector of structural parameters can be split as $\boldsymbol{\vartheta} = \left(\boldsymbol{\vartheta}_D', \boldsymbol{\vartheta}_S'\right)'$ where the block $\boldsymbol{\vartheta}_D$ governs the dynamic part -i.e., the lag polynomials- and $\boldsymbol{\vartheta}_S$, the static component of the model. Besides, we assume that $\boldsymbol{\vartheta}_D \in \mathbb{R}^{K_D}$ and $\boldsymbol{\vartheta}_S \in \mathbb{R}^{K_S}$ with $K_D + K_S = K$. In general,

$K \leq \overline{d} = d^2(p+q+1)$; when none external restrictions are imposed, then $K = \overline{d}$; while if some external restrictions are set over the structural parameters, $K < \overline{d}$.

With the block $\boldsymbol{\vartheta}_D$ we can construct AR and MA polynomials, that is $\boldsymbol{\Phi}(L, \boldsymbol{\vartheta}_D)$ and $\boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_D)$; while with the block $\boldsymbol{\vartheta}_S$, we construct the static component of the model, $\boldsymbol{B}(\boldsymbol{\vartheta}_S)$. We denote by $\boldsymbol{\vartheta}_0 = \left(\boldsymbol{\vartheta}'_{D,0}, \boldsymbol{\vartheta}'_{S,0}\right)'$ the true values of structural parameters, then the elements in equation (4.2) can be rewritten as $\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}(L, \boldsymbol{\vartheta}_{D,0})$, $\boldsymbol{\Theta}(L) = \boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_{D,0})$ and $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{\vartheta}_{S,0})$. Hence, the structural model in (4.2) can be expressed as follows: $\boldsymbol{\Phi}(L, \boldsymbol{\vartheta}_{D,0})\boldsymbol{Y}_t = \boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_{D,0})\boldsymbol{B}(\boldsymbol{\vartheta}_{S,0})\boldsymbol{\varepsilon}_t$. Therefore, in general, we can index the AR and MA polynomials as $\boldsymbol{\Phi}(L; \boldsymbol{\vartheta})$ and $\boldsymbol{\Theta}(L; \boldsymbol{\vartheta})$, respectively. The following assumption imposes some structure on lag polynomials, define $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, $\mathbb{T}_+ = \{z \in \mathbb{C} : |z| < 1\}$ and $\mathbb{T}_- = \{z \in \mathbb{C} : |z| > 1\}$:

Assumption 2.1. *Let p and q be the orders of AR and MA polynomials. We assume for any $\boldsymbol{\vartheta} \in \mathcal{V}$ that:*

(a) $\boldsymbol{\Phi}(z; \boldsymbol{\vartheta})$ *and* $\boldsymbol{\Theta}(z; \boldsymbol{\vartheta})$ *are left co-prime.*

(b) $\det\left(\boldsymbol{\Phi}_p(\boldsymbol{\vartheta})\right)\det\left(\boldsymbol{\Theta}_q(\boldsymbol{\vartheta})\right) \neq 0$*, where* $\boldsymbol{\Phi}_p(\boldsymbol{\vartheta})$ *and* $\boldsymbol{\Theta}_q(\boldsymbol{\vartheta})$ *are the matrix coefficients associated to the largest power of backward-shift operator, L, in AR and MA polynomials.*

(c) $\det(\boldsymbol{\Phi}(z; \boldsymbol{\vartheta}))\det(\boldsymbol{\Theta}(z; \boldsymbol{\vartheta})) \neq 0$ *for all* $z \in \mathbb{T} \cup \{\mathbf{0}\} \cup \{z : |z| \to \infty\}$.

(d) $\boldsymbol{\Phi}(z; \boldsymbol{\vartheta})$ *and* $\boldsymbol{\Theta}(z; \boldsymbol{\vartheta})$ *admit left canonical factorization with respect to* $\mathbb{T}$.

Assumptions 2.1.(a) and 2.1.(b) are the standard necessary and sufficient conditions for avoiding non-unique representations (see Hannan (1976); Mainassara and Francq (2011) for details). Assumption 2.1.(c) rules out the presence of unit roots in both the AR and MA polynomials. Besides, we discard roots at zero or infinity. This assumption is important for two technical reasons: first, AR unit roots prevent the possibility of having stationary solutions to difference equation (4.2) (see Giurcanu (2015) and the references therein for a formal proof) and make unfeasible to construct our dependence measure; second, MA unit roots makes unfeasible to recover shocks from a linear combination of observable variables. Additionally, it would not be feasible to apply matrix factorization of AR and MA polynomials (see Baggio and Ferrante (2018); I. Gohberg et al. (1982); Israel Gohberg et al. (2005) for some details on the factorization of rational matrix functions).

Assumption 2.1.(d) is critical for making it feasible to compute sample model residuals in the time domain. Lanne and Saikkonen (2013) applies a similar assumption. Their parameterization assumes that a fundamental part drives the AR lag polynomial.

In contrast, the non-fundamental part is explained by a lead polynomial with roots outside the unit circle. On the contrary, Funovits (2020) employs a general (Wiener-Hopf) factorization of a rational matrix polynomial of the form $\mathbf{\Phi}_+(z)\Lambda(z)\mathbf{\Phi}_-(z)$ where $\Lambda(z) = \operatorname{diag}(z^{\kappa_1},\ldots,z^{\kappa_d})$. Our factorization is a particular case, occurring when $\kappa_1 = \cdots = \kappa_d = 0$, and imposes that AR dynamics has two components: a fundamental, with roots of lag polynomial outside the unit circle; and a non-fundamental, with roots of lag polynomials inside the unit circle. Velasco (2022a); Velasco (2022b) do not impose this factorization because both article work in the frequency domain.

The left canonical factorization of AR and MA polynomials consists on $\mathbf{\Phi}(z;\boldsymbol{\vartheta}) = \mathbf{\Phi}_+(z;\boldsymbol{\vartheta})\mathbf{\Phi}_-(z;\boldsymbol{\vartheta})$ and $\mathbf{\Theta}(z;\boldsymbol{\vartheta}) = \mathbf{\Theta}_+(z;\boldsymbol{\vartheta})\mathbf{\Theta}_-(z;\boldsymbol{\vartheta})$, where $\mathbf{\Phi}_+(z;\boldsymbol{\vartheta})$ and $\mathbf{\Theta}_+(z;\boldsymbol{\vartheta})$ along with their inverses are analytic in $\mathbb{T}_+$; similarly, $\mathbf{\Phi}_-(z;\boldsymbol{\vartheta})$ and $\mathbf{\Theta}_-(z;\boldsymbol{\vartheta})$ and their inverses are analytic in $\mathbb{T}_-$. Furthermore, $\mathbf{\Phi}_\pm(z;\boldsymbol{\vartheta}) = \boldsymbol{I}_d - \sum_{j=1}^{p_\pm}\mathbf{\Phi}_j^{(\pm)}z^j$ and $\mathbf{\Theta}_\pm(z;\boldsymbol{\vartheta}) = \boldsymbol{I}_d + \sum_{j=1}^{q_\pm}\mathbf{\Theta}_j^{(\pm)}z^j$. p_+ and p_- denote the polynomial degrees of non-causal and causal AR components, respectively; q_+ and q_- represent de polynomial order for non-invertible and invertible MA components, respectively. Besides, notice that $p = p_+ + p_-$ and $q = q_+ - q_-$. Thus, given p and q, the location of roots is determined by the pair (p_+, q_+). For instance, if we had a 2-dimensional SVARMA$(1,0)$ with $p_+ = 0$, then the model is a causal SVAR(1). Additionally, the number of causal and non-causal roots are $dp_- = 2$ and $dp_+ = 0$, respectively. This example shows that our factorization implies that the number of causal (invertible) or non-causal (non-invertible) roots are multiples of d. In total, we have $d(p_+ + p_-)$ AR roots, and $d(q_+ + q_-)$ MA roots.

Given the assumption (2.1.d), we can refine the structure of structural parameters vector $\boldsymbol{\vartheta}$. In particular, $\boldsymbol{\vartheta}_D = \left(\boldsymbol{\phi}'_+, \boldsymbol{\phi}'_-, \boldsymbol{\theta}'_+, \boldsymbol{\theta}'_-\right)'$, where $\boldsymbol{\phi}_\pm$ are the coefficients for autoregressive polynomials; $\boldsymbol{\theta}_\pm$, the parameters of moving-average polynomials. When $K = \bar{d}$, then $\boldsymbol{\phi}_\pm = \operatorname{vec}\left(\begin{bmatrix}\mathbf{\Phi}_1^\pm & \cdots & \mathbf{\Phi}_{p_\pm}^\pm\end{bmatrix}\right)$, $\boldsymbol{\theta}'_\pm = \operatorname{vec}\left(\begin{bmatrix}\mathbf{\Theta}_1^\pm & \cdots & \mathbf{\Theta}_{q_\pm}^\pm\end{bmatrix}\right)$ and $\boldsymbol{\vartheta}_S = \operatorname{vec}(\boldsymbol{B})$.

For any $\boldsymbol{\vartheta}$, the model errors, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$, are defined as follows:

$$\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \stackrel{\text{def}}{=} \mathbf{\Psi}^{-1}(L;\boldsymbol{\vartheta})\boldsymbol{Y}_t = \boldsymbol{B}^{-1}(\boldsymbol{\vartheta})\mathbf{\Theta}^{-1}(L;\boldsymbol{\vartheta})\mathbf{\Phi}(L;\boldsymbol{\vartheta})\boldsymbol{Y}_t = \sum_{j=-\infty}^{\infty}\mathbf{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-j},$$

where $\mathbf{\Psi}(L;\boldsymbol{\vartheta}) = \mathbf{\Phi}^{-1}(L;\boldsymbol{\vartheta})\mathbf{\Theta}(L;\boldsymbol{\vartheta})\boldsymbol{B}(\boldsymbol{\vartheta}) = \sum_{j=-\infty}^{\infty}\mathbf{\Psi}_j(\boldsymbol{\vartheta})L^j$. Since we are ruling out the presence of unit roots for any value in parameter space, then it holds that the filters $\mathbf{\Phi}^{-1}(L;\boldsymbol{\vartheta})$ and $\mathbf{\Theta}^{-1}(L;\boldsymbol{\vartheta})$ are absolute summable for any value in the parameter space. That is $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\sum_{j=-\infty}^{\infty}\left\|\mathbf{\Phi}_j^{(-1)}(\boldsymbol{\vartheta})\right\| < \infty$ and $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\sum_{j=-\infty}^{\infty}\left\|\mathbf{\Theta}_j^{(-1)}(\boldsymbol{\vartheta})\right\| < \infty$. Then, without loss of generality, it can be specified that $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left\|\mathbf{\Phi}_j^{(-1)}(\boldsymbol{\vartheta})\right\| \le C|j|^{-\mu_0}$ for $|j| > 0$ and $\mu_0 > 2$. In Lemma 1 it is proven that given the absolute summability condition for filters $\mathbf{\Phi}^{-1}(L;\boldsymbol{\vartheta})$ (resp. $\mathbf{\Theta}^{-1}(L;\boldsymbol{\vartheta})$), then $\mathbf{\Psi}(L;\boldsymbol{\vartheta})$ and $\mathbf{\Psi}^{-1}(L;\boldsymbol{\vartheta})$ are absolute summable as well, and $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}} \left\|\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})\right\| \leq C|j|^{-\bar{\mu}_0}$ and $\sup_{\boldsymbol{\vartheta}\in\mathcal{V}} \left\|\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\right\| \leq C|j|^{-\bar{\mu}_0}$ hold with $\bar{\mu}_0 = 1-\mu_0$.

For identification purposes, we assume the following

Assumption 2.2. *There exists $\boldsymbol{\vartheta}_0 \in \mathring{\mathcal{V}}$, with $\mathring{\mathcal{V}}$ the interior set of $\mathcal{V}$, such that*

(a) $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}_0, L)\boldsymbol{Y}_t = \boldsymbol{\varepsilon}_t$.

(b) $\boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}, L)\boldsymbol{\Psi}(\boldsymbol{\vartheta}_0, L) \neq \boldsymbol{I}_d$ *for any* $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0 \in \mathcal{V}$.

Assumption 2.2 is crucial for achieving identification of structural parameters up to signed-permutation (see proof of Theorem (2.1) for some details). In standard SVAR literature, this assumption is not necessary because it is assumed AR polynomial has a stable/causal representation, which guarantees *dynamic identification*. Besides, if we were working with general linear, possibly non-fundamental models, Assumption 2.2.(b) rules out the case of getting a back/forward shift monomial. Since we are working with vector ARMA model, then $\boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}, L) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})L^j$ and $\boldsymbol{\Psi}(\boldsymbol{\vartheta}_0, L) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j(\boldsymbol{\vartheta}_0)L^j$ and their multiplication is $\boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}, L)\boldsymbol{\Psi}(\boldsymbol{\vartheta}_0, L) = \sum_{j_1,j_2} \boldsymbol{\Psi}_{j_1}^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{\Psi}_{j_2}(\boldsymbol{\vartheta}_0)L^{j_1+j_2}$. Thus, this will be a back/forward shift monomial if and only if $\boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}, L) = \boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}_0, L)L^k$ for some $k \neq 0 \in \mathbb{Z}$. Such a scenario is implausible in a VARMA context.

From Assumption 2.2.(a), model errors are equal to

$$\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \equiv \boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}; L)\boldsymbol{\Psi}(\boldsymbol{\vartheta}_0; L)\boldsymbol{\varepsilon}_t = \boldsymbol{\delta}(\boldsymbol{\vartheta}; L)\boldsymbol{\varepsilon}_t = \sum_{j=-\infty}^{\infty} \boldsymbol{\delta}_j(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-j}. \tag{2.3}$$

Equation (2.3) can be interpreted as follows: for any $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$, model residuals are a possibly infinite, two-sided filter of true structural innovations. Many filters produce serially correlated model errors, but others generate serially uncorrelated ones. These latter filters are known as *all-pass* filters. Hence, the dynamic identification problem can be summarized as follows: any filter equal to an all-pass filter is observationally equivalent.

2.2.5. Identification with independent structural shocks

In this section, it is shown the identification of parameters of the structural model in (4.2). Showing identification is crucial because it permits learning the true value of parameters from data. Typically, identification is operationalized as the existence of a unique minimizer of a suitable population loss function. Hence, it is important to detail how we construct our population loss function.

Loss Function

The primary input of the loss function is the generalized spectral density (GSD), first introduced in time-series literature by Hong (1999), who employs it for testing the serial white-noise property of a stationary variable. According to Hong (1999), the GSD is the Fourier transform of the autocovariance function of the transformation $e^{i\boldsymbol{\tau}'\mathbf{x}_t}$, where $\mathbf{x}_t$ is a strict stationary, ergodic random vector.

The following assumption makes explicit the structure about shocks $\boldsymbol{\varepsilon}_t$ and provides sufficient conditions for ensuring the existence of the GSD.

Assumption 2.3. *Let $\boldsymbol{\varepsilon}_t = \left(\varepsilon_{t,1}, \ldots, \varepsilon_{t,d}\right)'$, the vector of structural shocks, satisfy:*

(a) $\boldsymbol{\varepsilon}_t$ *is an independent identically distributed (i.i.d.) process.*

(b) $\varepsilon_{t,i}$ *follow a non-Gaussian distribution for each $i \in \{1, \ldots, d\}$, and $\boldsymbol{\varepsilon}_t$ possess mutually independent components for all $t \in \mathbb{Z}$.*

(c) $\mathbb{E}\boldsymbol{\varepsilon}_t = 0$, $\mathbb{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \boldsymbol{I}_d$.

(d) $\sum_{j=-\infty}^{\infty} \alpha_{\boldsymbol{\varepsilon}}(j)^{\frac{\nu-1}{\nu}} < \infty$ *for $\nu > 1$, where $\alpha_{\boldsymbol{\varepsilon}}(j)$ is the j-th strong mixing coefficient of process $\boldsymbol{\varepsilon}_t$.*

Assumption 2.3.(a) implies that $\boldsymbol{\varepsilon}_t$ is a strict stationary, ergodic process. This assumption involves that $\boldsymbol{Y}_t$ follows a *strong* SVARMA model. Assumption 2.3.(b) is the basis of Independent Component Analysis (ICA) studied by Comon (1994). In this paper, it is stated that if $\boldsymbol{\varepsilon}_t$ contains at most one Gaussian component, then a linear combination of $\boldsymbol{\varepsilon}_t$, i.e., $\boldsymbol{B}\boldsymbol{\varepsilon}_t$, is identified up to signed-permutation. Under the context of this article, on the contrary, it is required that all the components in the vector of structural disturbances be non-Gaussian distributed because we are dealing with the dynamic identification (see Chan et al. (2006); Funovits (2020); Gouriéroux et al. (2019); Velasco (2022a); Velasco (2022b) for similar assumptions).

It is usual in the literature to characterize non-Gaussianity by having, at least, distributions with a non-zero skewness coefficient, i.e., asymmetric shocks. On the contrary, Assumption 2.3 does not impose such characterization, and, as detailed in the Appendix, such requirement is sufficient but not necessary for identification. Finally, Assumption 2.3.(c) is standard in SVAR literature and implies we work with standardized centered innovations. Finally, Assumption 2.3.(d) is a technical requirement, also required in Hong (1999), and ensures the existence of the GSD for the null hypothesis of independence.

Let consider the following transformation of model errors, $\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right)$. The expected value of this transformation always exists for any $\boldsymbol{\vartheta} \in \mathcal{V}$ satisfying Assumption 2.1 and under Assumption 2.3, because is the characteristic function of model errors, $\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}) = \mathbb{E}[\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right)]$.

The joint characteristic function of the pair of model errors $\left(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}),\, \boldsymbol{\varepsilon}_{t-j}(\boldsymbol{\vartheta})\right)$ is:

$$\varphi_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \mathbb{E}[\exp\left(i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) + i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}(\boldsymbol{\vartheta})\right)],\ j = \ldots, -2, -1, 0, 1, 2, \ldots. \tag{2.4}$$

Therefore, our dependence measure is defined as

$$\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) := \varphi_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2). \tag{2.5}$$

This measure captures all the pairwise dynamics of the pair $\left(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}),\, \boldsymbol{\varepsilon}_{t-j}(\boldsymbol{\vartheta})\right)$.

Dependence measure in (2.5) is equivalent to

$$\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \mathbb{C}\mathrm{ov}\left(z_t(\boldsymbol{\tau}_1, \boldsymbol{\vartheta}), z_{t-j}(\boldsymbol{\tau}_2, \boldsymbol{\vartheta})\right), \tag{2.6}$$

where $z_t(\boldsymbol{\tau}, \boldsymbol{\vartheta}) = e^{i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} - \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau})$. This is why we called equation (2.5) as *generalized autocovariance* of order j. Notice that $\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ satisfies the following: (i) $\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(-\boldsymbol{\tau}_1, -\boldsymbol{\tau}_2) = \overline{\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)}$; (ii) $\sigma_{-j}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)$; and (iii) $\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = 0 \Leftrightarrow \varphi_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2)$. This latter implies independence of $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ and $\boldsymbol{\varepsilon}_{t-j}(\boldsymbol{\vartheta})$ if holds for every $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \in \mathbb{R}^{2d}$.

Using the previous dependence measure, the GSD is

$$s_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega; \boldsymbol{\tau}_1, \boldsymbol{\tau}_2) := \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) e^{-i\omega j}, \quad \omega \in [-\pi, \pi]. \tag{2.7}$$

Note that $s_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega; \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ is the Fourier transform of generalized autocovariance function, $\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$.

The cumulative spectral distribution (hereafter CSD) function for any interval of the form $[-\omega, \omega] \subset [-\pi, \pi]$ is defined as

$$\begin{aligned} F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega; \boldsymbol{\tau}_1, \boldsymbol{\tau}_2) &:= \int_{-\omega}^{\omega} s_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\lambda; \boldsymbol{\tau}_1, \boldsymbol{\tau}_2) d\lambda \\ &= \frac{\omega}{\pi}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) + \sum_{|j|\neq 0} \frac{\sin(j\omega)}{j\pi}\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2),\ \omega \in [0, \pi]. \end{aligned} \tag{2.8}$$

Now, under Assumption 2.3.(a), which imposes that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0)$ is an i.i.d. process, then $\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = 0$ for all $j \neq 0$. Thus, the form of $F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega; \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$, when the i.i.d. assumption is imposed, is equal to

$$F^{IID}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\frac{\omega}{\pi}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2).$$

And, if Assumption 2.3.(b) is additionally assumed, then the form of $F^{IID}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ for the restricted case of having i.i.d. structural shocks with mutual independent components is equivalent to

$$F^{IICA}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\frac{\omega}{\pi}\left[\prod_{m=1}^{d}\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{1m})\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{2m})\right]$$

where $\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})=\mathbb{E}\left[\exp(i\tau_{1m}\varepsilon_{t,m}(\boldsymbol{\vartheta}))\exp(i\tau_{2m}\varepsilon_{t,m}(\boldsymbol{\vartheta}))\right]$ and $\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{1m})=\mathbb{E}\exp(i\tau_{1m}\varepsilon_{t,m}(\boldsymbol{\vartheta}))$.

The **population loss function** when structural shocks are independent is the L^2-distance between the unrestricted CSD, $F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$, and its restricted form, imposing i.i.d. and ICA assumptions, $F^{IICA}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ over all frequencies $\omega\in[0,\pi]$.

$$\begin{aligned}\mathscr{L}_0(\boldsymbol{\vartheta})&=\int\int\left|F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-F^{IICA}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 d\omega W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2) \qquad (2.9)\\ &=\frac{\pi}{3}\int\left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)+\frac{1}{2\pi}\sum_{j=1}^{\infty}\frac{1}{j^2}\int\left|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\end{aligned}$$

with $\sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\prod_{m=1}^{d}\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{1m})\varphi_0^{\varepsilon_m(\boldsymbol{\vartheta})}(\tau_{2m})$. $W(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ is a continuous weighting function which belongs to the space of bounded variation functions. A complementary technical condition is to assume that $\int(\|\boldsymbol{\tau}_1\|^a+\|\boldsymbol{\tau}_2\|^a)W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2)<\infty$ for $a\le 5$. For finding closed form expressions of $\mathscr{L}_0(\boldsymbol{\vartheta})$, Gaussian $W(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=W_1(\boldsymbol{\tau}_1)W_2(\boldsymbol{\tau}_2)$ is employed, with $W_n(\boldsymbol{\tau}_n)=(2\pi)^{-d/2}\exp\left(\boldsymbol{\tau}_n'\boldsymbol{\tau}_n\right)$. Although, any $W(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=W_1(\boldsymbol{\tau}_1)W_2(\boldsymbol{\tau}_2)$ with this property can be employed.

Theorem 2.1. *Let $\boldsymbol{Y}_t$ generated by the structural model in 4.2. If Assumptions 2.1 and 2.3 holds, then*

(a) $\mathscr{L}_0(\boldsymbol{\vartheta}_0)=0$*;*

(b) $\mathscr{L}_0(\boldsymbol{\vartheta})>0$ *for all* $\boldsymbol{\vartheta}\neq\bar{\mathbf{P}}\boldsymbol{\vartheta}_0$.

where $\bar{\mathbf{P}}=\begin{bmatrix}\boldsymbol{I}_{d^2(p+q)} & \mathbf{0}_{d^2(p+q)\times d^2}\\ \mathbf{0}_{d^2\times d^2(p+q)} & \boldsymbol{PD}'\otimes\boldsymbol{I}_d\end{bmatrix}$*, with* $\boldsymbol{P}$ *a permutation matrix and* $\boldsymbol{D}$ *a diagonal sign matrix.*

Theorem 2.1 establishes that our population loss function $\mathscr{L}_0(\boldsymbol{\vartheta})$ is minimized only at points of the form $\boldsymbol{\vartheta}=\bar{\mathbf{P}}\boldsymbol{\vartheta}_0$. This result means our model is identified up to the signed permutation. $\mathscr{L}(\boldsymbol{\vartheta}) = 0$ implies that we have found a model, given by the filter $\tilde{\boldsymbol{\Psi}}(L, \boldsymbol{\vartheta})$, such that model errors, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$, are pairwise independent. We are aware that this is different from being fully independent. Nonetheless, the gap between pairwise and full independence is difficult to characterize. As far as our structural model is linear, this gap may not be critical. A way to shrink this gap is to consider dependence measure for higher tuples, such as $\sigma_{j,l}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3) = \mathbb{E}\left[z_t(\boldsymbol{\tau}_1, \boldsymbol{\vartheta})z_{t-j}(\boldsymbol{\tau}_2, \boldsymbol{\vartheta})z_{t-l}(\boldsymbol{\tau}_3, \boldsymbol{\vartheta})\right]$. Of course, the more higher-order tuples are included in the loss function, the smaller the gap.

Besides, it is worth highlighting the difference between Theorem 2.1 to Chan et al. (2006) identification result. In the work of Chan et al., identification of a very general stationary linear process requires that each component of structural disturbances, $\{\varepsilon_{t,m}\}_{m=1}^{d}$, has finite non-zero skewness coefficient. In our case, the identification of the structural model does not depend on such particular characterization of non-Gaussianity, but on a broader sense of non-Gaussian behavior. The intuition behind this could be the following. Complete independence implies that shocks are serially uncorrelated and orthogonal across their components when structural shocks are Gaussian distributed. There exists different types of models or parameter vector, $\tilde{\boldsymbol{\Psi}}(L, \boldsymbol{\vartheta})$, whose roots may lie inside or outside unit circle, that generates serially uncorrelated model errors. This feature means that the identified set would be equal to the space of all-pass filters ($\boldsymbol{\delta}(L, \boldsymbol{\vartheta})$). As argued in Hong (1999), when each structural shock is non-Gaussian distributed, being serially and marginal uncorrelated, it is not sufficient for obtaining independent model errors. Hence, the combination of complete independence and non-Gaussian behavior helps to discriminate among the different all-pass filters. Assuming pairwise independence characterizes full independence, the remaining set of filters is discrete because any sign permutation of the static component leads to marginal independent shocks.

Complete or Global Identification

Identification up to signed-permutation reduces the identified set to a discrete one, meaning that model is not point-identified. For achieving global or point identification, it is necessary to have a mechanism for selecting a specific permutation. In the literature, there exist several procedures for achieving this goal. In this paper, we follow the idea stated by Hallin and Mehta (2015), which is also employed by Lanne, Meitz, et al. (2017).

Hallin's procedure consists of re-defining the set where the contemporaneous matrix coefficient, $\boldsymbol{B}(\boldsymbol{\vartheta})$, belongs. In general, it is assumed that $\boldsymbol{B}(\boldsymbol{\vartheta}) \in \mathscr{M}^{d\times d}$, where $\mathscr{M}^{d\times d}$ denotes the space of square, real-valued, non-singular matrices of size d. Nonetheless, let consider the set $\ell(d) \subset \mathscr{M}^{d\times d}$, which is the set of all matrices $\boldsymbol{B}$ such that $\boldsymbol{D}_1^{\boldsymbol{B}}$, $\boldsymbol{P}^{\boldsymbol{B}}$ and $\boldsymbol{D}_2^{\boldsymbol{B}}$ exist and satisfy the following:

(i) $\boldsymbol{D}_1^{\boldsymbol{B}} = \text{diag}(d_{11}^B, \ldots, d_{1d}^B)$ with $d_{1m}^B > 0$, and $\boldsymbol{C}_1 := \boldsymbol{B}\boldsymbol{D}_1^{\boldsymbol{B}} = (\boldsymbol{c}_{11} \ldots \boldsymbol{c}_{1d})$ with $\|\boldsymbol{c}_{1m}\| = 1$ for all $m = 1, \ldots, d$;

(ii) $\boldsymbol{P}^{\boldsymbol{B}}$ is a permutation matrix such that $\boldsymbol{C}_P = \boldsymbol{C}_1^{\boldsymbol{B}}\boldsymbol{P}^{\boldsymbol{B}} = [c_{mj,P}]_{m,j=1}^d$ satisfies $|c_{mm,P}| > |c_{mj,P}|$ for $m < j \leq d$ and $m = 1, \ldots, d$;

(iii) $\boldsymbol{D}_2^{\boldsymbol{B}}$ is a diagonal matrix such $\boldsymbol{C}_2 := \boldsymbol{B}\boldsymbol{D}_1^{\boldsymbol{B}}\boldsymbol{P}^{\boldsymbol{B}}\boldsymbol{D}_2^{\boldsymbol{B}} = \left[C_{mj}\right]_{m,j=1}^d$ with $C_{mm,2} = 1$ for all $m = 1, \ldots, d$.

For the rest of the paper, we assume that the true static component $\boldsymbol{B}(\boldsymbol{\vartheta}_{S,0}) \in \ell(d)$, which allows us to employ this selection approach. Alternatively, Lanne, Meitz, et al. (2017) or Gouriéroux et al. (2019) suggest that permutation selection can be made by imposing a particular shape of the impulse response function.

2.2.6. Identification with dependent structural shocks

A major drawback of our previous identification result is that assumes $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence. This requirement is quite restrictive, especially in empirical macroeconomics and finance, where it is common to find that observed data exhibit volatility that changes over time. This characteristic implies that structural disturbances in the model may exhibit conditional heteroskedasticity, leading to discarding independence across time.

One inconvenience, in this case, is to make explicit the type of dependence in structural errors. We cannot assume the weak white noise condition on structural errors because, as explained above, this requirement does not provide sufficient information for overcoming the dynamic identification problem. However, we would like not to impose a specific dependence structure, e.g., ARCH or GARCH behavior of structural shocks, but to keep our problem as general as possible.

Assumption 2.4. *The structural shocks, $\boldsymbol{\varepsilon}_t$, in model (4.2) satisfies:*

(a) $\mathbb{E}\left[\boldsymbol{\varepsilon}_t \,|\, \mathscr{I}_{t-1}^{\boldsymbol{\varepsilon}}\right] = 0$, *where* $\mathscr{I}_{t-1}^{\boldsymbol{\varepsilon}} = \sigma(\{\boldsymbol{\varepsilon}_s : s \leq t-1\})$.

(b) $\boldsymbol{\varepsilon}_t$ *is a strict stationary, ergodic, strong mixing process with mixing coefficient* $\alpha(j) = O(|j|^{1-\mu_1})$, $\mu_1 > 1$ *and* $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\nu-4}{\nu}} < \infty$ *for* $\nu > 5$.

(c) $\boldsymbol{\varepsilon}_t = \left(\varepsilon_{1,t}, \ldots, \varepsilon_{d,t}\right)$ *has mutually independent components and* $\varepsilon_{m,t}$ *is non-Gaussian distributed for each* $m = 1, \ldots, d$. $\mathbb{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'] = \boldsymbol{I}_d$

(d) $\mathbb{E}[\boldsymbol{\varepsilon}_s \,|\boldsymbol{H}(L)\boldsymbol{\varepsilon}_t]$ *is a non-linear function for some* $s, t \in \mathbb{Z}$, *where* $\boldsymbol{H}(L)$ *is a possibly infinite, two sided linear filter.*

Assumption 2.4.(a) states the time dependence structure of $\boldsymbol{\varepsilon}_t$. It means that $\boldsymbol{\varepsilon}_t$ is a martingale difference sequence (m.d.s.). Being a m.d.s. implies milder restrictions than independence on the memory of the sequence of structural shocks; however, it is still stronger than the white noise condition. This assumption is a quite general dependence structure, which covers relevant cases common in empirical macroeconomics and finance, such as conditional heteroskedasticity (GARCH errors) or stochastic volatility innovations. To our knowledge of the current literature of structural VAR or VARMA models with non-Gaussian errors, identification with dependent shocks has not been explored unless the structural model is assumed to be fundamental. For instance, Lanne, Meitz, et al. (2017) obtain identification of the static component in a causal structural VAR model with only white noise shocks.

Assumption 2.4.(b) it is necessary since being a m.d.s. does not imply to be stationary. This condition is required due to technical reasons because it permits to *approximate* the behavior of a possibly infinite, two-sided moving average, such as the model errors $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$, by a truncated (finite), two-sided moving average, $\overline{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})$. Assumption 2.4.()c) is the ICA condition and imposes that each shock is non-Gaussian distributed. Besides, it states that shocks are standardized. Assumption 2.4.()d) is crucial for identification. It can be interpreted as our characterization of non-Gaussian behavior. Rosenblatt, 2000, Section 5.4, requires a similar condition for proving that a possibly non-fundamental univariate ARMA model is identified.

A sufficient condition for Assumption 2.4.(d) to hold is that each structural shock has at least a non-zero third-order cumulant. Chen et al., 2017 develop a procedure for testing the martingale difference property within the context of a non-fundamental structural model, and requires a stronger version of Assumption 2.4.(d). Specifically, they impose that all the structural shocks in the model have non-zero third-order cumulant, i.e., each structural error exhibits asymmetric behavior. This characterization is partially unsatisfactory since it rules out cases where non-Gaussian behavior comes from kurtosis's excess (or deficit). A similar assumption is made by Funovits, 2020; Gouriéroux et al., 2019. In our case, this condition is necessary and sufficient and rules out the possibility that $\boldsymbol{\varepsilon}_s$ and $\boldsymbol{\delta}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t$ are jointly distributed as a multivariate Pearson or multivariate t-student since such distributions produce linear conditional expectations (see Kotz et al. (2004, p.6-10) and Roth (2012) for more examples).

Loss Function

When $\boldsymbol{\varepsilon}_t$ is a m.d.s.[2], then in general $\sigma_j^{\varepsilon}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) \neq 0$ for any j and $(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) \in \mathbb{R}^{2d}$. Hence,

[2] $\mathbb{E}\left[\boldsymbol{\varepsilon}_t | \mathscr{I}_{t-1}^{\varepsilon}\right] = \mathbf{0}$ can be interpreted as a conditional mean independence assumption.

the loss function $\mathscr{L}_0(\boldsymbol{\vartheta})$ cannot be used for identification purposes because it is unknown the functional form of $\sigma_j^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ under a the m.d.s. assumption.

Let define $\Delta^{(r_1,r_2)}\sigma_j^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) := \frac{\partial^{r_1+r_2}}{(\partial\boldsymbol{\tau}_1')^{r_1}(\partial\boldsymbol{\tau}_2')^{r_2}}\sigma_j^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$. Setting $r_1 = 1$ and $r_2 = 0$ we obtain

$$\begin{aligned}
\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) := \frac{\partial}{\partial\boldsymbol{\tau}_1'}\sigma_{|j|}^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) &= \mathbb{C}\mathrm{ov}\left(\frac{\partial}{\partial\boldsymbol{\tau}_1'}z_t(\boldsymbol{\tau}_1),\, z_{t-|j|}(\boldsymbol{\tau}_2)\right) \\
&= i\,\mathbb{C}\mathrm{ov}\left\{\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t} - \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right],\, z_{t-|j|}(\boldsymbol{\tau}_2)\right\} \\
&= i\,\mathbb{C}\mathrm{ov}\left\{\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t},\, z_{t-|j|}(\boldsymbol{\tau}_2)\right\} = i\,\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\cdot z_{t-|j|}(\boldsymbol{\tau}_2)\right)
\end{aligned}$$

where last equality uses the fact that $\mathbb{E}\left[z_{t-|j|}(\boldsymbol{\tau}_2)\right] = 0$.

Evaluating the expression above at $\boldsymbol{\tau}_1 = \mathbf{0}$, we get $\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}}(\mathbf{0},\boldsymbol{\tau}_2) = \mathbb{E}\left(i\boldsymbol{\varepsilon}_t\cdot z_{t-|j|}(\boldsymbol{\tau}_2)\right)$. Therefore:

$$\mathbb{E}[\boldsymbol{\varepsilon}_t|\mathscr{I}_{t-1}^{\boldsymbol{\varepsilon}}] = 0 \Rightarrow \Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}}(\mathbf{0},\boldsymbol{\tau}_2) = 0\ \forall\ |j|>0,\ \boldsymbol{\tau}_2\in\mathbb{R}^d. \tag{2.10}$$

To be strict, $\mathbb{E}\left(i\boldsymbol{\varepsilon}_t\cdot z_{t-|j|}(\boldsymbol{\tau}_2)\right) = 0$ for all $\boldsymbol{\tau}_2\in\mathbb{R}^d$ if and only if $\mathbb{E}[\boldsymbol{\varepsilon}_t|\boldsymbol{\varepsilon}_{t-j}] = 0$, which is not exactly the definition of being a m.d.s.. Let call a process satisfying $\mathbb{E}[\boldsymbol{\varepsilon}_t|\boldsymbol{\varepsilon}_{t-j}] = 0$ as a *pairwise martingale difference sequence* (p.m.d.s.). The gap between being a p.m.d.s. and a m.d.s. is difficult to characterize but may be important. Nonetheless, as discussed in Hong and Lee (2005), the p.m.d.s. covers some interesting time series models. A more accurate approximation to a m.d.s. is achieved by considering triples, such that $\Delta^{(1,0,0)}\sigma_{|j|,|l|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\boldsymbol{\tau}_3)$, or even higher tuples. However, again, the curse of dimensionality becomes an important issue.

Like in the independence case, we define the generalized spectrum for the dependent case as the Fourier transform of covariance function $\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) = i\,\mathbb{C}\mathrm{ov}\left\{\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t},\, z_{t-|j|}(\boldsymbol{\tau}_2)\right\}$

$$\varsigma_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega,\boldsymbol{\tau}_2) := \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\left\{\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\}e^{-i\omega j},\ \omega\in[-\pi,\pi],$$

and its cumulative distribution function is

$$\begin{aligned}
F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}(\omega;\boldsymbol{\tau}_2) &:= \int_{-\omega}^{\omega}\varsigma_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\lambda,\boldsymbol{\tau}_2)d\lambda \\
&= \frac{\omega}{\pi}\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) + \sum_{|j|\neq 0}\frac{\sin(j\omega)}{j\pi}\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2),\ \omega\in[0,\pi].
\end{aligned} \tag{2.11}$$

Accordingly to (2.10), when $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ is a m.d.s., the restricted form of $F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}(\omega;\boldsymbol{\tau}_2)$ is equal to

$$\left[F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MD}(\omega;\boldsymbol{\tau}_2) = \frac{\omega}{\pi}\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)$$

and, if $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ also satisfies mutual independence components, $[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}]^{MD}(\omega;\boldsymbol{\tau}_2)$ collapses to

$$\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\omega;\boldsymbol{\tau}_2) = i\frac{\omega}{\pi}\begin{bmatrix} \mathbb{C}\mathrm{ov}\left(\varepsilon_{t,1}(\boldsymbol{\vartheta}),\, z_t(\tau_{2,1},\boldsymbol{\vartheta})\right)\prod_{m\neq 1}\varphi_m^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,m}) \\ \vdots \\ \mathbb{C}\mathrm{ov}\left(\varepsilon_{t,d}(\boldsymbol{\vartheta}),\, z_t(\tau_{2,d},\boldsymbol{\vartheta})\right)\prod_{m\neq d}\varphi_m^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,m}) \end{bmatrix}$$

where $z_{m,t}(\tau_{2,m},\boldsymbol{\vartheta}) = e^{i\tau_{2,m}\varepsilon_{t,m}(\boldsymbol{\vartheta})} - \mathbb{E}[e^{i\tau_{2,m}\varepsilon_{t,m}(\boldsymbol{\vartheta})}]$.

Similarly to the situation of a model with independent shocks, the loss function is the L^2-distance between the unrestricted CSD, $F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_2)$, and its restricted version when structural errors are a m.d.s. and have mutual independent components, $\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\omega;\boldsymbol{\tau}_2)$.

$$\mathscr{R}_0(\boldsymbol{\vartheta}) = \int\int\left\|F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_2) - \left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\omega;\boldsymbol{\tau}_2)\right\|^2 d\omega W(d\boldsymbol{\tau}_2) \tag{2.12}$$

$$= \frac{\pi}{3}\int\left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \begin{bmatrix} \mathbb{C}\mathrm{ov}\left(\boldsymbol{\varepsilon}_{t,1}(\boldsymbol{\vartheta}),\, z_t(\tau_{2,1},\boldsymbol{\vartheta})\right)\prod_{m\neq 1}\varphi_m^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,m}) \\ \vdots \\ \mathbb{C}\mathrm{ov}\left(\boldsymbol{\varepsilon}_{t,d}(\boldsymbol{\vartheta}),\, z_t(\tau_{2,d},\boldsymbol{\vartheta})\right)\prod_{m\neq d}\varphi_m^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,m}) \end{bmatrix}\right\|^2 W(d\boldsymbol{\tau}_2)$$

$$+\frac{1}{2\pi}\sum_{j=1}^{\infty}\frac{1}{j^2}\int\left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 W(d\boldsymbol{\tau}_2), \tag{2.13}$$

where $W(\boldsymbol{\tau}_2)$ is a differentiable weighting matrix, satisfying similar properties like the one specified in equation (2.9).

Using the loss function, $\mathscr{R}_0(\boldsymbol{\vartheta})$, and the dependence structure for structural errors, we state identification of the model described in (4.2).

Theorem 2.2. *If Assumptions 2.1, 2.2 and 2.4 hold, then*

(a) $\mathscr{R}_0(\boldsymbol{\vartheta}_0) = 0$*;*

(b) $\mathscr{R}_0(\boldsymbol{\vartheta}) > 0$ *for all* $\boldsymbol{\vartheta} \neq \overline{\mathbf{P}}\boldsymbol{\vartheta}_0$*;*

where $\overline{\mathbf{P}}$ *is any permutation matrix defined similarly like in Theorem 2.1.*

The result in Theorem 2.2 means that the structural model described by (4.2) is identified up to sign-permutations. For achieving this, we only require that structural shocks have finite second-order moments and exhibit non-Gaussian behavior in a broad sense (Assumption 2.4.(d)).

2.3. Estimation

This section details the procedure for estimating the structural parameters vector, $\boldsymbol{\vartheta}$, in both dependence structures, i.e., with i.i.d. errors and m.d.s. disturbances. We construct the sample counterpart of the population loss functions defined in the previous section. Furthermore, note that model errors, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-j}$, are not observable since we have a finite sample of the endogenous variables $\{\boldsymbol{Y}_t : t = 1,\dots,T\}$. Thus, we employ the sample model residuals, which are defined as follows:

$$\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) :- \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-j}\mathbb{I}\{1 \le t-j \le T\},$$

where $\mathbb{I}\{1 \le t-j \le T\}$ is equal to 1 if the lag/lead j is in the range $\{t-T,\dots,t-1\}$, and 0 otherwise.

Moreover, it is common that endogenous variables may have a deterministic component ($\boldsymbol{\mu}_t$), i.e., $\boldsymbol{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$. Typically the derterministic component is a non-zero mean, $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$, or linear trend, $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$. For our analysis, $\boldsymbol{\mu}_t$ is irrelevant and can be estimated by OLS in a previous step. For the rest of the paper $\boldsymbol{Y}_t$ denotes the de-meaned or de-trended observable data.

Given the model residuals, $\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})$, the sample counterparts of the joint characteristic function, the dependence measure and the CSD are

$$\hat{\varphi}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T} e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})}$$

$$\hat{\sigma}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T} \hat{z}_t(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\hat{z}_{t-|j|}(\boldsymbol{\tau}_2,\boldsymbol{\vartheta}) = \hat{\varphi}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{0})\hat{\varphi}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{0},\boldsymbol{\tau}_2)$$

$$\hat{F}_{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\omega;\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \frac{\omega}{\pi}\hat{\sigma}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) + 2\sum_{j=1}^{T-1}\hat{\sigma}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\frac{\sin(j\omega)}{j\pi},$$

with $\hat{z}_t(\boldsymbol{\tau}_1,\boldsymbol{\vartheta}) = \exp(i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})) - \hat{\varphi}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{0})$.

Although the GSD has been employed in Hong (1999) for testing serial dependence, its consistent estimator involves using a smoother kernel, $\mathcal{K}(j/h)$ with bandwidth h. The introduction of this smoother affects the rate of convergence of our estimators. In this sense, the CSD has a tremendous advantage over the GSD because it avoids the usage of a smoother factor.

2.3.1. Estimation with independent shocks

The sample loss function, $\mathscr{L}_T(\boldsymbol{\vartheta})$, is

$$\mathscr{L}_T(\boldsymbol{\vartheta}) = \frac{\pi}{3}\int \left|\hat{\sigma}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \hat{\sigma}_{0,IICA}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2) + \frac{1}{2\pi}\sum_{j=1}^{T-1}\left(1-\frac{|j|}{T}\right)\frac{1}{j^2}\int \left|\hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2)$$

where $\hat{\sigma}_{0,IICA}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \prod_{m=1}^d \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^d \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\tau_{1m},0)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(0,\tau_{2m})$.

The estimator of structural parameters vector, $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$, is defined as the minimizer of sample loss function $\mathscr{L}_T(\boldsymbol{\vartheta})$,

$$\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} = \arg\min_{\boldsymbol{\vartheta}\in\mathcal{V}} \mathscr{L}_T(\boldsymbol{\vartheta}). \tag{2.14}$$

Given the compactness of the parameters space, $\mathcal{V}$, and that $\mathscr{L}_T(\boldsymbol{\vartheta})$ is, at least, a twice differentiable function regarding $\boldsymbol{\vartheta}$, the existence of $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$ is guaranteed. Taken as known the values of overall degrees (p,q) and the order of non-fundamental polynomials (p_+,q_+), the estimator $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$ solves:

$$\mathrm{s}_T^{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) := \left.\frac{\partial \mathscr{L}_T(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}} = 0$$

where $\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta})$ is the *score function*, i.e. the gradient vector of loss function $\mathscr{L}_T(\boldsymbol{\vartheta})$.

Consistency

Before setting the consistency of our minimum distance estimator ($\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$), it is necessary to show that our sample loss function ($\mathscr{L}_T(\boldsymbol{\vartheta})$) converges uniformly to its population version.

Theorem 2.3. *If Assumptions 2.1-2.3 hold and using Hallin's permutation selection, then*

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}} |\mathscr{L}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})| \xrightarrow{p} 0, \textit{ as } T\to\infty$$

Theorem 2.3 shows that, given our parameterization and sufficient conditions that guarantee the existence of GSD, the sample loss function, $\mathscr{L}_T(\boldsymbol{\vartheta})$, converges uniformly to the population loss function, $\mathscr{L}_0(\boldsymbol{\vartheta})$.

Once the uniform convergence of the sample loss function is provided, consistency of our minimum distance estimator, $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$, is straightforward.

Theorem 2.4. *If Assumptions 2.1-2.3 hold and using Hallin's permutation selection, then*

$$\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} \xrightarrow{p} \boldsymbol{\vartheta}_0, \text{ as } T \to \infty$$

Theorem 2.4 establishes the consistency of our proposed estimator using sample loss function $\mathscr{L}_T(\boldsymbol{\vartheta})$. The assumptions guarantee that all the requirements in Amemiya (1985, Theorem 4.1) are satisfied. Furthermore, this result depends explicitly on the existence of second-order moments for the structural shocks, i.e., $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^2 < \infty$, which is standard in the literature. Velasco (2022b) shows that efficiency of $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$ can be improved using Newton-Raphson updating procedure but using the GSD.

Asymptotic Normality

Once the consistency of our minimum distance estimator is provided, another important asymptotic result is to show the convergence in the distribution of the estimator. However, as a preliminary step, we show that derivatives of the all-pass filter, $\boldsymbol{\delta}(L,\boldsymbol{\vartheta})$, are bounded given the structure of our structural model and our parameterization. For instance, when $K = \bar{d}$ and $p_+ > 0$, we can compute the following

$$\begin{aligned}
\frac{\partial}{\partial \phi_{+,i}} \boldsymbol{\delta}(L,\boldsymbol{\vartheta}) &= \frac{\partial}{\partial \phi_{+,i}} \left(\boldsymbol{\Phi}^{-1}(L,\boldsymbol{\vartheta})\boldsymbol{\Theta}(L,\boldsymbol{\vartheta})\boldsymbol{B}(\boldsymbol{\vartheta})\right)^{-1} \boldsymbol{\Psi}(L,\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{\vartheta})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\vartheta}) \left(\frac{\partial}{\partial \phi_{+,i}} \boldsymbol{\Phi}_+(L,\boldsymbol{\phi}_+)\right) \boldsymbol{\Phi}_-(L,\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L,\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{\vartheta})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\vartheta}) \left(\frac{\partial}{\partial \phi_{+,i}} \boldsymbol{\Phi}_+(L,\boldsymbol{\phi}_+)\right) \boldsymbol{\Phi}_-(L,\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L,\boldsymbol{\vartheta}_0).
\end{aligned}$$

Hence, given that $\boldsymbol{\Psi}(L,\boldsymbol{\vartheta})$ and $\boldsymbol{\Psi}^{-1}(L,\boldsymbol{\vartheta}_0)$ are absolute summable polynomials, and under an appropriate decay rate of their coefficients, it holds that the derivative polynomial $\frac{\partial}{\partial \phi_{+,i}} \boldsymbol{\delta}(L,\boldsymbol{\vartheta})$ satisfies absolute summability with coefficients decaying at a similar rate to $\boldsymbol{\Psi}(L,\boldsymbol{\vartheta})$ and $\boldsymbol{\Psi}^{-1}(L,\boldsymbol{\vartheta}_0)$. Then, unlike Velasco (2022a), assuming a specific behavior for these derivatives is not necessary.

Theorem 2.5. *Under Assumptions 2.1-2.3 hold and using Hallin's permutation selection. If $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^3 < \infty$, then*

(a) $\left.\frac{\partial^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} \mathscr{L}_T(\boldsymbol{\vartheta})\right|_{\boldsymbol{\vartheta}=\tilde{\boldsymbol{\vartheta}}_T} \xrightarrow{p} \boldsymbol{\Lambda}_{IICA}(\boldsymbol{\vartheta}_0) \equiv \mathbb{E}\left[\frac{\partial^2}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} \mathscr{L}_T(\boldsymbol{\vartheta}_0)\right]$ *for any* $\tilde{\boldsymbol{\vartheta}}_T \xrightarrow{p} \boldsymbol{\vartheta}_0$.

(b) The random vector $\sqrt{T}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$ *converges in distribution to a multivariate Normal with*

mean zero and variance $\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)$, i.e.

$$\sqrt{T}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)\right)$$

and

$$\sqrt{T}\left(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} - \boldsymbol{\vartheta}_0\right) \xrightarrow{d} \mathcal{N}_K\left(\mathbf{0}, \mathbf{\Lambda}_{IICA}^{-1}(\boldsymbol{\vartheta}_0)\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)\mathbf{\Lambda}_{IICA}^{-1}(\boldsymbol{\vartheta}_0)\right).$$

Asymptotic Standard Errors

$\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)$ denotes the asymptotic variance of $\sqrt{T}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$.

$$\begin{aligned} T^{1/2}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) = & -\left[\mathrm{vec}(\mathbf{S}_{-\mathrm{diag}}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))' \otimes \boldsymbol{I}_{\tilde{d}}\right]\left\{\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\tilde{e}_t^0 - \tilde{e}_t^{0,1} - \tilde{e}_t^{0,2} - \check{e}^{0,1} - \check{e}^{0,2} - \check{e}^{0,3}\right)\right\} \\ & - \frac{1}{T^{1/2}}\sum_{t=2}^{T}\left[X_{t-1}^0 e_t^0 + E_{t-1}^0 x_t^0\right] + o_p(1) \end{aligned}$$

where

$$\begin{aligned} \tilde{e}_t^{0,1} &= \mathrm{Re}\left(i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\tilde{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right] dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right) \\ \tilde{e}_t^{0,2} &= \mathrm{Re}\left(-i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\tilde{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes \boldsymbol{\tau}_1'\right]\right) \\ \tilde{e}_t^{0,3} &= \mathrm{Re}\left(-i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\mathrm{vec}\left[\left(\boldsymbol{I}_{\tilde{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes \boldsymbol{\tau}_2'\right] dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right) \\ \check{e}^{0,1} &= \mathrm{Re}\left(i\int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\tilde{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right] dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right) \\ \check{e}^{0,2} &= \mathrm{Re}\left(-i\int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\tilde{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes \boldsymbol{\tau}_1'\right]\right) \\ \check{e}^{0,3} &= \mathrm{Re}\left(-i\int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\tilde{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes \boldsymbol{\tau}_2'\right] dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right) \end{aligned}$$

and $z_t^0(\boldsymbol{\tau}) = z_t(\boldsymbol{\tau}, \boldsymbol{\vartheta}_0)$. Then, the asymptotic variance $\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)$ is

$$\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0) = \mathbf{C}(\boldsymbol{\vartheta}_0)\mathbb{E}\left[S_t S_t'\right]\mathbf{C}(\boldsymbol{\vartheta}_0)'$$

where $\mathbf{C}(\boldsymbol{\vartheta}_0) = \left[-\left[\mathrm{vec}(\mathbf{S}_{-\mathrm{diag}}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))' \otimes \boldsymbol{I}_{\tilde{d}}\right] \vdots\ \boldsymbol{I}\right]$, $\mathbb{E}\left[S_t S_t'\right] = \mathbb{E}\begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22}\end{bmatrix}$, $S_{11} = \mathbb{E}\left(\tilde{E}_t^0 \tilde{E}_t^{0\prime}\right)$ with $\tilde{E}_t^0 = \tilde{e}_t^0 - \tilde{e}_t^{0,1} - \tilde{e}_t^{0,2} - \left[\check{e}^{0,1} + \check{e}^{0,2} + \check{e}^{0,3}\right]$, $S_{12} = \mathbb{E}\left(\tilde{E}_t^0[X_{t-1}^0 e_t^0 + E_{t-1}^0 x_t^0]'\right)$ and $S_{22} = \mathbb{E}\left([X_{t-1}^0 e_t^0 + E_{t-1}^0 x_t^0][X_{t-1}^0 e_t^0 + E_{t-1}^0 x_t^0]'\right)$. The symbol $[\circ]$ denotes the penetrating face product, a generalization of Hadamard element-wise product (see Slyusar (1999) for technical details).

A consistent estimation of $\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)$ is obtained by using the sample counterparts of the elements. We employ a Gaussian weighting function, W, to simplify the calculus because it allows us to find closed-form solutions to the integrals. In Appendix (5.10.1), we can find the details of this computation procedure. Alternatively, $\mathbf{\Omega}_{IICA}(\boldsymbol{\vartheta}_0)$ can be consistently estimated by employing a non-parametric method, such as a bootstrap procedure with a bootstrap sampling that needs to preserve the dependence structure of observable variables, e.g., block bootstrapping.

2.3.2. Estimation with dependent shocks

The sample loss function, $\mathscr{R}_T(\boldsymbol{\vartheta})$, for the sample model residuals $\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})$ is

$$\mathscr{R}_T(\boldsymbol{\vartheta}) = \frac{\pi}{3}\int \left\| \Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \begin{bmatrix} \mathbb{C}\mathrm{ov}_T\left(\hat{\varepsilon}_{t,1}(\boldsymbol{\vartheta}),\, \hat{z}_t(\tau_{2,1},\boldsymbol{\vartheta})\right)\prod_{j\neq 1}\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,j}) \\ \vdots \\ \mathbb{C}\mathrm{ov}_T\left(\hat{\varepsilon}_{t,d}(\boldsymbol{\vartheta}),\, \hat{z}_t(\tau_{2,d},\boldsymbol{\vartheta})\right)\prod_{j\neq d}\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,j}) \end{bmatrix} \right\|^2 W(d\boldsymbol{\tau}_2)$$
$$+ \frac{1}{2\pi}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\frac{1}{j^2}\int \left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 W(d\boldsymbol{\tau}_2).$$

The estimator of structural parameters, $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}$, is defined as the minimizer of the sample loss function.

$$\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}} = \arg\min_{\boldsymbol{\vartheta}\in\mathscr{V}} \mathscr{R}_T(\boldsymbol{\vartheta}). \tag{2.15}$$

Like in the case of estimator $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$, compactness of $\mathscr{V}$ and differentiability of $\mathscr{R}_T(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\vartheta}$ guarantees the existence of the minimizer $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}$. Taken as known the overall order of lag polynomials and the order of non-fundamental ones, $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}$ is a solution to

$$\mathrm{s}_T^{\mathscr{R}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}) := \left.\frac{\partial \mathscr{R}_T(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}} = 0,$$

where $\mathbf{s}_T^{\mathscr{R}}(.)$ is the *score function*, that is the gradient of loss function $\mathscr{R}_T(.)$.

Consistency

Theorem 2.6. *If Assumptions 2.1, 2.2 and 2.4 hold, then*

$$\sup_{\boldsymbol{\vartheta}\in\mathscr{V}} |\mathscr{R}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta})| \xrightarrow{p} 0, \textit{ as } T\to\infty.$$

Theorem 2.6 shows the uniform convergence of our loss function to its population counterpart when the structural model is given by (4.2) and satisfies Assumptions 2.1 and 2.2, and structural shocks are dependent with structure given by Assumption (2.4). Given this result, the consistency of the estimator can be stated.

Theorem 2.7. *Under assumption 2.1, 2.2 and 2.4, and using Hallin's selection scheme, then*

$$\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}} \xrightarrow{p} \boldsymbol{\vartheta}_0, \text{ as } T \to \infty.$$

Like in the case of the independent shocks, Theorem 2.7 is a direct application of Theorem 4.1 in Amemiya, 1985. Besides, it is worth noticing that no additional requirements regarding moments' existence to those stated in Assumption 2.4 are needed. It is evident that if a sufficient condition similar to the one assumed in Chen et al. (2017) is imposed, Theorem 2.7 is still valid.

Asymptotic Normality

Theorem 2.8. *If Theorem 2.7 holds and assuming that $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^\nu < \infty$ for $\nu \geq 5$, then*

(a) $\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{R}_T(\boldsymbol{\vartheta})\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_T^\star} \xrightarrow{p} \boldsymbol{\Lambda}_{MDIC}(\boldsymbol{\vartheta}_0) \equiv \mathbb{E}\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{R}_T(\boldsymbol{\vartheta}_0)\right]$ *for any* $\tilde{\boldsymbol{\vartheta}}_T \xrightarrow{p} \boldsymbol{\vartheta}_0$.

(b) The vector $\sqrt{T}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}_0)$ converges in distribution to a multivariate Normal with mean zero and variance $\boldsymbol{\Omega}_{MDIC}(\boldsymbol{\vartheta}_0)$, i.e.

$$\sqrt{T}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}_0) \xrightarrow{d} \mathscr{N}(\mathbf{0}, \boldsymbol{\Omega}_{MDIC}(\boldsymbol{\vartheta}_0)), \text{ and}$$
$$\sqrt{T}\left(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}} - \boldsymbol{\vartheta}_0\right) \xrightarrow{d} \mathscr{N}_K\left(\mathbf{0}, \boldsymbol{\Lambda}_{MDIC}^{-1}(\boldsymbol{\vartheta}_0)\boldsymbol{\Omega}_{MDIC}(\boldsymbol{\vartheta}_0)\boldsymbol{\Lambda}_{MDIC}^{-1}(\boldsymbol{\vartheta}_0)\right).$$

2.3.3. Estimation of polynomial degrees (p_+, q_+)

Consistency of parameter estimators has been achieved by taking as fixed overall degrees (p,q) and the order of non-fundamental polynomials (p_+,q_+). Once these two pairs are known, the order of fundamental polynomials is given by $p_- = p - p_+$ and $q_- = q - q_+$. For applying the optimization problem, instead of minimizing over the whole parameter space $\mathscr{V}$, we split it into a finite number of partitions defined by all the possible values of duple (p_+,q_+) for a fixed pair (p,q). For instance, for $(p,q) = (1,1)$, we have four different subsets of $\mathscr{V}$ since $(p_+,q_+) \in \{(1,1);(1,0);(0,1);(0,0)\}$.

For each feasible pair (p_+,q_+), we find the estimator of structural parameters vector, $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}(p_+,q_+)$. After substituting the estimated parameter vector, the sample loss function will depend on the pair (p_+, q_+). Thus, the estimator of (p_+, q_+) is defined as:

$$(\hat{p}_+, \hat{q}_+) = \arg \min_{(p_+, q_+) \in \mathscr{D}(p,q)} \mathscr{L}_T^c(p_+, q_+) \tag{2.16}$$

where $\mathscr{D}(p,q) = \{0,1,\ldots,p\} \times \{0,1,\ldots,q\}$ and (p,q) the overall degrees of lag polynomials.

We consider the risk of over-parameterizing limited since the total number of parameters does not change at any pair in $\mathscr{D}(p,q)$. For instance, in the case of structural VARMA(1,0), the true number of parameters is $d(p+1)$. The proposed procedure above compares the loss function for causal and non-causal roots. In each case, the number of parameters is $d(p+1)$.

Estimation Algorithm

We detail how to obtain the estimator $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$ or $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$. Given the pair (p,q), $\mathscr{D}(p,q) = \{0,\ldots,p\} \times \{0,\ldots,q\}$ and $|\mathscr{D}(p,q)| = (p+1)(q+1)$.

1. Fix $(p_+, q_+) = (0,0)$ and $(p_-, q_-) = (p,q)$, i.e., it starts by fitting a causal and invertible structural VARMA model. The estimated parameter vector is $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}(0,0)$ or $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}(0,0)$. Besides, compute the causal and invertible roots. The overall number of roots is dp and dq, respectively.

2. Choose $(p_+, q_+) = (p', q') \neq 0,0$. Then $p_- = p - p'$ and $q_- = q - q'$.

   (i) Take dp_+ and dq_+ roots from the estimated causal and invertible roots, respectively, and flip them.

   (ii) Find an initial value of lag polynomials $\{\tilde{\boldsymbol{\Phi}}_+(L), \tilde{\boldsymbol{\Phi}}_-(L), \tilde{\boldsymbol{\Theta}}_+(L), \tilde{\boldsymbol{\Theta}}_-(L)\}$ that are coherent with the selected roots.

   (iii) Given the initial value of lag polynomials, find the estimator $\hat{\boldsymbol{\vartheta}}_T$ for the selected roots.

   (iv) The estimator $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}(p_+, q_+)$ ($\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}(p_+, q_+)$) is chosen from all the possible combinations of roots (dp_+, dp_-, dq_+, dq_-).

3. Do step 2 for each $(p', q') \in \mathscr{D}(p,q) \setminus \{(0,0)\}$.

4. Finally, the estimated structural parameters and the order of lag polynomials $(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}, \hat{p}_+, \hat{q}_+)$ $((\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}, \hat{p}_+, \hat{q}_+))$ is chosen such that it reports the smallest $\mathscr{L}_T(\hat{p}_+, \hat{q}_+)$ $(\mathscr{R}_T(\hat{p}_+, \hat{q}_+))$.

2.3.4. Lag Length Selection

Choosing the overall degrees (p,q) is a relevant problem, but it is beyond the scope of this paper. Standard procedures minimize information criteria, which are usually derived from likelihood. In our case, we follow the criteria proposed by Boubacar Mainassara (2012) for a weak VARMA model, i.e., a fundamental VARMA with serially uncorrelated model errors. This approach means that either in the case of independent shocks or with dependent ones, we choose (p,q) as the degrees of a fundamental VARMA approximation.

$$AIC_M^{MDS} = T\log(\hat{\Sigma}_\varepsilon) + \frac{T^2d^2}{Td - d^2(p+q)} + \frac{Td}{2(Td - d^2(p+q))}\text{tr}\left(\hat{\mathbf{\Omega}}_{11,T}\hat{\mathbf{\Lambda}}_{11,T}^{-1}\right)$$
$$AIC_M^{IID} = T\log(\hat{\Sigma}_\varepsilon) + Td + \frac{Td}{Td - d^2(p+q)}2(d^2(p+q))$$

where $\hat{\mathbf{\Omega}}_{11,T}$ is the square upper-block of $\hat{\mathbf{\Omega}}_T$ and $\hat{\mathbf{\Lambda}}_{11,T}$ is the upper block of $\hat{\mathbf{\Lambda}}_T$ with dimension $d^2(p+q)$ when shocks satisfy m.d.s property. $\hat{\Sigma}_\varepsilon$ is the estimated variance-covariance matrix of shocks.

2.4. Simulation Study

We evaluate the proposed methodology with the following simulation study. We use different data-generating processes (DGPs) and distributions for structural shocks.

2.4.1. Simple Case

The simulation setup is: $d = 2$, i.e. bi-dimensional vector, $p = 1$ and $q = 1$, i.e overall AR(1) and MA(1), respectively. The following tables summarize the proportion of successful location of the roots or the rate of correct estimation of the pair (p_+, q_+) conditional to the overall dynamic orders (p,q)).

Table 2.1 shows the rate of correct root location. The findings are mixed. In the case of simple structural causal VAR or invertible VMA, the rate of successful root location is around 91% and 95%, respectively. These rates represent around double the theoretical rate in the case of using only second-order moments[3], which is 50% for both possible representations. In the case of causal and invertible VARMA, the rate fluctuates between

[3]Remember, it is always possible to find a non-causal (non-invertible) model such that the residuals are serially uncorrelated. In case of a VAR(1), we have only two possible value for p_+, $\{0,1\}$. Then, the probability of having $p_+ = 0$ is 0,5 as well as the probability of having $p_+ = 1$. Similarly, for a VMA(1). In the case of a VARMA(1,1), we have four possible values for the pair (p_+, q_+), then the probability of each possible pair is 25%.

82 and 92 percent, which are high above the 25% when only second-order moments are used. However, in the case of a non-causal and non-invertible representation, the rates are lower and, in some cases, near the theoretical rates. In particular, the non-causal VAR is only identified correctly 51% of the times, using shocks distributed as chi-square with 4 and 5 degrees of freedom. This rate is barely above the 50% rate when second-order moments are used. For the simple non-invertible VMA, we obtain a 77% rate of correct root identification (1.5 times higher than the theoretical rate using only second-order moments). In the case of non-causal and non-invertible, the rate ranges from 41 to 51 percent, although it is almost twice the 25% theoretical rate. However, we expected to obtain a higher rate.

A possible explanation for the low rates of non-causal and (or) non-invertible models can be related to the numerical optimization procedure. The way we start the optimization procedure is to estimate a causal and (or) invertible VARMA model. Then, for obtaining non-causal and (or) non-invertible initial points, we invert the appropriated number of causal and (or) invertible roots to be consistent with the values of (p_+, q_+). That starting point is the one we use for finding the respective minimizer. However, as Dufour and Pelletier (2021) explain, unlike the univariate case, there does not exist a unique AR polynomial that produces some specified roots[4]. Since the optimization method we employ is based on derivatives, that is, a local minimizer; the solution is sensitive to the starting point. Besides, it is quite often that the higher the parameter space dimension, the more sensitive these methods become. Additionally, there could be more than one local minimum, even in the proper case.

Table 2.1: Rate of Successful Identification, $T = 250$.
(Local Solution)

Distribution	VARMA Models $(p=1,q=0)$ $(p_+=1)$	$(p=1,q=0)$ $(p_+=0)$	$(p=0,q=1)$ $(q_+=1)$	$(p=0,q=1)$ $(q_+=0)$	$(p=1,q=1)$ $(p_+=0,q_+=0)$	$(p=1,q=1)$ $(p_+=1,q_+=1)$
(χ^2_4, χ^2_5)	0.51	0.91	0.77	0.95	0.92	0.51
(t_4, t_5)	0.66	0.85	0.80	0.95	0.87	0.44
$(t_4, U(-2,2))$	0.40	0.60	0.61	0.90	0.82	0.41

Optimization procedure was performed through local approximation of Lagrangian, that includes barrier and non-linear constraints. The maximum number of number of iterations and function evaluations was set to 10^3 and 5×10^4, respectively.

Hence, it is better to employ a global optimization procedure instead of a local one. A possible alternative is to employ different random starting points; however, this could be very costly computationally. Alternatively, we can use a global optimization algorithm

[4]For instance, if our non-causal roots were $\lambda_1 = 1/2$ and $\lambda_2 = 1/3$. These can comes from $\mathbf{\Phi}(\lambda) = \begin{bmatrix} 1-2\lambda & 0 \\ 0 & 1-3\lambda \end{bmatrix}$ or $\mathbf{\Phi}(\lambda) = \begin{bmatrix} 1-2\lambda & 0 \\ 0.5 & 1-3\lambda \end{bmatrix}$

called *genetic algorithm*. This algorithm starts with an initial population, is created randomly, and then selects the points below the objective function's average value. Once the *first generation survivors* are selected, it creates a second generation of points based on the former survivors, called *children*, and also *cross-over* or *mutations* points. The procedure continues until obtaining convergence. The pros of this algorithm are: (i) it is less sensitive to starting points, and (ii) it is a derivative-free method. The cons are that, in comparison to local methods, it requires higher computing time and power. Table 2.2 shows the rate of correct root location using a genetic algorithm. The improvements can be seen easily in the most problematic cases.

Table 2.2: Rate of Successful Identification, $T = 250$.
(Global Solution)

Distribution	VARMA Models					
	$(p=1,q=0)$ $(p_+=1)$	$(p=1,q=0)$ $(p_+=0)$	$(p=0,q=1)$ $(q_+=1)$	$(p=0,q=1)$ $(q_+=0)$	$(p=1,q=1)$ $(p_+=0,q_+=0)$	$(p=1,q=1)$ $(p_+=1,q_+=1)$
(χ^2_4,χ^2_5)	0.98	0.97	0.83	0.97	0.94	0.81
(t_4,t_5)	0.84	0.96	0.70	0.96	0.86	0.65
$(t_4,U(-2,2))$	0.75	0.94	0.62	0.97	0.90	0.58

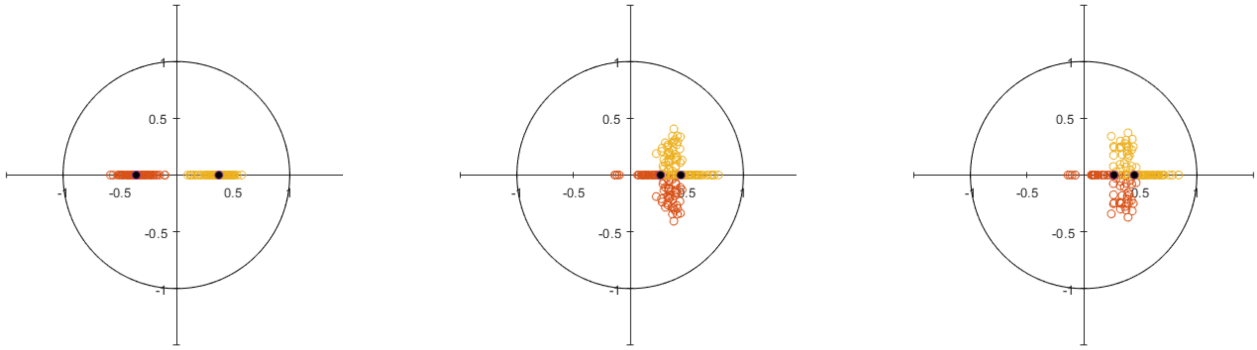
Optimization procedure was performed through global optimization. Genetic Algorithm was the selected optimizer. The initial population size is 400. The maximum time was set in 500 seconds.

In the case of a structural causal and invertible VARMA of order 1, the rate of successful location is around 94% when shocks are distributed as chi-square; it lowers to 89% when shocks are distributed as t-student. In both cases, the obtained rate is, at least, 3.5 times higher than the theoretical rate, which for this case is 25%. For a non-causal and non-invertible VARMA of order 1, the rate is 51% (2 times higher than the theoretical rate) when errors are distributed as chi-square, and 44% (1.76 times higher than the theoretical rate) when errors are distributed as t-student.

In the case of pure VAR with non-causal roots and pure VMA with non-invertible, the following graphs (Figures 2.1 and 2.2) shows how close the estimated roots are from the true ones. In both figures, we only plot the successful cases. We can notice that not only the order estimation is accurate using the global optimization procedure but also the specific location of the roots. Thus, identifying non-invertible roots is more complex than non-causal roots.

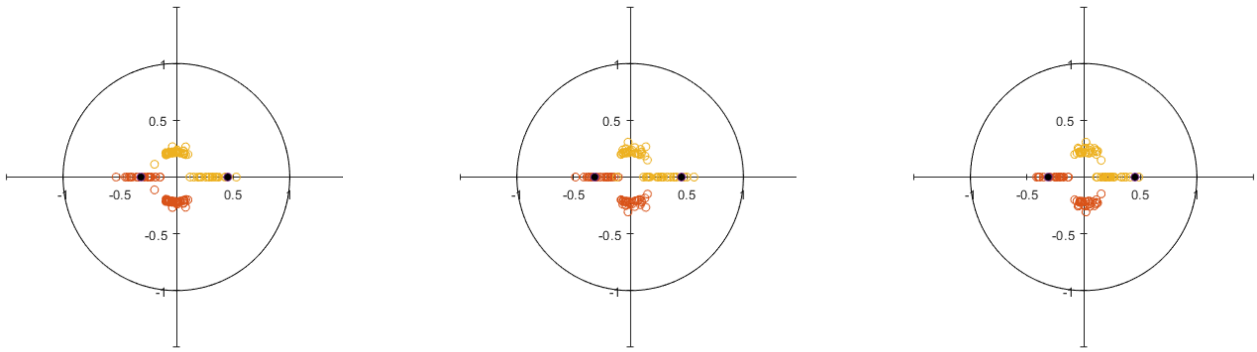
2.4.2. Static Model

To check if the method identifies the static component well, we simulate a simple static model, $p = q = 0$. In this case, we use $\boldsymbol{B} = \begin{bmatrix} 2.50 & -0.126 \\ 1.626 & 0.50 \end{bmatrix} = \begin{bmatrix} 1 & -0.25 \\ 0.65 & 1 \end{bmatrix} \begin{bmatrix} 2.5 & 0 \\ 0 & 0.5 \end{bmatrix}$.



(a) (b) (c)

Figure 2.1: **Pure non-causal VAR(1) Roots**. (a) $(\varepsilon_{t,1}, \varepsilon_{t,2}) \sim (\chi^2_4, \chi^2_5)$; (b) $(\varepsilon_{t,1}, \varepsilon_{t,2}) \sim (t_4, t_5)$; (c) $(\varepsilon_{t,1}, \varepsilon_{t,2}) \sim (t_4, U[-2,2])$. Black-filled dots represent true roots. Orange and yellow circles represent the estimated roots.



(a) (b) (c)

Figure 2.2: **Pure non-invertible VMA(1) Roots**. (a) $(\varepsilon_{t,1}, \varepsilon_{t,2}) \sim (\chi^2_4, \chi^2_5)$; (b) $(\varepsilon_{t,1}, \varepsilon_{t,2}) \sim (t_4, t_5)$; (c) $(\varepsilon_{t,1}, \varepsilon_{t,2}) \sim (t_4, U[-2,2])$. Black-filled dots represent true roots. Orange and yellow circles represent the estimated roots.

Figure 2.3 shows the mean, median, and confidence interval at 95% level and the estimated coefficient series. We report the coefficient in the diagonal. The simulation was done with 500 replications with $T = 250$. The selected distribution for innovations was (χ^2_3, t_4). As we can observe in Figure 2.3, the proposed method can identify the true contemporaneous effect quite well, on average.

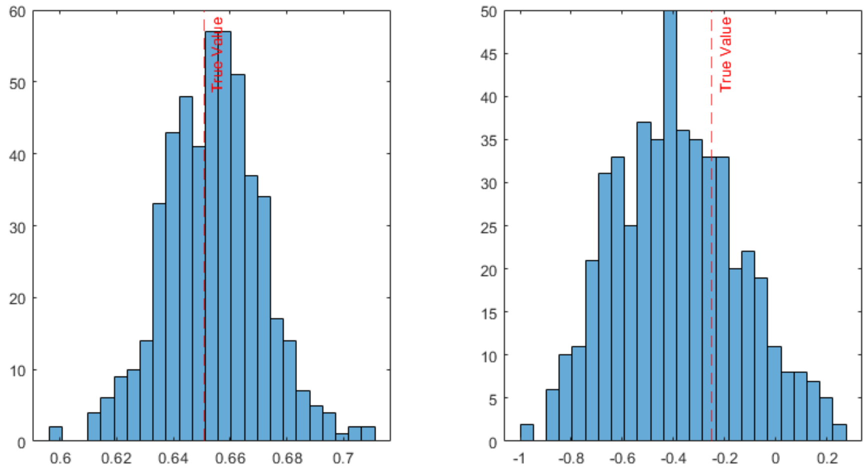


Figure 2.3: **Static VARMA**. (Left panel) Static response of $y_{1,t}$ to shock in $\varepsilon_{t,2}$; (Right panel) Static response of $y_{2,t}$ to shock in $\varepsilon_{t,1}$

2.4.3. Sensitivity Analysis

We explore how sensible the method is when the vector dimension is $d > 2$. We explore the case of $d = 3$. The computational cost may be significant for higher dimensions, but the accuracy would be maintained. Additionally, we explore distributions near Normal.

Higher Dimension Vector

An important drawback of structural VAR models is that the number of parameters grows with the vector dimension or the number of lags. For instance, if we pass from a d-dimensional SVAR(p) to $d+1$-dimensional SVAR(p), the number of parameters grows by $(p+1)(2d+1)$; in case we pass to a d-dimensional SVAR($p+1$), the number of parameters increases by d^2. This feature represents an issue because if the sample size, T, does not grow, then the accuracy of estimation decays. Structural VARMA models face the same problem.

In this case, we fix $T = 250$ and analyze the following two cases for $d = 3$: (i) Pure non-causal VAR(1); (ii) pure non-invertible VMA(1).

Table 2.3: Rate of Successful Identification.
(3-dimensional VARMA)

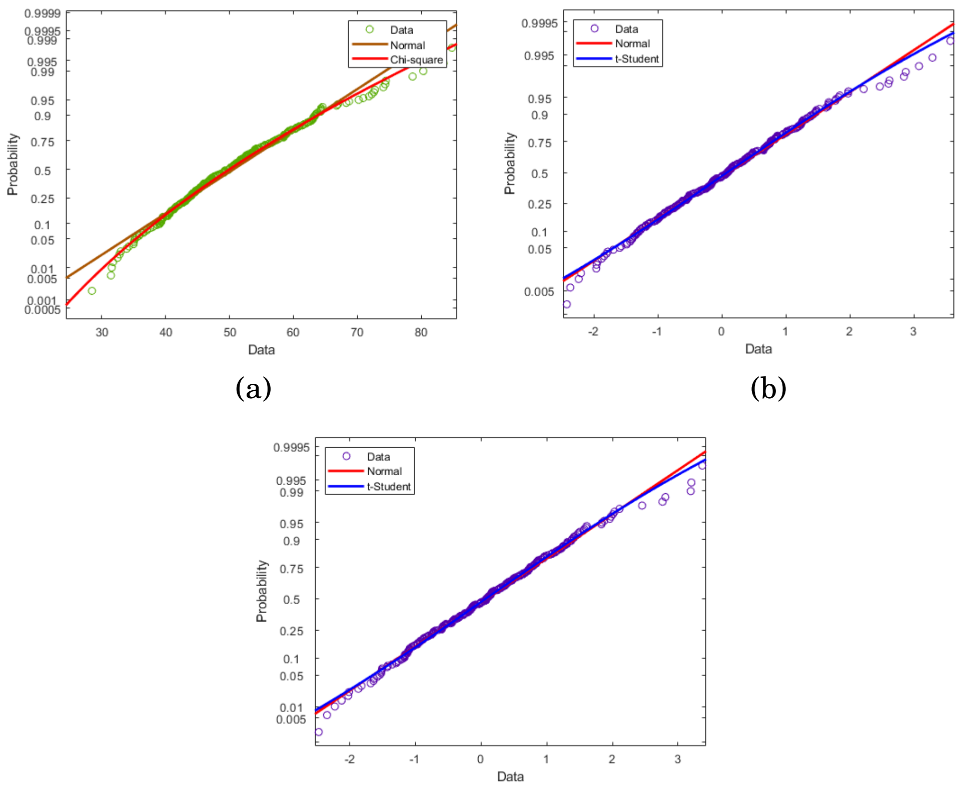
Distribution	VARMA Models	
	$(p = 1, q = 0)$ $(p_+ = 1)$	$(p = 0, q = 1)$ $(q_+ = 1)$
$(\chi^2_4, \chi^2_5, \chi^2_6)$	0.94	0.95
(t_4, t_5, t_6)	0.94	0.93

Optimization procedure was performed through global optimization. Genetic Algorithm was the selected optimizer. The initial population size is 400. The maximum time was set in 700 seconds.

Quasi Gaussian Distributions

We study how the results change in case of having an innovation with quasi-Gaussian behavior. Specifically, three scenarios are considered: (i) χ^2 with 50 degrees of freedom; (ii) t-student with 30 degrees of freedom; (iii) t-student with 100 degrees of freedom. To know how close to a Gaussian distribution are the three chosen cases, Figure 2.4 shows the probability plot for data generated with these distributions. Apart from some outliers, one can observe that most of the data are concentrated along the *Normal* line.

In total three cases, we combine these distributions by pairs and simulate two simple VARMA models. (i) pure non-causal VAR(1); (ii) pure non-invertible VMA(1). Table 2.4 shows the success rate of correct identification. It can be noticed that when we have a skewed shock as part of the innovation vector, the rate of successful identification of roots, although not as high as in the situations of Table 2.2, is significantly above the theoretical rate, which 50% for the models considered in the simulation. However, when both shocks are symmetric, the successful rate lowers, and it is around the theoretical 50% rate. This rate implies that, in practice, our methodology can be applied even though we have Gaussian innovations, and the method will provide the correct dynamic identification approximately half of the time. Additionally, this simulation was done with fixed T; we expect these rates to increase if T rises.



(a) (b) (c)

Figure 2.4: **Quasi-Gaussian Distributions**. (a) Data is generated from χ^2_{50}; (b) Data is generated from t_{30}; (c) Data is generated from t_{100}.

Table 2.4: Rate of Successful Identification.
(2-dimensional VARMA)

Distribution	VARMA Models			
	$(p=1, q=0)$ $(p_+=1)$	$(p=1, q=0)$ $(p_+=0)$	$(p=0, q=1)$ $(q_+=1)$	$(p=0, q=1)$ $(q_+=0)$
(χ^2_{50}, t_{30})	0.63	0.65	0.59	0.66
(χ^2_{50}, t_{100})	0.60	0.61	0.56	0.65
(t_{30}, t_{100})	0.49	0.48	0.47	0.60

Optimization procedure was performed through global optimization. Genetic Algorithm was the selected optimizer. The initial population size is 400. The maximum time was set in 700 seconds.

2.5. Empirical Application

We take the dataset employed in Blanchard and Quah (1989)[5] (hereafter BQ). The endogenous variables in this model are output growth, measured as the growth rate of Gross National Product, and the unemployment rate, i.e., $\boldsymbol{Y}_t = [\Delta\text{GNP}_t, \text{Unemp}_t]$. It is worth mentioning that output growth has been adjusted, considering the potential structural break in 1973. In Blanchard and Quah (1989) work, the selected structural model was an SVAR with 8 lags. However, since an SVARMA model is more parsimonious than an SVAR, we do not need to introduce many lags. In particular, we set the overall degrees for AR and MA polynomials at $p = 1$ and $q = 1$.

We fit two specifications, a restricted and an unrestricted structural model. The restricted case assumes that the roots of the AR polynomial lie outside the unit circle, i.e., the model has a causal representation but leaves the location of the roots of the MA polynomial to be determined from the data. The unrestricted SVARMA model does not impose causal roots in the AR polynomial, i.e., the location of roots for both polynomials is determined from the data.

In the first case, our proposed method selects an invertible MA polynomial. That is, the selected MA representation is a fundamental one. The estimated polynomials and the static components are:

$$\hat{\Phi}(L, \hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \boldsymbol{I}_2 - \begin{bmatrix} \underset{(0.0133)}{0.4273} & \underset{(0.0360)}{0.1352} \\ \underset{(0.0136)}{-0.3260} & \underset{(0.0430)}{1.0672} \end{bmatrix}, \quad \hat{\Theta}(L, \hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \boldsymbol{I}_2 + \begin{bmatrix} \underset{(0.0150)}{-0.2128} & \underset{(0.0930)}{-0.7015} \\ \underset{(0.0162)}{0.1935} & \underset{(0.0120)}{0.5339} \end{bmatrix}, \quad \hat{\boldsymbol{B}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \begin{bmatrix} \underset{0.0020}{0.5784} & \underset{0.0050}{0.3377} \\ \underset{0.0070}{-0.1598} & \underset{0.0030}{0.5243} \end{bmatrix}$$

From observing the estimated static component, $\hat{\boldsymbol{B}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}})$, we can observe that the effect of the first shock in the model has a positive effect over the growth rate of GNP, about 0.58

[5]The dataset was obtained from this link

percentage points, while a negative impact over unemployment rate about 0.16 percentage points.

Figure 2.5 shows the auto-correlation functions of the level and squares of model residuals. We observe that the restricted SVARMA(1,1) generates white-noise residuals (panel (a) in Figure 2.5), although the estimated residuals exhibit some higher order dependence, as it can be noted in the autocorrelation of squared residuals (panel (b) in Figure 2.5).

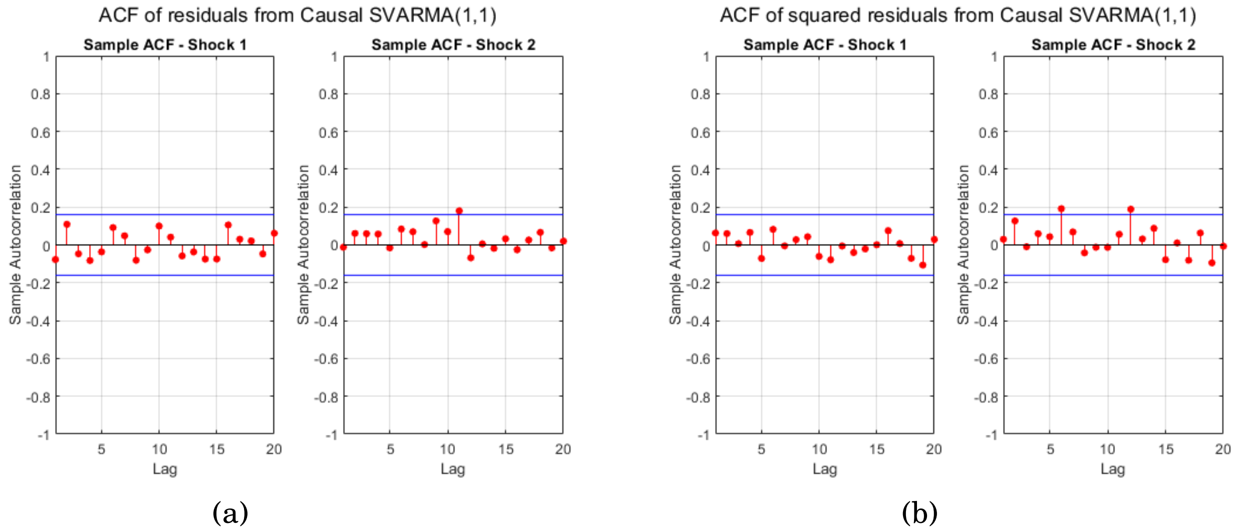
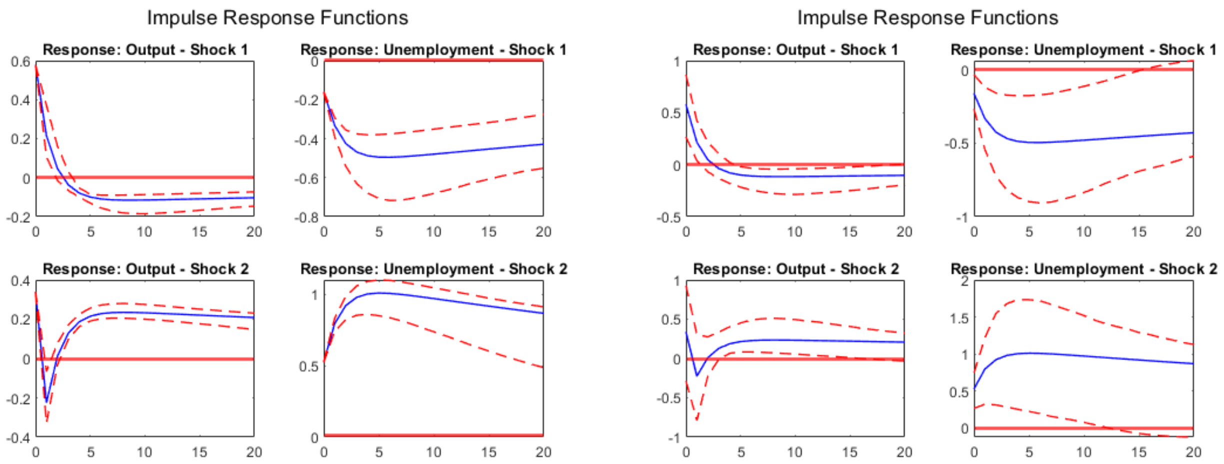


Figure 2.5: Sample auto-correlation function of estimated residuals. (a) Residuals; (b) Squared residuals.

Based on the estimated AR and MA polynomials and the static component, we construct the causal effects of both shocks over the endogenous variables in the structural model. From Figure 2.6, we observe that the shock labeled as *shock* 2 has a long-run effect on unemployment. In contrast, the effect of *shock* 1 has a transitory, although very persistent, effect over both endogenous variables. The former response is similar to the one identified in Blanchard and Quah (1989). Nonetheless, the persistent behavior of *shock* 1 differs from that identified by Blanchard and Quah (1989). It is worth mentioning that these effects are estimated without requiring any external identification restriction, such as the one imposed in the original work of BQ.



(a) (b)

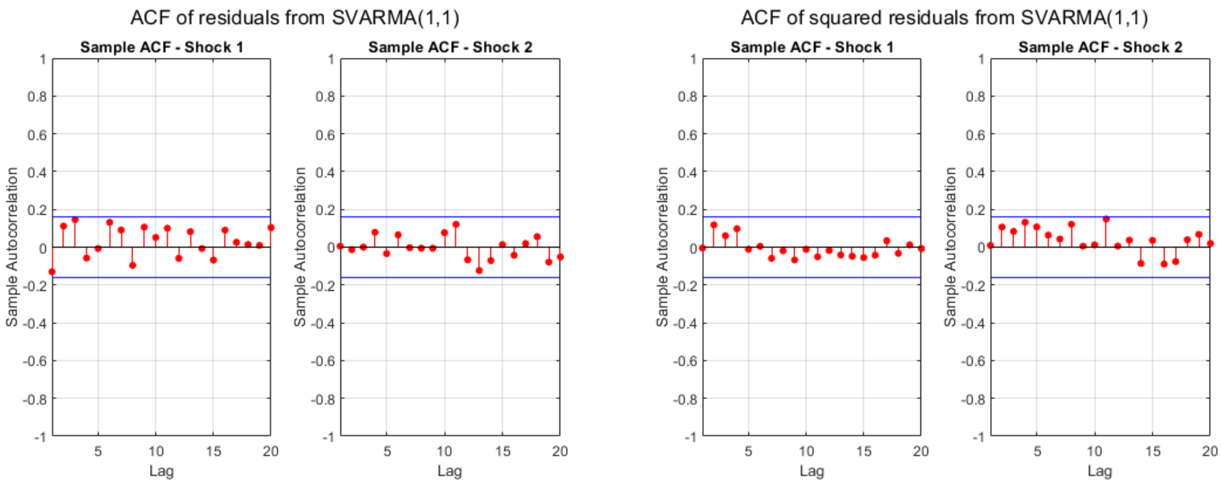
Figure 2.6: Estimated IRFs. (a) Analytical S.E.; (b) Bootstrap S.E.

When the restriction of having a causal representation in the AR polynomial is removed, our method estimates a structural VARMA(1,1) with AR roots inside the unit circle but MA roots outside it. That is, the estimated model is a non-causal but invertible SVARMA model. The estimated polynomials and the static component are listed below.

$$\hat{\boldsymbol{\Phi}}(L,\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \boldsymbol{I}_2 - \begin{bmatrix} \underset{(0.0383)}{2.5285} & \underset{(0.0348)}{0.5631} \\ \underset{(0.0091)}{-1.3390} & \underset{(0.0129)}{0.7919} \end{bmatrix}, \quad \hat{\boldsymbol{\Theta}}(L,\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) \ = \boldsymbol{I}_2 + \begin{bmatrix} \underset{(0.0377)}{-2.4498} & \underset{(0.0442)}{-5.9769} \\ \underset{0.0090}{1.2229} & \underset{(0.0430)}{2.9127} \end{bmatrix}, \quad \hat{\boldsymbol{B}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \begin{bmatrix} \underset{0.0018}{0.8108} & \underset{(0.0087)}{0.5061} \\ \underset{(0.0041)}{-0.3522} & \underset{(0.0057)}{0.7745} \end{bmatrix}$$

The behavior of model residuals is shown in Figure 2.7. It can be noticed (in panel (a) in Figure 2.7) that the auto-correlation of residuals shows that they behave like a white noise process. Moreover, regarding the higher order dependence, notice that squared residuals are also uncorrelated (panel (b) in Figure 2.7). This result is not a formal testing result, but it gives us intuition that model residuals are, at least, not dependent in the second order.

This result is new in the literature. Several works have estimated a non-fundamental model for the BQ dataset, but they identified a causal, non-invertible SVARMA model. The non-causal behavior of the AR polynomial makes intricate to estimate the causal effects of structural shocks because the infinite moving-average representation of this non-fundamental model is $\boldsymbol{Y}_t = \boldsymbol{\Phi}^{-1}(L)\boldsymbol{\Theta}(L)\boldsymbol{B}\boldsymbol{\varepsilon}_t = \left(\sum_{j=1}^{\infty}(-1)^j\boldsymbol{\Phi}_1^j L^{-j}\right)(\boldsymbol{I}+\boldsymbol{\Theta}_1 L)\boldsymbol{B}\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty}\boldsymbol{\delta}_j\boldsymbol{\varepsilon}_{t+j}$, implying that $\mathbb{E}[Y_{T+h}|Y_1,\ldots,Y_{T-1},\boldsymbol{\varepsilon}_T]$ is an unknown, non-linear function. Estimating $\mathbb{E}[Y_{T+h}|Y_1,\ldots,Y_{T-1},\boldsymbol{\varepsilon}_T]$ for a non-causal model remains to be an open question in the literature.



(a) (b)

Figure 2.7: Sample auto-correlation function of estimated residuals. (a) Residuals; (b) Squared residuals.

2.6. Concluding Remarks

Structural VAR models have become an essential tool in empirical macroeconometrics because they offer a simple framework for identifying and estimating the causal effects of economic shocks over macroeconomic or financial outcomes. The focus has been on identifying the contemporaneous effects matrix, for which various identification strategies have been proposed. Moreover, most of these approaches are based on using external sources of information for imposing identification restrictions.

In this chapter, we focus on SVARMA models, which exhibit several advantages over SVAR ones. First, VARMA models are more parsimonious than VAR representations. Besides, it is a more accurate representation of theoretical macroeconomic models. A growing literature shows that a fundamental SVAR model is quite limiting for addressing important characteristics of macroeconomic models. For instance, depending on the calibration, the solution to a DSGE model is better represented as a non-invertible SVARMA rather than a causal SVAR model. When fundamental representation is imposed, the estimates of the causal effects of economic shocks may be biased.

This chapter aims to design an estimation method that allows the researcher to identify the structural parameters of an SVARMA model without imposing the location of the roots of lag polynomials. Besides, the proposed method, unlike the current methods in the literature, is robust to whichever the joint density function of structural shocks is; and does not require the existence of a large number of higher-order moments. Our proposal exploits a general pairwise dependence measure constructed from the joint characteristic function of model errors.

Furthermore, our framework allows us to adapt the method for weakening the i.i.d. assumption. We consider a very general dependence structure of shocks, such as to assume they are m.d.s. This general characterization covers several relevant features of macroeconomic and financial data. Identification and estimation explicitly require the existence of the data's second-order moments and a broader characterization of non-Gaussian behavior. Besides, the asymptotic distribution only requires the existence of fifth-order moments.

Simulation results show that our method satisfactorily identifies the location of roots of AR and MA polynomials in an SVARMA model. However, the correct identification rate decreases when a local optimization procedure is employed, contrary to the findings when a global optimization method is used. This feature occurs due to the highly non-linear nature of our optimization problem, which makes it quite sensitive to initial point conditions. Additionally, the method works fine with vector dimensions higher than 2, although the results have yet to be tested in higher dimensional situations. The method should work fine, although, as pointed out in the SVAR literature, it entails a high computational cost for estimating the structural parameters because the number of parameters increases exponentially.

3. CHAPTER II: ROBUST ESTIMATION OF THE NON-GAUSSIAN DIMENSION IN LINEAR STRUCTURAL SYSTEMS

Statistical identification of possibly non-fundamental structural VARMA models requires that shocks satisfy: (i) to be an i.i.d process, (ii) to be mutually independent across components, and (iii) each shock must be non-Gaussian distributed. Therefore, provided the first two requirements, it is crucial to empirically evaluate the third requirement without requiring estimation of the structural model. We address this problem by relating the non-Gaussian dimension of structural errors to the rank of a matrix built from the higher-order spectrum of reduced-form errors. Our proposal is robust to the location of the roots of lag polynomials, generalizing the current procedures designed for the restricted case of a causal structural VAR model. Montecarlo's exercises show that our method satisfactorily estimates the number of non-Gaussian components.

3.1. Introduction

In the mainstream macroeconometrics literature of structural vector autoregressive (hereafter SVAR) models, Gaussian behavior of shocks is assumed either explicitly or implicitly, leading to exploiting only second-order information for identification and estimation purposes. This feature provokes the well-documented problem of the observational equivalence of any orthonormal rotation of structural disturbances. Several strategies have been proposed for overcoming such identification issues, resorting to different sources of external information. Some strategies impose zero, linear, or non-linear identification restrictions over the matrix of contemporaneous effects (e.g., Bernanke and Mihov (1998); Blanchard and Perotti (2002); Blanchard and Quah (1989); Gali (1992); Sims (1980), among others). Alternative strategies are: (i) establish agnostic restrictions over the impulse response functions, i.e., sign-restrictions (e.g., Arias et al. (2018); Canova and Pappa (2007); Uhlig (2005)); (ii) estimate a structural shock using extensive institutional information, known as *narrative identification approach* (C. D. Romer and D. H. Romer (1989); C. D. Romer and D. H. Romer (2010)); or (iii) use proxy measures as instrumental variables for some latent structural shock (e.g., Mertens and Ravn (2013); Mertens and Ravn (2014); Stock and Watson (2018)).

Despite the usefulness of these identification schemes, they suffer from some drawbacks. For instance, when external identification restrictions are imposed (zero, linear, non-linear, or sign restrictions), these cannot be empirically evaluated unless the model is overidentified. Besides, those restrictions may be pretty arguable in some cases. On the other hand, when instrumental or proxy variables are used as identification devices, only a subset of structural shocks can be identified and, as documented by Herwartz and Lütkepohl (2014), the identification power of this strategy relies on how plausible the relevance and exogeneity conditions.

Alternatively, the statistical identification strategy (hereafter SIS) has appeared on the scene as part of the *data-driven* identification approach and has gained a spotlight in recent years. The works of Guay (2021); Lanne, Meitz, et al. (2017); Maxand (2020) identify and estimate the matrix of contemporaneous effects in a causal SVAR model without imposing any external identification restriction because according to the result in Comon (1994) such matrix is identified (up to signed-permutation) whenever the three following identification assumptions hold: (i) structural shocks are an i.i.d. process; (ii) the structural errors are mutually independent across their components; and (iii) at most one structural disturbance in the model is Gaussian.

For more general stationary, linear processes, the third requirement in Comon (1994) is insufficient for assuring identification. Chan et al. (2006) show that identification of a general stationary, linear process needs -provided that shocks are an i.i.d. process with mutual independent components and some other regularity conditions- that each structural error in the system to be non-Gaussian distributed, with non-zero third order cumulant and finite fourth order moment. This result permits dealing with the identification of parameters in a possibly non-fundamental structural vector autoregressive moving-average (hereafter SVARMA) model, i.e., without imposing causality or invertibility (see Gouriéroux et al. (2020); Lanne, Lütkepohl, and Maciejowska (2010) for applications with likelihoods methods and Velasco (2022b) for an equivalent identification result through cumulant conditions). Notice that in this general context or the restricted case of a fundamental model, even though i.i.d. and mutually independent components conditions are met, non-Gaussian behavior of structural shocks is decisive for the feasibility of the SIS.

The main advantage of the SIS is that it makes any economically motivated restriction amenable to being empirically assessed. Thus, any identification constraint in the standard macroeconometrics literature for causal SVAR models may be tested. In order to perform these empirical exercises, it is central to determine the non-Gaussian dimension of structural shocks before applying the SIS. Otherwise, this strategy would be unfeasible and need external identification restrictions. Therefore, our ultimate goal in this paper is to propose a method for determining empirically this identification constraint. Assessing identification restrictions is only sometimes feasible. When models are overidentified, such a task is achievable. For just-identified models, economic arguments often support the assessment of identification restrictions. We do not neglect the relevance of such assessments; yet, in our context, they may be insufficient and vague for justifying the size of the non-Gaussian dimension. Fortunately, we manage to design a method for evaluating the identification requirement about the magnitude of non-Gaussian dimension before blindly applying the SIS.

There is a voluminous literature that copes with the problem of determining whether a random variable (or vector) is Gaussian distributed or not. Some procedures are based on the empirical distribution of data (Kolmogorov (1933); Massey Jr (1951); S. S. Shapiro and Wilk (1965); Smirnov (1948)); other approaches employ the characteristic function (Epps and Pulley (1983); P. Hall and Welsh (1983)); and, others exploit third and fourth centered moments (Bera and Jarque (1982); d'Agostino (1971); Lobato and Velasco (2004)). These approaches only assess the null hypothesis of joint Gaussianity, which, in case of no rejection, would make it infeasible to apply the statistical identification strategy of an SVARMA model. However, if joint Gaussianity were rejected, this would support only the existence of at least one non-Gaussian component in the system, which is only sufficient for employing the SIS if we were dealing with a fundamental SVARMA model with only two endogenous variables.

In contrast, the literature for estimating the non-Gaussian dimension in a random vector is limited. The proposals relate this dimension to the number of non-zero eigenvalues or the rank of some matrix constructed from third or fourth-order moments or cumulants. Nordhausen et al. (2017) proposes both an asymptotic and bootstrap sequential tests to estimate the Gaussian dimension of an unobservable random vector by analyzing the number of non-zero eigenvalues of the scatter matrix[6] constructed using information from an observable vector, an affine transformation of the unobservable vector of interest.[7] They show that when the random vector contains r Gaussian components, the scatter matrix has exactly r eigenvalues equal to 0.[8] Maxand (2020) adapts this strategy for determining the non-Gaussian dimension in the vector of structural shocks for a causal SVAR model. Her idea is to apply Nordhausen et al., 2017 approach directly to estimated structural shocks because the non-Gaussian block is identified even though there is more than one Gaussian shock. Alternatively, Guay (2021) relates the non-Gaussian dimension in a causal SVAR model to the rank of a rectangular array constructed employing third and fourth order cumulants of RF errors.[9]

[6]A scatter matrix contains fourth-order moments of a vector.

[7]Let $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ be an observable and unobservable d-dimensional random vectors, respectively. An affine transformation is define as $\mathbf{x} = \boldsymbol{b}_0 + \boldsymbol{B}\boldsymbol{\varepsilon}$, with $\boldsymbol{b}_0$ possibly non-zero vector and $\boldsymbol{B}$ time-invariant, full rank, square matrix.

[8]To be completely precise, there exists r eigenvalues equal to $d+2$, but Nordhausen et al. (2017) employ a normalized version of the scatter matrix.

[9]When third and fourth order moments exist, the third order cumulant is the same as the asymmetry coefficient, while the fourth order cumulant represents the excess of kurtosis. Gaussian distribution is the only one with all cumulants of third or higher order equal to zero. See Marcinkiewicz (1939) for technical details.

The current approaches described above are helpful but restrictive for several reasons. First, many structural macroeconomic models can be represented and fitted more accurately by an SVARMA model rather than an SVAR one. Second, as discussed in Gouriéroux et al. (2020); Velasco (2022b) and the references therein, the fundamentalness assumption is a restriction to avoid the dynamic identification problem -i.e., any stationary linear process has a fundamental and non-fundamental representation which are observationally equivalent when only second-order information is exploited. Besides, as surveyed by Alessi et al. (2011), non-fundamentalness may be consistent with many macroeconomic models and data features. Thus, it is paramount to design a method for determining the non-Gaussian dimension in an SVARMA model that does not require the root location of the dynamic polynomials as input. Notwithstanding, this task is quite challenging because when the location of roots is not imposed, the reduced-form (hereafter RF) errors from a fundamental VARMA approximation -call it RF VARMA model- might not be a simple static rotation of structural shocks, but a dynamic filter of those. This characteristic makes invalid Guay (2021) or Maxand (2020) approaches since the rank or the non-zero eigenvalues of the matrix constructed from contemporaneous fourth-order cumulants or moments of RF errors are not related to the non-Gaussian dimension unless unrealistic assumptions are made. This paper aims to fill this gap in the literature. We exploit third or fourth-order cumulant spectrums of RF errors for constructing a matrix whose rank unveils the non-Gaussian dimension in the vector of structural shocks. Our work can be seen as the extension of Lobato and Velasco (2004) approach to a multivariate context and the robustification of Guay (2021) work against possibly non-fundamentalness.

For estimating the rank of a matrix, there exist several strategies in the literature. We follow Kleibergen and Paap (2006) (hereafter KP) approach. The KP statistic is built from the singular value decomposition of the matrix of interest. The asymptotic distribution of the statistic is a standard chi-square whose degrees of freedom change depending on the null hypothesis. Unlike KP work, our context changes the asymptotic distribution of the test statistic for some particular rank values, specifically under joint Gaussianity. For other null hypotheses, the asymptotic distribution is chi-square, but the degrees of freedom are generally unknown. Thus, we propose a bootstrap strategy. This path implies another challenge: to impose the null hypothesis in the resampled data. Montecarlo exercises were performed to analyze the size and power of the bootstrap test; these show that our strategy estimates satisfactorily the non-Gaussian dimension. We apply our procedure to two well-known macroeconomic datasets. Our proposal detects a skewed structural shock in the system described by Blanchard and Quah (1989) (hereafter BQ). Using Blanchard and Perotti (2002) (hereafter BP), our approach detects at least two skewed and non-mesokurtic structural errors, unlike Guay (2021) whose procedure only could detect one non-mesokurtic structural shock. In the case of imposing roots outside the unit circle, this latter result implies that the SIS can be applied to BQ or BP datasets.

The remainder of this chapter is structured as follows: Section 2 describes our time series model and states the main assumptions. Section 3 shows the connection between the number of non-Gaussian structural shocks and the rank of an array constructed from third and fourth-order spectrums of reduced-form errors. In Section 4, the test procedure and estimation are detailed. Section 5 shows simulation results and the empirical application. We conclude in Section 6.

3.2. Model and Assumptions

Let $\boldsymbol{y}_t$ be a d-dimensional stationary, zero-mean, multivariate process generated from an SVARMA model described by

$$\boldsymbol{\Phi}(L)\boldsymbol{y}_t = \boldsymbol{\Theta}(L)\boldsymbol{B}\boldsymbol{\varepsilon}_t, \tag{3.1}$$

where $\boldsymbol{\Phi}(L) = \boldsymbol{I}_d - \sum_{j=1}^{p}\boldsymbol{\Phi}_j L^j$, $\boldsymbol{\Theta}(L) = \boldsymbol{I}_d + \sum_{j=1}^{q}\boldsymbol{\Theta}_j L^j$. L is the lag (back-shift) operator, i.e. $L^j\boldsymbol{y}_t = \boldsymbol{y}_{t-j}$. $\boldsymbol{\varepsilon}_t$ is a d-dimensional vector of structural shocks. $\boldsymbol{B}$ is a time-invariant, full-rank, squared matrix of contemporaneous effect. If $q = 0$, model in (3.1) represents a SVAR model. Representation in equation (3.1) is standard in the literature, see Gouriéroux et al. (2020); Mainassara and Francq (2011); Velasco (2022b) for equivalent expressions.

The autoregressive and moving-average polynomials, $\boldsymbol{\Phi}(z)$ and $\boldsymbol{\Theta}(z)$, satisfy

$$\det(\boldsymbol{\Phi}(z))\det(\boldsymbol{\Theta}(z)) \neq 0, \quad \forall\, z \in \mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}. \tag{3.2}$$

The condition in (3.2) only rules out unit roots in both auto-regressive (AR) or moving-average (MA) polynomials, otherwise it is not possible to find a stationary solution of equation (3.1). Thus, we are not imposing an exact location of the polynomial roots[10], meaning that the structural model in (3.1) permits for possibly non-fundamental representations.

The vector of structural shocks, $\boldsymbol{\varepsilon}_t$, satisfies:

Assumption 3.1.

[10]Compares this to alternative works such as Guay (2021); Maxand (2020), where they assume that condition in (3.2) rules out roots on and inside the complex unit circle.

(i) *It is an independent, identically distributed (i.i.d.) process;*

(ii) *the components of $\boldsymbol{\varepsilon}_t$ are mutually independent;*

(iii) *there exists $0 \le d_{ng} \le d$ non-Gaussian components in the vector of structural shocks, $\boldsymbol{\varepsilon}_t$;*

(iv) $\mathbb{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$, $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{I}_d$ *and* $\mathbb{E}\left(\|\boldsymbol{\varepsilon}_t\|^8\right) < \infty$ *for any* $t \in \mathbb{Z}$.

Assumption (3.1.i) is common in the literature[11]. It simplifies the analysis however, it is a quite restrictive requisite because it implies structural errors do not exhibit any linear or nonlinear dependence[12]. Assumption (3.1.ii) is imposed in the work of Comon (1994), the basis of *Independent Component Analysis* (hereafter ICA). This requirement is not as restrictive as it may seem, e.g., in macroeconomics, it is standard to assume that productivity shock is independent of monetary or fiscal policy shocks. Besides, Assumption (3.1.iv) imposes structural errors that are centered and standardized with finite moments up to order eight. The high number of required finite moments is needed to find the asymptotic distribution of the third and fourth-order cumulant spectrum. Finally, condition Assumption (3.1.iii) does not impose a particular number of non-Gaussian shocks.

For clearness, we group true values of structural parameters of the model in (3.1) inside a K-dimensional column vector, denoted by $\boldsymbol{\vartheta}_0$. Besides, we split this vector into two blocks: one related to dynamic behavior ($\boldsymbol{\vartheta}_{0,1}$) and another that governs the contemporaneous part ($\boldsymbol{\vartheta}_{0,2}$). Thus, we can write the elements in equation (3.1) as follows: $\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}(L, \boldsymbol{\vartheta}_{0,1})$, $\boldsymbol{\Theta}(L) = \boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_{0,1})$ and $\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})$. Therefore, the true structural model can be written as $\boldsymbol{\Phi}(L, \boldsymbol{\vartheta}_{0,1})\boldsymbol{y}_t = \boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_{0,1})\boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})\boldsymbol{\varepsilon}_t$. The moving-average representation of the observable vector is $\boldsymbol{y}_t = \boldsymbol{\Psi}(L, \boldsymbol{\vartheta}_{0,1})\boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\Psi}(L, \boldsymbol{\vartheta}_{0,1}) := \boldsymbol{\Phi}^{-1}(L, \boldsymbol{\vartheta}_{0,1})\boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_{0,1})$ is a possibly non-causal filter.[13]

Structural parameters $\boldsymbol{\vartheta}_0$ are not identified under Assumption (3.1). In particular, e.g., if AR and MA polynomials roots are known to be outside the unit circle, the block $\boldsymbol{\vartheta}_{0,2}$ is unidentified because $\boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})$ is observationally equivalent to any orthogonal rotation of it, unless, as showed by Comon (1994), under Assumption (3.1.ii) $d_{ng} \ge d - 1$. On the other hand, if only unit roots are discarded from AR and MA polynomials, both blocks $\boldsymbol{\vartheta}_{0,1}$ and $\boldsymbol{\vartheta}_{0,2}$ are unidentified. The lack of identification of $\boldsymbol{\vartheta}_{0,1}$ comes from the fact that

[11] See for instance Gouriéroux et al. (2020); Guay (2021); Lanne, Lütkepohl, and Maciejowska (2010); Lanne, Meitz, et al. (2017); Velasco (2022b).

[12] This condition rules out some characteristics that may be relevant in the empirical analysis of macroeconomic outcomes such as stochastic volatility or conditional heteroskedasticity.

[13] A non-causal filter is of the form $\boldsymbol{\Psi}(L) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Psi}_j L^j$ with non-zero $\boldsymbol{\Psi}_j$ for some $j < 0$.

there exist $\tilde{\boldsymbol{\vartheta}}$ such that $\boldsymbol{y}_t = \boldsymbol{\Psi}(L, \tilde{\boldsymbol{\vartheta}})\boldsymbol{e}_t$ with $\boldsymbol{e}_t$ a white noise process. In such general setup, identification requires Assumption (3.1) to be $d_{ng} = d$ (see Chan et al. (2006); Gouriéroux et al. (2020); Velasco (2022b)).

Nevertheless, the identification of $\boldsymbol{\vartheta}_0$ is not the aim of this work but to empirically verify the key identification restriction about the non-Gaussian dimension of the structural shocks vector. According to Wold's Decomposition Theorem (hereafter WDT), the stationary process $\boldsymbol{y}_t$ can be represented as a square summable, infinite, causal moving average of serially uncorrelated errors (see Anderson (2011) for more details), i.e., $\boldsymbol{y}_t = \tilde{\boldsymbol{\Psi}}(L)\boldsymbol{u}_t$ with $\tilde{\boldsymbol{\Psi}}(L) = \sum_{j=0}^{\infty} \tilde{\boldsymbol{\Psi}}_j L^j$ such that $\sum_{j=0}^{\infty} \left\| \tilde{\boldsymbol{\Psi}}_j \right\|^2 < \infty$, $\tilde{\boldsymbol{\Psi}}_0 = \boldsymbol{I}_d$ and $\boldsymbol{u}_t$ is a white noise process. Using our representation for a structural VARMA model, we could write the WDT causal filter as $\tilde{\boldsymbol{\Psi}}(L) = \tilde{\boldsymbol{\Phi}}^{-1}(L)\tilde{\boldsymbol{\Theta}}(L)$ with $\det\left(\tilde{\boldsymbol{\Phi}}(z)\right)\det\left(\tilde{\boldsymbol{\Theta}}(z)\right) \neq 0$ for all $|z| \leq 1$. Furthermore, we can write $\tilde{\boldsymbol{\Phi}}(L) = \boldsymbol{\Phi}(L, \boldsymbol{\vartheta}_f)$, $\tilde{\boldsymbol{\Theta}}(L) = \boldsymbol{\Theta}(L, \boldsymbol{\vartheta}_f)$ and $\tilde{\boldsymbol{\Psi}}(L) = \boldsymbol{\Psi}(L, \boldsymbol{\vartheta}_f)$, where $\boldsymbol{\vartheta}_f$ denotes the parameters vector that generates the causal filter of the WDT of $\boldsymbol{y}_t$. We call this the fundamental representation of $\boldsymbol{y}_t$, hence the subscript "f".

The serially uncorrelated RF errors, $\boldsymbol{u}_t$, are defined as

$$\boldsymbol{u}_t := \boldsymbol{\Psi}^{-1}(L, \boldsymbol{\vartheta}_f)\boldsymbol{y}_t = \boldsymbol{\Psi}^{-1}(L, \boldsymbol{\vartheta}_f)\boldsymbol{\Psi}(L, \boldsymbol{\vartheta}_{0,1})\boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})\boldsymbol{\varepsilon}_t = \boldsymbol{\delta}(L, \boldsymbol{\vartheta}_f, \boldsymbol{\vartheta}_0)\boldsymbol{\varepsilon}_t. \tag{3.3}$$

If the structural model in equation (3.1) were fundamental, then $\boldsymbol{\delta}(L, \boldsymbol{\vartheta}_f, \boldsymbol{\vartheta}_0) = \boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})$, implying that RF errors are a linear transformation of unobserved structural shocks. In general, when only unit roots are discarded from AR or MA polynomials in (3.1), the filter $\boldsymbol{\delta}(L, \boldsymbol{\vartheta}_f, \boldsymbol{\vartheta}_0)$ is non-causal. Hence, RF errors are not a static rotation of structural shocks but a dynamic combination of rotations of the structural errors. For simplicity, we write $\boldsymbol{\delta}(L, \boldsymbol{\vartheta}_f, \boldsymbol{\vartheta}_0) = \boldsymbol{\delta}(L, \boldsymbol{\vartheta}_f)$. Besides, serial no correlation of RF errors, $\boldsymbol{u}_t$, implies

$$\boldsymbol{\delta}(e^{i\lambda}, \boldsymbol{\vartheta}_f)\boldsymbol{\delta}^*(e^{i\lambda}, \boldsymbol{\vartheta}_f) = \boldsymbol{\Omega} \tag{3.4}$$

where $\boldsymbol{\Omega}$ is positive definite, symmetric constant matrix and $\boldsymbol{\delta}^*(e^{i\lambda}, \boldsymbol{\vartheta}_f)$ denotes the transposed, conjugate matrix polynomial of $\boldsymbol{\delta}(e^{i\lambda}, \boldsymbol{\vartheta}_f)$. Baggio and Ferrante (2018); Velasco (2022b) and the references therein denominate filters satisfying equation (3.4) as *all-pass filter*.[14]

[14]To be completely precise, an *all-pass* filter requires that $\boldsymbol{\Omega} = \boldsymbol{I}_d$. However, this is not a problem in our analysis because $\boldsymbol{\Omega}$ is a constant matrix.

3.3. Higher-order Spectrum and non-Gaussian Dimension

Equation (3.3) shows that the RF errors capture information of unobserved structural shocks. Thus, the non-Gaussian dimension in $\boldsymbol{\varepsilon}_t$ may be addressed by analyzing $\boldsymbol{u}_t$. Like Jarque and Bera (1987); Lobato and Velasco (2004), who interpret the non-Gaussian behavior as exhibiting asymmetric or non-mesokurtic nature, we focus on the third and fourth-order spectrum cumulant of RF errors to capture the asymmetry or the excess of kurtosis in structural shocks.

3.3.1. Cumulants of Random Vector

The characteristic function of structural shocks is $\phi_{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}) := \mathbb{E}(\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}))$ with $\boldsymbol{\tau} = (\tau_1, \dots, \tau_d)'$ a d-dimensional vector of constants. The cumulant generating function is $\boldsymbol{\kappa}_{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}) := \log\left(\phi_{\boldsymbol{\varepsilon}}(\boldsymbol{\tau})\right)$ and the cumulant for a k-tuple $(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_k,t})$ is

$$\mathbb{C}\text{um}\left(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_k,t}\right) = \frac{\partial^k}{(\partial\tau_1)^{k_1} \dots (\partial\tau_d)^{k_d}} \boldsymbol{\kappa}_{\boldsymbol{\varepsilon}}(\boldsymbol{\tau})\Big|_{\boldsymbol{\tau}=\mathbf{0}}, \tag{3.5}$$

where $\sum_{j=1}^d k_j = k \geq 1$ and $k_j \geq 0$, $j_m \in \{1, \dots, d\}$ for $m = 1, \dots, k$. When moments exist, Jammalamadaka et al. (2006) shows that the expression in (3.5) is equal to

$$\mathbb{C}\text{um}\left(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_k,t}\right) = \sum_{\mathfrak{p}} (|\mathfrak{p}| - 1)!(-1)^{|\mathfrak{p}|-1} \prod_{P \in \mathfrak{p}} \mathbb{E}\left(\prod_{j_m \in P} \varepsilon_{j_m,t}\right), \tag{3.6}$$

where $\mathfrak{p}$ is a partition of $\{j_1, \dots, j_k\}$, $|\mathfrak{p}|$ is the number of parts in partition $\mathfrak{p}$, and P a block of partition $\mathfrak{p}$. For instance, if $k = 2$ and we choose the duplet $(\varepsilon_{j_1,t}, \varepsilon_{j_2,t})$, $\mathbb{C}\text{um}(\varepsilon_{1,t}, \varepsilon_{2,t}) = \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}] = \mathbb{C}\text{ov}(\varepsilon_{j_1,t}, \varepsilon_{j_2,t})$. When $k = 3$ and the chosen triplet is $(\varepsilon_{j_1,t}, \varepsilon_{j_2,t}, \varepsilon_{j_3,t})$, then $\mathbb{C}\text{um}(\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}) = \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}\varepsilon_{j_3,t}]$. With $k = 4$ and the chosen 4-tuple, $(\varepsilon_{j_1,t}, \varepsilon_{j_2,t}, \varepsilon_{j_3,t}, \varepsilon_{j_4,t})$, we get $\mathbb{C}\text{um}\left(\varepsilon_{j_1,t}, \varepsilon_{j_2,t}, \varepsilon_{j_3,t}, \varepsilon_{j_4,t}\right) = \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}\varepsilon_{j_3,t}\varepsilon_{j_4,t}] - \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}]\mathbb{E}[\varepsilon_{j_3,t}\varepsilon_{j_4,t}] - \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_3,t}]\mathbb{E}[\varepsilon_{j_2,t}\varepsilon_{j_4,t}] - \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_4,t}]\mathbb{E}[\varepsilon_{j_2,t}\varepsilon_{j_3,t}]$. The m-th marginal cumulant of order k is $\kappa^{\varepsilon}_{k,m} = \mathbb{C}\text{um}\left(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_k,t}\right)$ with $j_1 = j_2 = \dots = j_k = m$.

Collecting all the cumulants for any k-tuple in a single-column vector, we have

$$\boldsymbol{\kappa}_k(\boldsymbol{\varepsilon}) = \left[\mathbb{C}\text{um}(\varepsilon_{1,t}, \dots, \varepsilon_{k,t})\right]_{(j_1,\dots,j_k) \in \boldsymbol{\sigma}_k(\{1,\dots,d\})} \tag{3.7}$$

where $\boldsymbol{\sigma}_k(\{1, \dots, d\})$ is the set of all permutations of length k formed with numbers in the set $\{1, \dots, d\}$. For instance, if $d = 2$ and $k = 2$, $\boldsymbol{\sigma}_2(\{1, \dots, 2\}) = \{\{1,1\}; \{1,2\}; \{2,1\}; \{2,2\}\}$. From this, it clear that dimension of $\boldsymbol{\kappa}_k(\boldsymbol{\varepsilon})$ is d^k, while the number of non-repeated elements in $\boldsymbol{\kappa}_k(\boldsymbol{\varepsilon})$ is $\binom{d+k-1}{k} = \frac{(d+k-1)!}{k!(d-1)!}$.

Let $\mathbf{v}_k^{\boldsymbol{\varepsilon}}$ be a $d^{k-1} \times d$ matrix such that $\boldsymbol{\kappa}_k(\boldsymbol{\varepsilon}) = \text{vec}(\mathbf{v}_k^{\boldsymbol{\varepsilon}})$. By assumption (3.1.ii-iv), all non-marginal cumulants of $\boldsymbol{\varepsilon}_t$ are zero, then $\mathbf{v}_k^{\boldsymbol{\varepsilon}}$ has the following structure

$$\mathbf{v}_k^{\boldsymbol{\varepsilon}} = \begin{bmatrix} \kappa_{k,1}\boldsymbol{e}_1^{\otimes k-1} & \kappa_{k,2}\boldsymbol{e}_2^{\otimes k-1} & \dots & \kappa_{k,d}\boldsymbol{e}_d^{\otimes k-1} \end{bmatrix}, \tag{3.8}$$

where $\boldsymbol{e}_j$ is the j-th canonical vector in $\mathbb{R}^d$.

3.3.2. Number of non-Gaussian shocks

From equation (3.8), it is feasible to identify the non-Gaussian dimension in $\boldsymbol{\varepsilon}$ by determining the non-zero third or fourth order marginal cumulants $\kappa_{k,m}^{\varepsilon}$. Besides, the number of non-zero marginal higher order cumulants is equal to rank($\mathbf{v}_k^{\varepsilon}$). Thus, if the structural errors were observable, rank$\left(\mathbf{v}_3^{\boldsymbol{\varepsilon}}\right)$ would be equal to the number of asymmetric shocks. In contrast, rank$\left(\mathbf{v}_4^{\boldsymbol{\varepsilon}}\right)$ would provide the number of non-mesokurtic shocks in the structural model. This relationship is behind the approach followed byMaxand (2020) work; since the structural model is restricted to a causal SVAR, structural shock estimates can be recovered. She employs Nordhausen et al. (2017) strategy, which relates the non-zero eigenvalues of the standardized fourth-centered moments of the vector (scatter matrix) with the non-Gaussian dimension.

In contrast, Guay (2021) approach does not employ $\mathbf{v}_k^{\boldsymbol{\varepsilon}}$ directly. Like Maxand (2018), his model is a causal SVAR, which implies that $(\boldsymbol{\vartheta}_{0,1})$ is identified by estimating an RF-VAR model. Therefore, as we explained above, the errors of the RF-VAR are a linear rotation of structural shocks, that is $\boldsymbol{u}_t = \boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})\boldsymbol{\varepsilon}_t$ where $\boldsymbol{\vartheta}_{0,2}$ governs the contemporaneous or static part of the SVAR model and is not identified. According to Jammalamadaka et al. (2006), $\boldsymbol{\kappa}_k^{\boldsymbol{u}} = \boldsymbol{B}^{\otimes k}\boldsymbol{\kappa}_k(\boldsymbol{\varepsilon})$ where $\boldsymbol{B}^{\otimes k}$ denotes the k order Kronecker power, i.e. the Kronecker product of $\boldsymbol{B}$ with itself k times. Then, similarly to $\mathbf{v}_k^{\boldsymbol{\varepsilon}}$ we define

$$\mathbf{v}_k^{\boldsymbol{u}} := \text{vec}^{-1}(\boldsymbol{\kappa}_k^{\boldsymbol{u}}) = \boldsymbol{B}^{\otimes k-1}\mathbf{v}_k^{\boldsymbol{\varepsilon}}\boldsymbol{B}'$$

Since $\boldsymbol{B} := \boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})$ is a squared full rank matrix, it follows that rank$\left(\mathbf{v}_k^{\boldsymbol{u}}\right) =$ rank$\left(\mathbf{v}_k^{\varepsilon}\right)$ and therefore, identification of structural shocks is not needed to determine the non-Gaussian dimension in the structural model.

Even so, under a more general structural model like the one considered in this paper, Maxand (2020) and Guay (2021) approaches are invalid. First, a possibly non-fundamental structural VARMA is only identified if all shocks are non-Gaussian, which we ignore. Second, the errors from the fundamental RF-VARMA model are a possibly non-causal filter of structural shocks. In this case:

$$
\begin{aligned}
\boldsymbol{\kappa}_k^{\boldsymbol{u}} &= \sum_{j=-\infty}^{\infty} \boldsymbol{\delta}_j^{\otimes k} \boldsymbol{\kappa}_k^{\boldsymbol{\varepsilon}} \\
\mathbf{v}_k^{\boldsymbol{u}} &= \sum_{j=-\infty}^{\infty} \boldsymbol{\delta}_j^{\otimes k-1} \mathbf{v}_k^{\boldsymbol{\varepsilon}} \boldsymbol{\delta}_j'
\end{aligned} \tag{3.9}
$$

From (3.9), it is evident that $\operatorname{rank}\left(\mathbf{v}_k^{\boldsymbol{u}}\right) \neq \operatorname{rank}\left(\mathbf{v}_k^{\boldsymbol{\varepsilon}}\right)$ in general. In fact, $\operatorname{rank}\left(\mathbf{v}_k^{\boldsymbol{u}}\right)$ could be smaller than $\operatorname{rank}\left(\mathbf{v}_k^{\boldsymbol{\varepsilon}}\right)$ even though each $\boldsymbol{\delta}_j$ were full rank.

We employ the cumulant spectrum of order k for RF errors to deal with this problem. According to Brillinger (2001), the k-th order of cumulant spectrum is:

$$
g_k^u(\boldsymbol{\lambda}) = (2\pi)^{k-1} \left(\boldsymbol{\delta}\left(e^{i \sum_{m=1}^{k-1} \lambda_m}; \boldsymbol{\vartheta}_f\right) \otimes \bigotimes_{j=1}^{k-1} \left\{ \boldsymbol{\delta}\left(e^{-i\lambda_{k-j}}; \boldsymbol{\vartheta}_f\right) \right\} \right) \boldsymbol{\kappa}_k(\boldsymbol{\varepsilon}), \tag{3.10}
$$

where $\bigotimes_{j=1}^{n} \boldsymbol{A}_j := \boldsymbol{A}_1 \otimes \boldsymbol{A}_2 \otimes \cdots \otimes \boldsymbol{A}_n$.

Analogously to the definition of $\mathbf{v}_k^{\boldsymbol{u}}$, we construct the $d^{k-1} \times d$ matrix $G_k^u(\boldsymbol{\lambda})$ such that $g_k^u(\boldsymbol{\lambda}) = \operatorname{vec}\left(G_k^u(\boldsymbol{\lambda})\right)$. Therefore

$$
G_k^u(\boldsymbol{\lambda}) = \left(\bigotimes_{j=1}^{k-1} \left\{ \boldsymbol{\delta}\left(e^{-i\lambda_{k-j}}; \boldsymbol{\vartheta}_f\right) \right\} \right) \mathbf{v}_k^{\boldsymbol{\varepsilon}} \left(\boldsymbol{\delta}'\left(e^{i \sum_{m=1}^{k-1} \lambda_m}; \boldsymbol{\vartheta}_f\right) \right) \tag{3.11}
$$

Since $\boldsymbol{\delta}(z, \boldsymbol{\vartheta}_f)$ is an all-pass filter and $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, provided that AC and BD are conformable, we define the $d \times d$ matrix $G_k^{u,2}(\lambda_k)$ as

$$
G_k^{u,2}(\lambda_k) := \left[G_k^u(\boldsymbol{\lambda})\right]^* G_k^u(\boldsymbol{\lambda}) = \boldsymbol{\delta}\left(e^{-i \sum_{m=1}^{k-1} \lambda_m}; \boldsymbol{\vartheta}_f\right) \left(\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime} \boldsymbol{\Omega}^{\otimes(k-1)} \mathbf{v}_k^{\boldsymbol{\varepsilon}} \right) \boldsymbol{\delta}'\left(e^{i \sum_{m=1}^{k-1} \lambda_m}; \boldsymbol{\vartheta}_f\right), \tag{3.12}
$$

where $\boldsymbol{A}^*$ denotes the conjugate transpose of the complex-valued matrix A and $\lambda_k = -\sum_{m=1}^{k-1} \lambda_m$. Notice that $G_k^{u,2}(\lambda_k)$ only depends on the scalar frequency $\lambda_k \in [-\pi, \pi]$, hence from now on we drop the lower-index k of the frequency.

From (3.12) and since $\boldsymbol{\delta}(z, \boldsymbol{\vartheta}_f)$ is full rank for any $|z| = 1$, it is clear that $\operatorname{rank}\left(G_k^{u,2}(\lambda)\right) = \operatorname{rank}\left(\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime} \boldsymbol{\Omega}^{\otimes k-1} \mathbf{v}_k^{\boldsymbol{\varepsilon}}\right) = \operatorname{rank}\left(\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime} \mathbf{v}_k^{\boldsymbol{\varepsilon}}\right)$. Thus, we relate the non-Gaussian dimension in $\boldsymbol{\varepsilon}_t$ to the rank of matrix $G_k^{u,2}(\lambda)$. The following proposition summarizes our main finding in this paper.

Proposition 3.1.

Let the structural model be described by equations (3.1) and (3.2), with structural shocks satisfying Assumption (3.1). Consider the arrays $G_3^{u,2}(\lambda)$, $G_4^{u,2}(\lambda)$ and $G_{34}^{u,2}(\lambda) = \begin{bmatrix} G_3^{u,2}(\lambda) \\ G_4^{u,2}(\lambda) \end{bmatrix}$, *therefore*

$$\begin{aligned} rank\left(G_3^{u,2}(\lambda)\right) &= rank\left(\mathbf{v}_3^{\varepsilon\prime}\mathbf{v}_3^{\varepsilon}\right) = d_3, \\ rank\left(G_4^{u,2}(\lambda)\right) &= rank\left(\mathbf{v}_4^{\varepsilon\prime}\mathbf{v}_4^{\varepsilon}\right) = d_4, \\ rank\left(G_{34}^{u,2}(\lambda)\right) &= d_{34}, \quad \forall\ \lambda \in [-\pi,\pi], \end{aligned}$$

where d_3 and d_4 are the number of skewed and non-mesokurtic structural shocks, respectively, and d_{34} is the number of asymmetric or non-mesokurtic shocks.

Proposition (3.1) can be interpreted as follows: the non-Gaussian dimension in the vector of structural shocks is equal to the rank of a matrix constructed from higher order cumulant spectrum of RF errors at a given frequency. In particular, if only the third-order cumulant spectrum is employed, we obtain the non-Gaussian dimension delivered by asymmetric non-Gaussian shocks. If only the fourth-order spectrum is utilized, we capture the non-Gaussian dimension spanned by non-mesokurtic shocks.

Besides, notice that in case the condition in (3.2) holds for all $z \in \mathbb{T}_+ = \{x \in \mathbb{C} \mid |x| \le 1\}$, higher order cumulant spectrums are constant. Hence, array $G_k^{u,2}(\lambda)$ is constant, i.e., the same for any frequency. In particular, $G_k^{u,2}(\lambda) = \boldsymbol{B}(\boldsymbol{\vartheta}_{0,2})\left(\mathbf{v}_k^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes k-1}\mathbf{v}_k^{\varepsilon}\right)\boldsymbol{B}'(\boldsymbol{\vartheta}_{0,2}) = \mathbf{v}_k^{\boldsymbol{u}\prime}\mathbf{v}_k^{\boldsymbol{u}}$ for any $\lambda \in [-\pi,\pi]$. This setup is the case analyzed in Guay, 2021. Under this situation, proposition (3.1) holds as well. In other words, if a fundamental structural VARMA model generates $\boldsymbol{y}_t$, a rectangular array constructed from cumulant spectrums of order 3 or 4 based on RF errors identifies the non-Gaussian dimension.

3.4. Estimating the number of non-Gaussian elements

From the discussion above, to determine the non-Gaussian dimension in the structural shocks vector, we need to find the rank of a rectangular array constructed from the cumulant spectrum of order $k = 3,4$. Thus, our empirical problem becomes the estimation of the rank of a matrix. In the literature, this problem has been dealt with in two approaches: sequential testing or estimation by information criteria (see Camba-Méndez and Kapetanios (2009) for a detailed review). We opt for the former approach because it may be problematic to construct a pseudo-likelihood function for a non-causal filter. Additionally, the parametric estimation of our matrix of interest is not feasible because the possibly non-fundamental filter $\boldsymbol{\delta}(z,\boldsymbol{\vartheta}_f,\boldsymbol{\vartheta}_{0,1}) = \boldsymbol{\Psi}^{-1}(z,\boldsymbol{\vartheta}_f)\boldsymbol{\Psi}(z,\boldsymbol{\vartheta}_{0,1})$ is not identified unless we impose fundamentalness. Consequently, we decide to estimate $g_k(\boldsymbol{\lambda})$ non-parametrically.

3.4.1. Estimation of non-Gaussian Dimension

Dealing with $G_k^{u,2}(\lambda)$ may be problematic since it implies working with complex-valued terms. For simplifying the analysis, we use $\text{Re}(G_k^2(\lambda))$ instead. At frequency zero, there is no problem since $G_k^{u,2}(0) = \text{Re}(G_k^{u,2}(0))$. At other frequencies, $\text{Re}(G_k^{u,2}(\lambda))$ is the sum of two positive semidefinite quadratic forms; thus, we cannot lose rank, and we gain rank only if both quadratic forms has linearly independent columns.

We now describe our estimation of the non-Gaussian dimension in structural shocks through a sequential hypothesis test. At step $s = 1, 2, \ldots, d$ in the sequential procedure, the null and alternative hypotheses are

$$H_{0,s} : \text{rank}\left(\text{Re}(G_k^{u,2}(\lambda))\right) = r_s$$
$$H_{1,s} : \text{rank}\left(\text{Re}(G_k^{u,2}(\lambda))\right) \geq r_s + 1,$$

where r_s denotes the rank under the null hypothesis at step s.

We start the sequential procedure (at step $s = 1$) by imposing $r_1 = 0$, i.e., joint Gaussianity of the structural shocks. If this is rejected, we continue with the next step, $s = 2$, and the null hypothesis is updated to $r_2 = 1$, i.e., only one asymmetric or non-mesokurtic structural shock. We continue this way until we cannot reject a null hypothesis or reach the final step $s = d$, where $r_d = d - 1$. If this is not rejected, the non-Gaussian dimension in the structural shocks vector is $d - 1$; on the contrary, the non-Gaussian dimension is d.

The literature on rank estimation via hypothesis testing is extensive (see Al-Sadoon (2017) for a complete survey). We center our approach on Kleibergen and Paap (2006) (hereafter KP) proposal. They employ the singular value decomposition (hereafter SVD) of a matrix of interest, $\mathbf{\Pi}$. The SVD of matrix $\mathbf{\Pi}_{m \times n}$ consists on finding squared orthonormal matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ of dimension m and n, respectively; and a quasi-diagonal rectangular matrix $\mathscr{L}_{m \times n}$ such that[15]

$$\boldsymbol{R}_1' \mathbf{\Pi} \boldsymbol{R}_2 = \mathscr{L}, \tag{3.13}$$

where $\mathscr{L}$ is a rectangular array with a squared block $\mathscr{L}_{\overline{m}} = \text{diag}(l_1, \ldots, l_{\overline{m}})$ with $\overline{m} = \min\{m, n\}$ and $l_1 \geq l_2 \geq \cdots \geq l_{\overline{m}} \geq 0$ are the singular values of $\mathbf{\Pi}$. If $\mathbf{\Pi}$ is squared, then $\mathscr{L} = \mathscr{L}_{\overline{m}}$. Assuming that $m > n$, $\mathscr{L} = \left[\mathscr{L}_{\overline{m}} \quad \mathbf{0}_{n \times (m-n)}\right]'$.

From decomposition in (3.13) and selecting an integer $r \in \{0, \ldots, \overline{m} - 1\}$, Kleibergen

[15] See Golub and Van Loan (2012) for more details on SVD of a matrix.

and Paap (2006) obtain that

$$\boldsymbol{\Pi} = \boldsymbol{C}_r\boldsymbol{D}_r + \boldsymbol{C}_{r,\perp}\mathscr{L}_r\boldsymbol{D}_{r,\perp} \tag{3.14}$$

where $\boldsymbol{C}_r\boldsymbol{D}_r = \underbrace{\begin{bmatrix}\boldsymbol{R}_{1,11}\\ \boldsymbol{R}_{1,21}\end{bmatrix}}_{m\times r}\underbrace{\mathscr{L}_1}_{r\times r}\underbrace{\begin{bmatrix}\boldsymbol{R}'_{2,11} & \boldsymbol{R}'_{2,21}\end{bmatrix}}_{r\times n}$; $\boldsymbol{C}_{r,\perp}\mathscr{L}_r\boldsymbol{D}_{r,\perp} = \underbrace{\begin{bmatrix}\boldsymbol{R}_{1,12}\\ \boldsymbol{R}_{1,22}\end{bmatrix}}_{m\times(m-r)}\underbrace{\mathscr{L}_2}_{(m-r)\times(n-r)}\underbrace{\begin{bmatrix}\boldsymbol{R}'_{2,12} & \boldsymbol{R}'_{2,22}\end{bmatrix}}_{(n-r)\times n}$. $\boldsymbol{C}_{r,\perp}$ and $\boldsymbol{D}_{r,\perp}$ are the orthogonal complements of $\boldsymbol{C}_r$ and $\boldsymbol{D}_r$, respectively. Thus, Kleibergen and Paap (2006) approach consists in decomposing the matrix $\boldsymbol{\Pi}$ into two linear independent parts: one that is made of the multiplication of two matrices that are full column rank, $\boldsymbol{C}_r\boldsymbol{D}_r$, and another with null rank $\boldsymbol{C}_{r,\perp}\mathscr{L}_r\boldsymbol{D}_{r,\perp}$. Therefore, $\text{rank}(\boldsymbol{\Pi}) = \text{rank}(\boldsymbol{C}_r\boldsymbol{D}_r) = r$. This result is only achieved if we place all the non-zero singular values into $\mathscr{L}_1$. Accordingly, the null hypothesis of $\text{rank}(\boldsymbol{\Pi}) = r$ is equivalent to $\mathscr{L}_r = \boldsymbol{0}$.

Since $\boldsymbol{\Pi}$ is not observable, the SVD is applied to its estimator $\widehat{\boldsymbol{\Pi}}$. Thus, we can write $\widehat{\boldsymbol{\Pi}} = \widehat{\boldsymbol{C}}_r\widehat{\boldsymbol{D}}_r + \widehat{\boldsymbol{C}}_{r,\perp}\widehat{\mathscr{L}}_r\widehat{\boldsymbol{D}}_{r,\perp}$. Under the assumption of $\sqrt{T}(\text{vec}[\widehat{\boldsymbol{\Pi}}] - \text{vec}[\boldsymbol{\Pi}]) \xrightarrow{d} \mathcal{N}_{mn}(\boldsymbol{0}, \Xi)$ with Ξ positive definite, then under the null hypothesis of $\text{rank}(\boldsymbol{\Pi}) = r$ the asymptotic distribution of $\widehat{\boldsymbol{\ell}}_r = \text{vec}\left(\widehat{\mathscr{L}}_r\right)$ is $\sqrt{T}\widehat{\boldsymbol{\ell}}_r \xrightarrow{d} \mathcal{N}_{(m-r)(n-r)}\left(\boldsymbol{0}, \tilde{\Xi}_r\right)$ where $\tilde{\Xi}_r = (\boldsymbol{D}_{r,\perp} \otimes \boldsymbol{C}'_{r,\perp})\Xi(\boldsymbol{D}_{r,\perp} \otimes \boldsymbol{C}'_{r,\perp})$. The KP-statistic is

$$KP_r^{(T)} = T\widehat{\boldsymbol{\ell}}'_r\widehat{\tilde{\Xi}}_r^{-1}\widehat{\boldsymbol{\ell}}_r \xrightarrow{d} \chi^2_{(m-r)(n-r)}, \tag{3.15}$$

where $\widehat{\tilde{\Xi}}_r$ is the consistent estimator of the asymptotic variance, $\tilde{\Xi}_r$.

In some dimensions, our problem departs from Kleibergen and Paap (2006) context. First, our matrix of interest $\boldsymbol{\Pi} = \text{Re}(G_k^{u,2}(\lambda))$ is symmetric, i.e. it contains repeated elements. Hence, the asymptotic variance of unrestricted estimator $\widehat{\boldsymbol{\Pi}}$, Ξ, is only positive semi-definite. Second, as we detail in the section where it is discussed the estimation of higher order spectrum, the asymptotic distribution of the unrestricted estimator of our matrix of interest changes under the null hypothesis of joint Gaussianity, i.e., under the null of $\text{rank}(\boldsymbol{\Pi}) = 0$. Finally, our convergence rates for the asymptotic distribution of the statistic are lower in comparison to the standard speed of convergence, $\sqrt{T}$.

Test Statistic at first step

For the first step in the sequential procedure, the null hypothesis is $\text{rank}(\boldsymbol{\Pi}) = 0$, i.e., the vector of structural shocks is an uncorrelated Gaussian process. The KP statistic under this null hypothesis is

$$KP_{r_1}^{(T)} = a_T^4\widehat{\boldsymbol{\ell}}'_0\widehat{\text{aVar}}\left(\left(\overline{\boldsymbol{D}}_{0,\perp} \otimes \overline{\boldsymbol{C}}'_{0,\perp}\right)\boldsymbol{Q}\right)^{\dagger}\widehat{\boldsymbol{\ell}}_0, \tag{3.16}$$

where $\boldsymbol{A}^\dagger$ denotes the Moore-Penrose inverse of matrix $\boldsymbol{A}$, $\boldsymbol{Q}$ is the vectorized form of a linear combination of an inverse Wishart distribution, which parameters depends on the asymptotic variance of spectrum estimates (see Section 4.3 below), and $\overline{\boldsymbol{D}}_{0,\perp}$ and $\overline{\boldsymbol{C}}'_{0,\perp}$ represent the limiting values of $\widehat{\boldsymbol{D}}_{0,\perp}$ and $\widehat{\boldsymbol{C}}_{0,\perp}$. a_T is the convergence rate of spectrum estimates.

Test Statistic under mixed cases

Now, when the null hypothesis is $\mathrm{rank}(\boldsymbol{\Pi}) = r$ with $r > 0$, i.e., the vector of structural shocks is an i.i.d process with mutually independent components and there are r non-Gaussian distributed shocks. In this case, the asymptotic distribution of $\widehat{\boldsymbol{\Pi}}$ is Normal, and the KP statistic adopts a similar form as in Kleibergen and Paap (2006),

$$KP_{r_s}^{(T)} = a_T^2 \widehat{\boldsymbol{\ell}}'_r \left(\left[(\widehat{\boldsymbol{D}}_{r,\perp} \otimes \widehat{\boldsymbol{C}}'_{r,\perp}) \widehat{\Xi} (\widehat{\boldsymbol{D}}_{r,\perp} \otimes \widehat{\boldsymbol{C}}'_{r,\perp})' \right] \right)^\dagger \widehat{\boldsymbol{\ell}}_r. \tag{3.17}$$

The statistic in (3.17) is asymptotically distributed as a chi-square with degrees of freedom $\nu = \mathrm{rank}\left[(\widehat{\boldsymbol{D}}_{r,\perp} \otimes \widehat{\boldsymbol{C}}'_{r,\perp}) \widehat{\Xi} (\widehat{\boldsymbol{D}}_{r,\perp} \otimes \widehat{\boldsymbol{C}}'_{r,\perp})' \right]$. Unfortunately, the exact value for ν is not easy to determine, because -although $\widehat{\boldsymbol{C}}_{r,\perp}$ and $\widehat{\boldsymbol{D}}'_{r,\perp}$ are full column rank matrices of dimension $d \times (d-r)$- the matrix $(\widehat{\boldsymbol{D}}_{r,\perp} \otimes \widehat{\boldsymbol{C}}'_{r,\perp})$ is full row rank matrix of dimension $(d-r)^2 \times d^2$ and the matrix $\widehat{\Xi}$ is only positive semidefinite, i.e. it is not full rank. From this discussion, we can bound the degrees of freedom, $1 \leq \nu \leq \min\{\mathrm{rank}(\Xi), (d-r_s)^2\}$.

3.4.2. Bootstrap Test

Because of the difficulties characterizing the asymptotic distributions of the KP statistic for each step in the sequential procedure, we opt for a bootstrap strategy. Nonetheless, this path entails other challenges. The most problematic is to accomplish that the bootstrap sample appropriately reflects the null hypothesis of each step. Otherwise, it may severely compromise the size and power of the test (see P. Hall and Wilson (1991); Portier and Delyon (2014)).

It is a well-known result from the SVD of a matrix $\boldsymbol{\Pi}_{m\times n}$ with $m > n$ and $\mathrm{rank}(\boldsymbol{\Pi}) = r$, that $\mathrm{null}(\boldsymbol{\Pi})$ is spanned by the last $n-r$ columns of $\boldsymbol{R}_2$ and the range $\mathrm{ran}(\boldsymbol{\Pi})$ is spanned by the first r columns of $\boldsymbol{R}_1$. Thus, when $\mathrm{rank}(\boldsymbol{\Pi}) = r$, the first r columns of $\boldsymbol{R}_1$ span the non-Gaussian dimension, while last $n-r$ columns of $\boldsymbol{R}_2$ span the Gaussian one. Thus, we follow Nordhausen et al. (2017) approach and use the estimates of $\boldsymbol{R}_2$ for a projection of residuals into Gaussian and non-Gaussian dimensions.

Given the sample $\{\boldsymbol{y}_t\}_{t=1}^T$, we estimate a RF-VARMA model and compute its RF residuals $\{\hat{\boldsymbol{u}}_t\}_{t=1}^T$. Based on these estimates, we construct our matrix of interest $\widehat{\boldsymbol{\Pi}} = \mathrm{Re}\left(\widehat{G}_k^{\hat{u},2,(T)}(\lambda)\right)$ and the matrices $\widehat{\mathscr{L}}$, $\widehat{\boldsymbol{R}}_1$ and $\widehat{\boldsymbol{R}}_2$. Under $H_{0,r}$: rank$(\boldsymbol{\Pi}) = r$ and according to Nordhausen et al. (2017), we construct $\widehat{\boldsymbol{R}}_{2,r} := \left[\hat{\mathbf{r}}_{2,r+1} \cdots \hat{\mathbf{r}}_{2,d}\right]$ and the projection matrix $\mathbf{M}_r = \boldsymbol{I} - \hat{\Sigma}_{\hat{u}}^{1/2}\widehat{\boldsymbol{R}}_{2,r}\widehat{\boldsymbol{R}}_{2,r}'\hat{\Sigma}_{\hat{u}}^{-1/2}$, where $\hat{\Sigma}_{\hat{u}}$ is variance of RF residuals vector. $\mathbf{M}_r$ is the projection matrix into the orthogonal space to the one spaned by $\widehat{\boldsymbol{R}}_{2,r}$, i.e., $\mathbf{M}_r$ is a projection into non-Gaussian space. Based on these inputs, at any step s in the sequential testing procedure, each bootstrap sample is created following Algorithm 1.

Algorithm 1 Bootstrap Sample under $H_{0,s}$: rank$(\boldsymbol{\Pi}) = r_s$

1. Obtain an unrestricted bootstrap sample of RF residuals $\{\tilde{\boldsymbol{u}}_t\}_{t=1}^T$.
2. Draw a sequence of size T of random vectors, $\{\boldsymbol{\eta}_t\}_{t=1}^T$, from multivariate standard Normal distribution, $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_{d-r_s})$.
3. Bootstrap sample of RF residuals consistent with $H_{0,s}$ is given by
$$\boldsymbol{u}_t^\star = \mathbf{M}_{r_s}\tilde{\boldsymbol{u}}_t + \hat{\Sigma}_{\hat{u}}^{1/2}\widehat{\boldsymbol{R}}_{2,r_s}\boldsymbol{\eta}_t.$$
4. The restricted bootstrap sample of $\{\boldsymbol{y}_t^\star\}_{t=1}^T$ is built parametrically, i.e. using $\boldsymbol{u}_t^\star$ and estimated parameters from RF-VARMA.

It is worth mentioning that the unrestricted bootstrap sample of RF residuals is not performed by the usual *independent bootstrap* procedure stated by Efron (1979), but by the proposed method in Politis and Romano (1994), called the *stationary bootstrap*. According to the authors, this procedure is suitable for stationary weakly dependent time series because, unlike other proposals such as Kunsch (1989); Politis and Romano (1992), it exhibits the desirable property that the resampled time series obtained are stationary conditional on the original data. This type of bootstrap sampling is necessary because RF errors are only serially uncorrelated, but they are not independent unless the fundamentalness of data is imposed or Gaussian is assumed. Besides, for obtaining the restricted bootstrap sample for observable data, we employ a parametric approach; otherwise is not possible to impose the null hypothesis in the data.

Once the restricted bootstrap sample has been obtained and we set a nominal size $\alpha \in (0,1)$ for all the steps in the sequence, the test proceeds as follows:

Algorithm 2 Bootstrap Test

0. Set the initial step $s = 1$.
1. Given the step s, set the null hypothesis as $H_{0,s}$: rank$(\mathbf{\Pi}) = r_s = s-1$ and compute the KP statistic with original sample, $KP^{(T)}_{r_s}$.
2. Use Algorithm 1 for obtaining B bootstrap samples. Compute statistic $KP^{(T),\star}_{r_s,b}$ for each bootstrap sample $b = 1,\dots,B$.
3. Compute the bootstrap p-value for H_{0,r_s} as $\widehat{pv}_{r_s} = \frac{1+\sum_{b=1}^{B}\mathbf{1}\left(KP^{(T)}_{r_s}\leq KP^{(T),\star}_{r_s,b}\right)}{1+B}$ and we decide in this fashion:
   - 3.1. If $\widehat{pv}_{r_s} \geq \alpha$, then $H_{0,s}$ is not rejected, the estimated rank is $\hat{r} = r_s$ and the procedure ends;
   - 3.2. else if $s < d-1$, update $s = s+1$ and repeat since stage 1;
   - 3.3. else, the procedure ends and the rank is $\hat{r} = d$.

3.4.3. Estimation of Cumulant Spectrum of order k

This part briefly discusses some details of estimating higher order cumulant spectrum. The discussion follows closely Brillinger and Rosenblatt (1967). The sample periodogram for a k-tuple of RF errors is

$$I^{(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^{k-1}T}\prod_{j=1}^{k} z^{(T)}_{c_j}(\lambda_{t_j}), \tag{3.18}$$

where $\boldsymbol{c} = (c_1,\dots,c_k) \in \sigma_k(\{1,\dots,d\})$. $\lambda_{t_j} = 2\pi\frac{t_j}{T}$ for $t_j = 1,\dots,T-1$, $j = 1,\dots,k-1$ and $\sum_{j=1}^{k}\lambda_{t_j} = 0[\mathrm{mod}(2\pi)]$. Besides, $z^{(T)}_{c_j}(\lambda_{t_j}) = \sum_{t=0}^{T-1} u_{c_j,t}e^{-i\lambda_{t_j}t}$ is the discrete Fourier transform (DFT) of c_j-th reduced-form error, $u_{c_j,t}$.

A consistent estimator of the cumulant spectrum of order k for a $\boldsymbol{c}$-tuple of RF errors is

$$\hat{g}^{\hat{u},(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda}) = \left(\frac{2\pi}{T}\right)^{k-1}\sum_{t_1,\dots,t_k=0}^{T-1} W_T\left(\lambda_1-\lambda_{t_1},\dots,\lambda_k-\lambda_{t_k}\right)I^{(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda}) \tag{3.19}$$

with $W_T(\boldsymbol{a}) = H_T^{-(k-1)}\sum_{\boldsymbol{j}=(j_1,\dots,j_k):\sum_{m=1}^{k}(j_m+a_m)=0} W\left(\frac{1}{H_T}(\boldsymbol{a}+2\pi\boldsymbol{j})\right)$ and H_T satisfies $\lim_{T\to\infty} H_T = 0$ and $\lim_{T\to\infty} TH_T^{k-1} = \infty$. The weighting function, $W(\boldsymbol{a})$, is symmetric around $\mathbf{0}$ and satisfies conditions stated in Brillinger and Rosenblatt (1967, Assumption II).

The consistent estimator of $g^u_k(\boldsymbol{\lambda})$ is the collection of all $\hat{g}^{\hat{u},(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda})$ in a vector, which we call $\hat{g}^{\hat{u},(T)}_k(\boldsymbol{\lambda})$. Thus, we can obtain the $d^{k-1}\times d$ matrix $\hat{G}^{\hat{u},(T)}_k(\boldsymbol{\lambda})$ such that $\hat{g}^{\hat{u},(T)}_k(\boldsymbol{\lambda}) =$

$\mathrm{vec}\left(\hat{G}_k^{\hat{u},(T)}(\boldsymbol{\lambda})\right)$. Besides, $\hat{G}_k^{\hat{u},2,(T)}(\lambda) = \left[\hat{G}_k^{\hat{u},(T)}(\boldsymbol{\lambda})\right]^* \hat{G}_k^{\hat{u},(T)}(\boldsymbol{\lambda})$. Finally, our matrix of interest is $\mathrm{Re}\left(\hat{G}_k^{\hat{u},2,(T)}(\lambda)\right)$.

3.4.4. Asymptotics of spectral estimators

Brillinger and Rosenblatt (1967) show that

$$\sqrt{H_T^{k-1}T}\left(\begin{bmatrix}\mathrm{Re}(\hat{g}_k^{\hat{u},(T)}(\boldsymbol{\lambda}))\\ \mathrm{Im}(\hat{g}_k^{\hat{u},(T)}(\boldsymbol{\lambda}))\end{bmatrix} - \begin{bmatrix}\mathrm{Re}(g_k^{u}(\boldsymbol{\lambda}))\\ \mathrm{Im}(g_k^{u}(\boldsymbol{\lambda}))\end{bmatrix}\right) \xrightarrow{d} \mathcal{N}_{2d^k}\left(\mathbf{0}, \mathcal{V}(\boldsymbol{\lambda})\right) \tag{3.20}$$

where $\mathcal{N}_{2d^k}$ represents a multivariate Gaussian random $2d^k$-dimensional column-vector with mean $\mathbf{0}$ and variance $\mathcal{V}(\boldsymbol{\lambda})$; $\mathcal{V}(\boldsymbol{\lambda}) = \begin{bmatrix}\mathrm{Re}(\boldsymbol{\Lambda}^k(\boldsymbol{\lambda},\boldsymbol{\lambda})) & -\mathrm{Im}(\boldsymbol{\Lambda}^k(\boldsymbol{\lambda},\boldsymbol{\lambda}))\\ \mathrm{Im}(\boldsymbol{\Lambda}^k(\boldsymbol{\lambda},\boldsymbol{\lambda})) & \mathrm{Re}(\boldsymbol{\Lambda}^k(\boldsymbol{\lambda},\boldsymbol{\lambda}))\end{bmatrix}$ and $\boldsymbol{\Lambda}^k(\boldsymbol{\lambda},\boldsymbol{\mu}) = \left[\Lambda^k_{(\boldsymbol{c},\boldsymbol{b})}(\boldsymbol{\lambda},\boldsymbol{\mu})\right]_{(\boldsymbol{c},\boldsymbol{b})\in[\sigma_k(\{1,\ldots,d\})]^2}$ with

$$\begin{aligned}\lim_{T\to\infty} H_T^{k-1}T\,\mathbb{C}\mathrm{ov}\left(\hat{g}_{\boldsymbol{c},k}^{\hat{u},(T)}(\boldsymbol{\lambda}), \hat{g}_{\boldsymbol{b},k}^{\hat{u},(T)}(\boldsymbol{\mu})\right) &= \Lambda^k_{(\boldsymbol{c},\boldsymbol{b})}(\boldsymbol{\lambda},\boldsymbol{\mu})\\ &= 2\pi \sum_{\sigma(\{1,\ldots,k\})}\prod_{m=1}^{k}\eta\left(\lambda_m - (\sigma\mu)_m\right) g_{2,(c_m,(\sigma b)_m)}(\lambda_m)\int W(\boldsymbol{\beta})W(\sigma\boldsymbol{\beta})\delta_D\left(\sum_{m=1}^{k}\beta_m\right)d\boldsymbol{\beta}\end{aligned}$$

where $\sigma\boldsymbol{b} = (b_{\sigma_1},\ldots,b_{\sigma_k})$, $\sigma = (\sigma_1,\ldots,\sigma_k) \in \sigma(\{1,\ldots,k\})$, $\sum_{m=1}^{k}\lambda_m = 0[\mathrm{mod}(2\pi)]$, $\eta(x) = \sum_{j=-\infty}^{\infty}\delta_D(x+2\pi j)$ and $\delta_D(x)$ is the delta-Dirac function.

The estimator of our matrix of interest is $\widehat{\mathbf{\Pi}} = \mathrm{Re}\left\{\hat{G}_k^{\hat{u},2,(T)}(\boldsymbol{\lambda})\right\}$. Besides, given that $\hat{G}_k^{\hat{u},(T)}(\boldsymbol{\lambda}) = \mathrm{vec}^{-1}\left(\hat{g}_k^{\hat{u},(T)}(\boldsymbol{\lambda})\right)$, $\widehat{\mathbf{\Pi}}$ is a function of real and imaginary parts of $\hat{g}_k^{\hat{u},(T)}(\boldsymbol{\lambda})$ as the following equation shows

$$\mathrm{vec}\left(\widehat{\mathbf{\Pi}}\right) = \mathrm{vec}\left(\mathrm{Re}\left\{\hat{G}_k^{\hat{u},2,(T)}(\boldsymbol{\lambda})\right\}\right) = \left[\left(I_d \otimes \hat{g}_{\mathrm{Re},k}^{\hat{u},(T)}(\boldsymbol{\lambda})\right)^{\otimes 2} + \left(I_d \otimes \hat{g}_{\mathrm{Im},k}^{\hat{u},(T)}(\boldsymbol{\lambda})\right)^{\otimes 2}\right]\mathrm{vec}(\mathbf{I}(d)),$$

where $\hat{g}_{\mathrm{Re},k}^{\hat{u},(T)}(\boldsymbol{\lambda})$ and $\hat{g}_{\mathrm{Im},k}^{\hat{u},(T)}(\boldsymbol{\lambda})$ denote the real and imaginary parts of vector $\hat{g}_k^{\hat{u},(T)}(\boldsymbol{\lambda})$, respectively; and $\mathbf{I}(d) = \mathrm{vec}(\boldsymbol{I}_d)\mathrm{vec}(\boldsymbol{I}_d)' \otimes \boldsymbol{I}_{d^{k-1}}$.

Consequently, the asymptotic distribution of $\mathrm{vec}\left(\widehat{\mathbf{\Pi}}\right)$ depends on the null hypothesis at step s, $H_{0,s}$. Particularly, at the first step $H_{0,1}$: $\mathrm{rank}(\mathbf{\Pi}) = 0$, i.e., under joint Gaussianity of the structural shocks, the population spectrum of order k is zero for $k = 3,4$. This result implies that the Jacobian of $\mathrm{vec}\left(\widehat{\mathbf{\Pi}}\right)$ at population values is null, making it not feasible to use the standard Delta method for finding the asymptotic distribution of $\mathrm{vec}\left(\widehat{\mathbf{\Pi}}\right)$. Proposition (3.2) states the asymptotic distribution for $\mathrm{vec}\left(\widehat{\mathbf{\Pi}}\right)$ at different stages in the sequential testing procedure.

Proposition 3.2. *Assuming that result in (3.20) holds, then:*

1. *When $s = 1$, under the null hypothesis $H_{0,1}$:*

$$H_T^{k-1} T \left(vec \left(\widehat{\boldsymbol{\Pi}} \right) \right) \xrightarrow{d} \boldsymbol{Q} \,, \tag{3.21}$$

2. *For $s \geq 2$, under the null hypotesis $H_{0,s}$:*

$$\sqrt{H_T^{k-1} T} \left(vec \left(\widehat{\boldsymbol{\Pi}} \right) - vec \left(\boldsymbol{\Pi} \right) \right) \xrightarrow{d} \mathcal{N}_{d^2} \left(\mathbf{0},\, \boldsymbol{J}'_{\hat{G}^2_k}(\boldsymbol{\lambda}) \mathcal{V}(\boldsymbol{\lambda}) \boldsymbol{J}_{\hat{G}^2_k}(\boldsymbol{\lambda}) \right), \tag{3.22}$$

In Proposition 3.2, $\boldsymbol{Q} := \frac{1}{2} \left(I_{d^2} \otimes \mathbf{X}' \right) \mathcal{H}_{\hat{G}_k^{\hat{u},2,(T)}}$ and $\mathcal{H}_{\hat{G}_k^{\hat{u},2,(T)}}$ is the vectorized Hessian[16] of $\text{vec}\left(\text{Re}\left\{ \hat{G}_k^{\hat{u},2,(T)}(\lambda) \right\} \right)$ evaluated at zero. $\mathbf{X} := \text{vec}[\mathcal{W}_1(\mathcal{V}(\boldsymbol{\lambda}))]$ with $\mathcal{W}_1(\mathcal{V}(\boldsymbol{\lambda}))$ denotes a Wishart distribution with parameters $\mathcal{V}(\boldsymbol{\lambda})$ and 1 degree of freedom. Besides, $\boldsymbol{J}_{\hat{G}^2_k}(\boldsymbol{\lambda})$ is the Jacobian of $\text{vec}\left(\text{Re}\left\{ \hat{G}_k^{2,(T)}(\lambda) \right\} \right)$ evaluated at population values of cumulant spectrum of order k.

3.4.5. Asymptotic Equivalence between spectrum estimators of $\boldsymbol{u}_t$ and $\hat{\boldsymbol{u}}_t$

The results above are obtained for unobserved RF errors, $\boldsymbol{u}_t$. However, we employ the estimated RF residuals using a finite sample of size T using the estimated RF parameters, $\hat{\boldsymbol{\vartheta}}_f$. In consequence, it is essential to show that asymptotic results remain valid when using estimated RF residuals, $\hat{\boldsymbol{u}}_t$. The spectrum of order k for the estimated RF residuals can be obtained easily replacing $\boldsymbol{u}_t$ by $\hat{\boldsymbol{u}}_t$ in (3.19), denoted by $\hat{g}_{\boldsymbol{c},k}^{\hat{u},(T)}(\boldsymbol{\lambda})$. The rest of the estimators can be obtained as explained above using $\hat{g}_{\boldsymbol{c},k}^{\hat{u},(T)}(\boldsymbol{\lambda})$. This analysis is omitted in Maxand (2020) and Guay (2021) works, maybe explained because the structural linear model is fundamental.

Proposition 3.3. *Given our model is determined by (3.1)-(3.2) and under Assumption 3.1, it holds*

$$\mathbb{E} \sqrt{H_T^{k-1} T} \left| \hat{g}_{\boldsymbol{c},k}^{u,(T)}(\boldsymbol{\lambda}) - \hat{g}_{\boldsymbol{c},k}^{\hat{u},(T)}(\boldsymbol{\lambda}) \right| \leq C H_T^{\frac{k-1}{2}}$$

Proposition 3.3 states the asymptotic equivalence between estimators of higher-order cumulant spectrum using unobserved RF errors and their sample counterparts. This implies that the discrepancy generated by using $\hat{g}_{\boldsymbol{c},k}^{\hat{u},(T)}(\boldsymbol{\lambda})$ decreases to zero at the rate of $C H_T^{\frac{k-1}{2}}$, which is slower than $\sqrt{T}$. This convergence rate comes from the fact that we are

[16]The Hessian of a vector-valued function $\mathbb{R}^d \mapsto \mathbb{R}^q$ is a 3-tensor. Intuitively, the Hessian matrix is 3-dimensional array of dimensions $d \times d \times q$.

using a non-parametric estimation for the higher order spectrum, which incorporates a kernel with band-with H_T.

Corollary 3.1.
Let structural model be described by (3.1) but condition in (3.2) holds for any point in $\mathbb{T}_+$*, then*

$$\mathbb{E}\sqrt{H_T^{k-1}T}\left|\hat{g}_{\boldsymbol{c},k}^{u,(T)}(\boldsymbol{\lambda})-\hat{g}_{\boldsymbol{c},k}^{\hat{u},(T)}(\boldsymbol{\lambda})\right|\leq CH_T^{\frac{k-1}{2}};$$
$$\mathbb{E}\sqrt{T}\left|\hat{\boldsymbol{\kappa}}_k^{\boldsymbol{u},(T)}-\hat{\boldsymbol{\kappa}}_k^{\hat{\boldsymbol{u}},(T)}\right|\leq CT^{-\frac{1}{2}}.$$

In Corollary 3.1, $\hat{\boldsymbol{\kappa}}_k^{\boldsymbol{u},(T)}$ and $\hat{\boldsymbol{\kappa}}_k^{\hat{\boldsymbol{u}},(T)}$ are the sample estimators for cumulant of order k based on unobserved RF errors and RF residuals, respectively.

This corollary states that in the case of working with a more restricted structural model, since the researcher imposes the location of roots, Proposition (3.3) still holds. The asymptotic equivalence remains if contemporaneous higher-order cumulants were employed instead of the higher-order spectrum. However, the convergence rate is faster than in the general setting. This latter result is because the estimation does not use kernel smoothing when using contemporaneous higher-order cumulants.

3.5. Simulation and Empirical Application

Show the validity of the restricted bootstrap sampling analytically is quite intricate. We believe that following procedures that have been proven to be consistent may assure the effectiveness of our approach, although this is not equivalent to formal proof. Additionally, we present some evidence from different Montecarlo exercises as additional support for the validity of our bootstrap sampling and test.

The data-generating process or the true structural model for all the exercises is a non-causal SVAR(1). We employ different distributions for the structural disturbances. Montecarlo and bootstrap repetitions are set to $m = 250$ and $B = 500$, respectively. The nominal significance level is set to $\alpha = 5\%$. The number of points for the DFT is set to a minimum even integer greater than or equal to the sample size T, and the size of the window is set to $\lfloor T/4 \rfloor$.

3.5.1. Simulation Results

Table 3.1 shows the rejection rates of our sequential procedure in the bivariate case ($d = 2$). Besides, the values in a box represent the size of the test. At panel (3.1a), we employ the zero frequency ($\lambda = 0$). The first row is associated with the Gaussian case. Its size is 4%. When we consider a mixed case (second row of the panel (3.1a)), the rejection rate of $H_{0,1}$ represents the power, 32%; while its size is 7%. Finally, when structural shocks are fully non-Gaussian distributed (third row in panel (3.1a)), both rejection rates of $H_{0,1}$ and $H_{0,2}$ represent the power at each step. These values are 42% and 26%, respectively.

Table 3.1: Estimating rank $\mathrm{Re}\left(G_3^2(\lambda_3)\right)$: Rejection Rates (Level: $\alpha = 5\%$, Sample: $T = 250$)

(a) Single Frequency ($\lambda = 0$)

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.04}$	0.01
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.21	$\boxed{0.07}$
$\left(E(0,1);\ \chi_2^2\right)$	0.42	0.26

(c) Grid of 11 frequencies

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.04}$	0.01
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.20	$\boxed{0.06}$
$\left(E(0,1);\ \chi_2^2\right)$	0.45	0.29

(b) Grid of 26 frequencies

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.02}$	0.00
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.22	$\boxed{0.06}$
$\left(E(0,1);\ \chi_2^2\right)$	0.42	0.28

Employ estimates at a single frequency of higher order cumulant spectrum may be noisy, leading to imprecise results, especially when the sample size is small. To make our procedure robust to this unpleasant characteristic, we use a grid of frequencies instead of only the zero frequency. Let $\mathscr{G}$ denote the finite grid of frequencies that are selected, instead of constructing our matrix of interest as $\boldsymbol{\Pi} = \sum_{\lambda_k \in \mathscr{F}} \mathrm{Re}\left(G_k^2(\lambda_k)\right)$ where $\mathscr{F}$ denotes a finite grid of frequencies, because it may entail the risk of gaining rank spuriously; we use the following statistic $\overline{KP}_r^{(T)} = \max_{\lambda \in \mathscr{F}} \left\{ KP_r^{(T)}(\lambda) \right\}$, where $KP_r^{(T)}(\lambda)$ denotes the KP-statistic for a particular frequency λ.

In panel (3.1c) of Table 3.1, a grid of length 11 is selected. For this grid, the test size when all disturbances are Gaussian distributed is 4%; when only one non-Gaussian shock is present, the size is around 6%. The power is quite similar to when a single frequency is employed. In panel (3.1b) of Table 3.1, a grid of 26 frequencies was selected. The size test when $\boldsymbol{\varepsilon}_t$ is a Gaussian process is 5%; and when we have only non-Gaussian shock, the size is around 6%.

Regarding power, the levels remain close to the previous exercises. Furthermore, in Table 3.2, we can observe the effect of increasing the sample size. Regarding the test size for each case, the results show that it remains close to the values obtained with $T = 250$. Moreover, concerning the power, we observe a significant increment in the power, especially in the case where only one single frequency is employed.

Table 3.2: Estimating rank $\left(G_3^2(\lambda)\right)$: Rejection Rates (Level: $\alpha = 5\%$, Sample: $T = 500$)

(a) Single frequency ($\lambda = 0$)

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.04}$	0.01
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.35	$\boxed{0.07}$
$\left(E(0,1);\ \chi_2^2\right)$	0.49	0.38

(b) Grid of 11 frequencies

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,0}$	$H_{0,1}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.03}$	0.00
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.31	$\boxed{0.06}$
$\left(E(0,1);\ \chi_2^2\right)$	0.41	0.32

Suppose we take a strategy for restricting the bootstrap sample as the one followed by Guay (2021), which employs the first r_s columns of $\boldsymbol{R}_1$ for projecting the residuals into the non-Gaussian dimension. In our case, since our matrix of interest is squared and symmetric, $\boldsymbol{R}_1 = \boldsymbol{R}_2$, thus we use the first r_s columns of $\boldsymbol{R}_2$. The rest of the components $d - r_s$ are drawn from a multivariate normal distribution with mean zero and variance $\boldsymbol{I}_{d-r_s}$. In Table (3.3), we compute the rejection rates using Guay's bootstrap sampling approach. The size of the test when having fully Gaussian shocks is, disregarding the length of the frequency grid, around 5%. The size when $\boldsymbol{\varepsilon}_t$ has one Gaussian component is around 7% for the different grid lengths. Nonetheless, the power using Nordhausen et al. (2017) bootstrap sampling is consistently higher than using Guay's approach.

Table 3.3: Estimating rank $\left(G_3^2(\lambda_3)\right)$: Rejection Rates (Level: $\alpha = 5\%$, Sample: $T = 250$) (Restricted Bootstrap sample built following Guay (2021))

(a) Single frequency ($\lambda_3 = 0$)

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.04}$	0.01
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.24	$\boxed{0.07}$
$\left(E(0,1);\ \chi_2^2\right)$	0.40	0.23

(c) Grid of 11 frequencies

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.04}$	0.01
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.21	$\boxed{0.07}$
$\left(E(0,1);\ \chi_2^2\right)$	0.39	0.24

(b) Grid of 26 frequencies

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$	$\boxed{0.05}$	0.01
$\left(\mathcal{N}(0,1);\ \chi_2^2\right)$	0.24	$\boxed{0.07}$
$\left(E(0,1);\ \chi_2^2\right)$	0.42	0.27

For larger dimensional models, we choose $d = 3, 4$. We perform these exercises with sample size $T = 250$ and the grid $\lambda_3 = 0$. The results can be observed in Table (3.4). The distribution MN_1 is a mixture of two normal distributions ($\mathcal{N}(10, 0.75)$ and $\mathcal{N}(-2, 4)$) such that the distribution has a positive skewness coefficient. It can be noted that, while the sample size is fixed, it is more difficult to reject the null hypothesis for larger r_s.

Table 3.4: Estimating rank $\left(G_3^2(\lambda_3)\right)$: Rejection Rates (Level: $\alpha = 5\%$, Sample: $T = 250$)

Distribution of $\boldsymbol{\varepsilon}_t$	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$	$H_{0,4}$
$\mathcal{N}(\mathbf{0}, \boldsymbol{I}_3)$	$\boxed{0.03}$	0.01	0.00	−−
$(E(0,1); MN_1; \mathcal{N}(\mathbf{0},1))$	0.36	0.21	$\boxed{0.08}$	−−
$\left(\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2); E(0,1); \chi_2^2\right)$	0.57	0.27	$\boxed{0.04}$	0.01

3.5.2. Empirical Application

We select two well-known data sets in empirical macroeconomics. The first data set is used in the seminal work of Blanchard and Quah (hereafter BQ), where they introduced the long-run restrictions as an identification scheme for causal structural VAR models. This database contains two endogenous variables: the US GNP's growth and the unemployment rate. The second data set is taken from the work of Blanchard and Perotti (2002) (hereafter BP). Their data includes three endogenous variables: tax revenues, government spending, and GDP (all in real terms). In both cases, we use the exact SVAR specification for each work, avoiding our conclusions from being affected by a specification bias.

Table 3.5 shows the results of applying our proposal to the BQ database. We can identify a single asymmetric structural shock using only third-order cumulant spectral density. When information in the fourth-order spectral cumulant is employed, the proposed method cannot identify any component with excess kurtosis. This result can seem contradictory, though the limited sample size ($T = 148$) might affect the precision of fourth-order estimates. On the other hand, Table 3.6 shows the results of applying our proposed method to the BP dataset. Using third-order information, we detect at most two asymmetric structural shocks at a significance level of 10%. This result is consistently found disregarding if only a single frequency or a grid of frequencies is employed. When fourth-order information is used, at 10% of significance, we can detect three non-mesokurtic structural shocks at zero frequency. When a grid of frequencies is employed instead, the method does detect two non-mesokurtic structural shocks. Besides, if these results are compared to those obtained in Guay (2021), our proposal can detect at least one extra non-mesokurtic structural shock.

Table 3.5: Number of non-Gaussian components in Blanchard and Quah (1989)

(a) $\text{rank}\left(\text{Re}\left(G_3^2(\lambda)\right)\right)$

Frequency Grid	Null Hypotheses	
	$H_{0,1}$	$H_{0,2}$
$\lambda = 0$	0.077	0.160
Grid of 11 freqs.	0.037	0.128
Grid of 26 freqs.	0.093	0.157

P-values reported, based on $B = 1000$ bootstrap samples.

(b) $\text{rank}\left(\text{Re}\left(G_4^2(\lambda)\right)\right)$

Frequency Grid	Null Hypotheses	
	$H_{0,1}$	$H_{0,2}$
$\lambda = 0$	0.659	$--$
Grid of 11 freqs.	0.723	$--$
Grid of 26 freqs.	0.741	$--$

P-values reported, based on $B = 500$ bootstrap samples.

Table 3.6: Number of non-Gaussian components in Blanchard and Perotti (2002)

(a) $\text{rank}\left(\text{Re}\left(G_3^2(\lambda)\right)\right)$

Frequency Grid	Null Hypotheses		
	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
$\lambda = 0$	0.077	0.021	0.377
Grid of 11 freqs.	0.066	0.041	0.729
Grid of 26 freqs.	0.038	0.064	0.450

P-values reported, based on $B = 1000$ bootstrap samples.

(b) $\text{rank}\left(\text{Re}\left(G_4^2(\lambda)\right)\right)$

Frequency Grid	Null Hypotheses		
	$H_{0,1}$	$H_{0,2}$	$H_{0,3}$
$\lambda = 0$	0.034	0.029	0.068
Grid of 11 freqs.	0.042	0.044	0.103
Grid of 26 freqs.	0.024	0.094	0.252

P-values reported, based on $B = 250$ bootstrap samples.

From the previous results, provided that structural shocks are independent across time and components, if the researcher imposes fundamentalness, applying the SIS to both BQ and BP datasets is feasible. A causal and invertible SVARMA model can be identified without imposing external identification restrictions or using proxy variables. Moreover, our results suggest that for the BP dataset, it is feasible to identify a possibly non-fundamental SVARMA model. In the case of the BQ dataset, the most limiting factor is the quite small sample size.

3.6. Concluding Remarks

This chapter aims to design a procedure to determine the number of non-Gaussian shocks in a structural linear VARMA model which is robust to the type of dynamic representation, i.e., whether the structural model is fundamental or non-fundamental. This objective is mainly motivated because knowing the number of non-Gaussian shocks in the structural model allows the researcher to implement an estimation procedure of structural parameters based on the statistical identification approach. In the fundamentalness of the structural model, the requirement is to have at most one Gaussian structural shock; if the researcher does not want to impose the root location, the requirement is that all the structural errors are non-Gaussian distributed.

We generalize the procedure in Guay (2021) by exploiting that the rank of a matrix constructed from the third-order cumulant spectrum of the RF errors reveals the number of skewed or asymmetric structural errors in the SVARMA model. Meanwhile, the rank of an array constructed from the fourth-order cumulant spectrum reveals the number of non-mesokurtic structural shocks in the system. Simulation results show that our procedure correctly estimates the non-Gaussian dimension. Additionally, from a practice point of view, our proposal is intensive computationally, especially if we employ the cumulant spectrum of fourth order or when the dimension of the structural model increases or the sample size is large.

4. CHAPTER III: IDENTIFICATION OF FUNDAMENTAL SVAR MODELS USING HIGHER ORDER CUMULANTS AND SIGN-RESTRICTIONS

The statistical identification strategy of causal structural vector autoregressive models has gained attention recently because it permits the identification of the causal effects of structural shocks without resorting to external identification restrictions. This strategy exploits higher-order information provided structural errors are mutually independent and non-Gaussian distributed. This approach only identifies the structural model up to the signed permutation, which implies a finite set of admissible models. The solution proposed in the literature for selecting a model relies on applying a mechanical procedure. This paper aims to study an alternative scheme based on imposing economic-motivated sign restrictions over the causal effects of structural shocks. Only the permutations matter when a single, strict sign restriction is applied to each shock. A sufficient condition for achieving global identification is that the matrix of contemporaneous effects has a generalized recursive structure. A weaker sufficient condition for accomplishing point identification of a single shock requires imposing as many sign restrictions as the number of endogenous variables in the model and that the system of sign restrictions is sign-solvable. On the other hand, when the model contains more than one Gaussian structural error, identification using higher-order cumulants ensures that the non-Gaussian block remains identified, although the order is unknown.

4.1. Introduction

Afterward the publication of *Macroeconomics and Reality* (Sims (1980)), structural vector autoregressive (hereafter SVAR) models have become one of the most popular toolkits in empirical macroeconomics because they allow to summarize and forecast data, as well as to identify the effects of meaningful economic shocks over outcomes of interest (Stock and Watson (2001)). This latter is one of the aims of several scholars and policymakers. Albeit, such a task is not straightforward. As it has been surveyed by Rubio-Ramirez et al. (2010), SVAR models are not identified since any orthogonal rotation of structural shocks generates the same variance structure of the reduced-form (hereafter RF) errors, i.e., structural shocks and their orthogonal rotations are observationally equivalent. Thus, several strategies relying on external identification restrictions have been proposed in the macroeconometrics literature. Sims (1980) suggests performing a Cholesky decomposition of the RF errors covariance matrix, called as *recursive identification* because it implies sorting the observables variables from the most exogenous to the most endogenous. However, this approach is sensitive to the chosen ordering of endogenous variables. Bernanke and Mihov (1998) proposes introducing zero restrictions in the contemporaneous effects matrix following some theoretical macroeconomic model. And Blanchard and Quah (1989) states a strategy by imposing zeros over the long-run effects of structural shocks. Gali (1992) combine both long-run and short-run zero restrictions.

An alternative identification strategy, proposed by Canova and De Nicolo (2002);

Peersman (2005); Uhlig (2005), is based on setting particular signs to the causal effects (also called responses) of a subset of structural errors over some endogenous variables at specific horizons. Such agnostic restrictions are usually extracted from a theoretical structural economic model[17]. A more economic-based strategy is the so-called *narrative identification approach*, proposed firstly by C. D. Romer and D. H. Romer (1989). Such a strategy is based on a profound analysis of how policies are conducted to measure the policy shock. Alternatively, Braun, Brüggemann, et al. (2017); Stock and Watson (2018) propose to employ such measures as instrumental variables for unobserved structural economic shocks.

Except for the narrative approach, the other identification strategies introduce external information to complement the conditions obtained from second-order moments of the data. The attention to second-order information is supported by explicitly or implicitly assuming the Gaussian behavior of structural shocks. Nonetheless, Gaussianity may not be justified by empirical evidence. Table 4.1 shows the p-values of individual and joint Normality tests for the estimated structural residuals from selected empirical macroeconomic works. The results point out that non-Gaussuian behavior might be supported by data. Theoretical foundations behind the non-Gaussian behavior of structural errors may be varied, e.g., non-linearities and time-varying volatility, among other causes. For instance, Fernández-Villaverde et al. (2018, pp. 5) find that financial frictions "*induce complex, non-linear behavior in aggregate time series, including bi-modalities, skewness, and fat tails, in the ergodic distributions of the variables of interest*".

Additionally, **linde2016challenges** find empirical evidence that forecast errors and some estimated innovations from an empirical model exhibit significant excess of kurtosis, specifically the shocks related to monetary policy and risk premium.

Therefore, although Gaussian behavior of structural errors simplifies the analysis, since it allows to focus only on their serial and cross-correlation, it may be reasonable to assume that economic shocks are non-Gaussian distributed. This feature enables the exploitation of higher-order information for statistically identifying structural parameters without incorporating external restrictions or information. For instance, Herwartz and Lütkepohl (2014); Lütkepohl (2005); Rigobon (2003) employ unconditional or conditional heteroskedasticity of structural shocks for obtaining additional moment conditions that are employed to attain identification up to sign-permutation of the parameters in SVAR model with time-invariant coefficients. On the other hand, Gouriéroux et al. (2017); Gouriéroux et al. (2020); Lanne, Meitz, et al. (2017); Maxand (2020) exploit the non-Gaussianity of structural shocks explicitly for identifying parameters of a causal SVAR

[17]For instance, it is common to obtain some sign-restrictions for some shocks using RBC or DSGE models

Table 4.1: Normality Tests of Residuals (p-values)

	Cramer-von-Mises	*Watson*	*Anderson-Darling*	*Jarque-Bera*
	Stock and Watson (2001)			
Cost shock	0.003	0.002	0.002	0.037
Demand shock	0.006	0.007	0.007	0.000
MP shock	0.000	0.000	0.0000	0.000
Joint Normality		0.000		
	Christiano, Eichenbaum and Evans (1996)			
Supply shock	0.624	0.579	0.639	0.817
Cost shock	0.938	0.929	0.976	0.935
Commodity shock	0.758	0.718	0.715	0.925
Monetary shock 1	0.033	0.026	0.030	0.000
Monetary shock 2	0.001	0.001	0.000	0.000
Monetary shock 3	0.001	0.001	0.000	0.000
Demand shock	0.001	0.001	0.001	0.026
Joint Normality		0.000		
	Blanchard and Quah (1989)			
Supply shock	0.187	0.212	0.073	0.012
Demand shock	0.002	0.002	0.001	0.027
Joint Normality		0.120		

P-values reported.

model. Their identification relies on the work of Comon (1994), who requires at most one Gaussian shock in the system for obtaining identification (up to signed permutation). Alternatively, Lanne and Luoto (2019); Velasco (2022b) characterize non-Gaussianity by its third and fourth-order moments, introducing additional moment conditions for identification purposes.

The statistical identification strategy (hereafter SIS) permits identifying a finite set of feasible signed permutations of the structural shocks that are observationally equivalent. This partial identification result affects the type of inference that can be performed. Identification up to signed permutation is enough for learning about elements invariant to permutations, such as the number of non-Gaussian elements or the rank of the contemporaneous effect matrix. Nevertheless, there are elements in a structural model that are not permutation invariant, e.g., the causal effects of a structural error over a subset of endogenous variables. Thus, testing these elements requires either having point identification of structural parameters or performing the testing exercise over all admissible permutations. Furthermore, in the context of a finite identified set, point identification requires selecting a particular model, such as choosing an admissible permutation. In the literature, some procedures have been proposed to accomplish this. For instance, Hallin and Mehta (2015) proposes a selection procedure based on assuming that the contemporaneous effects' matrix belongs to a broad class of matrices, which are equivalent under permutation and scaling.

When a d-dimensional SVAR model is statistically identified, the number of admissible signed permutations equals $2^d \times d!$, whose amount increases with the number of endogenous variables in the system. Thus, it is crucial to have a permutation selection device. This paper investigates whether sign restrictions over impulse responses can be applied to choosing a permutation instead of a mechanical procedure. Specifically, under which conditions sign restrictions may lead to point identification of the structural parameters in the model. We find that: when each shock is subject to a single, strict sign restriction, the number of admissible permutations reduces to $d!$. This result means the following: once the sign of structural shocks is fixed, we only need to care about permutations. A stronger sufficient condition is that the matrix of contemporaneous effects has a *generalized recursive* structure to obtain point identification with single sign restrictions to each shock. Besides, when all shocks are asymmetric, it is possible to relate the recursive structure of the matrix of contemporaneous effects. A weaker condition for point identification of a single structural shock can be obtained using multiple sign restrictions. In this case, we get that the number of strict sign restrictions must be, at least, as many as the endogenous variables in the model, and the inequality system implied by the sign restrictions has to be sign-solvable. For identifying multiple shocks, it is found that the condition for a single shock identification must hold recursively. Additionally, we discuss the necessary conditions for identifying multiple shocks with sign restrictions using only second-order information since the sufficient conditions established in the literature that assures the existence of identified sets for single shocks are not enough when multiple structural errors are involved. Besides, we discuss identifying a structural model in the system's presence of multiple Gaussian shocks.

This chapter is organized as follows: Section 4.2 presents the structural model and the basics of sign restrictions; Section 4.3 explains how cumulants are defined and states the main identification results based on higher-order cumulants. Also, the role of the mutual independence assumption is discussed. Section 4.4 provides the concluding remarks.

4.2. Model and Assumptions

Let $\boldsymbol{y}_t$ be a d-dimensional, observable random vector. $\boldsymbol{y}_t$ is the stationary solution of the following stochastic difference equation:

$$\boldsymbol{\Gamma}_0 \boldsymbol{y}_t = \boldsymbol{D} + \boldsymbol{\Gamma}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t. \tag{4.1}$$

The matrix coefficients $\{\boldsymbol{\Gamma}_i\}_{i=0}^p$ are time-invariant and squared; $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_p$ are non-singular; $\boldsymbol{D}$ is a d-dimensional column vector of constants. $\boldsymbol{\varepsilon}_t$ is a d-dimensional random vector, known as the structural shocks, and it is assumed they are economically meaningful. Equation (4.1) can be expressed as

$$\boldsymbol{y}_t = \boldsymbol{C} + \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\varepsilon}_t = \boldsymbol{C} + \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{y}_{t-k} + \boldsymbol{B}\boldsymbol{\varepsilon}_t, \tag{4.2}$$

where $\boldsymbol{A}_i = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_i$ for $i = 1,2,\dots p$, $\boldsymbol{C} = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{D}$ and $\boldsymbol{B} = \boldsymbol{\Gamma}_0^{-1}$. Throughout this paper, equation (4.2) is called the structural VAR(p) model. In this paper, we are not dealing with dynamic identification. Therefore we impose that the polynomial $\boldsymbol{A}(L) = \boldsymbol{I}_d - \sum_{k=1}^{p} \boldsymbol{A}_k L^k$ satisfies

$$\det(\boldsymbol{A}(z)) \neq 0 \quad \forall\ |z| \leq 1. \tag{4.3}$$

The condition in (4.3) implies that the structural model admits a causal or fundamental representation. This prerequisite implies that the Hilbert space generated by the past and current values of observables, $\mathscr{H}^t(\mathbf{y}) = \mathscr{H}\left\{\boldsymbol{y}_s\ :\ s \leq t\right\}$, is equivalent to the space spanned by the past and present values of unobserved structural shocks, $\mathscr{H}^t(\boldsymbol{\varepsilon}) = \mathscr{H}\{\boldsymbol{\varepsilon}_s\ :\ s \leq t\}$. The requirement in (4.3) assures the dynamic identification of the structural model. Otherwise, there may exist a non-fundamental representation that is observationally equivalent.

The description of the model in (4.2) is not complete until the structure of $\boldsymbol{\varepsilon}_t$ is established. The following assumption states the main structure of the structural errors. Let $\mathscr{I}$ denotes the set of indices $\{1,2,\dots,d\}$, $\mathscr{I}_{ng} \subseteq \mathscr{I}$ is the index set of non-Gaussian structural shocks with $\left|\mathscr{I}_{ng}\right| = d_{ng}$.

Assumption 4.1. *The structural shocks, $\boldsymbol{\varepsilon}_t$, in the model satisfy*

(a) $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ *is an ergodic, serially uncorrelated sequence.*

(b) $\boldsymbol{\varepsilon}_t$ *has mutually independent components for all t.*

(c) $Cum_n\left[\varepsilon_{i,t}\right] \neq 0$ *for some $n \in \{3,4\}$ and all $i \in \mathscr{I}_{ng}$, where $Cum_n(x)$ denotes the marginal cumulant of order n of variable x.*

(d) $\mathbb{E}[\boldsymbol{\varepsilon}_t] = 0$, $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \boldsymbol{I}_d$, $\mathbb{E}[\|\boldsymbol{\varepsilon}_t\|^4] < \infty$ *for all* t.

Assumption 4.1.(a) is the minimum requirement for structural errors. Although this assumption allows for potential higher-order dependence of the structural shocks, this does not represent a significant threat to identification, given that the dynamic identification problem is not a concern. Assumption 4.1.(b) is necessary since Gaussianity of structural shocks is abandoned, otherwise even though errors are not contemporaneously correlated, higher-order dependence may exist, and non-linear forms of the others could predict some structural error. On the other hand, Assumption 4.1.(c) states that at least one non-Gaussian structural shock exists in the system with finite third or fourth-order cumulant. This assumption is insufficient for achieving statistical identification, though below, a refinement of this condition will be done to assure statistical identification. Finally, Assumption 4.1.(d) is relatively standard in the literature but includes the finiteness of at most fourth-order moments because, as it will be shown later, the identification of structural parameters is based on higher-order cumulant conditions.

Calling $\boldsymbol{u}_t = \boldsymbol{B}\boldsymbol{\varepsilon}_t$, the structural model in (4.2) lead to the RF-VAR(p) model:

$$\boldsymbol{A}(L)\boldsymbol{y}_t = \boldsymbol{C} + \boldsymbol{u}_t. \tag{4.4}$$

The vector $\boldsymbol{u}_t$ is known as the RF error. $\boldsymbol{A}(L) = \boldsymbol{I}_d - \sum_{k=1}^{p} \boldsymbol{A}_k L^k$ is the AR polynomial, where L is the lag or back-shift operator, i.e. $L^s \boldsymbol{y}_t = \boldsymbol{y}_{t-s}$. Condition in (4.3) implies that the parameters of the RF model, $\{\boldsymbol{A}_j\}_{j=1}^{p}$ and $\mathbb{V}\mathrm{ar}(\boldsymbol{u}_t)$, are identified.

4.2.1. Impulse Response Analysis

Given the structural model in (4.2) and fundamentalness condition (4.3), $\boldsymbol{y}_t$ admits a infinite, causal moving-average representation as follows

$$\boldsymbol{y}_t = \boldsymbol{\Psi}(L)(\boldsymbol{C} + \boldsymbol{u}_t) = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{B}\boldsymbol{\varepsilon}_{t-j}, \tag{4.5}$$

where $\boldsymbol{\Psi}(L) = \boldsymbol{A}^{-1}(L) = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_j L^j$ with $\boldsymbol{\Psi}_0 = \boldsymbol{I}_d$, and $\boldsymbol{\mu} = \boldsymbol{\Psi}(1)\boldsymbol{C}$. Based on equation (4.5), the causal effect of a change in the l-th structural shock over the j-th endogenous variables at horizon $t+h$, also known as the impulse response (hereafter IR) of $y_{j,t+h}$ to $\varepsilon_{l,t}$, is equal to

$$\mathrm{IR}(y_{j,t+h}|\varepsilon_{l,t}) := \mathbb{E}[y_{j,t+h}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1},\varepsilon_{l,t}=1] - \mathbb{E}[y_{j,t+h}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1},\varepsilon_{l,t}=0] = [\boldsymbol{\Psi}_h]_{(j)}\boldsymbol{b}_l = \boldsymbol{e}_j'\boldsymbol{\Psi}_h\boldsymbol{B}\boldsymbol{e}_l, \tag{4.6}$$

where $\boldsymbol{e}_j'$ and $\boldsymbol{e}_l$ select the j-th row and l-th column of a matrix, respectively.

When an orthogonal rotation is applied to the vector of structural innovations, $\boldsymbol{\varepsilon}_t^\star = \boldsymbol{Q}\boldsymbol{\varepsilon}_t$, then the matrix of contemporaneous effects is rotated as well, $\boldsymbol{B}^\star = \boldsymbol{B}\boldsymbol{Q}'$. Rotation matrix $\boldsymbol{Q}$ satisfies $\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}_d$. Thus, an equivalent expression to (4.6) for a change in the l-th rotated structural shock is

$$\mathrm{IR}(y_{j,t+h}|\varepsilon_{l,t}^\star) = [\boldsymbol{\Psi}_h]_{(j)}\boldsymbol{b}_l^\star = \boldsymbol{e}_j'\boldsymbol{\Psi}_h\boldsymbol{B}\boldsymbol{Q}'\boldsymbol{e}_l = \boldsymbol{e}_j'\boldsymbol{\Psi}_h\boldsymbol{B}\boldsymbol{q}_l, \tag{4.7}$$

where $\boldsymbol{q}_l$ is the l-th column of $\boldsymbol{Q}'$.

By comparing expressions in (4.6) and (4.7), it is evident that the causal effect or impulse response function (hereafter IRF) of structural shocks is not invariant to orthogonal rotations. Besides, from this comparison, the static identification problem can be interpreted as follows: although any orthogonal rotation of structural shocks generates the same variance structure of the RF errors because $\mathbb{V}\mathrm{ar}(\boldsymbol{u}_t) = \mathbb{V}\mathrm{ar}(\boldsymbol{B}\boldsymbol{\varepsilon}_t) = \mathbb{V}\mathrm{ar}(\boldsymbol{B}^\star\boldsymbol{\varepsilon}_t^\star)$, each orthogonal rotation implies different paths for IRFs of the structural shocks, generally.

Sign Restrictions

As Uhlig (2005) and Arias et al. (2018) explain, sign restrictions do not impose a particular value but a direction to the IR in (4.6). For example, it could be imposed that the causal effect of the l-th structural shock over the first endogenous variable at horizon $t+2$ is non-positive, then

$$\mathrm{IR}(y_{1,t+2}|\varepsilon_{l,t}) = e_1'\boldsymbol{\Psi}_2\boldsymbol{B}\boldsymbol{e}_l \leq 0 \quad \Leftrightarrow \quad (-1)\left(e_1'\boldsymbol{\Psi}_2\boldsymbol{B}\boldsymbol{e}_l\right) \geq 0.$$

When the researcher imposes more than a single sign-restriction over the responses to changes in $\varepsilon_{l,t}$, the sign-restrictions form a system of inequalities, which can be represented as follows: let $\mathscr{C}(l)$ the set of indexes (h,j) of endogenous variables and time horizons whose responses to the l-th structural shock are restricted. Specifically, $\mathscr{C}(l)$ is defined as $\mathscr{C}: \{1,\ldots,d\} \rightrightarrows \{\mathbb{N}\cup\{0\}\}\times\{1,\ldots,d\}$[18] and its cardinality is $|\mathscr{C}(l)| = r_l < \infty$ represents the number of responses to the l-th structural error that are sign-restricted. Since $\mathscr{C}(l)$ is a finite discrete set, we can index each pair as (h_i, j_i) for $i = 1,\ldots,r_l$. Hence, for a pair $(h_i, j_i) \in \mathscr{C}(l)$, the sign-restriction to the response of j_i-th endogenous variable at

[18] $\mathscr{C}(l)$ is a subset of $\{\{\mathbb{N}\cup\{0\}\}\times\{1,\ldots,d\}\}$. In general, $\mathscr{C}(l)$ may have infinite elements. However, in this work, we take finite $\mathscr{C}(l)$ for each l structural shock.

horizon $t+h_i$ to a change in

$$\mathrm{IR}(y_{j_m,t+h_m}|\varepsilon_{l,t}) = s_m\left(e'_{j_m}\boldsymbol{\Psi}_{h_m}\boldsymbol{B}\boldsymbol{e}_l\right) \geq 0 \quad \forall\,(h_m,j_m)\in\mathscr{C}(l),\; m=1,\dots,r_l,$$

where $s_m \in \{-1,1\}$, with $m \in \{1,\dots,r_l\}$, represents the direction for the sign-restriction. For instance, if $s_m = -1$, we are imposing the effect of structural shock $\varepsilon_{l,t}$ is negative over the response of $y_{j_m,t+h_m}$. Gathering each of these restrictions in a vector, we can write the previous system more compactly

$$\mathrm{IR}(y_{\boldsymbol{j}(l),t+\boldsymbol{h}(l)}|\varepsilon_{l,t}) = \begin{bmatrix} \mathrm{IR}(y_{j_1,t+h_1}|\varepsilon_{l,t}) \\ \mathrm{IR}(y_{j_2,t+h_2}|\varepsilon_{l,t}) \\ \vdots \\ \mathrm{IR}(y_{j_{r_l},t+h_{r_l}}|\varepsilon_{l,t}) \end{bmatrix} = \begin{bmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_{r_l} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}'_{j_1}\boldsymbol{\Psi}_{h_1} \\ \boldsymbol{e}'_{j_2}\boldsymbol{\Psi}_{h_2} \\ \vdots \\ \boldsymbol{e}'_{j_{r_l}}\boldsymbol{\Psi}_{h_{r_l}} \end{bmatrix} \boldsymbol{B}\boldsymbol{e}_l = \mathbf{S}_l\,\{\boldsymbol{\Psi}(l)\}'\,\boldsymbol{b}_l \geq 0, \tag{4.8}$$

where $\mathbf{S}_l = \mathrm{diag}(s_1,\dots,s_{r_l})$ contains the signs of the restrictions, $\boldsymbol{\Psi}(l) = [\psi_1,\dots,\psi_{r_l}]$ with $\psi_i = \boldsymbol{\Psi}'_{h_i}\boldsymbol{e}_{j_i}$ and $\boldsymbol{\Psi}_{h_i}$ is the h_i-th matrix coefficient from the infinite MA representation in (4.5).

If orthogonal rotated structural shocks are considered, the sign restrictions for the l-th structural shock are

$$\mathbf{S}_l\,\{\boldsymbol{\Psi}(l)\}'\boldsymbol{B}\boldsymbol{Q}'\boldsymbol{e}_l = \mathbf{S}_l\,\{\boldsymbol{\Psi}(l)\}'\boldsymbol{B}\boldsymbol{q}_l \geq 0,$$

where $\boldsymbol{q}_l$ is the l-th column of an orthogonal rotation matrix $\boldsymbol{Q}'$.

Now, in case multiple structural shocks are subject to sign restrictions, we can express the system of sign restrictions as follows: let $\mathscr{I}_{SR} \subset \{1,\dots,d\}$ denotes the index set of structural errors subject to sign-restrictions, with $|\mathscr{I}_{SR}| = d_{SR} \leq d$ represents the total number of sign-restricted shocks. Each shock is $\mathscr{I}_{SR}$ has associated a set of sign-restrictions like in (4.8) and collecting each of them in a vector, the system of sign-restrictions is

$$\begin{bmatrix} \mathrm{IR}(y_{\boldsymbol{j}(l_1),t+\boldsymbol{h}(l_1)}|\varepsilon_{l_1,t}) \\ \mathrm{IR}(y_{\boldsymbol{j}(l_2),t+\boldsymbol{h}(l_2)}|\varepsilon_{l_2,t}) \\ \vdots \\ \mathrm{IR}(y_{\boldsymbol{j}(l_{d_{SR}}),t+\boldsymbol{h}(l_{d_{SR}})}|\varepsilon_{l_{d_{SR}},t}) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{l_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{l_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{S}_{l_{d_R}} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}'_{l_1}\otimes\{\boldsymbol{\Psi}(l_1)\}' \\ \boldsymbol{e}'_{l_2}\otimes\{\boldsymbol{\Psi}(l_2)\}' \\ \vdots \\ \boldsymbol{e}'_{l_{d_{SR}}}\otimes\{\boldsymbol{\Psi}(l_{d_{SR}})\}' \end{bmatrix} \mathrm{vec}(\boldsymbol{B}) = \mathbf{S}_{SR}\overline{\boldsymbol{\Psi}}'\,\mathrm{vec}(\boldsymbol{B}) \geq 0, \tag{4.9}$$

where $\mathbf{S}_{SR} = \mathrm{diag}\left(\mathbf{S}_{l_1} \quad \dots \quad \mathbf{S}_{l_{d_{SR}}}\right)$ is a block diagonal matrix, whose blocks contain the signs for restrictions to l-th shock, and $\overline{\boldsymbol{\Psi}} = \left[\boldsymbol{e}_{l_1} \otimes \boldsymbol{\Psi}(l_1) \quad \dots \quad \boldsymbol{e}_{l_{d_{SR}}} \otimes \boldsymbol{\Psi}(l_{d_{SR}})\right]$.

According to Uhlig (2005) or Arias et al. (2018); Rubio-Ramirez et al. (2010), a valid value for matrix $\boldsymbol{B}$ must satisfy the system of sign-restrictions in (4.9). Identification with sign-restrictions works as follows: if an initial value $\boldsymbol{B}$ violates the system in (4.9), this is rotated, i.e., $\boldsymbol{B}^\star = \boldsymbol{B}\boldsymbol{Q}'$, such that system (4.9) holds. Therefore, it is clear that sign-restrictions identify many different models, each of them associated with a particular rotated $\boldsymbol{B}^\star$, whose are consistent with the sign-restrictions to shocks in $\mathscr{I}_{SR}$.

Additionally, the sign restrictions are a linear system of inequalities of the form $\varsigma := \{\mathbf{A}'\mathbf{x} \leq 0\}$. The solution set of this system is $F(\varsigma)$ and we define $F^{\perp}(\varsigma) := \{\mathbf{x} \mid \mathbf{x}'\mathbf{y} = 0 \text{ for some } \mathbf{y} \in F(\varsigma)\}$. Besides, $F^{\circ}(\varsigma)$ denotes the interior set of $F(\varsigma)$. For instance, the sign-restrictions associated to shock $\varepsilon_{l,t}$ is $\varsigma_l^{SR} = \{-\boldsymbol{S}_l\boldsymbol{\Psi}'(l)\boldsymbol{B}\mathbf{x} \leq 0\}$ and $F(\varsigma_l^{SR})$ represents the set of vectors consistent with the sign-restrictions. In case of restrictions to multiple shocks, the system of inequalities is $\varsigma^{SR} = \{-\mathbf{S}_{SR}\overline{\boldsymbol{\Psi}}' \operatorname{vec}(\boldsymbol{B}) \leq 0\}$. Also, let $\mathscr{A}_1 \times \mathscr{A}_2$ be the Cartesian product of sets $\mathscr{A}_1$ and $\mathscr{A}_2$; then $GS(\mathscr{A}_1 \times \mathscr{A}_2)$ is the set of tuples obtained by applying Gram-Schmidt algorithm to each element of $\mathscr{A}_1 \times \mathscr{A}_2$. For instance, consider that $(\mathbf{x}_1, \mathbf{x}_2) \in \mathscr{A}_1 \times \mathscr{A}_2$, then $(\mathbf{x}_1, \tilde{\mathbf{x}}_2)$ is the element of $GS(\mathscr{A}_1 \times \mathscr{A}_2)$ with $\tilde{\mathbf{x}}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\|^2}\mathbf{x}_1$.

An alternative way to state that a matrix $\boldsymbol{B}$ satisfies system (4.9) follows Granziera et al. (2018). Let $\mathscr{G} : \mathbb{R}^{d^2} \mapsto \mathbb{R}_+$ be a distance mapping, defined as

$$\mathscr{G}(\boldsymbol{B}) = \min_{\boldsymbol{v} \in \mathbb{R}_+^{\overline{d}}} \left\| \mathbf{S}_{SR}\overline{\boldsymbol{\Psi}}' \operatorname{vec}(\boldsymbol{B}) - \boldsymbol{v} \right\|^2,$$

where $\overline{d} = \sum_{j=1}^{d_{SR}} r_{l_j}$, denotes the slackness vector of the system of sign-restrictions in (4.9). Thus, it is evident that $\boldsymbol{B}$ satisfies the system of sign-restrictions in (4.9) if and only if $\mathscr{G}(\boldsymbol{B}) = 0$.

4.3. Identification using Cumulants conditions

This section discusses the identification of the structural model using cumulant conditions. Also, it is discussed the conditions under which sign restrictions may lead to achieving point identification. Nonetheless, before continuing, let us introduce some notation and concepts to avoid confusion. $\mathscr{B}(d)$ is the space of real-valued, squared, non-singular matrices of size d. $\boldsymbol{B}_0$ denotes the "true" value of contemporaneous effects matrix; while $\boldsymbol{B}_1$, a solution to the population cumulant conditions. In general, $\boldsymbol{B}_0 \neq \boldsymbol{B}_1$. $\mathscr{B}_{0,n}$ denotes the set of all solutions to the population cumulant conditions up to order n, while $\mathscr{B}_{0,n}^{Q}$ is equivalent to $\mathscr{B}_{0,n}$, but generated by orthogonal rotations $\boldsymbol{Q}$ of $\boldsymbol{B}_1$. Besides, $\mathscr{B}_{0,n}^{Q,SR}$ denotes the sign-restricted set, i.e., the set of all orthogonal rotations of any $\boldsymbol{B}_1$

that satisfy the sign-restrictions system in (4.9). It is clear that $\mathscr{B}^{Q,SR}_{0,n} \subset \mathscr{B}^{Q}_{0,n} \subseteq \mathscr{B}_{0,n}$. Finally, since cumulants are employed for identification purposes, let us explain what is a cumulant of order n for a random element.

4.3.1. Cumulants

Let $\phi_\varepsilon(\boldsymbol{\tau}) = \mathbb{E}\left[e^{i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t}\right]$ be the characteristic function of a strict stationary, random vector $\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_d)' \in \mathbb{R}^d$ is a constant vector. Unlike the moment generating function, $\phi_\varepsilon(\boldsymbol{\tau})$ is well defined for any $\boldsymbol{\tau}$ even though $\boldsymbol{\varepsilon}_t$ does not possess any finite moment. The cumulant generating function is defined as

$$\boldsymbol{\kappa}_\varepsilon(\boldsymbol{\tau}) = \log(\phi_\varepsilon(\boldsymbol{\tau})). \tag{4.10}$$

A cumulant for an n-tuple of elements of vector $\boldsymbol{\varepsilon}_t$, or a cumulant of order n, is

$$\operatorname{Cum}(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_n,t}) = \left.\frac{\partial^n \boldsymbol{\kappa}_\varepsilon(\boldsymbol{\tau})}{\partial\tau_{j_1}\dots\partial\tau_{j_n}}\right|_{\boldsymbol{\tau}=0}, \tag{4.11}$$

where $(j_1, \dots, j_n) \in \sigma_n(\{1, \dots, d\})$ and $\sigma_n(\{1, \dots, d\})$ represent the set of all permutations of the elements in the index set $\{1, \dots, d\}$ taken in n-tuples.[19] This implies there are d^n cumulants of order n constructed with the elements of $\boldsymbol{\varepsilon}_t$. Since $\operatorname{Cum}(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_n,t})$ is invariant to permutations. Thus, repeated values exist among the d^n cumulants of order n. The number of non-repetated cumulants of order n is equal to $\frac{(d+n-1)!}{n!(d-1)!}$. Moreover, it can be distinguished between marginal and non-marginal cumulants. Specifically, the marginal cumulant of order n for the j-th element in the vector $\boldsymbol{\varepsilon}_t$ is $\kappa^\varepsilon_{n,j} := \operatorname{Cum}(\varepsilon_{j,t}, \dots, \varepsilon_{j,t})$.

When $n = 3$, the values $\{\kappa^\varepsilon_{3,j}\}_{j=1}^d$ are known as the skewness coefficients; while for $n = 4$, $\{\kappa^\varepsilon_{4,j}\}_{j=1}^d$ are called the excess-kurtosis coefficients. Marcinkiewicz (1939) shows that the Gaussian distribution is the only one whose cumulant generating function is a quadratic polynomial. Thus, non-Gaussian distributions have, in general, non-zero cumulants of order $n \geq 3$. Petrochilos and Comon (2006) demonstrate that it is possible to find non-Gaussian distributions with zero-cumulant up to order 3 or 4, although higher-order cumulants must be different from 0. Jammalamadaka et al. (2006) shows that a cumulant of order n can be expressed in terms of moments when these latter exist.

$$\mathbb{C}\mathrm{um}\left(\varepsilon_{j_1,t}, \dots, \varepsilon_{j_n,t}\right) = \sum_{\mathfrak{P}} (|\mathfrak{P}| - 1)!(-1)^{|\mathfrak{P}|-1} \prod_{\mathfrak{p}\in\mathfrak{P}} \mathbb{E}\left(\prod_{j_m\in\mathfrak{p}} \varepsilon_{j_m,t}\right), \tag{4.12}$$

[19]For instance, let the set of numbers $\{1,2\}$. $\sigma_3(\{1,2\}) = \{(1,1,1);(1,1,2);(1,2,1);(1,2,2);(2,1,1);(2,1,2);(2,2,1);(2,2,2)\}$.

where $\mathfrak{P}$ is a partition of $\{j_1,\ldots,j_n\}$, $|\mathfrak{P}|$ is the number of parts in partition $\mathfrak{P}$, and $\mathfrak{p}$ is a block of partition $\mathfrak{P}$. For instance, if $n = 2$ and we choose the duplet $(\varepsilon_{j_1,t},\varepsilon_{j_2,t})$, $\mathbb{C}\mathrm{um}(\varepsilon_{1,t},\varepsilon_{2,t}) = \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}] = \mathbb{C}\mathrm{ov}(\varepsilon_{j_1,t},\varepsilon_{j_2,t})$. When $n = 3$ and the chosen triplet is $(\varepsilon_{j_1,t},\varepsilon_{j_2,t},\varepsilon_{j_3,t})$, $\mathbb{C}\mathrm{um}(\varepsilon_{1,t},\varepsilon_{2,t},\varepsilon_{3,t}) = \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}\varepsilon_{j_3,t}]$. However, with $n = 4$ and the 4-tuple is $(\varepsilon_{j_1,t},\varepsilon_{j_2,t},\varepsilon_{j_3,t},\varepsilon_{j_4,t})$, $\mathbb{C}\mathrm{um}\left(\varepsilon_{j_1,t},\varepsilon_{j_2,t},\varepsilon_{j_3,t},\varepsilon_{j_4,t}\right) = \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}\varepsilon_{j_3,t}\varepsilon_{j_4,t}] - \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_2,t}]\mathbb{E}[\varepsilon_{j_3,t}\varepsilon_{j_4,t}] - \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_3,t}]\mathbb{E}[\varepsilon_{j_2,t}\varepsilon_{j_4,t}] - \mathbb{E}[\varepsilon_{j_1,t}\varepsilon_{j_4,t}]\mathbb{E}[\varepsilon_{j_2,t}\varepsilon_{j_3,t}]$.

We can gather all cumulants of order n in a column vector $\boldsymbol{\kappa}_n^{\boldsymbol{\varepsilon}} = \left[\mathbb{C}\mathrm{um}\left(\varepsilon_{j_1,t},\ldots,\varepsilon_{j_n,t}\right)\right]_{(j_1,\ldots,j_n)\in\sigma_n(\{1,\ldots,d\})}$, whose dimension is, evidently, d^n. When $\boldsymbol{\varepsilon}_t$ satisfies Assumption 4.1 (a) and (b), then all the non-marginal cumulants are zero. The vector $\boldsymbol{\kappa}_n^{\boldsymbol{\varepsilon}}$ has a particular structure. Let $\mathbf{v}_n^{\boldsymbol{\varepsilon}}$ be a $d^{n-1}\times d$ matrix, such that $\boldsymbol{\kappa}_n^{\boldsymbol{\varepsilon}} = \mathrm{vec}(\mathbf{v}_n^{\boldsymbol{\varepsilon}})$ with

$$\mathbf{v}_n^{\boldsymbol{\varepsilon}} = \left[\kappa_{n,1}^{\varepsilon}\boldsymbol{e}_1^{\otimes n-1}, \kappa_{n,2}^{\varepsilon}\boldsymbol{e}_2^{\otimes n-1}, \ldots, \kappa_{n,d}^{\varepsilon}\boldsymbol{e}_d^{\otimes n-1}\right] \quad \text{for } n \geq 2,$$

where $\boldsymbol{e}_j^{\otimes n-1}$ denote the $(n-1)$-th Kronecker power. To see a simple example of this structure, consider a bivariate SVAR $d = 2$ and third-order cumulants, $n = 3$, thus

$$\mathbf{v}_3^{\boldsymbol{\varepsilon}} = \begin{bmatrix} \kappa_{3,1}^{\varepsilon} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \kappa_{3,2}^{\varepsilon} \end{bmatrix}.$$

Since the RF errors are linear combinations of structural shocks, i.e., $\boldsymbol{u}_t = \boldsymbol{B}\boldsymbol{\varepsilon}_t$, and according to Jammalamadaka et al. (2006), the cumulants of $\boldsymbol{u}_t$ and $\boldsymbol{\varepsilon}_t$ are connected as follows

$$\boldsymbol{g}_n\left(\boldsymbol{B}, \{\kappa_{s,j}^{\boldsymbol{\varepsilon}}\}_{s=3,j=1}^{n,d}\right) := \boldsymbol{\kappa}_n^{\boldsymbol{u}} - \boldsymbol{B}^{\otimes n}\boldsymbol{\kappa}_n^{\boldsymbol{\varepsilon}} = 0,\ n \geq 2. \tag{4.13}$$

In this paper, we characterize non-Gaussian behavior through possibly non-zero third or fourth-order cumulants, i.e., $n \leq 4$ because it might be troublesome to provide economic interpretation for 5-th or higher-order cumulants. The system in (4.13) is called the *population cumulant conditions* (hereafter PCC) up to order n. This system is satisfied, evidently, by $\boldsymbol{B}_0$ and the generic solution $\boldsymbol{B}_1$. Also, the cumulant conditions in (4.13) only consider the marginal higher order cumulants since $\boldsymbol{\kappa}_2^{\boldsymbol{\varepsilon}} = \mathrm{vec}(\boldsymbol{I}_2)$. The higher order non-zero, marginal cumulants of non-Gaussian shocks in (4.13) for $n \geq 3$ can be retrieved in terms of $\boldsymbol{B}$ and $\boldsymbol{\kappa}_n^{\boldsymbol{u}}$. Thus, we can concentrate $\boldsymbol{g}_n(.)$ in terms of $\boldsymbol{B}$ only, $\tilde{\boldsymbol{g}}_n(\boldsymbol{B})$. Now, let us define a population criteria using the PCC up to order n:

$$\mathscr{M}^n(\boldsymbol{B}) = \frac{1}{2}\sum_{m=2}^{n}\tilde{\boldsymbol{g}}_m'(\boldsymbol{B})W_m\tilde{\boldsymbol{g}}_m(\boldsymbol{B}), \tag{4.14}$$

where W_m is positive definite weighting matrix. It is clear that $(\boldsymbol{B}_1)$ solves the PCC if and only if $\tilde{\mathscr{M}}^n(\boldsymbol{B}_1) = 0$.

4.3.2. Identification with Second Order Cumulants

The set of all matrices $\boldsymbol{B}$ that solves the PCC in (4.13) up to order $n = 2$,

$$\mathscr{B}_{0,2} = \left\{\boldsymbol{B} \in \mathscr{B}(d) \,|\, \tilde{\mathscr{M}}^2(\boldsymbol{B}) = 0\right\}, \tag{4.15}$$

also known as the *identified set*, contains a continuum of admissible values for $\boldsymbol{B}$, implying that is an infinite set. A prominent element of this set is Σ_C, the Cholesky decomposition[20] of $\Sigma_u = \mathbb{E}[\boldsymbol{u}_t\boldsymbol{u}_t']$. Setting $\boldsymbol{B}_1 = \Sigma_C$, the rest of elements of $\mathscr{B}_{0,2}$ can be found as $\boldsymbol{B}^\star = \boldsymbol{B}_1\boldsymbol{Q}'$ with $\boldsymbol{Q}$ an orthonormal matrix. Thus, the identified set using orthogonal rotations is (see Appendix **??** for a formal proof of this statement):

$$\mathscr{B}_{0,2}^{\boldsymbol{Q}} = \left\{\boldsymbol{B} \,|\, \tilde{\mathscr{M}}^2(\boldsymbol{B}) = 0 \text{ for } \boldsymbol{B} = \Sigma_C\boldsymbol{Q}', \boldsymbol{Q} \in \mathscr{O}(d)\right\},$$

where $\mathscr{O}(d)$ is the space of orthonormal square matrices of size d.

From this discussion, it is evident that $\mathscr{B}_{0,2}^{\boldsymbol{Q}} = \mathscr{B}_{0,2}$. Also, $\mathscr{B}_{0,2}^{\boldsymbol{Q}}$ can be seen as the image set of the linear transformation $\mathscr{T} : \mathscr{O}(d) \to \mathscr{B}(d)$ with $\mathscr{T} = \Sigma_C\boldsymbol{Q}'$. Then, $\mathscr{B}_{0,2}^{\boldsymbol{Q}}$ is just an homeomorphism of the set $\mathscr{O}(d)$ and inherits some of its properties, such compactness. Finally, the identification problem in the standard macroeconometrics literature can be seen in this way: using only second-order information captured by the cumulants of second order, the identified set contains any orthogonal rotation of Σ_C. This result means the RF errors are identified, not structural shocks.

Sign restrictions can shrink the set $\mathscr{B}_{0,2}^{\boldsymbol{Q}}$. The identified set using second-order cumulants and sign restrictions over the responses to structural shocks in $\mathscr{I}_{SR}$ is

$$\mathscr{B}_{0,2}^{\boldsymbol{Q},SR} = \left\{\boldsymbol{B} \,|\, \tilde{\mathscr{M}}^2(\boldsymbol{B}) = 0,\, \mathbf{S}_{SR}\overline{\boldsymbol{\Psi}}'\operatorname{vec}(\boldsymbol{B}) \geq \mathbf{0}, \text{ for } \boldsymbol{B} = \Sigma_C\boldsymbol{Q}' \text{ with } \boldsymbol{Q} \in \mathscr{O}(d)\right\} = \mathscr{B}_{0,2}^{\boldsymbol{Q}} \cap F(\varsigma^{SR}), \tag{4.16}$$

where $\varsigma^{SR} = \left\{-\mathbf{S}_{SR}\overline{\boldsymbol{\Psi}}'\operatorname{vec}(\Sigma_C\boldsymbol{Q}') \leq 0\right\}$. In case a sign-restrictions is imposed only for identifying $\varepsilon_{l,t}$, the system of inequalities that are employed is ς_l^{SR}. Under this scenario, the rest of the shocks are unrestricted. Therefore, the attention is on under which conditions $F(\varsigma_l^{SR})$ is non-empty. However, when multiple structural shocks are subject to sign restrictions, the existence of each $F(\varsigma_l^{SR})$ does not guarantee the existence of identified set $\mathscr{B}_{0,2}^{\boldsymbol{Q},SR}$. Additionally, we can employ the distance function $\mathscr{G}(.)$ for measuring if the

[20]By Assumption 4.1 is a finite, positive definite matrix.

orthogonal rotations satisfy the sign restrictions, then

$$\mathscr{B}_{0,2}^{Q,SR} = \left\{\boldsymbol{B} \mid \tilde{\mathscr{M}}^2(\Sigma_C \boldsymbol{Q}') = 0,\ \mathscr{G}(\boldsymbol{Q}) = 0 \text{ for } \boldsymbol{Q} \in \mathscr{O}(d)\right\},$$

with $\mathscr{G}(\boldsymbol{Q}) = \min_{\boldsymbol{\nu} \geq 0} \left\| \mathbf{S}_{SR} \overline{\boldsymbol{\Psi}}' \operatorname{vec}(\Sigma_C \boldsymbol{Q}') - \boldsymbol{\nu} \right\|^2$.

Lemma 4.1 (Identification by second order cumulants and sign-restrictions). *Consider the structural model in (4.2) satisfying condition (4.3). Under Assumption 4.1 and imposing sign-restrictions over the structural shocks in $\mathscr{I}_{SR}$, then*

(a) *The solution set $F(\varsigma_l^{SR})$ exists, and it is not a singleton if and only if*
$$\nexists\ \boldsymbol{q}_l \in \mathbb{R}_{++}^d \text{ s.t. } \mathbf{S}_l \boldsymbol{\Psi}'(l) \boldsymbol{B} \boldsymbol{q}_l = 0$$

(b) *$F(\varsigma_l^{SR})$ is unbounded, closed and convex.*

(c) *$\mathscr{B}_{0,2}^{Q,SR}$ is:*

  i. *Not empty if and only if $\times_{j \in \mathscr{I}_{SR}} F(\varsigma_j^{SR}) \cap GS\left(\times_{j \in \mathscr{I}_{SR}} F(\varsigma_j^{SR})\right) \neq \emptyset$.*

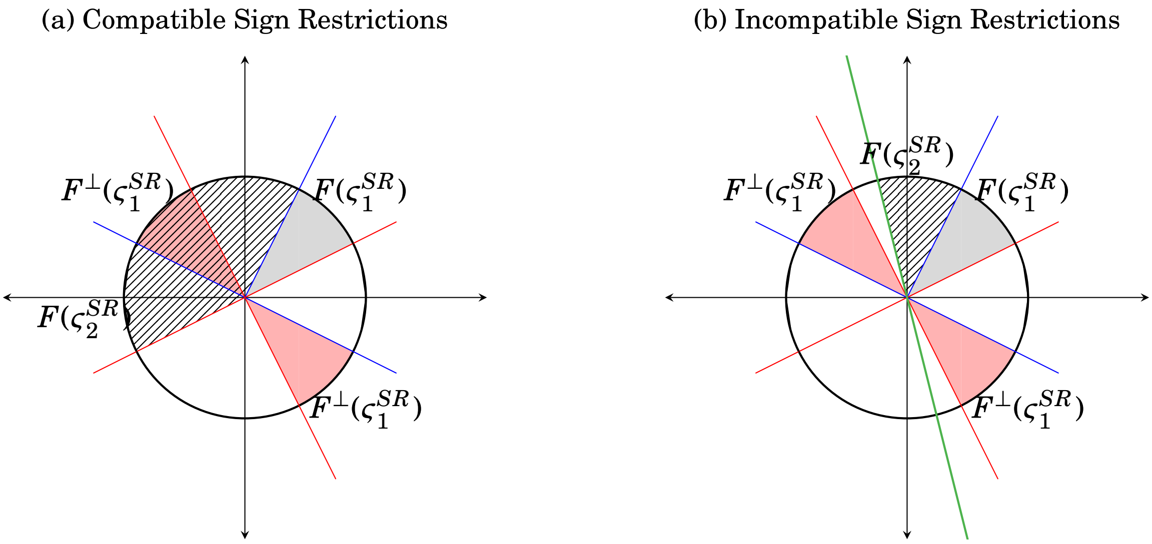
  ii. *Not a singleton if and only if $\times_{j \in \mathscr{I}_{SR}} F^{\circ}(\varsigma_j^{SR}) \cap GS\left(\times_{j \in \mathscr{I}_{SR}} F^{\circ}(\varsigma_j^{SR})\right) \neq \emptyset$.*

When $|\mathscr{I}_{SR}| = 1$, i.e. only one structural shock is signed restricted, Lemma 4.1 is equivalent to Proposition 1 in Granziera et al. (2018). When $|\mathscr{I}_{SR}| > 1$ and $r_l = 1$ for all $l \in \mathscr{I}_{SR}$, the results in Lemma 4.1 is straightforward, since such restriction splits the space into two parts. Lemma 4.1.(a) is a direct application of the existence result in R Tyrrell Rockafellar (1970) for cones formed by linear inequalities. Besides, Lemma 4.1 makes evident that the identified set with sign-restrictions and second-order cumulants, $\mathscr{B}_{0,2}^{Q,SR}$, is, in general, a continuum of admissible permutations of Σ_C, implying that researcher is identifying an infinite number of models whose are only coherent with the sign restrictions over a subset of shocks. At the same time, the rest of the responses are unrestricted. This feature makes inference a challenging task (see Granziera et al. (2018) and the references therein for details on inference with frequentist tools) because it is difficult to separate the model from the sampling uncertainty. This is similar to the criticism made by Fry and Pagan (2011).

Lemma 4.1.(c) is relevant when the researcher looks to identify multiple structural shocks since even though $F(\varsigma_l^{SR})$ exists for each $l \in \mathscr{I}_{SR}$, the existence of $F(\varsigma^{SR})$ is not assured since it is necessary to obtain a d-tuple of mutually orthogonal vectors. In a simple bi-variate context, the condition in Lemma 4.1.(c.i) is equivalent to require that set $F^{\perp}(\varsigma_1^{SR})$ intersects $F(\varsigma_2^{SR})$. Let us exemplify the result in Lemma 4.1.(c) with a simple example. Let structural model with $d = 2$ and $\boldsymbol{B}_0$ such that $\boldsymbol{B}_{0,jm} > 0 > \boldsymbol{B}_{0,j'm}$ for all $j < j'$

and m. If we impose $r_l = 2$ for $l \in \{1,2\}$ over the contemporaneous effects matrix such that $\boldsymbol{S}_1 = \text{diag}(+1,-1)$ and $\boldsymbol{S}_2 = \text{diag}(-1,-1)$. Nonetheless, if a third sign restriction over the second shock is added such that $\boldsymbol{S}_2 = \text{diag}(-1,-1,+1)$ and it is a non-convex combination of the edges of $F(\varsigma_1^{SR})$ (represented by the thicker green line in Figure 4.1), then there are no compatible solutions. Thus, the signs associated with the restrictions, contained in $\boldsymbol{S}_l$, play a crucial role in the existence of the identified set.

Figure 4.1: Identified sets and Sign Restrictions



4.3.3. Identification with Higher Order Cumulants

When structural errors in the model are non-Gaussian distributed, information from higher-order cumulants can be exploited for identification purposes, e.g., if all structural shocks exhibit asymmetric behavior ($\kappa_{3,j}^{\varepsilon} \neq 0$ for $j = 1,\ldots,d$), there are $\frac{d(d+1)(d+2)}{6}$ extra linear independent cumulant conditions. Therefore, the identified set using cumulants up to order n is

$$\mathscr{B}_{0,n} = \left\{\boldsymbol{B} \in \mathscr{B} \,|\, \tilde{\mathscr{M}}^n(\boldsymbol{B}) = 0\right\}, \quad n = 3,4.$$

When $n = 3$, third-order cumulants are employed for identifying the parameters of the structural model. In case the third-order cumulants were informationless, e.g., when structural errors are Gaussian or non-Gaussian but symmetric shocks, then $\mathscr{B}_{0,3} = \mathscr{B}_{0,2}$. In that situation, cumulants of order $n = 4$ should be employed.

Since non-zero third or fourth-order cumulants have characterized non-Gaussian behavior, it is possible to relate the number of non-Gaussian shocks in the system to the number of asymmetric or excess-kurtosis structural errors. This amount can be related to the rank of the matrix $\mathbf{v}_3^{\varepsilon}$ or $\mathbf{v}_4^{\varepsilon}$. Let $d_3 = \text{rank}\left(\mathbf{v}_3^{\varepsilon}\right)$, $d_4 = \text{rank}\left(\mathbf{v}_4^{\varepsilon}\right)$ and $d_{34} = \text{rank}\left(\mathbf{v}_{34}^{\varepsilon}\right)$

with $\mathbf{v}_{34}^{\varepsilon} = \left[\mathbf{v}_3^{\varepsilon\prime}, \mathbf{v}_4^{\varepsilon\prime}\right]'$. Notice that $d_3 + d_4 \geq d_{34} \geq \min\{d_3, d_4\}$. If $\mathbf{v}_4^{\varepsilon}$ and $\mathbf{v}_3^{\varepsilon}$ are linear independent block, then $d_{34} = d_3 + d_4$. In case third-order cumulants are informationless, $d_{34} = d_4$.

Assumption 4.2. *Let d_{ng} be the number of non-Gaussian shocks.*

(a) When $n = 3$, it is assumed that $d_{ng} = d_3 \geq d - 1$.

(b) When $n = 4$, it is assumed that $d_{ng} = d_{34} \geq d - 1$.

The implications of Assumption 4.2 are twofold. On the one hand, when information up to third order is exploited, Assumption 4.2.(a) imposes that the number of non-Gaussian shocks is equal to the number of asymmetric errors and this is greater or equal to $d - 1$, i.e., there is at most one symmetric shock in the structural model. On the other hand, Assumption 4.2.(b) states that when information up to fourth order is used, the number of non-Gaussian shocks is equal to the rank of $\mathbf{v}_{34}^{\varepsilon}$, which reflects the amount of asymmetric or non-mesokurtic errors.

Let $\boldsymbol{B}_1 \in \mathscr{B}(d)$ be a non-singular matrix such that $\tilde{\mathscr{M}}_0^3(\boldsymbol{B}_1) = 0$, i.e. solves both second and third order cumulant conditions. It is worth mentioning that, in general, when third-order cumulants are informative, $\boldsymbol{B}_1 \neq \Sigma_c$. Thus, like in the previous section, under Assumption 4.2 the identified set using up to cumulants of order n is

$$\mathscr{B}_{0,n} = \mathscr{B}_{0,n}^{Q} := \left\{\boldsymbol{B} \,|\, \tilde{\mathscr{M}}^n(\boldsymbol{B}) = 0, \boldsymbol{B} = \boldsymbol{B}_1 Q', \, Q' \in \mathscr{P}(d)\right\}, \tag{4.17}$$

with $\mathscr{P}(d) \subset \mathscr{O}(d)$ denoting the space of all signed-permutation matrices of size d. In Appendix **??**, we prove that this set is invariant to the choice of $\boldsymbol{B}_1$.

Intuitively, the expression in (4.17) may be interpreted as follows: when structural shocks are mutually independent and, at least, $d-1$ are asymmetric or show an excess of kurtosis, the matrix of contemporaneous effects is identified up to signed-permutation. It is important to remark that no external identification restrictions have been necessary to achieve such a result. Besides, like in the case of $\mathscr{B}_{0,2}^{Q}$, the true value $\boldsymbol{B}_0$ belongs to $\mathscr{B}_{0,n}^{Q}$. Also, notice that $\mathscr{B}_{0,n}^{Q}$ is a finite set with cardinality equal to $\left|\mathscr{B}_{0,n}^{Q}\right| = 2^d \times d!$. The power of the SIS for narrowing the identified set is unequivocal, but this still implies that there is no unique minimizer for the population criterion, though a finite set of isolated minimizers. In this sense, the SIS only provides a *local identification* of the SVAR model.

4.3.4. Role of Mutual Independence Assumption

Before discussing the permutation selection using sign restrictions, let us review the role of the mutual independence assumption, i.e., Assumption 4.1.(b), on identifying structural parameters. For simplicity, consider a tri-dimensional SVAR model with the vector of shocks $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}]'$ and assume that only fourth-order cumulants are informative. The number of fourth order cumulants is $3^4 = 81$, though only 15 of them are unique as follows:

(i) Of the type $\mathbb{E}[\varepsilon_{i,t}^4] = \kappa_{4,i}^{\varepsilon} + 3$ for $i \in \{1,2,3\}$ (3 unique coefficients);

(ii) of the type $\mathbb{E}[\varepsilon_{i,t}^2 \varepsilon_{j,t}\varepsilon_{s,t}] = 0$ for $i \neq j \neq s$, $i,j,s \in \{1,2,3\}$ (3 unique coefficients);

(iii) of the type $\mathbb{E}[\varepsilon_{i,t}^3 \varepsilon_{j,t}] = 0$ for $i \neq j$, $i,j \in \{1,2,3\}$ (6 unique coefficients); and

(iv) of the type $\mathbb{E}[\varepsilon_{i,t}^2 \varepsilon_{j,t}^2] = 1$ for $i \neq j$, $i,j \in \{1,2,3\}$ (3 unique coefficients).

Given that structural errors are independent, the identification conditions from second-order information are $\frac{d(d+1)}{2} = 6$, then it is necessary at least $\frac{d(d-1)}{2} = 3$ extra identification restrictions. Under Assumption 4.1.(b), type-(i) kurtosis condition does not help to identify $\boldsymbol{B}$, because $\kappa_{4,i}^{\varepsilon}$ is unknown. However, type-(ii), type-(iii), and type-(iv) conditions are informative. Therefore, provided the non-Gaussian behavior of shocks, Assumption 4.1.(b) gives the extra identification conditions. Only a subset of them is necessary for identifying the structural model, implying that it could be possible to relax the mutual independence.

This feature is pointed out and exploited by Lanne and Luoto (2021) to relax the mutual independence assumption. According to these authors, statistical identification can be achieved without assuming mutual independence and exploiting only type-(iv) cumulant conditions, while the rest could be non-zero. However, more is needed to solve the identification problem up to the signed permutation. They consider half of the type-(iii) conditions to solve this local identification result. Therefore, the model would be globally identified if the rest of the cumulant restrictions are non-zero. Let us see the mechanics of our reasoning with a simple example. Considering only type-(iii) conditions, then the matrix of fourth-order moments are

$$H = \begin{bmatrix} \mathbb{E}[\varepsilon_1^4] & \mathbb{E}[\varepsilon_1^3\varepsilon_2] & \mathbb{E}[\varepsilon_1^3\varepsilon_3] \\ \mathbb{E}[\varepsilon_2^3\varepsilon_1] & \mathbb{E}[\varepsilon_2^4] & \mathbb{E}[\varepsilon_2^3\varepsilon_3] \\ \mathbb{E}[\varepsilon_3^3\varepsilon_1] & \mathbb{E}[\varepsilon_3^3\varepsilon_2] & \mathbb{E}[\varepsilon_3^4] \end{bmatrix}.$$

Furthermore, let us impose the following identification restrictions:

$$H = \begin{bmatrix} \mathbb{E}[\varepsilon_1^4] & 0 & 0 \\ \mathbb{E}[\varepsilon_2^3\varepsilon_1] & \mathbb{E}[\varepsilon_2^4] & 0 \\ \mathbb{E}[\varepsilon_3^3\varepsilon_1] & \mathbb{E}[\varepsilon_3^3\varepsilon_2] & \mathbb{E}[\varepsilon_3^4] \end{bmatrix},$$

the elements on and below the main diagonal are not zero. This identification discards many permutations of structural shocks because many of these permutations may lead to an incorrect H matrix in the sense that the order of elements would not be the correct one. For instance, consider $\boldsymbol{\varepsilon}^\star = P\boldsymbol{\varepsilon}$ with

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then, the permuted vector of shocks is $\boldsymbol{\varepsilon}^\star = [\varepsilon_1^\star, \varepsilon_2^\star, \varepsilon_3^\star]' = [\varepsilon_2, \varepsilon_1, \varepsilon_3]'$ and its associated H matrix is:

$$H^\star = \begin{bmatrix} \mathbb{E}[\varepsilon_{\star,1}^4] & \mathbb{E}[\varepsilon_{\star,1}^3\varepsilon_{\star,2}] & \mathbb{E}[\varepsilon_{\star,1}^3\varepsilon_{\star,3}] \\ \mathbb{E}[\varepsilon_{\star,2}^3\varepsilon_{\star,1}] & \mathbb{E}[\varepsilon_{\star,2}^4] & \mathbb{E}[\varepsilon_{\star,2}^3\varepsilon_{\star,3}] \\ \mathbb{E}[\varepsilon_{\star,3}^3\varepsilon_{\star,1}] & \mathbb{E}[\varepsilon_{\star,3}^3\varepsilon_{\star,2}] & \mathbb{E}[\varepsilon_{\star,3}^4] \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\varepsilon_2^4] & \mathbb{E}[\varepsilon_2^3\varepsilon_1] & 0 \\ 0 & \mathbb{E}[\varepsilon_1^4] & 0 \\ \mathbb{E}[\varepsilon_3^3\varepsilon_2] & \mathbb{E}[\varepsilon_3^3\varepsilon_1] & \mathbb{E}[\varepsilon_3^4] \end{bmatrix}.$$

Since H and $H^\star$ do not have the same structure, this permutation would be invalid. Therefore, under the structure of H, the structural model is globally identified. Nevertheless, since the researcher only observes the structural errors after applying the identification restrictions, when these are imposed over the model errors, the identification up to the signed permutation is not avoided because model errors are a permutation of the true structural shocks. Thus, if marginal independence is relaxed, it is necessary to incorporate information about the ordering and which structural errors are marginally independent.

4.3.5. Permutation Selection with Sign Restrictions

According to many researchers (see Gouriéroux et al. (2017); Herwartz and Lütkepohl (2014); Lanne and Lütkepohl (2008); Lanne, Meitz, et al. (2017); Rigobon (2003), among others), identification up to signed-permutation of the SVAR model is sufficient to perform inference (e.g., testing), in particular regarding IR functions of structural shocks. However, this is partially right. Inference about parameters or measures invariant to signed permutations, such as the rank of $\boldsymbol{B}$ or the number of non-Gaussian shocks, can be implemented with a model identified up to signed permutation. Nevertheless, as is shown in (4.7), the IR functions are not invariant to rotations or permutations. Therefore, inference about these objects depends on the chosen permutation. Lanne, Meitz, et al. (2017) employs a mechanical procedure for selecting a permutation proposed by Hallin and Mehta (2015). On the other hand, Gouriéroux et al. (2017) propose to perform any testing procedure about the impulse responses or the elements of $\boldsymbol{B}_0$ over all the admissible permutations in the identified set $\mathscr{B}^{Q}_{0,n}$. According to them, the hypothesis $\boldsymbol{B}_0 = \boldsymbol{B}^{\star}$ implies the following $H_0 : \boldsymbol{B}_0 = \boldsymbol{B}^{\star}P'$ with $P \in \mathscr{P}(d)$. For example, suppose the researcher would like to assess whether the contemporaneous response of policy rate to a demand shock is positive using a structural model with 3 endogenous variables. In that case, it is necessary to consider $2^3 \times 3! = 48$ admissible models. Moreover, to reject the null hypothesis of no contemporaneous effect, it would be necessary to do it in each of the 48 scenarios. Then, the problem of labeling structural innovations may affect the power of such a procedure. In that sense, we can introduce "*meaningful economic restrictions*" that help us to label structural innovations.

Like in the case of using second-order information combined with sign restrictions, let us define the identified set using higher-order cumulants and sign restrictions:

$$\mathscr{B}^{Q,SR}_{0,n} = \left\{\boldsymbol{B} \,|\, \tilde{\mathscr{M}}^n(\boldsymbol{B}_1 \boldsymbol{Q}') = 0,\, \mathscr{G}(\boldsymbol{Q}) = 0 \text{ for } \boldsymbol{Q} \in \mathscr{P}(d)\right\}.$$

Proposition 4.1 (Identification Higher Order Cumulants and a single sign-restriction).
Suppose Assumptions 4.1-4.2 hold, $1 \leq |\mathscr{I}_{SR}| = d_{SR} \leq d$ *and* $r_l = 1$ *for all* $l \in \mathscr{I}_{SR}$*. Then:*

(a) $2^{d-d_{SR}} \times d! \leq \left|\mathscr{B}^{Q,SR}_{0,n}\right| \leq 2^d \times d!\left(1 - \frac{d_{SR}}{2d}\right) < 2^d \times d! = \left|\mathscr{B}^{Q}_{0,n}\right|.$

Moreover, when sign restrictions hold strictly:

(b) $2^{d-d_{SR}} \leq \left|\mathscr{B}^{Q,SR}_{0,n}\right| \leq 2^{d-d_{SR}} \times d!.$

Additionally, When $\boldsymbol{\Psi}(l) = \boldsymbol{\Psi}$ *for all* $l \in \mathscr{I}_{SR}$ *and rank* $\left(diag(\boldsymbol{\Psi}'\boldsymbol{B}_1)\right) = m$*:*

(c) $\left|\mathscr{B}^{Q,SR}_{0,n}\right| = d! \times 2^{d-m}.$

Propostion 4.1.(a) states that when a weak sign restriction is imposed over each of the system's structural shocks, the total number of admissible permutations reduces to at least half of the number of admissible models without any sign restriction. On the other hand, Proposition 4.1.(c) states that when the sign restriction is strict, the number of possible models reduces to $d!$. This latter result means that only permutations matter once the sign of each shock is fixed. Based on Proposition 4.1.(c), a strict sign-restriction can be imposed for labeling each structural shock in the system. For instance, it is a consensus in the macroeconomic literature that a monetary policy shock decreases inflation contemporaneously or that a government spending shock increases the real output. Therefore, we can incorporate a *shock definition* as part of $\boldsymbol{B}_0$. This feature is stated in the following assumption.

Assumption 4.3. ***(Shock definition)***
The elements of the diagonal of $\boldsymbol{B}_0$ are strictly positive, that is $B_{0,ii} > 0$ for all $i = 1, \ldots, d$.

Assumption 4.3 states that any structural innovation increases, contemporaneously, its endogenous correspondent variable. It is evident that when this assumption holds, the identified set $\mathscr{B}^Q_{0,n}$ reduces the number of admissible permutations to $d!$. This result means that if the structural model satisfies all the assumptions from 4.1 to 4.3, the SVAR model is statistically identified up to permutation. However, these types of restrictions do not yet provide global identification. In what follows, we will discuss the necessary and sufficient conditions for achieving global identification.

Point Identification with Single Strict Sign Restriction

Let us consider a simple example. Suppose we have a trivariate SVAR, i.e., $d = 3$, whose matrix of contemporaneous effects satisfies Assumption 4.3. Therefore, a solution $\boldsymbol{B}_1 = \left[B_{1,ij}\right]_{i,j=1,2,3}$ to the PCC should hold such restrictions, meaning that $B_{1,jj} > 0$ for $j = 1,2,3$. The condition in Assumption 4.3 can be interpreted as a strict sign-restriction over $\text{IR}(y_{j,t}|\varepsilon_{j,t})$. Thus, the conditions implied by *Shock-definitions* are $\text{IR}(y_{j,t}|\varepsilon_{j,t}) > 0$ for $j = 1,2,3$[21]. The key question is to find under what conditions there is no possibility to find $\boldsymbol{B} = \boldsymbol{B}_1 P$ with $P \neq \boldsymbol{I}_3$ such that $\text{IR}(y_{j,t}|\varepsilon^\star_{j,t}) > 0$ for $j = 1,2,3$, where $\varepsilon^\star_{j,t}$ is a permuted structural shock. The sign-restriction is

$$\text{IR}(y_{j,t}|\varepsilon^\star_{j,t}) = \boldsymbol{e}_j' \boldsymbol{B}_1 \boldsymbol{p}_j = \begin{bmatrix} B_{1,j1} & B_{1,j2} & B_{1,j3} \end{bmatrix} \boldsymbol{p}_j > 0, \; j = 1,2,3.$$

If $\boldsymbol{B}_1$ satisfies Assumption 4.3, an obvious solution to the former restriction is $\boldsymbol{p}_j = \boldsymbol{e}_j$ for $j = 1,2,3$, that is $P = \boldsymbol{I}_3$ is an admissible permutation. Nonetheless, if $B_{1,12}B_{1,23}B_{1,31} \neq$

[21] If $\boldsymbol{B}_1$ does not satisfy *innovation definition*, we can apply a permutation matrix $\bar{P}$ such that Assumption 4.3 is satisfied and we can do the analisys with $\boldsymbol{B}_1\bar{P}$

0, then the permutation matrix

$$P_1 = \begin{bmatrix} 0 & 0 & \text{sgn}(B_{1,31}) \\ \text{sgn}(B_{1,12}) & 0 & 0 \\ 0 & \text{sgn}(B_{1,23}) & 0 \end{bmatrix}$$

is an admissible permutation, that is $\boldsymbol{B}_1 P_1$ satisfies the sign-restriction implied by Assumption 4.3. Therefore, the set of admissible permutations coherent with Assumption 4.3 is formed by any combination of three vectors taken from the sets

$$\begin{aligned}
&\text{For } j=1: \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \text{sgn}(B_{1,12}) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \text{sgn}(B_{1,13}) \end{bmatrix} \right\} \\
&\text{For } j=2: \left\{ \begin{bmatrix} \text{sgn}(B_{1,21}) \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \text{sgn}(B_{1,23}) \end{bmatrix} \right\} \\
&\text{For } j=3: \left\{ \begin{bmatrix} \text{sgn}(B_{1,31}) \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \text{sgn}(B_{1,32}) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}
\end{aligned}$$

that is linearly independent. The number of admissible permutations is 6. To obtain $P = \boldsymbol{I}_3$ as the unique admissible permutation, it is sufficient to impose, for instance, that $B_{1,21} = B_{1,23} = B_{1,31} = 0$. However, this is not the only configuration that may be sufficient for assuring a unique admissible permutation coherent with Assumption 4.3. Let $\complement\mathscr{I}' := \mathscr{I} \setminus \mathscr{I}'$ with $\mathscr{I} = \{1,2,3\}$ and $\mathscr{I}' \subset \mathscr{I}$. The 6 possible cases of zero restrictions over $\boldsymbol{B}_1$ can be summarized as follows:

$$\textit{For } j \in \mathscr{I} \text{ and } s \in \complement\{j\} : B_{1,jk} = B_{1,sk'} = 0 \text{ for all } k \in \complement\{j\},\ k' \in \complement\{j,s\}$$

The following assumption summarizes the previous discussion and describes the condition above for a generic $\boldsymbol{B}_1$ of dimension d.

Assumption 4.4. ***(Generalized Recursive Structure)***
*Let $\boldsymbol{B}_0$ be a $d \times d$ non-singular matrix and $\mathscr{I} = \{1,2,\dots,d\}$, an index set. It is said $\boldsymbol{B}_0$ has a **generalized recursive structure** if:*

$$\exists!\ s_0 \in \mathscr{I}, s_1 \in \complement(\{s_0\}), \dots, s_{d-2} \in \complement\left(\bigcup_{k=0}^{d-3} \{s_k\}\right) \textit{ such that } \boldsymbol{B}_{0,s_j m} = 0\ \ \forall\, m \in \complement\left(\bigcup_{k=0}^{j} \{s_k\}\right) \textit{ and } j = 0,\dots,d-2.$$

Some noticeable examples of $\boldsymbol{B}_0$ satisfying Assumption 4.4 are the following: when

$s_j = j+1$ for $j = 0,1,\dots,d-2$, $\boldsymbol{B}_0$ is lower triangular; when $s_j = d-j$ for $j = 0,1,\dots,d-2$, $\boldsymbol{B}_0$ is upper triangular matrix.

Proposition 4.2. *Let Assumptions 4.1-4.2 hold. If $\boldsymbol{B}_0$ satisfies Assumptions 4.3 and 4.4, then*

$$\left|\mathscr{B}^{Q,SR}_{0,n}\right| = 1.$$

Proposition 4.2 states that when structural shocks in the model in (4.2) are mutually independent and at least $d-1$ are skewed or non-mesokurtic, the parameters of the structural model are point identified once the contemporaneous effects matrix follows a generalized recursive structure, i.e., if $\boldsymbol{B}_0 = P_L^\star \Sigma_c P_R^\star$ with P_L and P_R some permutation matrices. In Appendix E, we explore some consequences of Cholesky (or some right and left permutation) being a solution to $\mathscr{B}_{0,n}$. The result in Proposition 4.2 is unsatisfactory because recursive systems are very restrictive structures that may arise only in special cases. Moreover, if Assumption 4.4 holds, the model would be identified using only second-order cumulants (see Rubio-Ramirez et al. (2010) for details). Thus, third-order cumulants introduce overidentifying restrictions.

There exists a relation between the recursive structure of matrix $\boldsymbol{B}$ and the form of array $\mathbf{v}_n^{\boldsymbol{u}}$. For illustrate the idea, let consider $n=3$ and $d=3$; as it has been showed above: $\mathbf{v}_3^u = \boldsymbol{B}^{\otimes 2}\mathbf{v}_3^\varepsilon \boldsymbol{B}'$. Wlog, assuming that $\boldsymbol{B}$ has a Cholesky structure. Then,

$$\boldsymbol{B}^{\otimes 2} = \begin{bmatrix} B_{11}\boldsymbol{B} & \mathbf{0} & \mathbf{0} \\ B_{21}\boldsymbol{B} & B_{22}\boldsymbol{B} & \mathbf{0} \\ B_{31}\boldsymbol{B} & B_{32}\boldsymbol{B} & B_{33}\boldsymbol{B} \end{bmatrix}, \text{ and } \mathbf{v}_3^\varepsilon = \begin{bmatrix} \kappa_{3,1}\boldsymbol{e}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \kappa_{3,2}\boldsymbol{e}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \kappa_{3,3}\boldsymbol{e}_3 \end{bmatrix}.$$

Thus

$$\boldsymbol{B}^{\otimes 2}\mathbf{v}_3^\varepsilon = \begin{bmatrix} B_{11}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & \mathbf{0} & \mathbf{0} \\ B_{21}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & B_{22}\kappa_{3,2}\boldsymbol{B}\boldsymbol{e}_2 & \mathbf{0} \\ B_{31}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & B_{32}\kappa_{3,2}\boldsymbol{B}\boldsymbol{e}_2 & B_{33}\kappa_{3,3}\boldsymbol{B}\boldsymbol{e}_3 \end{bmatrix}.$$

Since $\boldsymbol{B}' = \begin{bmatrix} B_{11} & B_{21} & B_{31} \\ 0 & B_{22} & B_{32} \\ 0 & 0 & B_{33} \end{bmatrix}$, this implies the following:

1. The first block of $\boldsymbol{B}^{\otimes 2}\mathbf{v}_3^\varepsilon$, i.e. $\begin{bmatrix} B_{11}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & \mathbf{0} & \mathbf{0}\end{bmatrix}$, multiplied by $\boldsymbol{B}'$ is

$$\begin{bmatrix} B_{11}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{B}' = B_{11}\kappa_{3,1}\begin{bmatrix} B_{11} \\ B_{21} \\ B_{31}\end{bmatrix}\begin{bmatrix} B_{11} & B_{21} & B_{31}\end{bmatrix} = B_{11}\kappa_{3,1}\boldsymbol{b}_1\boldsymbol{b}_1',$$

Moreover, its rank is equal to 1.

2. The second block of $\boldsymbol{B}^{\otimes 2}\mathbf{v}_3^{\varepsilon}$, i.e. $\begin{bmatrix} B_{21}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & B_{22}\kappa_{3,2}\boldsymbol{B}\boldsymbol{e}_2 & \mathbf{0} \end{bmatrix}$, multiplied by $\boldsymbol{B}'$ is

$$\begin{bmatrix} B_{21}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & B_{22}\kappa_{3,2}\boldsymbol{B}\boldsymbol{e}_2 & \mathbf{0} \end{bmatrix}\boldsymbol{B}' = B_{21}\kappa_{3,1}\boldsymbol{b}_1\boldsymbol{b}_1' + B_{22}\kappa_{3,2}\boldsymbol{b}_2\boldsymbol{b}_2',$$

   and its rank is equal to 2.

3. The last block of $\boldsymbol{B}^{\otimes 2}\mathbf{v}_3^{\varepsilon}$, i.e. $\begin{bmatrix} B_{31}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & B_{32}\kappa_{3,2}\boldsymbol{B}\boldsymbol{e}_2 & B_{33}\kappa_{3,3}\boldsymbol{B}\boldsymbol{e}_3 \end{bmatrix}$, multiplied by $\boldsymbol{B}'$ is

$$\begin{bmatrix} B_{31}\kappa_{3,1}\boldsymbol{B}\boldsymbol{e}_1 & B_{32}\kappa_{3,2}\boldsymbol{B}\boldsymbol{e}_2 & B_{33}\kappa_{3,3}\boldsymbol{B}\boldsymbol{e}_3 \end{bmatrix}\boldsymbol{B}' = B_{31}\kappa_{3,1}\boldsymbol{b}_1\boldsymbol{b}_1' + B_{32}\kappa_{3,2}\boldsymbol{b}_2\boldsymbol{b}_2' + B_{33}\kappa_{3,3}\boldsymbol{b}_3\boldsymbol{b}_3'$$

   and its rank is equal to 3.

Thus, it is clear that the rank of the $d \times d$ blocks of $\mathbf{v}_3^{\boldsymbol{u}}$ can be employed for discovering whether $\boldsymbol{B}$ has a recursive structure or not. For generalizing this idea in the case of using third or fourth-order cumulants, let us introduce some definitions previously. Based on the set of indexes $\mathscr{I} = \{1, \ldots, d\}$, let define an one-to-one, ordering function $f_o : \mathscr{I} \mapsto \mathscr{I}$. For instance, when $d = 3$, the identity ordering function is $f_o(j) = j$, or the inverse ordering function is $f_o(j) = d + 1 - j$. Also, if $\boldsymbol{A}$ is a $d^m \times d$, $\boldsymbol{A}_j$ denotes the j-th block of size $d^{m-1} \times d$.

Proposition 4.3. *Let the structural model described by (4.2) satisfying condition (4.3) and structural shocks meet Assumptions 4.1-4.2.*

- *When $d_3 = d$, $\boldsymbol{B}$ has a (generalized) recursive structure if and only if for some one-to-one ordering mapping, $f_o(j)$*

$$rank\left(\mathbf{v}_{3,f_o(j)}^{\boldsymbol{u}}\right) = j, \text{ for } j = 1, \ldots, d.$$

- *When $d_4 = d$, $\boldsymbol{B}$ has a (generalized) recursive structure if and only if for some one-to-one ordering mapping, $f_o(j)$*

$$rank\left(\mathbf{v}_{4,f_o(j)}^{\boldsymbol{u}}\right) = j, \text{ for } j = 1, \ldots, d.$$

Point Identification with Multiple Sign Restrictions

Here, it is discussed whether adding more non-redundant sign restrictions may lead to global identification without requiring the very restrictive sufficient condition in Assumption 4.4. As we have detailed above, the system of sign restrictions ς_l^{SR} is not empty but contains a continuum of elements. Besides, in case structural shocks are non-Gaussian distributed, the identified set is a mapping from the space of all signed-permutation matrices; the structural model is identified up to sign-permutation. The set of admissible permutations for each shock is the set $\mathscr{E} := \mathscr{E}^+ \cup \mathscr{E}^-$, where $\mathscr{E}^+ := \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ and $\mathscr{E}^- := \{-\boldsymbol{e}_1, \ldots, -\boldsymbol{e}_d\}$, while the set of matrices of signed permutations is any collection of d vectors from $\mathscr{E}$ that are linearly independent. Furthermore, for a signed-restricted single shock $\varepsilon_{l,t}$, the set of admissible permutations consistent with the sign restrictions is $\mathscr{E} \cap \varsigma_l^{SR}$.

According to F. Hall and Li (2007), a *sign-pattern matrix* is an array whose entries are $\{+1, -1, 0\}$. If $\boldsymbol{A}$ is some real-valued matrix, $\mathbb{A} := \text{sgn}(\boldsymbol{A})$ denote the associated sign-pattern matrix to A. A *non-negative* elemnt-wise array has $\mathbb{A}$ with only $\{0, +1\}$ entries; a *non-positive* element-wise matrix has $\mathbb{A}$ with only $\{0, -1\}$ entries. A *sign-pattern class* or *qualitative class* associated to a $d \times d$ matrix $\boldsymbol{A}$ is defined as:

$$\mathscr{Q}(\boldsymbol{A}) = \{\boldsymbol{B} \in \mathbb{R}^{d \times d} \ : \text{sgn}(\boldsymbol{B}) = \text{sgn}(\boldsymbol{A})\}.$$

That is, $\mathscr{Q}(\boldsymbol{A})$ contains all the matrices with a similar sign pattern to $\boldsymbol{A}$. According to Brualdi and Shader (2009) the equations system $\{\boldsymbol{A}\mathrm{x} = \boldsymbol{b}\}$ is said to be *sign-solvable* if and only if

1. For each $\tilde{\boldsymbol{A}} \in \mathscr{Q}(\boldsymbol{A})$ and $\tilde{\boldsymbol{b}} \in \mathscr{Q}(\boldsymbol{b})$: the system $\tilde{\boldsymbol{A}}\mathbf{x} = \tilde{\boldsymbol{b}}$ is solvable.
2. For some $\tilde{\boldsymbol{A}} \in \mathscr{Q}(\boldsymbol{A})$ and $\tilde{\boldsymbol{b}} \in \mathscr{Q}(\boldsymbol{b})$: the solution set of $\tilde{\boldsymbol{A}}\mathbf{x} = \tilde{\boldsymbol{b}}$ is one qualitative class.

We say $\boldsymbol{a}$ is a *mono-sign* if $\text{sgn}(\boldsymbol{a})$ contains only values in either $\{+1, 0\}$ or $\{-1, 0\}$, i.e. either $\boldsymbol{a}$ is non-negative or non-positive element-wise. It is called strictly mono-sign if $\text{sgn}(\boldsymbol{a}) = \boldsymbol{\iota}$ or $\text{sgn}(\boldsymbol{a}) = -\boldsymbol{\iota}$ where $\boldsymbol{\iota}$ is a vector of ones. Additionally, following Brualdi and Shader (2009), a matrix $\boldsymbol{A}$ is called an L-matrix if all the members of its qualitative class $\mathscr{Q}(\boldsymbol{A})$ have linear independent rows. Also, $\boldsymbol{D}$ is a *row sign-change* matrix if it is a diagonal matrix with $D_{ii} \in \{+1, -1\}$. Theorem 2.1.1 in Brualdi and Shader (2009) states that: $\boldsymbol{A}$ is an L-matrix if and only if for each *row sign-change matrix* $\boldsymbol{D}$, $\boldsymbol{DA}$ has a *mono-sign* column for any $\boldsymbol{A} \in \mathscr{Q}(\boldsymbol{A})$. Finally, $\mathbb{S}^d$ denotes the unit sphere[22] in $\mathbb{R}^d$.

Lemma 4.2. *Let $\mathbf{x} \in \mathbb{S}^d$, and $\boldsymbol{A} \in \mathbb{R}^{d \times r}$ such that rank$(\boldsymbol{A}) = r \leq d$. Consider the homogenous system of inequalities $\varsigma = \{\mathbf{A}'\mathbf{x} \leq \mathbf{0}\}$ with non-empty, non-singleton solution set $F(\varsigma)$. Then,*

$$\mathscr{E} \cap F(\varsigma) \neq \emptyset \ \Leftrightarrow \ \mathbf{A}' \text{ is an } L\text{-matrix.}$$

[22] $\mathbb{S}^d$ denotes the sphere of radius 1 in $\mathbb{R}^d$, i.e. $\mathbb{S}^d := \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| = 1\}$

Lemma 4.2 provides a sufficient condition for the existence of $\mathscr{E}\cap F(\varsigma)$. Being an L-matrix implies that at least one of the columns of $\boldsymbol{A}'$ does not belong to the same qualitative class of the remaining columns. Moreover, the number of mono-sign columns may decline with the number of linearly independent rows in the L-matrix, implying that the number of elements in $\mathscr{E}\cap F(\varsigma)$ declines as well. For determining the cardinality of the set $\mathscr{E}\cap F(\varsigma)$ we use the following elements: let $\boldsymbol{A}' = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_d\end{bmatrix}$, $\mathbb{1}\left(\boldsymbol{a}_j \neq 0\right)$ denote a binary vector taking 1 if the element is non-zero and $\boldsymbol{\iota}'\boldsymbol{a}_j$ is the sum of elements in column $\boldsymbol{a}_j$. Thus, $\boldsymbol{\iota}'\mathbb{1}\left(\boldsymbol{a}_j \neq 0\right)$ is equal to amount of non-zero elements in $\boldsymbol{a}_j$, and $\boldsymbol{\iota}'\,\mathrm{sgn}(\boldsymbol{a}_j)$ is the sum of elements of the sign-pattern of $\boldsymbol{a}_j$. It is clear that $\boldsymbol{a}_j$ is a mono-sign column vector whenever $\boldsymbol{\iota}'\mathbb{1}\left(\boldsymbol{a}_j \neq 0\right) = \left|\boldsymbol{\iota}'\,\mathrm{sgn}\left(\boldsymbol{a}_j\right)\right|$. Let denote $v_j = \boldsymbol{\iota}'\mathbb{1}\left(\boldsymbol{a}_j \neq 0\right)$ and $o_j = \left|\boldsymbol{\iota}'\,\mathrm{sgn}(\boldsymbol{a}_j)\right|$.

Lemma 4.3.
Given a system of inequalities $\varsigma = \{\boldsymbol{A}'\mathbf{x} \leq 0\}$ with $\mathbf{A}'$ an L-matrix.

(a) If $v_j > 0$, then $\left|\{\boldsymbol{e}_j,\, -\boldsymbol{e}_j\}\cap F(\varsigma)\right| = 1 \Leftrightarrow v_j - o_j = 0$.

(b) If $v_j = 0$, then $\{\boldsymbol{e}_j, -\boldsymbol{e}_j\} \subset F(\varsigma)$ and $\left|\{\boldsymbol{e}_j,\, -\boldsymbol{e}_j\}\cap F(\varsigma)\right| = 2$.

Lemma 4.3.(a) implies that when the column $\boldsymbol{a}_j$ of matrix $\boldsymbol{A}'$ is non-zero, then if this column is mono-sign, then either $\boldsymbol{e}_j$ or $-\boldsymbol{e}_j$ is an admissible signed-permutation consistent with the inequalities system ς. Otherwise, if $\boldsymbol{a}_j$ is a null column, both signed permutations are admissible and satisfy the inequalities in ς. Finally, let denote $\tilde{m}_j = \left|\{\boldsymbol{e}_j, -\boldsymbol{e}_j\} \subset F(\varsigma)\right|$ and the vector $\tilde{\boldsymbol{m}} = (\tilde{m}_1, \ldots, \tilde{m}_d)'$. Therefore, it is clear that $|\mathscr{E}\cap F(\varsigma)| = \sum_{j=1}^{d} \tilde{m}_j$.

The following assumption incorporates the sufficient condition stated in Lemma 4.2 as a feature of the system of inequalities implied by the sign restrictions over a structural shock $\varepsilon_{l,t}$.

Assumption 4.5.
*Let the sign-restrictions over the IR functions of l-th structural shock, $\varsigma_l^{SR} := \{\boldsymbol{S}_l\boldsymbol{\Psi}'(l)\boldsymbol{B}_1\mathbf{x} \geq \mathbf{0}\}$, where $\boldsymbol{\Psi}'(l)$ is a $d \times r_l$ full column rank matrix and $r_l \leq d$, $\boldsymbol{S}_l$ be a diagonal sign-pattern matrix and $\boldsymbol{B}_1 \in \mathscr{B}_{0,3}^{Q}$. It is **assumed** that $\boldsymbol{\Psi}'(l)\boldsymbol{B}_1$ is **non-zero element wise** and an L-matrix.*

Based on this assumption, we state the following.

Proposition 4.4.
Consider the structural model in (4.2) satisfying conditions in (4.3) and Assumption 4.3;

the shocks in the model meet Assumptions 4.1 and 4.2; and the sign-restrictions system for identifying shock $l^\star$, $\varsigma^{SR}_{l^\star} = \{\boldsymbol{S}_{l^\star}\boldsymbol{\Psi}'(l^\star)\boldsymbol{B}_1\mathbf{x} \geq \mathbf{0}\}$. If rank$(\boldsymbol{\Psi}'(l^\star)\boldsymbol{B}_1) = r_{l^\star} = d$ and $\boldsymbol{\Psi}'(l^\star)\boldsymbol{B}_1$ satisfies Assumption 4.5, then:

$$\left|\mathscr{E} \cap F(\varsigma^{SR}_{l^\star})\right| = 1.$$

Proposition 4.4 can be interpreted as follows: if a shock-definition restriction were imposed on all structural errors, i.e., if we fixed the sign of each shock, and we add $d-1$ sign-restrictions over only a specific structural error $l^\star$ such that $\boldsymbol{\Psi}(l^\star)'\boldsymbol{B}_1$ is an L-matrix, then such a structural shock is point identified. Although, the rest of the shocks remain identified only up to permutation. Therefore, the number of admissible permutations reduces to $(d-1)!$. Notice that the previous result is sufficient for point identification in the case of a bivariate SVAR model. Nonetheless, to identify a subset of structural errors by sign-restrictions, i.e., $|\mathscr{I}_{SR}| = d_{SR} > 1$, the result in Proposition 4.4 applied to each shock separately is not sufficient.

Notice that $\mathscr{E} \cap F(\varsigma^{SR}_l)$ is the identified set for a particular shock $\varepsilon_{l,t}$ subject to sign-restrictions. By the previous result, this set is a singleton, e.g., $\mathscr{E} \cap F(\varsigma^{SR}_l) = \{\boldsymbol{e}_{j(l)}\}$. The point identification of a pair of shocks, say $(\varepsilon_{l,t}, \varepsilon_{l',t})$, requires that the set $\mathscr{E} \cap F(\varsigma^{SR}_{l'})$ is non-empty, containing at most 2 elements and one of them is linear independent from $\boldsymbol{e}_{j(l)}$. Under the general case in which $\boldsymbol{\Psi}'(l) \neq \boldsymbol{\Psi}'(l')$, it is complicated to find a sufficient condition for assuring the point identification of the pair of shocks $(\varepsilon_{l,t}, \varepsilon_{l',t})$. However, if $\boldsymbol{\Psi}'(l) = \boldsymbol{\Psi}'$ for all l, it is feasible to provide a sufficient and necessary condition. Furthermore, let $\boldsymbol{A}$ be a $d \times d$ matrix and the set of indexes $\{j_1, \ldots, j_k\}$ of length k, $[\boldsymbol{A}]_{-\{j_1,\ldots,j_k\}}$ denote a $(d-k)\times(d-k)$ matrix obtained by eliminating columns and rows indexed by $\{j_1, \ldots, j_k\}$. For instance, $[\boldsymbol{A}]_{-\{2,3\}}$ is the $(d-2)\times(d-2)$ submatrix obtained by dropping second and third rows and columns. Moreover, we take the convention that $[\boldsymbol{A}]_{-\emptyset} = \boldsymbol{A}$.

Assumption 4.6. ***(Coherent Restrictions)***
For the subset of sign-restricted structural shocks, $\mathscr{I}_{SR}$, the sequence of sign-restrictions $\{\varepsilon^{SR}_l\}_{l\in\mathscr{I}_{SR}}$, with $\boldsymbol{\Psi}(l) = \boldsymbol{\Psi}$ for all $l \in \mathscr{I}_{SR}$, are coherent if the following recursion holds for $j = 1, \ldots, d_{SR}$

(a) Given j and $l_j \in \complement\left(\mathscr{I}_{SR} \setminus \bigcup_{k=0}^{j-1}\{l_k\}\right)$, the matrix $\left[\boldsymbol{\Psi}'\boldsymbol{B}_1\right]_{-\bigcup_{k=0}^{j-1}\{l_k\}}$ is an L-matrix.

Assumption 4.6 implies the existence of a sequence $\{\boldsymbol{q}_l\}_{l\in\mathscr{I}_{SR}} \in \bigotimes_{l\in\mathscr{I}_{SR}} \{\mathscr{E} \cap F(\varsigma^{SR}_l)\}$ whose elements are mutually orthogonal.

Proposition 4.5.
Consider the structural model in (4.2) satisfying conditions in (4.3) and Assumption 4.3;

the shocks in the model meet assumptions 4.1 and 4.2. Assuming that the sign-restrictions for shocks in $\mathcal{I}_{SR}$ satisfy Assumption 4.6 and $r_l = d$, then

$$\left|\mathcal{B}_{0,n}^{Q,SR}\right| = (d - d_{SR})!$$

From Proposition 4.5, it is clear that when $\mathcal{I}_{SR} = \mathcal{I}$, then $d = d_{SR}$ and point identification is achieved, i.e. $\left|\mathcal{B}_{0,n}^{Q,SR}\right| = 1$.

4.3.6. Identification with more than one Gaussian error

Now, we explore what are the implications of relaxing 4.2, i.e. do not assume that $d_{ng} \geq d-1$, but $1 \leq d_{ng} \leq d-2$. In other words, there is more than one Gaussian structural error in the SVAR model in (4.2). From Comon (1994), it is clear that structural parameters are not statistically identified. However, the following proposition states a partial identification result.

Proposition 4.6.
Consider the structural model in (4.2), which satisfies condition (4.3), and the structural shocks satisfy Assumption 4.1. If $1 \leq d_{ng} \leq d-2$, then

$$\mathcal{B}_{0,n}^{Q} = \left\{\boldsymbol{B} \,|\, \tilde{\mathcal{M}}^n(\boldsymbol{B}) = 0, \boldsymbol{B} = \boldsymbol{B}_1 Q\right\}, \; n \in \{3,4\},$$

where

$$\begin{bmatrix} P_{d_{ng}\times d_{ng}} & \mathbf{0}_{d_{ng}\times(d-d_{ng})} \\ \mathbf{0}_{(d-d_{ng})\times d_{ng}} & \tilde{Q}_{(d-d_{ng})\times(d-d_{ng})} \end{bmatrix},$$

and $P_{d_{ng}\times d_{ng}} \in \mathcal{P}(d_{ng})$ is a signed-permutation matrix and $\tilde{Q} \in \mathcal{O}(d-d_{ng})$ any orthogonal matrix of size $d-d_{ng}$.

Proposition 4.6 expresses that, even though the number of Gaussian structural errors in the SVAR system is more than one, the non-Gaussian block remains still identified up to the signed permutation. Nevertheless, since the order is unknown, it is necessary to impose an order. In Proposition 4.6, it is assumed implicitly that the first d_{ng} components in the vector of structural errors are non-Gaussian distributed, while the rest are Gaussian shocks. This result is similar to what is found by Maxand (2020). Like in the case of having at most one Gaussian structural error, to reduce the total number of admissible permutations for the non-Gaussian block, $2^{d_{ng}} \times d_{ng}!$, shock-definition restrictions may be imposed. This result is relevant, especially in empirical analysis using SVAR models, because it provides identification of a subset of structural shocks. However, a significant limitation is the necessity of prior knowledge about ordering Gaussian and non-Gaussian errors. Furthermore, for estimation purposes under this scenario, if we're not interested in the Gaussian block and assuming the researcher knows the valid order of shocks, the estimation of structural parameters of non-Gaussian block fits in the literature of estimation of partially identified models, where the parameters of the Gaussian block and the ordering of errors act like reduced form and nuisance parameters.

4.4. Concluding Remarks

This chapter investigates if sign restrictions over impulse response functions of structural shocks can be employed to select a particular permutation. Fixing the sign of response of each shock deals with the sign of permutations and permits focusing only on no-signed permutations. However, the sign restrictions systems may satisfy the sign-solvable property to achieve point identification. Such a requirement is not immediately satisfied for any model. Therefore, sign restrictions impose identification restrictions that data may not support.

Additionally, we find that in the case of having all skewed structural shocks in the model, it is possible to relate the recursive structure of the contemporaneous effects matrix to the rank of third-order moments of reduced form errors. Finally, in the case of having non-fully identified models, because of the existence of more than one Gaussian structural error, the non-Gaussian block remains identified. However, the ordering between Gaussian and non-Gaussian shocks matters.

BIBLIOGRAPHY

Alessi, L., Barigozzi, M., & Capasso, M. (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, *79*(1), 16–47 (cit. on p. 47).

Al-Sadoon, M. M. (2017). A unifying theory of tests of rank. *Journal of Econometrics*, *199*(1), 49–62 (cit. on p. 55).

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press. (Cit. on pp. 27, 30).

Anderson, T. W. (2011). *The statistical analysis of time series*. John Wiley & Sons. (Cit. on p. 50).

Angelini, G., Cavaliere, G., & Fanelli, L. (2022). An identification and testing strategy for proxy-svars with weak proxies. *arXiv preprint arXiv:2210.04523* (cit. on p. 4).

Arias, J. E., Rubio-Ramírez, J. F., & Waggoner, D. F. (2018). Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications. *Econometrica*, *86*(2), 685–720 (cit. on pp. 3, 44, 74, 76).

Baggio, G., & Ferrante, A. (2018). Parametrization of minimal spectral factors of discrete-time rational spectral densities. *IEEE Transactions on Automatic Control*, *64*(1), 396–403 (cit. on pp. 14, 50).

Bera, A. K., & Jarque, C. M. (1982). Model specification tests: A simultaneous approach. *Journal of econometrics*, *20*(1), 59–82 (cit. on p. 46).

Bernanke, B. S., & Mihov, I. (1998). Measuring monetary policy. *The quarterly journal of economics*, *113*(3), 869–902 (cit. on pp. 3, 44, 68).

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons. (Cit. on p.).

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *the Quarterly Journal of economics*, *117*(4), 1329–1368 (cit. on pp. x, 3, 44, 47, 65, 66).

Blanchard, O., & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, *79*(4), 655–673 (cit. on pp. x, 3, 8, 39, 40, 44, 47, 65, 66, 68).

Boubacar Mainassara, Y. (2012). Selection of weak varma models by modified akaike's information criteria. *Journal of Time Series Analysis*, *33*(1), 121–130 (cit. on p. 32).

Braun, R., Brüggemann, R., et al. (2017). *Identification of svar models by combining sign restrictions with external instruments* (tech. rep.). Department of Economics, University of Konstanz. (Cit. on p. 69).

Brillinger, D. R. (2001). *Time series: Data analysis and theory*. SIAM. (Cit. on p. 53).

Brillinger, D. R., & Rosenblatt, M. (1967). Asymptotic theory of estimates of kth-order spectra. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(2), 206 (cit. on pp. 59, 60).

Brualdi, R. A., & Shader, B. L. (2009). *Matrices of sign-solvable linear systems* (Vol. 116). Cambridge University Press. (Cit. on p. 90).

Camba-Méndez, G., & Kapetanios, G. (2009). Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling. *Econometric Reviews*, *28*(6), 581–611 (cit. on p. 54).

Canova, F., & De Nicolo, G. (2002). Monetary disturbances matter for business fluctuations in the g-7. *Journal of Monetary Economics*, *49*(6), 1131–1159 (cit. on p. 68).

Canova, F., & Pappa, E. (2007). Price differentials in monetary unions: The role of fiscal shocks. *The Economic Journal*, *117*(520), 713–737 (cit. on p. 44).

Chan, K.-S., Ho, L.-H., & Tong, H. (2006). A note on time-reversibility of multivariate linear processes. *Biometrika*, *93*(1), 221–227 (cit. on pp. 5, 17, 20, 45, 50).

Chen, B., Choi, J., & Escanciano, J. C. (2017). Testing for fundamental vector moving average representations. *Quantitative Economics*, *8*(1), 149–180 (cit. on pp. 22, 30).

Christiano, L. J., Eichenbaum, M., & Evans, C. (1996). The effects of monetary policy shocks: Evidence from the flow of funds. *Review of Economics and Statistics*, *78*(1), 16–34 (cit. on p. 3).

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*(3), 287–314 (cit. on pp. 17, 45, 49, 70, 93).

d'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, *58*(2), 341–348 (cit. on p. 46).

Dedecker, J., Doukhan, P., Lang, G., Rafael, L. R. J., Louhichi, S., & Prieur, C. (2007). Weak dependence. In *Weak dependence: With examples and applications* (pp. 9–20). Springer. (Cit. on p.).

Deistler, M., & Hannan, E. (1981). Some properties of the parameterization of arma systems with unknown order. *Journal of Multivariate Analysis*, *11*(4), 474–484 (cit. on p. 10).

Deistler, M., & Hannan, E. (1988). The statistical theory of linear systems. (Cit. on p. 10).

Doukhan, P., & Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, *84*(2), 313–342 (cit. on p.).

Dufour, J.-M., & Pelletier, D. (2021). Practical methods for modeling weak varma processes: Identification, estimation and specification with a macroeconomic application. *Journal of Business & Economic Statistics*, 1–13 (cit. on p. 33).

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 1–26 (cit. on p. 58).

Epps, T. W., & Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, *70*(3), 723–726 (cit. on p. 46).

Fernández-Villaverde, J., Hurtado, S., & Nuño, G. (2018). Financial frictions and the wealth distribution (cit. on p. 69).

Fry, R., & Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, *49*(4), 938–60 (cit. on p. 80).

Funovits, B. (2020). Identifiability and estimation of possibly non-invertible svarma models: A new parametrisation. *arXiv preprint arXiv:2002.04346* (cit. on pp. 15, 17, 22).

Gali, J. (1992). How well does the is-lm model fit postwar us data? *The Quarterly Journal of Economics*, *107*(2), 709–738 (cit. on pp. 3, 44, 68).

Giurcanu, M. C. (2015). A simulation algorithm for non-causal varma processes. *Statistics & Probability Letters*, *98*, 65–72 (cit. on p. 14).

Gohberg, I. [I.], Lancaster, P., & Rodman, L. (1982). *Matrix polynomials*. Society for Industrial; Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104). https://books.google.es/books?id=KwEItnMvwbgC. (Cit. on p. 14)

Gohberg, I. [Israel], Lancaster, P., & Rodman, L. (2005). *Matrix polynomials*. Springer. (Cit. on p. 14).

Golub, G. H., & Van Loan, C. F. (2012). *Matrix computations* (Vol. 3). JHU press. (Cit. on p. 55).

Gouriéroux, C., Monfort, A., & Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics*, *196*(1), 111–126 (cit. on pp. 6, 69, 84, 85).

Gouriéroux, C., Monfort, A., & Renne, J.-P. (2019). Identification and estimation in non-fundamental structural varma models. *The Review of Economic Studies* (cit. on pp. 6, 17, 21, 22).

Gouriéroux, C., Monfort, A., & Renne, J.-P. (2020). Identification and estimation in non-fundamental structural varma models. *The Review of Economic Studies*, *87*(4), 1915–1953 (cit. on pp. 45, 47–50, 69).

Granziera, E., Moon, H. R., & Schorfheide, F. (2018). Inference for vars identified with sign restrictions. *Quantitative Economics*, *9*(3), 1087–1121 (cit. on pp. 76, 80).

Guay, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics*, *225*(1), 27–46 (cit. on pp. x, 6, 45–49, 52, 54, 61, 64, 65, 67).

Hall, F., & Li, Z. (2007). Sign pattern matrices, handbook of linear algebra, simon and hall. (Cit. on p. 90).

Hall, P., & Welsh, A. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, *70*(2), 485–489 (cit. on p. 46).

Hall, P., & Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, 757–762 (cit. on p. 57).

Hallin, M., & Mehta, C. (2015). R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, *110*(509), 218–232 (cit. on pp. 20, 70, 85).

Hamilton, J. D. (1994). *Time series analysis*. Princeton university press. (Cit. on p. 13).

Hannan, E. (1976). The identification and parameterization of armax and state space forms. *Econometrica: Journal of the Econometric Society*, 713–723 (cit. on pp. 10, 14).

Herwartz, H., & Lütkepohl, H. (2014). Structural vector autoregressions with markov switching: Combining conventional with statistical identification of shocks. *Journal of Econometrics*, *183*(1), 104–116 (cit. on pp. 44, 69, 84).

Hoffmann-Jørgensen, J. (1985). Necessary and sufficient condition for the uniform law of large numbers. In *Probability in banach spaces v* (pp. 258–272). Springer. (Cit. on p.).

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association*, *94*(448), 1201–1220 (cit. on pp. 7, 17, 20, 25).

Hong, Y., & Lee, Y.-J. (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *The Review of Economic Studies*, *72*(2), 499–541 (cit. on p. 23).

Jammalamadaka, S. R., Rao, T. S., & Terdik, G. (2006). Higher order cumulants of random vectors and applications to statistical inference and time series. *Sankhyā: The Indian Journal of Statistics*, 326–356 (cit. on pp. 51, 52, 77, 78).

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163–172 (cit. on p. 51).

Jordá, Ó. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, *95*(1), 161–182 (cit. on p. 12).

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, *133*(1), 97–126 (cit. on pp. 47, 55–57).

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, *4*, 83–91 (cit. on p. 46).

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2004). *Continuous multivariate distributions, volume 1: Models and applications*. John Wiley & Sons. (Cit. on p. 22).

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, 1217–1241 (cit. on p. 58).

Lanne, M., & Luoto, J. (2019). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, (just-accepted), 1–35 (cit. on p. 70).

Lanne, M., & Luoto, J. (2021). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, *39*(1), 69–81 (cit. on pp. 5, 6, 83).

Lanne, M., & Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking*, *40*(6), 1131–1149 (cit. on p. 84).

Lanne, M., Lütkepohl, H., & Maciejowska, K. (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control*, *34*(2), 121–131 (cit. on pp. 45, 49).

Lanne, M., Meitz, M., & Saikkonen, P. (2017). Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics*, *196*(2), 288–304 (cit. on pp. 5, 6, 20–22, 45, 49, 69, 84, 85).

Lanne, M., & Saikkonen, P. (2013). Noncausal vector autoregression. *Econometric Theory*, 447–481 (cit. on pp. 6, 14).

Lobato, I. N., & Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, *20*(4), 671–689 (cit. on pp. 46, 47, 51).

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media. (Cit. on pp. 10, 69).

Mainassara, Y. B., & Francq, C. (2011). Estimating structural varma models with uncorrelated but non-independent error terms. *Journal of Multivariate Analysis*, *102*(3), 496–505 (cit. on pp. 14, 48).

Marcinkiewicz, J. (1939). Sur une propriété de la loi de gauss. *Mathematische Zeitschrift*, *44*(1), 612–618 (cit. on pp. 46, 77).

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68–78 (cit. on p. 46).

Maxand, S. (2018). Identification of independent structural shocks in the presence of multiple gaussian components. *Econometrics and Statistics* (cit. on p. 52).

Maxand, S. (2020). Identification of independent structural shocks in the presence of multiple gaussian components. *Econometrics and Statistics*, *16*, 55–68 (cit. on pp. 45–48, 52, 61, 69, 93).

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American economic review*, *103*(4), 1212–47 (cit. on pp. 4, 44).

Mertens, K., & Ravn, M. O. (2014). A reconciliation of svar and narrative estimates of tax multipliers. *Journal of Monetary Economics*, *68*, S1–S19 (cit. on p. 44).

Mountford, A., & Uhlig, H. (2009). What are the effects of fiscal policy shocks? *Journal of applied econometrics*, *24*(6), 960–992 (cit. on p. 3).

Nobel, A., & Dembo, A. (1993). A note on uniform laws of averages for dependent processes. *Statistics & Probability Letters*, *17*(3), 169–172 (cit. on p.).

Nordhausen, K., Oja, H., Tyler, D. E., & Virta, J. (2017). Asymptotic and bootstrap tests for the dimension of the non-gaussian subspace. *IEEE Signal Processing Letters*, *24*(6), 887–891 (cit. on pp. 46, 52, 57, 58, 64).

Olea, J. L. M., Stock, J. H., & Watson, M. W. (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, *225*(1), 74–87 (cit. on p. 4).

Peersman, G. (2005). What caused the early millennium slowdown? evidence based on vector autoregressions. *Journal of Applied Econometrics*, *20*(2), 185–207 (cit. on p. 69).

Peskir, G., & Weber, M. (1994). Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Proc. Funct. Anal. IV (Dubrovnik 1993)*, *43*, 165–190 (cit. on p.).

Petrochilos, N., & Comon, P. (2006). A zero-cumulant random variable and its applications. *Signal processing*, *86*(11), 3334–3338 (cit. on p. 77).

Plagborg-Møller, M., & Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, *89*(2), 955–980 (cit. on p. 12).

Politis, D. N., & Romano, J. P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *The Annals of Statistics*, 1985–2007 (cit. on p. 58).

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*(428), 1303–1313 (cit. on p. 58).

Portier, F., & Delyon, B. (2014). Bootstrap testing of the rank of a matrix via least-squared constrained estimation. *Journal of the American Statistical Association*, *109*(505), 160–172 (cit. on p. 57).

Ramey, V. A. (2011). Identifying government spending shocks: It's all in the timing. *The Quarterly Journal of Economics*, *126*(1), 1–50 (cit. on p. 4).

Ramey, V. A., & Shapiro, M. D. (1998). Costly capital reallocation and the effects of government spending. *Carnegie-Rochester conference series on public policy*, *48*, 145–194 (cit. on p. 4).

Ravenna, F. (2007). Vector autoregressions and reduced form representations of dsge models. *Journal of monetary economics*, *54*(7), 2048–2064 (cit. on p. 4).

Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Science & Business Media. (Cit. on p. 10).

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, *85*(4), 777–792 (cit. on pp. 69, 84).

Rockafellar, R. T. [R Tyrrell]. (1970). *Convex analysis* (Vol. 28). Princeton university press. (Cit. on p. 80).

Rockafellar, R. T. [Ralph T]. (1969). The elementary vectors of a subspace of rˆ n combinatorial mathematics and its applications. *Proc. of the Chapel Hill Conf.*, 104–127 (cit. on p.).

Romer, C. D., & Romer, D. H. (1989). Does monetary policy matter? a new test in the spirit of friedman and schwartz. *NBER macroeconomics annual*, *4*, 121–170 (cit. on pp. 4, 44, 69).

Romer, C. D., & Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, *100*(3), 763–801 (cit. on pp. 4, 44).

Rosenblatt, M. (2000). *Gaussian and non-gaussian linear time series and random fields*. Springer Science & Business Media. (Cit. on pp. 5, 22).

Roth, M. (2012). *On the multivariate t distribution*. Linköping University Electronic Press. (Cit. on p. 22).

Rubio-Ramirez, J. F., Waggoner, D. F., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, *77*(2), 665–696 (cit. on pp. 3, 68, 76, 88).

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611 (cit. on p. 46).

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48 (cit. on pp. 3, 44, 68).

Slyusar, V. (1999). A family of face products of matrices and its properties. *Cybernetics and Systems Analysis*, *35*(3), 379–384 (cit. on p. 28).

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, *19*(2), 279–281 (cit. on p. 46).

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, *15*(4), 101–115 (cit. on p. 68).

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, *128*(610), 917–948 (cit. on pp. 4, 12, 44, 69).

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, *52*(2), 381–419 (cit. on pp. 3, 44, 69, 74, 76).

Velasco, C. (2020). Identification and estimation of structural varma models using higher order dynamics. *arXiv preprint arXiv:2009.04428* (cit. on p.).

Velasco, C. (2022a). Estimation of time series models using residuals dependence measures. *The Annals of Statistics*, *50*(5), 3039–3063 (cit. on pp. 5, 6, 15, 17, 27).

Velasco, C. (2022b). Identification and estimation of structural varma models using higher order dynamics. *Journal of Business & Economic Statistics*, (just-accepted), 1–36 (cit. on pp. 5, 6, 15, 17, 27, 45, 47–50, 70).

5. APPENDIX - CHAPTER I

5.1. Proof Theorem (2.1)

The population loss function, defined in section (2.2.5), is

$$\begin{aligned}\mathscr{L}_0(\boldsymbol{\vartheta}) &= \frac{\pi}{3}\int \left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2) + \frac{1}{2\pi}\sum_{j=1}^{\infty}\frac{1}{j^2}\int \left|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2)\\
&= \frac{\pi}{3}\int \left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \left\{\prod_{i=1}^{d}\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},\tau_{2i}) - \prod_{i=1}^{d}\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i})\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2i})\right\}\right|^2 W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2)\\
&\quad + \frac{1}{2\pi}\sum_{j=1}^{\infty}\frac{1}{j^2}\int \left|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 W(d\boldsymbol{\tau}_1,d\boldsymbol{\tau}_2),\end{aligned}$$

it is evident that $\mathscr{L}_0(\boldsymbol{\vartheta}) \geq 0$ for any $\boldsymbol{\vartheta} \in \mathcal{V}$. According to our definition of model residuals, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ and the identification Assumption (2.2), then $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0) = \boldsymbol{\varepsilon}_t$ and, therefore, $\mathscr{L}_0(\boldsymbol{\vartheta}_0) = 0$.

Now, the key question is whether there exists $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0 \in \mathcal{V}$ such that $\mathscr{L}_0(\boldsymbol{\vartheta}) = 0$ or not. Given that our loss function is a distance, this implies that $\mathscr{L}(\boldsymbol{\vartheta}) = 0$ if and only if

$$\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \left\{\prod_{i=1}^{d}\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},\tau_{2i}) - \prod_{i=1}^{d}\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i})\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2i})\right\} = 0 \tag{5.1}$$

$$\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = 0, \text{ for all } j \in \mathbb{Z}\setminus\{0\} \tag{5.2}$$

thus, our problem is equivalent to look if there exists $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_0 \in \mathcal{V}$, such that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) = \boldsymbol{\Psi}^{-1}(\boldsymbol{\vartheta}_1;L)\boldsymbol{\Psi}(\boldsymbol{\vartheta}_0;L)\boldsymbol{\varepsilon}_t$ satisfies conditions (5.1)-(5.2). Chan et al. (2006) demonstrated that when the random vector $\boldsymbol{\varepsilon}_t$ (and each of its components) follows a non-Gaussian distribution with third-order cumulants different from zero, the equivalence holds if and only if $\boldsymbol{\delta}_j(\boldsymbol{\vartheta}_1) = \boldsymbol{\delta}_{j-m}(\boldsymbol{\vartheta}_0)\boldsymbol{P}$ for some integer m and $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) = \boldsymbol{P}\boldsymbol{\varepsilon}_{t-m}(\boldsymbol{\vartheta}_0)$, where $\boldsymbol{P}$ is a $d \times d$ permutation matrix. In this proof we follow a different strategy. We show that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) = \boldsymbol{\delta}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_t$, and two situations arise

(i) $\boldsymbol{\delta}(L;\boldsymbol{\vartheta})$ is possibly two sided filter;

(ii) $\boldsymbol{\delta}(L;\boldsymbol{\vartheta})$ is a constant matrix.

Let analyze the first case. For gaining intitution, we study first the following situation

$\boldsymbol{\delta}(\boldsymbol{\vartheta};L)=\sum_{l=-k_1}^{k_1}\boldsymbol{\delta}_l(\boldsymbol{\vartheta})L^l$ with $k_1>0$, then

$$\begin{aligned}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)&=\boldsymbol{\delta}_{-k_1}(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t+k_1}+\cdots+\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_t+\cdots+\boldsymbol{\delta}_{k_1}(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-k_1}\\ \boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta}_1)&=\boldsymbol{\delta}_{-k_1}(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-|j|+k_1}+\cdots+\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-|j|}+\cdots+\boldsymbol{\delta}_{k_1}(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-|j|-k_1}\end{aligned}$$

when $|j|>2k_1$, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\perp\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})$, implying that $i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\perp i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})$ and $\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=0$. Now, when $|j|\le 2k_1$ we have to check if it is possible to find $\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=0$.

1. We start taking $k_1=1$, i.e. $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)=\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t+1}+\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_t+\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-1}$. Computing $\sigma_1^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ and $\sigma_2^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$:

$$\begin{aligned}\sigma_1^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=&\,\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_2'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\{\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)+\boldsymbol{\tau}_2'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)+\boldsymbol{\tau}_2'\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1))\\&-\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_2'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_2'\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1))\}\\ \sigma_2^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=&\,\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_2'\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_2'\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1))\{\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)+\boldsymbol{\tau}_2'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\\&-\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_1'\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1))\varphi^{\boldsymbol{\varepsilon}}(\boldsymbol{\tau}_2'\boldsymbol{\delta}_{-1}(\boldsymbol{\vartheta}_1))\}\end{aligned}$$

   and these two dependence measures are zero if and only if, for any four vectors $\mathbf{x}_1,\mathbf{x}_2\in\mathbb{R}^d$

$$\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1'+\mathbf{x}_2')=\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1')\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_2') \tag{5.3}$$

$$\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1'+\mathbf{x}_2')\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_3'+\mathbf{x}_4')=\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1')\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_2')\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_3')\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_4') \tag{5.4}$$

   These condition holds in the following cases:

(a) Trivially, when either $\mathbf{x}_1=\mathbf{0}$ or $\mathbf{x}_2=\mathbf{0}$; and either $\mathbf{x}_3=\mathbf{0}$ or $\mathbf{x}_4=\mathbf{0}$.

(b) If $\boldsymbol{\varepsilon}_t$ were Gaussian distributed with $\mathbb{E}(\boldsymbol{\varepsilon}_t)=\mathbf{0}$ and $\mathbb{V}\mathrm{ar}(\boldsymbol{\varepsilon}_t)=\boldsymbol{I}_d$, its characteristic function is $\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1'+\mathbf{x}_2')=e^{-\frac{1}{2}(\mathbf{x}_1+\mathbf{x}_2)'(\mathbf{x}_1+\mathbf{x}_2)}$ and it would be equal to $\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1')\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_2')$ if and only if $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal.

(c) If $\boldsymbol{\varepsilon}_t$ were Gaussian distributed, with $\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_1'+\mathbf{x}_2')=e^{-\frac{1}{2}(\mathbf{x}_1+\mathbf{x}_2)'(\mathbf{x}_1+\mathbf{x}_2)}$ and $\varphi^{\boldsymbol{\varepsilon}}(\mathbf{x}_3'+\mathbf{x}_4')=e^{-\frac{1}{2}(\mathbf{x}_3+\mathbf{x}_4)'(\mathbf{x}_3+\mathbf{x}_4)}$, then (5.4) holds when $\mathbf{x}_1'\mathbf{x}_2+\mathbf{x}_3'\mathbf{x}_4=\mathbf{0}$ or, equivalently, when $\begin{bmatrix}\mathbf{x}_1' & \mathbf{x}_3'\end{bmatrix}'$ is orthogonal to $\begin{bmatrix}\mathbf{x}_2' & \mathbf{x}_4'\end{bmatrix}'$.

Since only Gaussian distributions produce characteristic function to be $e^{-\frac{1}{2}(\mathbf{x}_1+\mathbf{x}_2)'(\mathbf{x}_1+\mathbf{x}_2)}$, non-Gaussian distributions will satisfy the condition (5.3) only in case (a). Now, in the Gaussian case, if $\delta(L,\boldsymbol{\vartheta}_1)$ were not an *all-pass* filter, equation (5.3) would not hold as well.

On the contrary, when $\delta(L,\boldsymbol{\vartheta}_1)$ generates uncorrelated model residuals, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)$, then

$$\mathbb{E}\left[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}'_{t-2}(\boldsymbol{\vartheta}_1)\right] = \boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_{-1}(\boldsymbol{\vartheta}_1) = \mathbf{0}$$

$$\mathbb{E}\left[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}'_{t-1}(\boldsymbol{\vartheta}_1)\right] = \boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_{-1}(\boldsymbol{\vartheta}_1) + \boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_0(\boldsymbol{\vartheta}_1) = \begin{pmatrix}\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1) & \boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}'_{-1}(\boldsymbol{\vartheta}_1)\\ \\ \boldsymbol{\delta}'_0(\boldsymbol{\vartheta}_1)\end{pmatrix} = \mathbf{0}$$

which would imply that

$$\boldsymbol{\tau}'_1\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_{-1}(\boldsymbol{\vartheta}_1)\boldsymbol{\tau}_2 = 0$$

$$\boldsymbol{\tau}'_1\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_{-1}(\boldsymbol{\vartheta}_1)\boldsymbol{\tau}_2 + \boldsymbol{\tau}'_1\boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_0(\boldsymbol{\vartheta}_1)\boldsymbol{\tau}_2 = \boldsymbol{\tau}'_1\begin{pmatrix}\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1) & \boldsymbol{\delta}_1(\boldsymbol{\vartheta}_1)\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}'_{-1}(\boldsymbol{\vartheta}_1)\\ \\ \boldsymbol{\delta}'_0(\boldsymbol{\vartheta}_1)\end{pmatrix}\boldsymbol{\tau}_2 = 0$$

making $\sigma_1^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \sigma_2^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = 0$.

A similar idea can be applied when $1 < k_1 < \infty$. In such case, we need to analyze up to $\sigma_{k+1}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$. When $k_1 \to \infty$, we take adavantage of the functional form of characteristic function for a Gaussian process and the fact that $\boldsymbol{\gamma}_h^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_1)} = \sum_{j=-\infty}^{\infty}\boldsymbol{\delta}_j(\boldsymbol{\vartheta}_1)\boldsymbol{\delta}'_{j-h}(\boldsymbol{\vartheta}_1) = \mathbf{0}$ for all $h \in \mathbb{Z}$.

Now, when $\boldsymbol{\delta}(\boldsymbol{\vartheta}_1;L) = \boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)$, it is clear that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) \perp \boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta}_1)$ for all $|j| > 0$, then $\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = 0$. Hence, in this case, any rotation of structural innovations, $\boldsymbol{\varepsilon}_t$, preserves the pairwise independence. Now, let analyze if condition (5.1) under this case. Since model residuals must be standardized errors, $\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_1) = \boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)$ is an orthonormal matrix, then j-th and i-th element of $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$, with $i \neq j$, are

$$\varepsilon_{t,m_1} = [\boldsymbol{\delta}_0(\boldsymbol{\vartheta})]_{m_1}\boldsymbol{\varepsilon}_t$$
$$\varepsilon_{t,m_2} = [\boldsymbol{\delta}_0(\boldsymbol{\vartheta})]_{m_2}\boldsymbol{\varepsilon}_t$$

In Gaussian case it would be enough that $[\boldsymbol{\delta}_0(\boldsymbol{\vartheta})]_{m_1}[\boldsymbol{\delta}_0(\boldsymbol{\vartheta})]'_{m_2} = 0$ for condition (5.1) to be fulfilled. However, when innovations are non-Gaussian, the condition $[\boldsymbol{\delta}_0(\boldsymbol{\vartheta})]_{m_1}[\boldsymbol{\delta}_0(\boldsymbol{\vartheta})]'_{m_2} = 0$ does not imply marginal independence, but only assures $\mathbb{C}\text{ov}(\varepsilon_{t,m_1},\boldsymbol{\varepsilon}_{t,m_2}) = 0$, thus condition (5.2) does not hold.

Therefore, under ICA, it is necessary and sufficient that

$$\boldsymbol{\varepsilon}_{t,m_1} = \delta^{(0)}_{m_1n_1}(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t,n_1}$$
$$\boldsymbol{\varepsilon}_{t,m_2} = \delta^{(0)}_{m_2n_2}(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t,n_2}$$

where $m_1 \neq m_2 \neq n_1 \neq n_2$, and $\delta^{(0)}_{mn}(\boldsymbol{\vartheta})$ is an sign-scaling factor. Thus, $\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)$ should be a sign permutation matrix. This implies that elements of m_1-th row of $\boldsymbol{\delta}_0(\boldsymbol{\vartheta})$ is

$$\delta^{(0)}_{ij}(\boldsymbol{\vartheta}) = \begin{cases} \delta, & \text{for some } m_1 \\ 0, & \text{otherwise} \end{cases} \tag{5.5}$$

with $\delta \in \{-1, 1\}$.

Given our parameterization, we look for that $\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_1) = \boldsymbol{P}\boldsymbol{D}^{-1}$, with $\boldsymbol{D}$ an sign-scaling, diagonal matrix and $\boldsymbol{P}$ a permutation matrix. Since, $\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_1) = \boldsymbol{\Psi}^{-1}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)$ and $\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}) = \boldsymbol{\Phi}^{-1}(L;\boldsymbol{\vartheta})\boldsymbol{\Theta}(L;\boldsymbol{\vartheta})\boldsymbol{B}(\boldsymbol{\vartheta})$, then

$$\begin{aligned}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_1) &= \boldsymbol{\Psi}^{-1}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\ &= \boldsymbol{B}^{-1}(\boldsymbol{\vartheta}_1)\boldsymbol{\Theta}^{-1}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\Phi}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\Phi}^{-1}(L;\boldsymbol{\vartheta}_0)\boldsymbol{\Theta}(L;\boldsymbol{\vartheta}_0)\boldsymbol{B}(\boldsymbol{\vartheta}_0)\end{aligned}$$

From what we discuss in the previous part, $\boldsymbol{\Theta}^{-1}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\Phi}(L;\boldsymbol{\vartheta}_1)\boldsymbol{\Phi}^{-1}(L;\boldsymbol{\vartheta}_0)\boldsymbol{\Theta}(L;\boldsymbol{\vartheta}_0) = \boldsymbol{I}_d$, otherwise pairwise independence does not hold. Thus:

$$\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_1) = \boldsymbol{B}^{-1}(\boldsymbol{\vartheta}_1)\boldsymbol{B}(\boldsymbol{\vartheta}_0) = \boldsymbol{P}\boldsymbol{D}^{-1}$$

from this it is evident that

$$\boldsymbol{B}(\boldsymbol{\vartheta}_1) = \boldsymbol{B}(\boldsymbol{\vartheta}_0)\boldsymbol{D}\boldsymbol{P}'$$

By the definition of vector $\boldsymbol{\vartheta}_1$, and calling $[\boldsymbol{\vartheta}_1]_{k:k'}$ the elements from position k to k' of vector $\boldsymbol{\vartheta}_1$ and $\boldsymbol{b}_0 = \text{vec}(\boldsymbol{B}(\boldsymbol{\vartheta}_0))$, we have:

$$[\boldsymbol{\vartheta}_1]_{d^2(p+q)+1:d^2(p+q+1)} = \text{vec}(\boldsymbol{B}(\boldsymbol{\vartheta}_1)) = [\boldsymbol{P}\boldsymbol{D} \otimes \boldsymbol{I}_d]\,\boldsymbol{b}_0 = [\boldsymbol{P}\boldsymbol{D} \otimes \boldsymbol{I}_d] \cdot [\boldsymbol{\vartheta}_0]_{d^2(p+q)+1:d^2(p+q+1)}$$

therefore, $= \mathscr{L}_0(\boldsymbol{\vartheta}_1) = 0$ if and only if

$$\boldsymbol{\vartheta} = \bar{\boldsymbol{P}}\boldsymbol{\vartheta}_0 \tag{5.6}$$

where $\bar{\boldsymbol{P}} = \begin{bmatrix} \boldsymbol{I}_{d^2(p+q)} & \boldsymbol{0}_{d^2(p+q),d^2} \\ \boldsymbol{0}_{d^2,d^2(p+q)} & \boldsymbol{P}\boldsymbol{D} \otimes \boldsymbol{I}_d \end{bmatrix}$, and $\boldsymbol{P} \in \mathscr{P}(d)$, the space of permutation matrices of size d. $\square$

5.2. Proof Theorem (2.2)

We proceed as in Appendix (5.1). From the loss function, defined in section (2.2.6), we have that $\mathscr{R}_0(\boldsymbol{\vartheta}) = 0$ if and only if

$$\Delta^{(1,0)}\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) = \mathbf{0},\ j > 0 \tag{5.7}$$

$$\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \begin{bmatrix} \mathbb{C}\mathrm{ov}\left(\boldsymbol{\varepsilon}_{t,1}(\boldsymbol{\vartheta}),\ z_t(\tau_{2,1}, \boldsymbol{\vartheta})\right) \prod_{j\neq 1} \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,i}) \\ \vdots \\ \mathbb{C}\mathrm{ov}\left(\boldsymbol{\varepsilon}_{t,d}(\boldsymbol{\vartheta}),\ z_t(\tau_{2,d}, \boldsymbol{\vartheta})\right) \prod_{j\neq d} \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2,i}) \end{bmatrix} = \mathbf{0} \tag{5.8}$$

by our parameterization and the identification assumption, conditions (5.7)-(5.8) are satisfied for $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. Then, it is straightforward to see that $\mathscr{R}_0(\boldsymbol{\vartheta}_0) = 0$.

Now, the question is if there exists $\boldsymbol{\vartheta}_1 \neq \boldsymbol{\vartheta}_0 \in \mathcal{V}$ such that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) = \boldsymbol{\delta}(L; \boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_t$ satisfies conditions (5.7)-(5.8). We start by analyzing condition (5.7).

If $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)$ satisfied equation (5.7), then it should hold $\mathbb{E}[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) \mid \boldsymbol{\varepsilon}_{t-j}(\boldsymbol{\vartheta}_1)] = 0$ for $j > 0$. Let analyze this latter condition.

$$\begin{aligned} \mathbb{E}[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) \mid \boldsymbol{\varepsilon}_{t-j}(\boldsymbol{\vartheta}_1)] &= \mathbb{E}\left[\sum_{l=-\infty}^{\infty} \boldsymbol{\delta}_l(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-l} \,\middle|\, \sum_{l=-\infty}^{\infty} \boldsymbol{\delta}_l(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-j-l} \right] \\ &= \sum_{l=-\infty}^{\infty} \boldsymbol{\delta}_l(\boldsymbol{\vartheta}_1)\mathbb{E}\left[\boldsymbol{\varepsilon}_{t-l} \,\middle|\, \sum_{l=-\infty}^{\infty} \boldsymbol{\delta}_l(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-j-l} \right] \end{aligned}$$

Taking a similar argument like Rosenblatt (2000) applied. The former conditional expectation satisfies the p.m.d.s property if and only if $\mathbb{E}\left[\boldsymbol{\varepsilon}_{t-l} \,\middle|\, \sum_{l=-\infty}^{\infty} \boldsymbol{\delta}_l(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-j-l}\right]$ is a linear function. When $\{\boldsymbol{\varepsilon}_t\}$ is a Gaussian process, then this expectation is linear. In case of non-Gaussian innovations, the answer depends. There exists some multivariate non-Gaussian distributions, such as the multivariate Pearson distribution, which marginals are Gaussian. Rosenblatt (2000) establish the sufficient condition of having, at least, third order cumulant different from zero for all the components in the non-Gaussian innovation vector. A similar characterization of non-Gaussianity is made by Chen et al. (2017), who assumes that all the third order cumulants must be different from zero. Velasco (2020) characterizes a non-Gaussian vector by having at least third or fourth order cumulants different from zero. In our case, we only require that necessary condition holds, non-linearity of $\mathbb{E}\left[\boldsymbol{\varepsilon}_{t-l} \,\middle|\, \sum_{l=-\infty}^{\infty} \boldsymbol{\delta}_l(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_{t-j-l}\right]$ is a linear function. When $\{\boldsymbol{\varepsilon}_t\}$.

Now, we analyze the case when $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1) = \boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)\boldsymbol{\varepsilon}_t$. In this situation, it is obvious that $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_1)$ satisfies p.m.d.s property. Thus, condition (5.7) holds for $j > 0$. Since residuals must be standardized, $\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1)$ is an orthonormal matrix. Like in the case of independence sequences, condition (5.8) holds if $\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_1) = \boldsymbol{P}\boldsymbol{D}$, where $\boldsymbol{P}$ a permutation matrix and $\boldsymbol{D}$ a diagonal sign-matrix. Thus, $\boldsymbol{B}(\boldsymbol{\vartheta}_1) = \boldsymbol{B}(\boldsymbol{\vartheta}_0)\boldsymbol{D}\boldsymbol{P}'$. Since, we are using Hallin's permutation selection method, and as it was proved by Lanne, Meitz, et al., 2017, we can pick a unique $\boldsymbol{P}$ and $\boldsymbol{D}$. □

5.3. Proof Theorem (2.3)

By triangle inequality

$$\sup_{\boldsymbol{\vartheta}} |\mathscr{L}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})| \le \sup_{\boldsymbol{\vartheta}} \left|\mathscr{L}_T(\boldsymbol{\vartheta}) - \overline{\mathscr{L}}_T(\boldsymbol{\vartheta})\right| + \sup_{\boldsymbol{\vartheta}} \left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right|$$

where $\overline{\mathscr{L}}_T(\boldsymbol{\vartheta})$ is the same as $\mathscr{L}_T(\boldsymbol{\vartheta})$ but employs population model residuals, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$, instead of sample residuals, $\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})$.

By using results stated in sections (5.3.1) and (5.3.2), it follows

$$\mathbb{E} \sup_{\boldsymbol{\vartheta}} |\mathscr{L}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})| \le o(1) \tag{5.9}$$

Therefore, Theorem (2.3) holds. □

5.3.1. Convergence of $\sup_{\boldsymbol{\vartheta}} \left|\mathscr{L}_T(\boldsymbol{\vartheta}) - \overline{\mathscr{L}}_T(\boldsymbol{\vartheta})\right|$

We have

$$\begin{aligned}\mathscr{L}_T(\boldsymbol{\vartheta}) - \overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) = \frac{\pi}{3} \int \Bigg\{ & \left| \hat{\sigma}_0^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \left\{ \prod_{m=1}^{d} \hat{\varphi}_0^{\varepsilon(\vartheta)}(\tau_{1m}, \tau_{2m}) - \prod_{m=1}^{d} \hat{\varphi}_0^{\varepsilon(\vartheta)}(\tau_{1m}, 0) \hat{\varphi}_0^{\varepsilon(\vartheta)}(0, \tau_{2m}) \right\} \right|^2 \\ & - \left| \overline{\sigma}_0^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \left\{ \prod_{m=1}^{d} \overline{\varphi}_0^{\varepsilon(\vartheta)}(\tau_{1m}, \tau_{2m}) - \prod_{m=1}^{d} \overline{\varphi}_0^{\varepsilon(\vartheta)}(\tau_{1m}, 0) \overline{\varphi}_0^{\varepsilon(\vartheta)}(0, \tau_{2m}) \right\} \right|^2 \Bigg\} W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2) \\ & + \frac{1}{2\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int \left\{ \left| \hat{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 - \left| \overline{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 \right\} W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2)\end{aligned}$$

Now, notice for $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{C}$ that $|\mathbf{z}_1|^2 - |\mathbf{z}_2|^2 = (|\mathbf{z}_1| - |\mathbf{z}_2|)(|\mathbf{z}_1| + |\mathbf{z}_2|)$. By triangle inequality $|\mathbf{z}_1| - |\mathbf{z}_2| \le |\mathbf{z}_1 - \mathbf{z}_2|$ and $|\mathbf{z}_1| + |\mathbf{z}_2| \le |\mathbf{z}_1 - \mathbf{z}_2| + 2|\mathbf{z}_2|$; thus $|\mathbf{z}_1|^2 - |\mathbf{z}_2|^2 \le |\mathbf{z}_1 - \mathbf{z}_2|^2 + 2|\mathbf{z}_1 - \mathbf{z}_2| \cdot |\mathbf{z}_2|$. Hence

$$\left| \left| \hat{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 - \left| \overline{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 \right| \le \left| \hat{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \overline{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 + 2 \left| \hat{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \overline{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right| \cdot \left| \overline{\sigma}_{|j|}^{\varepsilon(\vartheta)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|$$

From Lemma (5.9.3), we obtain

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\overline{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|\leq C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}(T-|j|)^{-1},\ j\neq 0 \tag{5.10}$$

besides, note that

$$\begin{aligned}\left|\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\prod_{m=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},0)\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2m})\right\}\right|^2 &\leq \left|\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2+\left|\hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\\ &\quad+2\left|\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\cdot\left|\hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\\ \left|\overline{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\prod_{m=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},0)\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2m})\right\}\right|^2 &\leq \left|\overline{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2+\left|\overline{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\\ &\quad+2\left|\overline{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\cdot\left|\overline{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\end{aligned}$$

where $\hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\prod_{m=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},0)\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2m})$ and $\overline{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\prod_{m=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},0)\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2m})$.

Thus, applying a similar procedure than in previous parts, we get:

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\Bigg|\left|\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\prod_{i=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},\tau_{2i})-\prod_{i=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},0)\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2i})\right\}\right|^2&\\ -\left|\overline{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\prod_{i=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},\tau_{2i})-\prod_{i=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},0)\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2i})\right\}\right|^2\Bigg|&\leq C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}T^{-1}\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\mathscr{L}_T(\boldsymbol{\vartheta})-\overline{\mathscr{L}}_T(\boldsymbol{\vartheta})\right| &\leq \frac{\pi}{3}\int\left(C[\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2]T^{-2}\right)W(d\boldsymbol{\tau}_1)W(d\boldsymbol{\tau}_2)\\ &\quad+\frac{1}{2\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\int\left(C[\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2](T-|j|)^{-2}\right)W(d\boldsymbol{\tau}_1)W(d\boldsymbol{\tau}_2)\\ &\leq C\cdot T^{-2}+C\sum_{j=1}^{T-1}\frac{1}{(T-|j|)^2j^2}\leq C\left\{\frac{1}{T^2}+\frac{\ln(T-1)}{T^3}\right\}\end{aligned}$$

5.3.2. Convergence of $\sup_{\boldsymbol{\vartheta}} \left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right|$

Like in part (5.3.1), we have

$$\sup_{\boldsymbol{\vartheta}} \left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right| \leq \sup_{\boldsymbol{\vartheta}} \left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{L}}_T(\boldsymbol{\vartheta})\right| + \sup_{\boldsymbol{\vartheta}} \left|\tilde{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right|$$

where $\tilde{\mathscr{L}}_T(\boldsymbol{\vartheta})$ is similar to $\overline{\mathscr{L}}_T(\boldsymbol{\vartheta})$ but defined using the population centered $\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ instead of its sample counterparts. Thus, it is sufficient if we prove the uniform convergence of each term in the right hand side. Using the results stated in sections (5.3.2)-(5.3.2), therefore

$$\mathbb{E} \sup_{\boldsymbol{\vartheta}} \left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right| \leq o(1) \tag{5.11}$$

Uniform convergence of $\left|\tilde{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right|$

By definition

$$\begin{aligned}\tilde{\mathscr{L}}_T(\boldsymbol{\vartheta}) = \frac{\pi}{3} \int \left| \sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \left\{ \prod_{m=1}^{d} \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m}, \tau_{2m}) - \prod_{m=1}^{d} \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m}) \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2m}) \right\} \right|^2 W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2) \\ + \frac{1}{2\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int \left| \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2)\end{aligned}$$

then

$$\tilde{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta}) = \frac{1}{2\pi} \left\{ \sum_{j=T}^{\infty} \frac{1}{j^2} \int \left| \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 W(d\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right\}$$

and

$$\begin{aligned}\sup_{\boldsymbol{\vartheta}} \left|\tilde{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \mathscr{L}_0(\boldsymbol{\vartheta})\right| &\leq \frac{1}{2\pi} \left\{ \sum_{j=T}^{\infty} \frac{1}{j^2} \int \sup_{\boldsymbol{\vartheta}} \left| \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right|^2 W(d\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right\} \\ &\leq \frac{1}{2\pi} \left\{ \sum_{j=T}^{\infty} \frac{1}{j^2} \int C \left(\|\boldsymbol{\tau}_1\|^2 + \|\boldsymbol{\tau}_2\|^2 \right) |j|^{2-2\mu_0} W(d\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right\} \\ &\leq C \sum_{j=T}^{\infty} \frac{1}{j^2} |j|^{2-2\mu_0} = C \sum_{j=T}^{\infty} |j|^{-2\mu_0} \\ &\leq C T^{1-2\mu_0} \leq C T^{-3}\end{aligned}$$

where the second inequality employs Lemma (5.9.2) and last line follows from the fact that we are taking $\mu_0 > 2$.

Uniform convergence of $\left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{L}}_T(\boldsymbol{\vartheta})\right|$

Note that

$$\begin{aligned}\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{L}}_T(\boldsymbol{\vartheta}) = \frac{\pi}{3}\int\left\{\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 - \left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right\} W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2)\\ + \frac{1}{2\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\int\left\{|\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right\} W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2)\end{aligned}$$

where $\sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \prod_{m=1}^d \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^d \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m})\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{2m})$. Thus

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{L}}_T(\boldsymbol{\vartheta})\right| \leq \frac{\pi}{3}\int \mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 - \left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right| W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2)\\ + \frac{1}{2\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\int \mathbb{E}\sup_{\boldsymbol{\vartheta}}\left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right| W(d\boldsymbol{\tau}_1, d\boldsymbol{\tau}_2)\end{aligned}$$

using the results below, we get:

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\overline{\mathscr{L}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{L}}_T(\boldsymbol{\vartheta})\right| &\leq \frac{\pi}{3}\int C(\|\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2\|)T^{-1/2}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\\ &\quad + \sum_{j=1}^{T-1}\frac{1}{j^2}\int C(\|\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2\|)\{(T-|j|)^{-1/2} + (T-|j|)^{-1/2}|j|^{2-2\mu_0}\}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\\ &\leq CT^{-1/2} + \sum_{j=1}^{T-1}\frac{1}{j^2}\frac{1}{(T-j)^{1/2}} + \sum_{j=1}^{T-1}\frac{1}{j^{2\mu_0}}\frac{1}{(T-j)^{1/2}} \leq CT^{-1/2} + \sum_{j=1}^{T-1}\frac{1}{j^2}\frac{1}{(T-j)^{1/2}}\\ &\leq CT^{-1/2} + C\frac{\ln(\sqrt{T}) + \ln(\sqrt{T}+1)}{T\sqrt{T}} = o(1)\end{aligned}$$

where before last inequality uses $\mu_0 > 2$.

Convergence of $\sup_{\boldsymbol{\vartheta}}\left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right|$

Using $\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$, which uses population centered $z_t(\boldsymbol{\tau},\boldsymbol{\vartheta})$ instead of its sample counterpart, and applying triangle inequality

$$\left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right| \leq \left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right| + \left||\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right|$$

For the first term in the right hand side (RHS), we have

$$\left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2 - |\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right| \leq \left|\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 + 2\left|\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| \cdot \left|\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|$$

using lemmas (5.9.4) and (5.9.5), and Hölder inequality, we get:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2-|\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right|\leq C\|\boldsymbol{\tau}_1\|(T-|j|)^{-1/2}$$

where we take advantage that $\sup_{\boldsymbol{\vartheta}}\left|\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\leq 4$.

Now, the second term in right hand side can be expressed as

$$\left||\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2-|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right|\leq\left|\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2+2\left|\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\cdot\left|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|$$

using lemmas (5.9.2) and (5.9.5), we get:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left||\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2-|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right|\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)\left((T-|j|)^{-1/2}+(T-|j|)^{-1/2}|j|^{2-2\mu_0}\right)$$

Hence

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left||\bar{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2-|\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)|^2\right|\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)\left((T-|j|)^{-1/2}+(T-|j|)^{-1/2}|j|^{2-2\mu_0}\right)$$

Convergence of $\sup_{\boldsymbol{\vartheta}}\left|\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|$

Like above

$$\begin{aligned}\left|\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|&\\ \leq\left|\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|&\\ +\left|\left|\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|&\end{aligned}$$

Analyzing the first term in the RHS:

$$\begin{aligned}&\left|\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|\\ &\leq\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}\right|^2\\ &+2\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}\right|\cdot\left|\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\end{aligned}$$

notice that

$$\begin{aligned}\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}\right|^2&\leq C\left|\bar{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\\ &+C\left|\bar{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\end{aligned}$$

Notice

$$\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=-\left\{\prod_{m=1}^{d}\bar{\varphi}^{\varepsilon(\boldsymbol{\vartheta})}_0(\tau_{1m},0)-\prod_{m=1}^{d}\varphi^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m})\right\}\left\{\prod_{m=1}^{d}\bar{\varphi}^{\varepsilon(\boldsymbol{\vartheta})}_0(0,\tau_{2m})-\prod_{m=1}^{d}\varphi^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{2m})\right\}$$

hence

$$\left|\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\leq 2\left|\left\{\prod_{m=1}^{d}\bar{\varphi}^{\varepsilon(\boldsymbol{\vartheta})}_0(\tau_{1m},0)-\prod_{m=1}^{d}\varphi^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m})\right\}\right|^2$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\leq C\|\boldsymbol{\tau}_1\|T^{-1/2}$$

where last inequality comes from applying Lemma (5.9.4).

Thus

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}\right|^2\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)T^{-1/2}$$

and taking advantage of uniformly boundedness of $\left|\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2$, therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left|\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)T^{-1/2}$$

Now, the second term in RHS

$$\begin{aligned}&\left|\left|\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|\\
&\qquad\leq\left|\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}\right|^2\\
&\qquad+2\left|\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\left\{\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}\right|\cdot\left|\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\end{aligned}$$

since $\left|\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|$ is bounded for each $\boldsymbol{\vartheta}\in\mathscr{V}$ and after applying Lemma (5.9.4), we get

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left|\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|\leq C(\|\boldsymbol{\tau}_1+\boldsymbol{\tau}_2\|)T^{-1/2}$$

Therefore, taking the convergence results for the two terms in RHS

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left|\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\bar{\sigma}^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2-\left|\sigma^{\varepsilon(\boldsymbol{\vartheta})}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\sigma^{\varepsilon(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right|\leq C(\|\boldsymbol{\tau}_1+\boldsymbol{\tau}_2\|)T^{-1/2}$$

$\square$

5.4. Proof Theorem (2.4)

Assumption 2.1 guarantees that space parameter is compact. Besides, Assumption 2.2 joint with the Hallin and Mehta (2015) procedure assure that $\mathscr{R}_0(\boldsymbol{\vartheta})$ is uniquely minimized at $\boldsymbol{\vartheta}_0$. Now, notice that $\mathscr{R}_T(\boldsymbol{\vartheta})$ is a measurable function for each $\boldsymbol{\vartheta} \in \mathscr{V}$. Moreover, given the transformation we use, $e^{i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}$, which is an smooth function with all derivatives well defined. $\mathscr{L}_T(\boldsymbol{\vartheta})$ is continuous for a provided sample $\{\boldsymbol{Y}_t\}_{t=1}^T$.

Thus, all the requirements of Theorem 4.1 in Amemiya (1985) are fulfilled. Additionally, the potential multiple local minimum we can find in the optimization process, something typical in non-linear optimization problems, comes from the fact that the order (p_+, q_+) is not determined. Then, it is possible to find minimizers in causal/invertible and non-causal/non-invertible situations. However, given the correct order (p_+, q_+), we have a unique minimizer. Thus, applying Theorem 4.1 in Amemiya (1985), we have that $\mathbb{P}\left(\|\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} - \boldsymbol{\vartheta}_0\| > \epsilon\right) \to 0$ as $T \to \infty$. $\square$

5.5. Proof Theorem (2.5)

5.5.1. Convergence of Hessian Matrix

The Hessian matrix is $\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{L}_T(\tilde{\boldsymbol{\vartheta}}_T)$.

Before proceed, let explain the intuition behind the proof. Assume that $f_T(\vartheta)$ is our objective function. Assuming that $\tilde{\vartheta}_T \xrightarrow{p} \vartheta_0$. If we want to assure the convergence of $\ddot{f}_T(\tilde{\vartheta}_T)$ we can do a first order Taylor approximation of $\ddot{f}_T(\tilde{\vartheta}_T)$ around ϑ_0. We would get

$$\ddot{f}_T(\tilde{\vartheta}_T) \approxeq \ddot{f}_T(\vartheta_0) + \dddot{f}_T(\vartheta_0)\cdot\left(\tilde{\vartheta}_T - \vartheta_0\right)$$

Hence, $|\ddot{f}_T(\tilde{\vartheta}_T) - \ddot{f}_T(\vartheta_0)| \leq C|\dddot{f}_T(\vartheta_0)|\cdot|\left(\tilde{\vartheta}_T - \vartheta_0\right)|$. If $|\dddot{f}_T(\vartheta_0)| < \infty$ and given that $|\tilde{\vartheta}_T - \vartheta_0| = o_p(1)$. We would obtain that $|\ddot{f}_T(\tilde{\vartheta}_T) - \ddot{f}_T(\vartheta_0)| \leq o_p(1)$.

With this, and given $\mathbb{E}\ddot{f}_T(\vartheta_0)$, we will get that $\ddot{f}_T(\tilde{\vartheta}_T) \xrightarrow{p} \mathbb{E}\ddot{f}_T(\vartheta_0)$.

Proof:

In our case, given that we have a parameter vector instead of a scalar parameter, we have K^3 third-order derivatives of $\mathscr{L}_T(\boldsymbol{\vartheta})$, thus we take the maximum of these derivatives.

Then

$$\left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \mathscr{L}_T(\tilde{\boldsymbol{\vartheta}}_T) - \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \mathscr{L}_T(\boldsymbol{\vartheta}_0) \right\| \leq C \max_{(k,l,m)\in\{1,\dots,K\}^3} \sup_{\boldsymbol{\vartheta}} \left| \frac{\partial^3}{\partial \vartheta_k \partial \vartheta_l \partial \vartheta_m} \mathscr{L}_T(\boldsymbol{\vartheta}) \right| \| \tilde{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0 \|$$

For simplicity, we work with the model residuals and not with the sample model residuals. This do not affect the proof, but keeps it simple.

Now, notice that:

$$\begin{aligned}
\frac{\partial}{\partial \vartheta_m} \mathscr{L}_T(\boldsymbol{\vartheta}) &= \frac{\pi}{3} \int 2\mathrm{Re}\left\{ \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right) \frac{\partial}{\partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right)} \right\} dW \\
&+ \frac{2}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int 2\mathrm{Re}\left\{ \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \frac{\partial}{\partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)} \right\} dW \\
\frac{\partial^2}{\partial \vartheta_l \partial \vartheta_m} \mathscr{L}_T(\boldsymbol{\vartheta}) &= \frac{\pi}{3} \int 2\mathrm{Re}\left\{ \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right) \frac{\partial^2}{\partial \vartheta_l \partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right)} \right\} dW \\
&+ \frac{\pi}{3} \int 2\mathrm{Re}\left\{ \frac{\partial}{\partial \vartheta_l} \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right) \frac{\partial}{\partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) - \hat{\sigma}^{0,ICA}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right)} \right\} dW \\
+ \frac{2}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int 2\mathrm{Re}&\left\{ \frac{\partial}{\partial \vartheta_l} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \frac{\partial}{\partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)} + \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \frac{\partial^2}{\partial \vartheta_l \partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)} \right\} dW
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^3}{\partial \vartheta_k \partial \vartheta_l \partial \vartheta_m} \mathscr{L}_T(\boldsymbol{\vartheta}) = &\frac{\pi}{3} \int 2\mathrm{Re}\left\{ \frac{\partial}{\vartheta_k} \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right) \frac{\partial^2}{\partial \vartheta_l \partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right)} \right\} dW \\
&+ \frac{\pi}{3} \int 2\mathrm{Re}\left\{ \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right) \frac{\partial^3}{\partial \vartheta_k \partial \vartheta_l \partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right)} \right\} dW \\
&+ \frac{\pi}{3} \int 2\mathrm{Re}\left\{ \frac{\partial^2}{\partial \vartheta_k \partial \vartheta_l} \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right) \frac{\partial}{\partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right)} \right\} dW \\
&+ \frac{\pi}{3} \int 2\mathrm{Re}\left\{ \frac{\partial}{\partial \vartheta_l} \left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right) \frac{\partial^2}{\partial \vartheta_k \partial \vartheta_m} \overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} - \hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right)} \right\} dW \\
&+ \frac{2}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int 2\mathrm{Re}\left\{ \frac{\partial^2}{\partial \vartheta_k \vartheta_l} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \frac{\partial}{\partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}} + \frac{\partial}{\partial \vartheta_k} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \frac{\partial^2}{\partial \vartheta_l \partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}} \right\} dW \\
&+ \frac{2}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int 2\mathrm{Re}\left\{ \frac{\partial}{\partial \vartheta_l} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \frac{\partial^2}{\partial \vartheta_k \partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}} + \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \frac{\partial^3}{\partial \vartheta_k \partial \vartheta_l \partial \vartheta_m} \overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}} \right\} dW
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial}{\partial\vartheta_m}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) &= \frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\cdot\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\cdot\frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\\
\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(.) &= \frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.) - \frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\cdot\hat{\varphi}_{|j|}(.) - \hat{\varphi}_{|j|}(.)\cdot\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\\
&\quad - \frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial}{\partial\vartheta_l}\hat{\varphi}_{|j|}(.) - \frac{\partial}{\partial\vartheta_l}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\\
\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(.) &= \frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.) - \frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\cdot\hat{\varphi}_{|j|}(.) - \hat{\varphi}_{|j|}(.)\cdot\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\\
&\quad - \frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial}{\partial\vartheta_k}\hat{\varphi}_{|j|}(.) - \frac{\partial}{\partial\vartheta_k}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\\
&\quad - \frac{\partial^2}{\partial\vartheta_k\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial}{\partial\vartheta_l}\hat{\varphi}_{|j|}(.) - \frac{\partial^2}{\partial\vartheta_k\partial\vartheta_l}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\\
&\quad - \frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_l}\hat{\varphi}_{|j|}(.) - \frac{\partial}{\partial\vartheta_l}\hat{\varphi}_{|j|}(.)\cdot\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_m}\hat{\varphi}_{|j|}(.)
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) &= \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(m)}(\boldsymbol{\vartheta})\right)\right\}\\
\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) &= \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\ddot{\boldsymbol{\varepsilon}}_t^{(l,m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\ddot{\boldsymbol{\varepsilon}}_{t-|j|}^{(l,m)}(\boldsymbol{\vartheta})\right)\right\}\\
&\quad+\frac{(-1)}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(m)}(\boldsymbol{\vartheta})\right)\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(l)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(l)}(\boldsymbol{\vartheta})\right)\right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) &= \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\dddot{\boldsymbol{\varepsilon}}_t^{(k,l,m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\dddot{\boldsymbol{\varepsilon}}_{t-|j|}^{(k,l,m)}(\boldsymbol{\vartheta})\right)\right\}\\
&\quad+\frac{(-1)}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(k)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(k)}(\boldsymbol{\vartheta})\right)\left(\boldsymbol{\tau}_1'\ddot{\boldsymbol{\varepsilon}}_t^{(l,m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\ddot{\boldsymbol{\varepsilon}}_{t-|j|}^{(l,m)}(\boldsymbol{\vartheta})\right)\right\}\\
&\quad+\frac{(-i)}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\prod_{a\in\{k,l,m\}}\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(a)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(a)}(\boldsymbol{\vartheta})\right)\right\}\\
&\quad+\frac{(-1)}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\ddot{\boldsymbol{\varepsilon}}_t^{(k,m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\ddot{\boldsymbol{\varepsilon}}_{t-|j|}^{(k,m)}(\boldsymbol{\vartheta})\right)\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(l)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(l)}(\boldsymbol{\vartheta})\right)\right\}\\
&\quad+\frac{(-1)}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)\right]\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(m)}(\boldsymbol{\vartheta})\right)\left(\boldsymbol{\tau}_1'\ddot{\boldsymbol{\varepsilon}}_t^{(k,l)}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\ddot{\boldsymbol{\varepsilon}}_{t-|j|}^{(k,l)}(\boldsymbol{\vartheta})\right)\right\}
\end{aligned}$$

and

$$\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta}) = \frac{\partial}{\partial\vartheta_m}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \partial\boldsymbol{\delta}^{(m)}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t = \sum_{s=-\infty}^{\infty}\partial\boldsymbol{\delta}_s^{(m)}(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-s}$$
$$\ddot{\boldsymbol{\varepsilon}}_t^{(l,m)}(\boldsymbol{\vartheta}) = \frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \ddot{\boldsymbol{\delta}}^{(l,m)}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t = \sum_{s=-\infty}^{\infty}\ddot{\boldsymbol{\delta}}_s^{(l,m)}(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-s}$$
$$\dddot{\boldsymbol{\varepsilon}}_t^{(k,l,m)}(\boldsymbol{\vartheta}) = \frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \dddot{\boldsymbol{\delta}}^{(k,l,m)}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t = \sum_{s=-\infty}^{\infty}\dddot{\boldsymbol{\delta}}_s^{(k,l,m)}(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-s}$$

Since $K = \bar{d}^2(p+q+1)$ and given the structure of $\boldsymbol{\vartheta} = [\boldsymbol{\phi}'_+, \boldsymbol{\phi}'_-, \boldsymbol{\theta}'_+, \boldsymbol{\theta}'_-, \boldsymbol{b}']$. By the results obtained in section (5.10), we have that

$$\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta}) = \begin{cases} \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil}, & \vartheta_i \in \boldsymbol{\phi}_+ \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil}, & \vartheta_i \in \boldsymbol{\phi}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil}, & \vartheta_i \in \boldsymbol{\theta}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil}, & \vartheta_i \in \boldsymbol{\theta}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\delta}(L,\boldsymbol{\vartheta}), & \vartheta_i \in \boldsymbol{b} \end{cases}$$

based on this:

(i) when $\vartheta_i \in \boldsymbol{\phi}_+$

$$\ddot{\boldsymbol{\delta}}^{(j,i)}(L,\boldsymbol{\vartheta}) = \begin{cases} \boldsymbol{0}, & \vartheta_j \in \boldsymbol{\phi}_+ \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\phi}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{b} \end{cases}$$

(ii) when $\vartheta_i \in \boldsymbol{\phi}_-$

$$\ddot{\boldsymbol{\delta}}^{(j,i)}(L,\boldsymbol{\vartheta}) = \begin{cases} \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\phi}_+ \\ \boldsymbol{0}, & \vartheta_j \in \boldsymbol{\phi}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{b} \end{cases}$$

(iii) when $\vartheta_i \in \boldsymbol{\theta}_+$

$$\ddot{\boldsymbol{\delta}}^{(j,i)}(L,\boldsymbol{\vartheta}) = \begin{cases} -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\phi}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\phi}_- \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left\{\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\right. & \\ \qquad \left.+\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\right\}\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_+ \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\times & \\ \qquad \boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{b} \end{cases}$$

(iv) when $\vartheta_i \in \boldsymbol{\theta}_-$

$$\ddot{\boldsymbol{\delta}}^{(j,i)}(L,\boldsymbol{\vartheta}) = \begin{cases} -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\phi}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\phi}_- \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\times & \\ \qquad \boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_+ \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left\{\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\right. & \\ \qquad \left.+\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\right\}\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil+\lceil j/d^2\rceil}, & \vartheta_j \in \boldsymbol{\theta}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{b} \end{cases}$$

(v) when $\vartheta_i \in \boldsymbol{\theta}_-$

$$\ddot{\boldsymbol{\delta}}^{(j,i)}(L,\boldsymbol{\vartheta}) = \begin{cases} -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\partial\boldsymbol{\delta}^{(j)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{\phi}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\partial\boldsymbol{\delta}^{(j)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{\phi}_- \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\partial\boldsymbol{\delta}^{(j)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{\theta}_+ \\ -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\partial\boldsymbol{\delta}^{(j)}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{\theta}_- \\ \boldsymbol{B}^{-1}(\boldsymbol{b})\left\{\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)+\left(\boldsymbol{e}_{l(j)}\boldsymbol{e}'_{m(j)}\right)\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\right\}\boldsymbol{\delta}(L,\boldsymbol{\vartheta}), & \vartheta_j \in \boldsymbol{b} \end{cases}$$

From these expressions, it is clear that

$$\sup_{\boldsymbol{\vartheta}}\left\|\partial\boldsymbol{\delta}^{(i)}(1,\boldsymbol{\vartheta})\right\| \le \sup_{\boldsymbol{\vartheta}}\begin{cases} \left\|\boldsymbol{B}^{-1}(\boldsymbol{b})\right\| \left\|\boldsymbol{\Theta}^{-1}(1,\boldsymbol{\theta})\right\| \left\|\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right\| \left\|\boldsymbol{\Phi}_{-}(1;\boldsymbol{\phi}_{-})\right\| \left\|\boldsymbol{\Psi}(1;\boldsymbol{\vartheta}_0)\right\|, & \vartheta_i \in \boldsymbol{\phi}_{+} \\ \left\|\boldsymbol{B}^{-1}(\boldsymbol{b})\right\| \left\|\boldsymbol{\Theta}^{-1}(1,\boldsymbol{\theta})\right\| \left\|\boldsymbol{\Phi}_{+}(1;\boldsymbol{\phi}_{+})\right\| \left\|\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right\| \left\|\boldsymbol{\Psi}(1;\boldsymbol{\vartheta}_0)\right\|, & \vartheta_i \in \boldsymbol{\phi}_{-} \\ \left\|\boldsymbol{B}^{-1}(\boldsymbol{b})\right\| \left\|\boldsymbol{\Theta}^{-1}(1,\boldsymbol{\theta})\right\| \left\|\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right\| \left\|\boldsymbol{\Theta}_{+}^{-1}(1,\boldsymbol{\theta}_{+})\right\| \left\|\boldsymbol{\Phi}(1;\boldsymbol{\phi})\right\| \left\|\boldsymbol{\Psi}(1;\boldsymbol{\vartheta}_0)\right\|, & \vartheta_i \in \boldsymbol{\theta}_{+} \\ \left\|\boldsymbol{B}^{-1}(\boldsymbol{b})\right\| \left\|\boldsymbol{\Theta}_{-}^{-1}(1,\boldsymbol{\theta}_{-})\right\| \left\|\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right\| \left\|\boldsymbol{\Theta}^{-1}(1,\boldsymbol{\theta})\right\| \left\|\boldsymbol{\Phi}(1;\boldsymbol{\phi})\right\| \left\|\boldsymbol{\Psi}(1;\boldsymbol{\vartheta}_0)\right\|, & \vartheta_i \in \boldsymbol{\theta}_{-} \\ \left\|\boldsymbol{B}^{-1}(\boldsymbol{b})\right\| \left\|\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right\| \left\|\boldsymbol{\delta}(1,\boldsymbol{\vartheta})\right\|, & \vartheta_i \in \boldsymbol{b} \end{cases}$$

by assumption (2.1), we do not have unit roots for any $\boldsymbol{\vartheta} \in \mathcal{V}$, and the model comes from a linear VARMA representation, then any of the filters $\boldsymbol{\Phi}(L,\boldsymbol{\vartheta})$, $\boldsymbol{\Theta}(L,\boldsymbol{\vartheta})$, $\boldsymbol{\Psi}(L,\boldsymbol{\vartheta})$, and their inverses are absolutely summable. Besides, $\left\|\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right\| = 1$, considering different matrix norms (Frobenius, $\|\cdot\|_p$ or $\|\cdot\|_\infty$). Thus

$$\sup_{\boldsymbol{\vartheta}}\left\|\partial\boldsymbol{\delta}^{(i)}(1,\boldsymbol{\vartheta})\right\| < \infty$$

In case of $\ddot{\boldsymbol{\delta}}^{(j,i)}(L,\boldsymbol{\vartheta})$ and $\dddot{\boldsymbol{\delta}}^{(k,j,i)}(L,\boldsymbol{\vartheta})$, since they are multiplication of absolutely summable filters for each (j,i) and (k,j,i), respectively. Then

$$\sup_{\boldsymbol{\vartheta}}\left\|\ddot{\boldsymbol{\delta}}^{(j,i)}(1,\boldsymbol{\vartheta})\right\| < \infty$$
$$\sup_{\boldsymbol{\vartheta}}\left\|\dddot{\boldsymbol{\delta}}^{(k,j,i)}(1,\boldsymbol{\vartheta})\right\| < \infty$$

With this, the following results hold under the assumption of $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^3 < \infty$ and for any $r \in [1,3]$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})\right\|^r < \infty, \quad \forall m$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\ddot{\boldsymbol{\varepsilon}}_t^{(l,m)}(\boldsymbol{\vartheta})\right\|^r < \infty, \quad \forall l,m$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\dddot{\boldsymbol{\varepsilon}}_t^{(k,l,m)}(\boldsymbol{\vartheta})\right\|^r < \infty, \quad \forall k,l,m$$

Besides, for $r \in [1,3]$ and applying Hölder's inequality:

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial}{\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\|^r &\le \frac{C}{T-|j|}\left\{\sum_{t=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})\|^r + \sum_{t=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(m)}(\boldsymbol{\vartheta})\|^r\right\} \\
&\le C(\|\boldsymbol{\tau}_1\|^r + \|\boldsymbol{\tau}_2\|^r)\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})\right\|^r \\
&\le C(\|\boldsymbol{\tau}_1\|^r + \|\boldsymbol{\tau}_2\|^r)\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^r < \infty
\end{aligned}$$

Now, for $r \in [1, 3/2]$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\|^r \leq \frac{C}{T-|j|}\left\{\sum_{t=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\tau}_1'\ddot{\boldsymbol{\varepsilon}}_t^{(m)}(\boldsymbol{\vartheta})\|^r + \sum_{t=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\tau}_2'\ddot{\boldsymbol{\varepsilon}}_{t-|j|}^{(m)}(\boldsymbol{\vartheta})\|^r\right\}$$
$$\frac{C}{T-|j|}\left\{\sum_{t=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t^{(m)}(\boldsymbol{\vartheta})\|^{2r} + \sum_{t=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\tau}_2'\partial\boldsymbol{\varepsilon}_{t-|j|}^{(m)}(\boldsymbol{\vartheta})\|^{2r}\right\}$$
$$\leq C(\|\boldsymbol{\tau}_1\|^{2r} + \|\boldsymbol{\tau}_2\|^{2r})\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^{2r} < \infty$$

By applying triangle and Hölder's inequalities together, and using the previous results, we obtain:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\| \leq C(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^3 < \infty$$

With this, we have

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial}{\partial\vartheta_m}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\|^3 \leq C(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^2}{\partial\vartheta_l\partial\vartheta_m}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\|^{3/2} \leq C(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\| \leq C(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)$$

and

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}}\right\| \leq 2\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}}\right\| < C(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial}{\partial\vartheta_l}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_m}\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}}\right\| \leq \left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial}{\partial\vartheta_l}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right\|^3\right)^{1/3}\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_m}\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}}\right\|^{3/2}\right)^{2/3}$$
$$\leq C(\|\boldsymbol{\tau}_1\| + \|\boldsymbol{\tau}_2\|)(\|\boldsymbol{\tau}_1\|^2 + \|\boldsymbol{\tau}_2\|^2) < C(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)$$

As consequence, by triangle inequality and the previous results:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_l\partial\vartheta_m}\mathscr{L}_T(\boldsymbol{\vartheta})\right\| \leq C\left\{\int(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) + \sum_{j=1}^{T-1}\frac{1}{j^2}\int(\|\boldsymbol{\tau}_1\|^3 + \|\boldsymbol{\tau}_2\|^3)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\}$$
$$\leq C\left\{1 + \frac{T-2}{T-1}\right\} \leq C$$

thus, the third derivative of sample loss is bounded in L^1. Therefore, since $\tilde{\boldsymbol{\vartheta}}_T \xrightarrow{p} \boldsymbol{\vartheta}_0$, we have:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{L}_T(\tilde{\boldsymbol{\vartheta}}_T)-\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{L}_T(\boldsymbol{\vartheta}_0)\right\|=o(1) \tag{5.12}$$

5.5.2. Asymptotic Distribution of $\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$

We want to establish the asymptotic distribution of $\sqrt{T}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$. Since we have proved the asymptotic equivalence between using sample model residuals and population model residuals, we use $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ instead of $\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})$. Now, notice that

$$\begin{aligned}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta})=\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathscr{L}_T(\boldsymbol{\vartheta})=\frac{\pi}{3}\int 2\mathrm{Re}\left\{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)}\right\}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\\ +\frac{2}{\pi}\sum_{t=1}^{T-1}\frac{1}{j^2}\int 2\mathrm{Re}\left\{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right\}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2),\end{aligned}$$

where

$$\begin{aligned}\partial\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)&\equiv\frac{\partial}{\partial\boldsymbol{\vartheta}}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\partial\hat{z}_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)\hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)+\hat{z}_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)\partial\hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)\right]\\ \partial\hat{z}_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)&\equiv\frac{\partial}{\partial\boldsymbol{\vartheta}}\hat{z}_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)=e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\left(i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right)-\frac{1}{T-|j|}\sum_{s=1+|j|}^{T}e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_s(\boldsymbol{\vartheta})}\left(i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_s(\boldsymbol{\vartheta})\right)\\ \partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})&\equiv\frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})=\left[\tfrac{\partial}{\partial\vartheta_1}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\quad\tfrac{\partial}{\partial\vartheta_2}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\quad\cdots\quad\tfrac{\partial}{\partial\vartheta_{\bar{d}}}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]\\ \frac{\partial}{\partial\vartheta_i}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})&=\frac{\partial}{\partial\vartheta_i}\left[\boldsymbol{\delta}(L;\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t\right]=\partial\boldsymbol{\delta}^{(i)}(L;\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t=\sum_{j=-\infty}^{\infty}\frac{\partial}{\partial\vartheta_i}\left(\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right)\boldsymbol{\varepsilon}_{t-j}\\ \partial\boldsymbol{\delta}(L;\boldsymbol{\vartheta})&=\left[\partial\boldsymbol{\delta}^{(1)}(L;\boldsymbol{\vartheta})\quad\partial\boldsymbol{\delta}^{(2)}(L;\boldsymbol{\vartheta})\quad\cdots\quad\partial\boldsymbol{\delta}^{(\bar{d})}(L;\boldsymbol{\vartheta})\right]\\ \partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})&=\partial\boldsymbol{\delta}(L;\boldsymbol{\vartheta})\left(\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_t\right)\end{aligned}$$

where $\partial\boldsymbol{\delta}^{(1)}(L;\boldsymbol{\vartheta})=\sum_{j=-\infty}^{\infty}\frac{\partial}{\partial\vartheta_1}\left(\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right)L^j$ and $\frac{\partial}{\partial\vartheta_1}\left(\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right)$ is a $d\times d$ matrix. Using numerator layout we are using, $\frac{\partial}{\partial\boldsymbol{\vartheta}}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ is a $1\times\bar{d}$ vector and $\frac{\partial}{\partial\boldsymbol{\vartheta}}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ is a $d\times\bar{d}$ matrix.

Let $f(\tau_1,\tau_2,\alpha)=e^{i(\tau_1h_1(\alpha)+\tau_2h_2(\alpha))}=\cos(\tau_1h_1(\alpha)+\tau_2h_2(\alpha))+i\sin(\tau_1h_1(\alpha)+\tau_2h_2(\alpha))$, its conjugate $\overline{f(\tau_1,\tau_2,\alpha)}=\cos(\tau_1h_1(\alpha)+\tau_2h_2(\alpha))-i\sin(\tau_1h_1(\alpha)+\tau_2h_2(\alpha))$. Since $\sin(x)$ is odd and $\cos(x)$ is even, then $\overline{f(\tau_1,\tau_2,\alpha)}=f(-\tau_1,-\tau_2,\alpha)$. Additionally, $\overline{z_1+z_2}=\overline{z_1}+\overline{z_2}$ for all $z_1,z_2\in\mathbb{C}$. Thus:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)}&=\frac{\partial}{\partial\boldsymbol{\vartheta}}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\\ \frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}&=\frac{\partial}{\partial\boldsymbol{\vartheta}}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\end{aligned}$$

For obtaining this asymptotic distribution we procedure in two steps. First, we prove $\sqrt{T}(\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)-\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)) \xrightarrow{p} 0$, where $\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$ is defined using population centered sample autocovariance $\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)$ and $\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$. Second, we have to find the limit distribution of $\sqrt{T}\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$

Convergence of $\sqrt{T}\{\mathbf{s}_T(\boldsymbol{\vartheta}_0)-\tilde{\mathbf{s}}_T(\boldsymbol{\vartheta}_0)\}$

Notice

$$\begin{aligned}\mathbf{s}_T(\boldsymbol{\vartheta}_0)-\tilde{\mathbf{s}}_T(\boldsymbol{\vartheta}_0) &= \frac{2\pi}{3}\int \mathrm{Re}\left\{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\hat{s}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}^{IICA}}\right)-\left(\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\}dW\\ &+\frac{4}{\pi}\sum_{j=1}^{T-1}\left(1-\frac{|j|}{T}\right)\frac{1}{j^2}\int \mathrm{Re}\left\{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\}dW\end{aligned}$$

where we omit $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ because of space constraints.

Analyzing the last term, notice

$$\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)=\left[\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right]\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)+\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right]$$

thus

$$\mathbb{E}\left\|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|\le\mathbb{E}\left\|\left[\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right]\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|+2\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|$$

where this inequality comes from the fact that $\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right|\le 2$. And, by applying Hölder's inequality, we get:

$$\begin{aligned}\mathbb{E}\left\|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\| &\le\left\{\mathbb{E}\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right|^2\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|^2\right\}^{1/2}+\\ &\quad 2\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|\end{aligned}$$

and by Lemma (5.9.5), we obtain:

$$\mathbb{E}\left\|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|\le C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)((T-|j|)^{-1}) \tag{5.13}$$

thus

$$\begin{aligned}\mathbb{E}\left\|\mathrm{Re}\left\{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\}\right\| &\le\mathbb{E}\left\|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|\\ &= C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)((T-|j|)^{-1}) \end{aligned}\tag{5.14}$$

therefore

$$\mathbb{E}\left\|\frac{4}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\int \mathrm{Re}\left\{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\}dW\right\| \leq \frac{C}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}(T-|j|)^{-1}$$
$$\leq \frac{C}{\pi}\left(\frac{2}{T^2}\ln(T-1)-\frac{1}{T}\frac{T-2}{T-1}\right)=O(T^{-1}) \tag{5.15}$$

Now, for analyzing the first term of the difference $\mathbf{s}_T(\boldsymbol{\vartheta}_0)-\tilde{\mathbf{s}}_T(\boldsymbol{\vartheta}_0)$ it is better to split it into two terms. First, let analyze $\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)$.

By the previous part, we have

$$\mathbb{E}\left\|\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|=C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)^{1/2}(T^{-1}) \tag{5.16}$$

Second, let analyze

$$\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)=\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right]$$
$$+\left[\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right]\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)$$

For any $\boldsymbol{\vartheta}$, $\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}$ is equivalent to

$$\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=-\left[\prod_{l=1}^{d}\hat{\varphi}_0(\tau_{1l},0)-\prod_{l=1}^{d}\varphi(\tau_{1l})\right]\left[\prod_{l=1}^{d}\hat{\varphi}_0(0,\tau_{2l})-\prod_{l=1}^{d}\varphi(\tau_{2l})\right]$$

and, since $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0)=\boldsymbol{\varepsilon}_t$, then both factors in the right-hand side are independent. Thus:

$$\mathbb{E}\left|\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|=\mathbb{E}\left|\prod_{l=1}^{d}\hat{\varphi}_0(\tau_{1l},0)-\prod_{l=1}^{d}\varphi(\tau_{1l})\right|\mathbb{E}\left|\prod_{l=1}^{d}\hat{\varphi}_0(0,\tau_{2l})-\prod_{l=1}^{d}\varphi(\tau_{2l})\right|$$

Besides, we know that:

$$\prod_{l=1}^{d}\hat{\varphi}_0(\tau_{1l},0)-\prod_{l=1}^{d}\varphi(\tau_{1l})=\sum_{i=0}^{d-1}\sum_{c=1}^{n_i}\left\{\prod_{j\in\mathscr{I}_{i,c}}(\hat{\varphi}_0(\tau_{1j},0)-\varphi(\tau_{1j}))\prod_{l\in\complement\mathscr{I}_{i,c}}\varphi(\tau_{1l})\right\}$$
$$\prod_{l=1}^{d}\hat{\varphi}_0(0,\tau_{2l})-\prod_{l=1}^{d}\varphi(\tau_{2l})=\sum_{i=0}^{d-1}\sum_{c=1}^{n_i}\left\{\prod_{j\in\mathscr{I}_{i,c}}(\hat{\varphi}_0(0,\tau_{2j})-\varphi(\tau_{2j}))\prod_{l\in\complement\mathscr{I}_{i,c}}\varphi(\tau_{2l})\right\}$$

thus

$$\mathbb{E}\left|\prod_{l=1}^{d}\hat{\varphi}_0(\tau_{1l},0)-\prod_{l=1}^{d}\varphi(\tau_{1l})\right| \le \sum_{i=0}^{d-1}\sum_{c=1}^{n_i}\left\{\prod_{j\in\mathscr{I}_{i,c}}\mathbb{E}\left|\hat{\varphi}_0(\tau_{1j},0)-\varphi(\tau_{1j})\right|\left|\prod_{l\in\complement\mathscr{I}_{i,c}}\varphi(\tau_{1l})\right|\right\}=O(T^{-1/2})$$

$$\mathbb{E}\left|\prod_{l=1}^{d}\hat{\varphi}_0(0,\tau_{2l})-\prod_{l=1}^{d}\varphi(\tau_{2l})\right| \le \sum_{i=0}^{d-1}\sum_{c=1}^{n_i}\left\{\prod_{j\in\mathscr{I}_{i,c}}\mathbb{E}\left|\hat{\varphi}_0(0,\tau_{2j})-\varphi(\tau_{2j})\right|\left|\prod_{l\in\complement\mathscr{I}_{i,c}}\varphi(\tau_{2l})\right|\right\}=O(T^{-1/2})$$

hence

$$\mathbb{E}\left|\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|=O(T^{-1}) \tag{5.17}$$

Now, analyzing $\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)$.

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right) &= \sum_{l=1}^{d}\partial\hat{\varphi}_0(\tau_{1l},\tau_{2l})\prod_{j\neq l}\hat{\varphi}_0(\tau_{1j},\tau_{2j})\\
&\quad-\sum_{l=1}^{d}\left\{\partial\hat{\varphi}_0(\tau_{1l},0)\hat{\varphi}_0(0,\tau_{2l})+\hat{\varphi}_0(\tau_{1l},0)\partial\hat{\varphi}_0(0,\tau_{2l})\right\}\prod_{j\neq l}\hat{\varphi}_0(\tau_{1j},0)\hat{\varphi}_0(0,\tau_{2j})\\
\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right) &= \sum_{l=1}^{d}\partial\hat{\varphi}_0(\tau_{1l},\tau_{2l})\prod_{j\neq l}\hat{\varphi}_0(\tau_{1j},\tau_{2j})\\
&\quad-\sum_{l=1}^{d}\left\{\partial\varphi(\tau_{1l})\hat{\varphi}_0(0,\tau_{2l})+\varphi(\tau_{1l})\partial\hat{\varphi}_0(0,\tau_{2l})\right\}\prod_{j\neq l}\varphi_0(\tau_{1j})\hat{\varphi}_0(0,\tau_{2j})\\
&\quad-\sum_{l=1}^{d}\left\{\partial\hat{\varphi}_0(\tau_{1l},0)\varphi(\tau_{2l})+\hat{\varphi}_0(\tau_{1l},0)\partial\varphi(\tau_{2l})\right\}\prod_{j\neq l}\hat{\varphi}_0(\tau_{1j},0)\varphi(\tau_{2j})\\
&\quad+\sum_{l=1}^{d}\left\{\partial\varphi(\tau_{1l})\varphi(\tau_{2l})+\varphi(\tau_{1l})\partial\varphi(\tau_{2l})\right\}\prod_{j\neq l}\varphi(\tau_{1j})\varphi(\tau_{2j})
\end{aligned}$$

hence

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right) &= \sum_{l=1}^{d}\left\{\partial\varphi(\tau_{2l})\prod_{j\neq l}\varphi(\tau_{2j})-\partial\hat{\varphi}_0(0,\tau_{2l})\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\}\left[\prod_{l=1}^{d}\hat{\varphi}_0(\tau_{1l},0)-\prod_{l=1}^{d}\varphi(\tau_{1l})\right]\\
&\quad+\sum_{l=1}^{d}\left\{\partial\varphi(\tau_{1l})\prod_{j\neq l}\varphi(\tau_{1j})-\partial\hat{\varphi}_0(\tau_{1l},0)\prod_{j\neq l}\hat{\varphi}_0(\tau_{1j},0)\right\}\left[\prod_{l=1}^{d}\hat{\varphi}_0(0,\tau_{2l})-\prod_{l=1}^{d}\varphi(\tau_{2l})\right]
\end{aligned}$$

notice that

$$\partial\varphi(\tau_{2l})\prod_{j\neq l}\varphi(\tau_{2j})-\partial\hat{\varphi}_0(0,\tau_{2l})\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})=-\left(\partial\varphi(\tau_{2l})-\partial\hat{\varphi}_0(0,\tau_{2l})\right)\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})-\partial\varphi(\tau_{2l})\left[\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})-\prod_{j\neq l}\varphi(\tau_{2j})\right]$$

then

$$\mathbb{E}\left\|\partial\varphi(\tau_{2l})\prod_{j\neq l}\varphi(\tau_{2j})-\partial\hat{\varphi}_0(0,\tau_{2l})\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\|\leq\mathbb{E}\left\|\{\partial\hat{\varphi}_0(0,\tau_{2l})-\partial\varphi(\tau_{2l})\}\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\|$$
$$+\mathbb{E}\left\|\partial\varphi(\tau_{2l})\left\{\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})-\prod_{j\neq l}\varphi(\tau_{2j})\right\}\right\|$$

by Cauchy-Schwarz, we have:

$$\mathbb{E}\left\|\{\partial\hat{\varphi}_0(0,\tau_{2l})-\partial\varphi(\tau_{2l})\}\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\|\leq\left\{\mathbb{E}\left\|\partial\hat{\varphi}_0(0,\tau_{2l})-\partial\varphi(\tau_{2l})\right\|^2\mathbb{E}\left\|\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\|^2\right\}^{1/2}$$
$$\mathbb{E}\left\|\partial\varphi(\tau_{2l})\left\{\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})-\prod_{j\neq l}\varphi(\tau_{2j})\right\}\right\|\leq\left\{\mathbb{E}\left\|\partial\varphi(\tau_{2l})\right\|^2\mathbb{E}\left\|\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})-\prod_{j\neq l}\varphi(\tau_{2j})\right\|^2\right\}^{1/2}$$

since

$$\mathbb{E}\left\|\partial\hat{\varphi}_0(0,\tau_{2l})-\partial\varphi(\tau_{2l})\right\|^2\leq O(T^{-1})$$
$$\mathbb{E}\left\|\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\|^2\leq C$$
$$\mathbb{E}\left\|\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})-\prod_{j\neq l}\varphi(\tau_{2j})\right\|^2\leq O(T^{-1})$$

thus

$$\mathbb{E}\left\|\partial\varphi(\tau_{2l})\prod_{j\neq l}\varphi(\tau_{2j})-\partial\hat{\varphi}_0(0,\tau_{2l})\prod_{j\neq l}\hat{\varphi}_0(0,\tau_{2j})\right\|\leq C(\|\boldsymbol{\tau}_1\|)T^{-1/2}$$

as consequence:

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)\right\|\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)T^{-1}$$

This implies that

$$\mathbb{E}\left\|\mathrm{Re}\left\{\left(\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\hat{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)-\left(\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right)\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}-\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\}\right\|\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)T^{-1}$$

therefore

$$\mathbb{E}\left\|\sqrt{T}\left(\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)-\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)\right)\right\|=O(T^{-1/2})=o(1) \tag{5.18}$$

Weak Convergence of $\sqrt{T}\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$

Now, $\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$ is defined as follows:

$$\begin{aligned}\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) = \frac{\pi}{3}\int 2\mathrm{Re}\left\{\left(\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right)\right\}\dots \\ + \frac{4}{\pi}\sum_{t=1}^{T-1}\frac{1}{j^2}\int 2\mathrm{Re}\left\{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right\}\end{aligned}$$

we have proved

$$\begin{aligned}\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)} - \mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right\| &\le C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)(T-|j|)^{-1}, \quad |j|\le T-1 \\ \mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)} - \mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right\| &\le C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)T^{-1}\end{aligned}$$

Thus

$$\begin{aligned}\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) = \frac{\pi}{3}\int 2\mathrm{Re}\left\{\left(\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right)\right\}\dots \\ + \frac{4}{\pi}\sum_{t=1}^{T-1}\frac{1}{j^2}\int 2\mathrm{Re}\left\{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right\} + o_p(1)\end{aligned}$$

since $\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)} = \partial\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)$ and $\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)} = \partial\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)$. Then

$$\begin{aligned}\mathbb{E}\partial\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2) &= -i\left(\boldsymbol{\tau}_1'\partial\boldsymbol{\delta}_{|j|}(\boldsymbol{\vartheta}_0)\boldsymbol{I}_{\tilde{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2) + \boldsymbol{\tau}_2'\partial\boldsymbol{\delta}_{-|j|}(\boldsymbol{\vartheta}_0)\boldsymbol{I}_{\tilde{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\right) \\ \mathbb{E}\partial\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2) &= -i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0)\left(\boldsymbol{I}_{\tilde{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right) + i\left\{\boldsymbol{\tau}_1'\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0)\left(\boldsymbol{I}_{\tilde{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right. \\ &\quad \left. + \boldsymbol{\tau}_2'\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0)\left(\boldsymbol{I}_{\tilde{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right\} \\ \mathbb{E}\partial\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2) &= \sum_{l=1}^{d}\partial\varphi_0^{(\boldsymbol{\vartheta}_0)}(-\tau_{1l},-\tau_{2l})\prod_{j\neq l}\varphi_0^{(\boldsymbol{\vartheta}_0)}(-\tau_{1j},-\tau_{2j}) \\ &\quad - \sum_{l=1}^{d}\left\{\partial\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{1l})\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{2l}) + \varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{1l})\partial\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{2l})\right\}\prod_{j\neq l}\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{1j})\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{2j}) \\ \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right] &= \begin{bmatrix}\mathbb{E}[\varepsilon_{t,1}e^{-i\tau_{11}\varepsilon_{t,1}}]\prod_{j\neq 1}\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{1j}) \\ \vdots \\ \mathbb{E}[\varepsilon_{t,d}e^{-i\tau_{1d}\varepsilon_{t,d}}]\prod_{j\neq d}\varphi^{(\boldsymbol{\vartheta}_0)}(-\tau_{1j})\end{bmatrix}\end{aligned}$$

The term $\mathbb{E}\partial\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)-\mathbb{E}\partial\tilde{\sigma}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)=\mathbf{A}_0=\mathbf{A}_{01}-\mathbf{A}_{02}-\mathbf{A}_{03}$, where

$$
\begin{aligned}
\mathbf{A}_{01} &= -i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\left\{\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0)\right\}\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)\\
\mathbf{A}_{02} &= -i\boldsymbol{\tau}_1'\left\{\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0)\right\}\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\mathbb{E}\left[e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)\\
\mathbf{A}_{03} &= -i\boldsymbol{\tau}_2'\left\{\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0)\right\}\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\mathbb{E}\left[e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)
\end{aligned}
$$

where $\mathbf{S}_{-\text{diag}}=\mathbf{1}_{d\times d}-\boldsymbol{I}_d=\begin{bmatrix}0&1&1&\dots&1\\1&0&1&\dots&1\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&1&1&\dots&0\end{bmatrix}$ and $[\circ]$ denotes the penetrating face product (see Slyusar (1999)). Besides, calling $\mathbf{A}_j=\mathbb{E}\partial\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)$ for $j=1,\dots,T-1$. Then:

$$
\begin{aligned}
\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) &= \frac{1}{T}\sum_{t=1}^{T}\text{Re}\left\{\int\frac{2\pi}{3}\left(z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_1)z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)-\sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)\mathbf{A}_0 dW\right\}\\
&+\frac{1}{T}\sum_{t=2}^{T}\text{Re}\left\{\int z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_1)\sum_{j=1}^{t-1}\frac{4}{\pi j^2}z_{t-j}(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)\mathbf{A}_j dW\right\}+o_p(T^{-1/2})
\end{aligned}
\tag{5.19}
$$

therefore

$$
\begin{aligned}
T^{1/2}\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) &= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\text{Re}\left\{\int\frac{2\pi}{3}\left(z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_1)z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)-\sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)\mathbf{A}_0 dW\right\}\\
&+\frac{1}{T^{1/2}}\sum_{t=2}^{T}\text{Re}\left\{\int z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_1)\sum_{j=1}^{t-1}\frac{4}{\pi j^2}z_{t-j}(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)\mathbf{A}_j dW\right\}+o_p(1)
\end{aligned}
\tag{5.20}
$$

note $\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$ is an $1\times\bar{d}$ row vector. For simplicity, we take the the column form $\text{vec}\left(\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)\right)$ and it is denoted in the same way, as $\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)$. Besides, notice

$$
\begin{aligned}
\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)} &= -\frac{i}{T-j}\sum_{t=1+j}^{T}z_t^0(\boldsymbol{\tau}_1)z_{t-j}^0(\boldsymbol{\tau}_2)\left[\boldsymbol{\tau}_1'\partial\boldsymbol{\delta}_{|j|}(\boldsymbol{\vartheta}_0)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\mathbb{E}\left[e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]\right)\right.\\
&\left.+\boldsymbol{\tau}_2'\partial\boldsymbol{\delta}_{-|j|}(\boldsymbol{\vartheta}_0)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_{t-j}e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]\mathbb{E}\left[e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)\right]
\end{aligned}
$$

then, the vectorized form of the RHS is

$$
\begin{aligned}
&-\frac{i}{T-j}\sum_{t=1+j}^{T}\left\{\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes z_t^0(\boldsymbol{\tau}_1)\boldsymbol{\tau}_1'\right]\text{vec}\left(\partial\boldsymbol{\delta}_{|j|}(\boldsymbol{\vartheta}_0)\right)z_{t-j}^0(\boldsymbol{\tau}_2)\mathbb{E}\left[e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]\right.\\
&\left.+\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_{t-j}e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]'\right)\otimes z_{t-j}^0(\boldsymbol{\tau}_2)\boldsymbol{\tau}_2'\right]\text{vec}\left(\partial\boldsymbol{\delta}_{-|j|}(\boldsymbol{\vartheta}_0)\right)z_t^0(\boldsymbol{\tau}_1)\mathbb{E}\left[e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right\}
\end{aligned}
$$

and, remember that if $\mathbf{X}_{n\times n}\boldsymbol{y}_{n\times 1}$ is a $n\times 1$ column vector, then $\mathbf{X}\boldsymbol{y} = \text{vec}(\mathbf{X}\boldsymbol{y}) = (\boldsymbol{y}'\otimes \boldsymbol{I}_n)\text{vec}(\mathbf{X})$. Then, the term above is equivalent to

$$-\frac{i}{T-j}\sum_{t=1+j}^{T}\left\{\left(\text{vec}\left(\partial\boldsymbol{\delta}_{|j|}(\boldsymbol{\vartheta}_0)\right)'\otimes \boldsymbol{I}_{\bar{d}}\right)\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes \boldsymbol{\tau}_1'\right] z_t^0(\boldsymbol{\tau}_1)z_{t-j}^0(\boldsymbol{\tau}_2)\mathbb{E}\left[e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]\right.$$
$$\left.+\left(\text{vec}\left(\partial\boldsymbol{\delta}_{-|j|}(\boldsymbol{\vartheta}_0)\right)'\otimes \boldsymbol{I}_{\bar{d}}\right)\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_{t-j} e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]'\right)\otimes \boldsymbol{\tau}_2'\right] z_{t-j}^0(\boldsymbol{\tau}_2)z_{t}^0(\boldsymbol{\tau}_1)\mathbb{E}\left[e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_{t}}\right]\right\}$$

Thus

$$\text{vec}\left(\int \text{Re}\left\{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}\right\}\right) = -\frac{1}{T-j}\sum_{t=1+j}^{T}\left[\left(\text{vec}\left(\partial\boldsymbol{\delta}_{|j|}(\boldsymbol{\vartheta}_0)\right)'\otimes \boldsymbol{I}_{\bar{d}}\right)e_t^0 x_{t-j}^0\right.$$
$$\left.+\left(\text{vec}\left(\partial\boldsymbol{\delta}_{-|j|}(\boldsymbol{\vartheta}_0)\right)'\otimes \boldsymbol{I}_{\bar{d}}\right)e_{t-j}^0 x_t^0\right]$$

where $e_t^0 = \frac{1}{i}\int \text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes \boldsymbol{\tau}_1'\right] z_t^0(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1)$, and $x_{t-j}^0 = \int z_{t-j}^0(\boldsymbol{\tau}_2)\mathbb{E}\left[e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-j}}\right]dW(\boldsymbol{\tau}_2)$.

For the contemporaneous term, first notice

$$\text{vec}\left(\frac{2\pi}{3}\frac{1}{T}\sum_{t=1}^{T}\text{Re}\left\{\int\left(z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_1)z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)-\sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW\right\}\right) =$$
$$\text{vec}\left(\frac{2\pi}{3}\frac{1}{T}\sum_{t=1}^{T}\text{Re}\left\{\int z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_1)z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW\right\}\right)$$
$$-\text{vec}\left(\frac{2\pi}{3}\text{Re}\left\{\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW\right\}\right)$$

and notices that

$$\text{vec}(\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)) = \text{vec}(\mathbf{A}_{01})-\text{vec}(\mathbf{A}_{02})-\text{vec}(\mathbf{A}_{03})$$
$$\text{vec}(\mathbf{A}_{01}) = -i\left[\text{vec}(\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes \boldsymbol{I}_{\bar{d}}\right]\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right]$$
$$\text{vec}(\mathbf{A}_{02}) = -i\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\left[\text{vec}(\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes \boldsymbol{I}_{\bar{d}}\right]\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes \boldsymbol{\tau}_1'\right]$$
$$\text{vec}(\mathbf{A}_{03}) = -i\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\left[\text{vec}(\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes \boldsymbol{I}_{\bar{d}}\right]\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes \boldsymbol{\tau}_2'\right]$$

thus

$$\text{vec}(\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)) = -i\left[\text{vec}(\mathbf{S}_{-\text{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes \boldsymbol{I}_{\bar{d}}\right]\left\{\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right]\right.$$
$$\left.-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes \boldsymbol{\tau}_1'\right]-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\text{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes \boldsymbol{\tau}_2'\right]\right\}$$

then

$$\frac{2\pi}{3}\frac{1}{T}\sum_{t=1}^{T}\mathrm{Re}\left\{\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\mathrm{vec}(\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2))dW(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_2)\right\}=$$
$$-\frac{2\pi}{3}\frac{1}{T}\sum_{t=1}^{T}\left[\mathrm{vec}(\mathbf{S}_{-\mathrm{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes\boldsymbol{I}_{\bar{d}}\right]\left(\tilde{e}_t^{0,1}+\tilde{e}_t^{0,2}+\tilde{e}_t^{0,3}\right)$$
$$\mathrm{vec}\left(\frac{2\pi}{3}\mathrm{Re}\left\{\int\sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathbf{A}_0(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW\right\}\right)=$$
$$-\frac{2\pi}{3}\left[\mathrm{vec}(\mathbf{S}_{-\mathrm{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes\boldsymbol{I}_{\bar{d}}\right]\left(\check{e}^{0,1}+\check{e}^{0,2}+\check{e}^{0,3}\right)$$

where

$$\tilde{e}_t^{0,1}=\mathrm{Re}\left(i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right]dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)$$
$$\tilde{e}_t^{0,2}=\mathrm{Re}\left(-i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes\boldsymbol{\tau}_1'\right]\right)$$
$$\tilde{e}_t^{0,3}=\mathrm{Re}\left(-i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes\boldsymbol{\tau}_2'\right]dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)$$
$$\check{e}^{0,1}=\mathrm{Re}\left(i\int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right]dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)$$
$$\check{e}^{0,2}=\mathrm{Re}\left(-i\int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes\boldsymbol{\tau}_1'\right]\right)$$
$$\check{e}^{0,3}=\mathrm{Re}\left(-i\int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes\boldsymbol{\tau}_2'\right]dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)$$

Therefore

$$\begin{aligned}T^{1/2}\tilde{\mathbf{s}}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0)&=-\left[\mathrm{vec}(\mathbf{S}_{-\mathrm{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes\boldsymbol{I}_{\bar{d}}\right]\left\{\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\tilde{e}_t^{0,1}+\tilde{e}_t^{0,2}+\tilde{e}_t^{0,3}-\check{e}^{0,1}-\check{e}^{0,2}-\check{e}^{0,3}\right)\right\}\\&\quad-\frac{1}{T^{1/2}}\sum_{t=2}^{T}\left[X_{t-1}^0e_t^0+E_{t-1}^0x_t^0\right]\\&=\left[-\left[\mathrm{vec}(\mathbf{S}_{-\mathrm{diag}}[\circ]\partial\boldsymbol{\delta}_0(\boldsymbol{\vartheta}_0))'\otimes\boldsymbol{I}_{\bar{d}}\right]\quad-\boldsymbol{I}\right]\frac{1}{T^{1/2}}\sum_{t=2}^{T}\begin{bmatrix}S_{t,0}\\S_{t,1}\end{bmatrix}\end{aligned}$$

where $X_{t-1}^0=\sum_{j=1}^{t-1}\frac{4}{j^2\pi}\left(\mathrm{vec}\left(\partial\boldsymbol{\delta}_{|j|}(\boldsymbol{\vartheta}_0)\right)'\otimes\boldsymbol{I}_{\bar{d}}\right)x_{t-j}^0$, $E_{t-1}^0=\sum_{j=1}^{t-1}\frac{4}{j^2\pi}\left(\mathrm{vec}\left(\partial\boldsymbol{\delta}_{-|j|}(\boldsymbol{\vartheta}_0)\right)'\otimes\boldsymbol{I}_{\bar{d}}\right)e_{t-j}^0$, and $S_t=\begin{bmatrix}S_{t,0}\\S_{t,1}\end{bmatrix}$.

By assumption (2.2), when $\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0$, S_t is formed by components that are independent between them. Then, they formed a martingale difference sequence. Denoting $\mathscr{I}_{t-1}$ the information set up to period $t-1$. We have $\mathbb{E}[S_t\,|\,\mathscr{I}_{t-1}]=\mathbf{0}$. For applying the CLT for

MDS, we have:

$$\sqrt{T}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0},\, \boldsymbol{\Omega}_0(\boldsymbol{\vartheta}_0)) \tag{5.21}$$

Now, applying mean value theorem of the score function at $\boldsymbol{\vartheta}_0$:

$$\mathbf{s}_T^{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0) + \left.\left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta})\right)\right|_{\boldsymbol{\vartheta}=\tilde{\boldsymbol{\vartheta}}_T}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} - \boldsymbol{\vartheta}_0)$$

where $\left\|\tilde{\boldsymbol{\vartheta}}_T - \boldsymbol{\vartheta}_0\right\| \le \left\|\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} - \boldsymbol{\vartheta}_0\right\|$ and $\tilde{\boldsymbol{\vartheta}}_T \xrightarrow{p} \boldsymbol{\vartheta}_0$. Since $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$ is a zero of the score function (and local minimizer of $\mathscr{L}_T(\boldsymbol{\vartheta})$), then $\mathbf{s}_T^{\mathscr{L}}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}) = \mathbf{0}$ and assuming that $\left.\left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta})\right)\right|_{\boldsymbol{\vartheta}=\tilde{\boldsymbol{\vartheta}}_T}$ is non-singular. Thus

$$\sqrt{T}(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}} - \boldsymbol{\vartheta}_0) = \left[-\left.\left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta})\right)\right|_{\boldsymbol{\vartheta}=\tilde{\boldsymbol{\vartheta}}_T}\right]^{-1}\sqrt{T}\mathbf{s}_T^{\mathscr{L}}(\boldsymbol{\vartheta}_0).$$

□

5.6. Proof Theorem (2.6)

Like in the proof of Theorem (2.3), we proceed as follows:

$$\sup_{\boldsymbol{\vartheta}}\|\mathscr{R}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta})\| \le \sup_{\boldsymbol{\vartheta}}\left\|\mathscr{R}_T(\boldsymbol{\vartheta}) - \bar{\mathscr{R}}_T(\boldsymbol{\vartheta})\right\| + \sup_{\boldsymbol{\vartheta}}\left\|\bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta})\right\| + \sup_{\boldsymbol{\vartheta}}\left\|\tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta})\right\|$$

where $\bar{\mathscr{R}}_T(\boldsymbol{\vartheta})$ and $\tilde{\mathscr{R}}_T(\boldsymbol{\vartheta})$ are similarly defined as $\mathscr{R}_T(\boldsymbol{\vartheta})$ but employing $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ and $\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$, respectively.

Hence

$$\begin{aligned}
\mathscr{R}_T(\boldsymbol{\vartheta}) - \bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) &= \frac{2}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\int\left\{\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2\right\}dW \\
&+ \frac{\pi}{3}\int\left\{\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2\right\}dW \\
\bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) &= \frac{2}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\int\left\{\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2\right\}dW \\
&+ \frac{\pi}{3}\int\left\{\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2\right\}dW \\
\tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta}) &= -\frac{2}{\pi}\sum_{j=T}^{\infty}\frac{1}{j^2}\int\left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2)\right\|^2 dW
\end{aligned}$$

Using the results below, we obtain

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\mathscr{R}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta})\| = o(1)$$

5.6.1. Convergence of $\sup_{\boldsymbol{\vartheta}} \left\| \mathscr{R}_T(\boldsymbol{\vartheta}) - \bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) \right\|$

Note that

$$\sup_{\boldsymbol{\vartheta}} \left\| \mathscr{R}_T(\boldsymbol{\vartheta}) - \bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) \right\| \leq \frac{2}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int \sup_{\boldsymbol{\vartheta}} \left\| \left\{ \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \right\} \right\| dW$$
$$+ \frac{\pi}{3} \int \sup_{\boldsymbol{\vartheta}} \left\| \left\{ \left\| \Delta^{(1,0)} \hat{\sigma}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}^{(1,0)}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \right\} \right\| dW$$

and notice for two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$: $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x} - \mathbf{y}\|^2 + 2\langle \mathbf{x} - \mathbf{y}, \mathbf{y} \rangle + \|\mathbf{y}\|^2$. Thus:

$$\left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 = \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2$$
$$+ 2 \langle \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2), \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \rangle$$

Applying results in Lemma (5.9.7), we obtain

$$\mathbb{E} \int \sup_{\boldsymbol{\vartheta}} \left\| \left\{ \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \right\} \right\| dW \leq C(T - |j|)^{-1}$$

Besides, notice that

$$\left\| \Delta^{(1,0)} \hat{\sigma}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}^{(1,0)}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \leq C \left\| \Delta^{(1,0)} \hat{\sigma}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \Delta^{(1,0)} \hat{\sigma}_0^{\boldsymbol{\varepsilon}}\right.$$
$$+ C \left\| \left[\hat{F}^{(1,0)}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right]^{MD}\right.$$

by applying Lemma (5.9.7) again

$$\mathbb{E} \sup_{\boldsymbol{\vartheta}} \left\| \Delta^{(1,0)} \hat{\sigma}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \Delta^{(1,0)} \hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \leq C T^{-1}$$
$$\mathbb{E} \sup_{\boldsymbol{\vartheta}} \left\| \left[\hat{F}^{(1,0)}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \leq C T^{-1}$$

thus

$$\begin{aligned} \mathbb{E} \sup_{\boldsymbol{\vartheta}} \left\| \mathscr{R}_T(\boldsymbol{\vartheta}) - \bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) \right\| &\leq \frac{C\pi}{3} T^{-1} + \frac{2C}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} (T - |j|)^{-1} \\ &\leq \frac{C\pi}{3} \frac{1}{T} + \frac{2C}{\pi} \frac{1}{T} \left\{ \sum_{j=1}^{T-1} \frac{1}{j^2} + \sum_{j=1}^{T-1} \frac{1}{|j|(T - |j|)} \right\} \\ &\leq \frac{C\pi}{3} \frac{1}{T} + \frac{2C}{\pi} \frac{1}{T} \left\{ \frac{1}{T-1} + \frac{2}{T} \ln(T-1) \right\} \leq C T^{-1} \end{aligned}$$

5.6.2. Convergence of $\sup_{\boldsymbol{\vartheta}} \left\| \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta}) \right\|$

Note that

$$\sup_{\boldsymbol{\vartheta}} \left\| \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta}) \right\| \leq \frac{2}{\pi} \sum_{j=T}^{\infty} \frac{1}{j^2} \int \sup_{\boldsymbol{\vartheta}} \left\| \Delta^{(1,0)} \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 dW$$

By Lemma (5.9.2), we have

$$\sup_{\boldsymbol{\vartheta}} \left\| \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \mathscr{R}_0(\boldsymbol{\vartheta}) \right\| \leq \frac{2}{\pi} \sum_{j=T}^{\infty} \frac{1}{j^2} \int C \left\| \boldsymbol{\tau}_2 \right\|^2 j^{1/2-\mu_0} dW \leq \frac{2C}{\pi} \sum_{j=T}^{\infty} \frac{1}{j^2} j^{1/2-\mu_0} \leq C \sum_{j=T}^{\infty} j^{-3/2-\mu_0}$$
$$\leq C \int_{T-1}^{\infty} j^{-3/2-\mu_0} dj \leq C(T-1)^{-1/2-\mu_0} = O(T^{-1/2-\mu_0}) \leq CT^{-1}, \ \mu_0 \geq 1/2$$

5.6.3. Convergence of $\sup_{\boldsymbol{\vartheta}} \left\| \bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) \right\|$

Since

$$\bar{\mathscr{R}}_T(\boldsymbol{\vartheta}) - \tilde{\mathscr{R}}_T(\boldsymbol{\vartheta}) = \frac{4}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int \left\{ \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \right\} dW$$
$$+ \frac{2}{3}\pi \int \left\{ \left\| \Delta^{(1,0)} \hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \left[F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \right\} dW$$

using the results below, and choosing $m = [j/2]$:

$$\frac{4}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \int \mathbb{E} \sup_{\boldsymbol{\vartheta}} \left\{ \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \right\} dW$$
$$\leq \frac{4}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \left\{ \frac{C}{(T-|j|)^{1/2}} \left\{ O(1) + O(j^{1/2-\mu_0}) + (1 + O(j^{1/2-\mu_0/2}))(T-1-|j| \right. \right.$$
$$\leq \frac{C}{T^{1/2}} \sum_{j=1}^{T-1} \frac{1}{j^2} \left\{ O(1) + O(j^{1/2-\mu_0}) + (1 + O(j^{1/2-\mu_0/2}))(T-1-|j|)^{-\mu_0/2} \right\} = ($$

Uniform convergence of $\left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 - \left\| \Delta^{(1,0)} \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2$

We use the following equivalence for the difference of two modules for any pair of d-dimensional complex vectors

$$\left\| \mathbf{x} \right\|^2 - \left\| \mathbf{z} \right\|^2 = \langle \mathbf{x} - \mathbf{z}, \overline{\mathbf{x}} - \overline{\mathbf{z}} \rangle + 2 \langle \mathbf{x} - \mathbf{z}, \overline{\mathbf{z}} \rangle = \left\| \mathbf{x} - \mathbf{z} \right\|^2 + 2 \langle \mathbf{x} - \mathbf{z}, \overline{\mathbf{z}} \rangle$$

and

$$\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 = \left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 + \left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\sigma\right.$$

from results below:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\right| \leq \frac{C}{(T-|j|)^{1/2}}\Big\{O(1)+O(m^{1/2-\mu_0}) + (1+O(m^{1/2-\mu_0/2}))(T-1-|j|)^{-\mu_0/2}\Big\} \qquad (5.22)$$

Uniform convergence of $\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2$

Note that

$$\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 - \left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 = \left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 + 2\left\langle \Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2), \overline{\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\right\rangle$$

since $\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) = \frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\Big|_{\boldsymbol{\tau}_1=0}\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)$ and $\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) =$ $\frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\Big|_{\boldsymbol{\tau}_1=0}\varphi(\boldsymbol{\tau}_2) - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0}\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2) + \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0}\varphi(\boldsymbol{\tau}_2)$.
Thus:

$$\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) = -\left(\frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0}\right)\left[\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2) - \varphi(\boldsymbol{\tau}_2)\right]$$

$$\frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0} = \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - i\,\mathbb{E}[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})] = \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$$

where last line comes from the fact that $\{\boldsymbol{\varepsilon}_t\}_{t\in\mathbb{Z}}$ is an m.d.s.. Then:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 = \frac{1}{(T-|j|)^4}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\|^2\sup_{\boldsymbol{\vartheta}}\left|\sum_{t=1+|j|}^{T}z_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)\right|^2 \leq \frac{C}{(T-|j|)^2}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\|^2$$

Now, note that

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\|^2 = \sum_{t=1+|j|}^{T}\sum_{t'=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\langle\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}),\overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})}\rangle$$

where

$$
\begin{aligned}
\langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})} \rangle &= \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}, m), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})} \rangle + \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}, m), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})} \rangle \\
&= \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}, m), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})} - \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}, m)} \rangle + \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}, m), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}, m)} \rangle \\
&\quad + \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}, m), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})} - \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}, m)} \rangle + \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}, m), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}, m)} \rangle \\
&= \langle \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle + \langle \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle \\
&\quad + \langle \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle + \langle \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle \\
&= \langle \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle + 2 \langle \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle \\
&\quad + \langle \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle
\end{aligned}
$$

and notices that

$$
\begin{aligned}
\mathbb{E} \sup_{\boldsymbol{\vartheta}} \langle \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle &= \mathbb{E} \sup_{\boldsymbol{\vartheta}} \sum_{|j|>m} \sum_{|j'|>m} \boldsymbol{\delta}_j'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}' \boldsymbol{\varepsilon}_{t'-j'} \boldsymbol{\delta}_{j'}(\boldsymbol{\vartheta}) \\
&\le C \sum_{|j|>m} \sum_{|j'|>m} |j|^{-\mu_0} \left|j'\right|^{-\mu_0} \mathbb{E} \left\| \boldsymbol{\varepsilon}_{t-j}' \right\| \mathbb{E} \left\| \boldsymbol{\varepsilon}_{t'-j'} \right\| \le C m^{2-2\mu_0} \\
\mathbb{E} \sup_{\boldsymbol{\vartheta}} \langle \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle &= \mathbb{E} \sup_{\boldsymbol{\vartheta}} \sum_{|j|>m} \sum_{|j'|\le m} \boldsymbol{\delta}_j'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}' \boldsymbol{\varepsilon}_{t'-j'} \boldsymbol{\delta}_{j'}(\boldsymbol{\vartheta}) \\
&\le \mathbb{E} \sup_{\boldsymbol{\vartheta}} \sum_{|j|>m} \boldsymbol{\delta}_j'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}' \boldsymbol{\varepsilon}_{t'} \boldsymbol{\delta}_0(\boldsymbol{\vartheta}) + \mathbb{E} \sup_{\boldsymbol{\vartheta}} \sum_{|j|>m} \sum_{0<|j'|\le m} \boldsymbol{\delta}_j'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}' \boldsymbol{\varepsilon}_{t'-j'} \boldsymbol{\delta}_{j'}(\boldsymbol{\vartheta}) \\
&\le C m^{1-\mu_0} + C m^{2-2\mu_0} \\
\mathbb{E} \sup_{\boldsymbol{\vartheta}} \langle \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}, \sum_{|j|\le m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'-j} \rangle &= \mathbb{E} \sup_{\boldsymbol{\vartheta}} \sum_{|j|\le m} \sum_{|j|'\le m} \boldsymbol{\delta}_j'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j}' \boldsymbol{\varepsilon}_{t'-j'} \boldsymbol{\delta}_{j'}(\boldsymbol{\vartheta}) \\
&\le C \left\{ \left|t-t'\right|^{1-\mu_0} \right\}, \ \left|t-t'\right| > 0
\end{aligned}
$$

thus

$$
\mathbb{E} \sup_{\boldsymbol{\vartheta}} \langle \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}), \overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})} \rangle \le \begin{cases} C m^{2-2\mu_0} + C m^{1-\mu_0} + C \left\{ \left|t-t'\right|^{1-\mu_0} \right\}, & \left|t-t'\right| > 0 \\ C m^{2-2\mu_0} + C m^{1-\mu_0} + C, & \left|t-t'\right| = 0 \end{cases}
$$

Hence

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\|^2 = \sum_{t=1+|j|}^{T}\sum_{t'=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\langle\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}),\overline{\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})}\rangle$$

$$\le (T-|j|)C\left[Cm^{2-2\mu_0}+Cm^{1-\mu_0}+C\right]+\sum_{|t-t'|=1}^{T-1-|j|}C\left[Cm^{2-2\mu_0}+Cm^{1-\mu_0}+C\left\{\left|t-t'\right|^{1-\mu_0}\right\}\right]$$

$$\le (T-|j|)C+C(T-1-|j|)m^{2-2\mu_0}+C(T-1-|j|)m^{1-\mu_0}+C(T-1-|j|)^{2-\mu_0}$$

therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 = \frac{C}{T-|j|}\left\{1+m^{1-\mu_0}+(T-1-|j|)^{1-\mu_0}\right\}$$

Now

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\langle\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2),\Delta^{(1,0)}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\rangle\right| \le \left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\right\|^2\right)^{1/2}\times$$

$$\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\right)^{1/2}$$

since $\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\right\|^2<\infty$ and

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\langle\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2),\Delta^{(1,0)}\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\rangle\right| \le C(T-|j|)^{-1/2}\left\{1+m^{\frac{1-\mu_0}{2}}+(T-1-|j|)^{\frac{1-\mu_0}{2}}\right\}$$

therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\right| \le C(T-|j|)^{-1/2}\left\{1+m^{\frac{1-\mu_0}{2}}+(T-1-|j|)^{\frac{1-\mu_0}{2}}\right\} \tag{5.23}$$

Uniform convergence of $\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2$
Notice that:

$$\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 = \left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2$$

$$+2\langle\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2),\Delta^{(1,0)}\overline{\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\rangle$$

since $\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) = \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0}\varphi(\boldsymbol{\tau}_2)$, we have:

$$\begin{aligned}\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) &= \frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} \\ &\quad - \left[\frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0}\right]\varphi(\boldsymbol{\tau}_2) \\ &\quad - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0}\left[\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2) - \varphi(\boldsymbol{\tau}_2)\right]\end{aligned}$$

since $\frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi(\boldsymbol{\tau}_1)\Big|_{\boldsymbol{\tau}_1=0} = i\,\mathbb{E}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$ and

$$\frac{\partial}{\partial\boldsymbol{\tau}_1}\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} - \frac{\partial}{\partial\boldsymbol{\tau}_1}\varphi_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\Big|_{\boldsymbol{\tau}_1=0} = \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\zeta_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_2),$$

where $\zeta_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) = \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\exp(i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})) - \mathbb{E}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\exp(i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta}))$.

Then

$$\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) = \frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\zeta_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) - \left[\frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]\varphi(\boldsymbol{\tau}_2)$$

and

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 &\le \left(\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left[\frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]\varphi(\boldsymbol{\tau}_2)\right\|^2\right\}^{1/2} + \right. \\ &\quad \left.\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\tilde{\boldsymbol{\varepsilon}}_t^{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)\right\|^2\right\}^{1/2}\right)^2\end{aligned}$$

Since $\left|\varphi(\boldsymbol{\tau}_2)\right|^2 = O(1)$ uniformly in $\boldsymbol{\vartheta}$, thus

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left[\frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]\varphi(\boldsymbol{\tau}_2)\right\|^2 &\le C\,\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right\|^2 \\ &\le \frac{C}{T-|j|}\left\{1+m^{1-\mu_0}+(T-1-|j|)^{1-\mu_0}\right\}\end{aligned}$$

Assuming $\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{i}{T-|j|}\sum_{t=1+|j|}^{T}\tilde{\boldsymbol{\varepsilon}}_t^{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)\right\|^2 = O((T-|j|)^{-\eta})$ with $\eta \ge 1$, consequently

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 \le \frac{C}{T-|j|}\left\{1+m^{1-\mu_0}+(T-1-|j|)^{1-\mu_0}+(T-|j|)^{1-\eta}\right\}$$

Therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\right|\leq C(T-|j|)^{-1/2}\left\{1+m^{\frac{1-\mu_0}{2}}+(T-1-|j|)^{\frac{1-\mu_0}{2}}+(T-|j|)^{\frac{1-\eta}{2}}\right\} \tag{5.24}$$

Uniform convergence of $\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2$

As above:

$$\begin{aligned}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2-\left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2 &=\\ \left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2-\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2&\\ +\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2-\left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2&\end{aligned}$$

with the results below, we have:

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}&\left|\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2-\left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2\right|\\ &=\frac{C}{T^{1/2}}\left\{1+m^{\frac{1-\mu_0}{2}}+(T-1)^{\frac{1-\mu_0}{2}}\right\}\\ &=o(1)\end{aligned}$$

if m grows at smaller rate than $O(T)$, the last line follows immediately.

Uniform Convergence of $\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2$

We have

$$\begin{aligned}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2-\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 &=\\ \left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left\{\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\}\right\|^2&\\ +2\,\langle\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left\{\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\},&\\ \overline{\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)}\rangle&\end{aligned}$$

By Minkowski inequality:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left\{\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\}\right\|^2\leq$$
$$\left(\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\right\}^{1/2}+\right.$$
$$\left.\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\right\}^{1/2}\right)^2$$

Using results above:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2=\frac{C}{T}\left(O(1)+O(m^{1-2\mu_0})+(1+O(m^{1-\mu_0}))(T-1)^{-\mu_0}\right)$$

Now, notice

$$\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)=i\begin{bmatrix}\frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,1}(\boldsymbol{\vartheta})\hat{z}_{t,1}(\boldsymbol{\vartheta},\tau_{2,1})\prod_{j\neq 1}\hat{\varphi}_0(0,\tau_{2,j})\\ \vdots\\ \frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,d}(\boldsymbol{\vartheta})\hat{z}_{t,d}(\boldsymbol{\vartheta},\tau_{2,d})\prod_{j\neq d}\hat{\varphi}_0(0,\tau_{2,j})\end{bmatrix}$$
$$\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)=i\begin{bmatrix}\frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,1}(\boldsymbol{\vartheta})z_{t,1}(\boldsymbol{\vartheta},\tau_{2,1})\prod_{j\neq 1}\hat{\varphi}_0(0,\tau_{2,j})\\ \vdots\\ \frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,d}(\boldsymbol{\vartheta})z_{t,d}(\boldsymbol{\vartheta},\tau_{2,d})\prod_{j\neq d}\hat{\varphi}_0(0,\tau_{2,j})\end{bmatrix}$$

hence

$$\begin{aligned}\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)&=i\begin{bmatrix}\frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,1}(\boldsymbol{\vartheta})\left(\hat{z}_{t,1}(\boldsymbol{\vartheta},\tau_{2,1})-z_{t,1}(\boldsymbol{\vartheta},\tau_{2,1})\right)\prod_{j\neq 1}\hat{\varphi}_0(0,\tau_{2,j})\\ \vdots\\ \frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,d}(\boldsymbol{\vartheta})\left(\hat{z}_{t,d}(\boldsymbol{\vartheta},\tau_{2,d})-z_{t,d}(\boldsymbol{\vartheta},\tau_{2,d})\right)\prod_{j\neq d}\hat{\varphi}_0(0,\tau_{2,j})\end{bmatrix}\\&=i\begin{bmatrix}\frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,1}(\boldsymbol{\vartheta})\left(\hat{\varphi}_0(0,\tau_{2,1})-\varphi(\tau_{2,1})\right)\prod_{j\neq 1}\hat{\varphi}_0(0,\tau_{2,j})\\ \vdots\\ \frac{1}{T}\sum_{t=1}^T\boldsymbol{\varepsilon}_{t,d}(\boldsymbol{\vartheta})\left(\hat{\varphi}_0(0,\tau_{2,d})-\varphi(\tau_{2,d})\right)\prod_{j\neq d}\hat{\varphi}_0(0,\tau_{2,j})\end{bmatrix}\end{aligned}$$

and

$$\left\| \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2) \right\|^2 = \sum_{i=1}^{d} \left| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{t,i}(\boldsymbol{\vartheta})\left(\hat{\varphi}_0(0,\tau_{2,i}) - \varphi(\tau_{2,i})\right)\prod_{j\neq i}\hat{\varphi}_0(0,\tau_{2,j})\right|^2$$

$$= \sum_{i=1}^{d} \left| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{t,i}(\boldsymbol{\vartheta})\left(\hat{\varphi}_0(0,\tau_{2,i}) - \varphi(\tau_{2,i})\right)\right|^2 \left|\prod_{j\neq i}\hat{\varphi}_0(0,\tau_{2,j})\right|^2$$

$$= \sum_{i=1}^{d} \left| \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{t,i}(\boldsymbol{\vartheta})\right|^2 \left| \frac{1}{T}\sum_{t=1}^{T} z_t(\boldsymbol{\vartheta},\tau_{2,i})\right|^2 \left|\prod_{j\neq i}\hat{\varphi}_0(0,\tau_{2,j})\right|^2$$

since $\left|\frac{1}{T}\sum_{t=1}^{T} z_t(\boldsymbol{\vartheta},\tau_{2,i})\right|^2$ and $\left|\prod_{j\neq i}\hat{\varphi}_0(0,\tau_{2,j})\right|^2$ are uniformly bounded in $\boldsymbol{\vartheta}$ by some constant. And $\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{t,i}(\boldsymbol{\vartheta})\right|^2$ is the univariate version of $\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{t}(\boldsymbol{\vartheta})\right\|^2$. Therefore:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\| \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2) \right\|^2 = \frac{dC}{T}\left\{O(1) + O(m^{1-2\mu_0}) + (1+O(m^{1-\mu_0}))(T-1)^{-\mu_0}\right\}$$

therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2 - \left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2\right|$$
$$= \frac{C}{T^{1/2}}\left\{O(1) + O(m^{1/2-\mu_0}) + (1+O(m^{1/2-\mu_0/2}))(T-1)^{-\mu_0/2}\right\}$$

Uniform Convergence of $\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2 - \left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2$

We have

$$\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2 - \left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2 =$$
$$\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left\{\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.) - \left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\}\right\|^2$$
$$+ 2\left\langle \Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left\{\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.) - \left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\}, \right.$$
$$\left. \overline{\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.) - \left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)}\right\rangle$$

using previous results, we have

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2-\left\|\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},.)\right\|^2\right|$$
$$=\frac{C}{T^{1/2}}\left\{O(1)+O(m^{1/2-\mu_0})+(1+O(m^{1/2-\mu_0/2}))(T-1)^{-\mu_0/2}\right\}=o(1)$$

5.7. Proof Theorem (2.7)

Assumption (2.1) guarantees that space parameter is compact. Besides, Assumption (2.2) joint with the Hallin and Mehta (2015) procedure assure that $\mathscr{R}_0(\boldsymbol{\vartheta})$ is uniquely minimized at $\boldsymbol{\vartheta}_0$. Now, notice that $\mathscr{R}_T(\boldsymbol{\vartheta})$ is a measurable function for each $\boldsymbol{\vartheta}\in\mathscr{V}$. Moreover, the transformation we use, $e^{i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}$, which is an smooth function with all derivatives well defined, and the linearity of our model residuals, $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})=\boldsymbol{\delta}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t$, the function $\mathscr{R}_T(\boldsymbol{\vartheta})$ is continuous for a provided sample $\{\boldsymbol{Y}_t\}_{t=1}^T$.

Thus, all the requirements of Theorem 4.1 in Amemiya (1985) are fulfilled. Additionally, the potential multiple local minimum we can find in the optimization process, something typical in non-linear optimization problems, comes from the fact that the order (p_+,q_+) is not determined. Then, it is possible to find minimizers in causal/invertible and non-causal/non-invertible situations. However, given the correct order (p_+,q_+), we have a unique minimizer. Thus, applying Theorem 4.1 in Amemiya (1985), we have that $\mathbb{P}\left(\|\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}-\boldsymbol{\vartheta}_0\|>\epsilon\right)\to 0$ as $T\to\infty$.

5.8. Proof Theorem (2.8)

To prove the asymptotic distribution of minimum distance estimator $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}$, we use a similar approach that we employed for showing the asymptotic distribution of $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$. Firstly, we will show the uniform boundedness of the derivative of the score $\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta})$.

5.8.1. Uniform Convergence of $\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta})$

For simplicity, we employ model residuals instead of sample model residuals. Notice $\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta})$:

$$\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta})=\frac{2}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\left(1-\frac{|j|}{T}\right)\int\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 dW(\boldsymbol{\tau}_2)+\frac{\pi}{3}\int\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,\boldsymbol{\tau}_2)\right\|^2 dW($$

where

$$\frac{\partial}{\partial \boldsymbol{\vartheta}} \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) \right\|^2 = \left[\frac{\partial}{\partial \vartheta_l} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) \right\|^2 \right]_{l=1,\ldots,K}$$

$$\frac{\partial}{\partial \boldsymbol{\vartheta}} \left\| \Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2) \right\|^2 = \left[\frac{\partial}{\partial \vartheta_l} \left\| \Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2) \right\|^2 \right]_{l=1,\ldots,K}$$

and

$$\frac{\partial}{\partial \vartheta_l} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) \right\|^2 = 2\mathrm{Re}\left\{ \left(\Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) \right)' \left(\frac{\partial}{\partial \vartheta_l} \overline{\Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)} \right) \right\}$$

$$\begin{aligned}
\frac{\partial}{\partial \vartheta_l} \left\| \Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2) \right\|^2 &= 2\mathrm{Re}\left\{ \left(\Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2) \right)' \right. \\
&\quad \left. \left(\frac{\partial}{\partial \vartheta_l} \overline{\Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2)} \right) \right\} \\
\frac{\partial}{\partial \vartheta_l} \overline{\Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)} &= \frac{\partial}{\partial \vartheta_l} \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2) \\
\frac{\partial}{\partial \vartheta_l} \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) &= \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left\{ \frac{\partial}{\partial \vartheta_l} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) + \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \frac{\partial}{\partial \vartheta_l} \hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) \right\} \\
\frac{\partial}{\partial \vartheta_l} \overline{\Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2)} &= \frac{\partial}{\partial \vartheta_l} \Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2) - \frac{\partial}{\partial \vartheta_l} \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,-\boldsymbol{\tau}_2) \\
\frac{\partial}{\partial \vartheta_l} \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2) &=
\end{aligned}$$

$$\begin{bmatrix}
\left(\frac{\partial}{\partial \vartheta_l} \Delta^{(1,0)} \hat{\sigma}_0^{\varepsilon_1(\boldsymbol{\vartheta})}(0,\tau_{2,1}) \right) \prod_{j\neq 1} \hat{\varphi}_0^{\varepsilon_j(\boldsymbol{\vartheta})}(0,\tau_{2,j}) + \Delta^{(1,0)} \hat{\sigma}_0^{\varepsilon_1(\boldsymbol{\vartheta})}(0,\tau_{2,1}) \left(\frac{\partial}{\partial \vartheta_l} \prod_{j\neq 1} \hat{\varphi}_0^{\varepsilon_j(\boldsymbol{\vartheta})}(0,\tau_{2,j}) \right) \\
\vdots \\
\left(\frac{\partial}{\partial \vartheta_l} \Delta^{(1,0)} \hat{\sigma}_0^{\varepsilon_d(\boldsymbol{\vartheta})}(0,\tau_{2,d}) \right) \prod_{j\neq d} \hat{\varphi}_0^{\varepsilon_j(\boldsymbol{\vartheta})}(0,\tau_{2,j}) + \Delta^{(1,0)} \hat{\sigma}_0^{\varepsilon_d(\boldsymbol{\vartheta})}(0,\tau_{2,d}) \left(\frac{\partial}{\partial \vartheta_l} \prod_{j\neq d} \hat{\varphi}_0^{\varepsilon_j(\boldsymbol{\vartheta})}(0,\tau_{2,j}) \right)
\end{bmatrix}$$

Now

$$\frac{\partial}{\partial \boldsymbol{\vartheta}} \mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}) = \left[\frac{\partial}{\partial \vartheta_m} \mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}) \right]_{m=1,\ldots,K} = \left[\frac{\partial}{\partial \vartheta_m \partial \vartheta_l} \mathscr{R}_T(\boldsymbol{\vartheta}) \right]_{m=1,\ldots,K}^{l=1,\ldots,K}$$

$$\begin{aligned}
\frac{\partial}{\partial \vartheta_m \partial \vartheta_l} \mathscr{R}_T(\boldsymbol{\vartheta}) &= \frac{2}{\pi} \sum_{j=1}^{T-1} \frac{1}{j^2} \left(1 - \frac{|j|}{T} \right) \int \frac{\partial}{\partial \vartheta_m \partial \vartheta_l} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2) \right\|^2 dW(\boldsymbol{\tau}_2) + \\
&\quad \frac{\pi}{3} \int \frac{\partial}{\partial \vartheta_m \partial \vartheta_l} \left\| \Delta^{(1,0)} \hat{\sigma}_{0}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)} \right]^{MDIC} (0,\boldsymbol{\tau}_2) \right\|^2 dW(\boldsymbol{\tau}_2)
\end{aligned}$$

with

$$\frac{\partial^2}{\partial\vartheta_m\partial\vartheta_l}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\hat{z}_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)\right\|^2 = 2\mathrm{Re}\left\{\left(\frac{\partial}{\partial\vartheta_m}\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right)'\left(\frac{\partial}{\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)}\right)\right\}+ 2\mathrm{Re}\left\{\left(\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right)'\left(\frac{\partial}{\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)}\right)\right\}$$

$$\frac{\partial^2}{\partial\vartheta_m\partial\vartheta_l}\left\|\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)\right\|^2 = 2\mathrm{Re}\left\{\frac{\partial}{\partial\vartheta_m}\left(\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)\right)' \left(\frac{\partial}{\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)}\right)\right\}+ 2\mathrm{Re}\left\{\left(\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)\right)' \left(\frac{\partial}{\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)}\right)\right\}$$

Following the same strategy as in the case of i.i.d. errors

$$\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathbf{s}^{\mathscr{R}}_T(\hat{\boldsymbol{\vartheta}}_T)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathbf{s}^{\mathscr{R}}_T(\boldsymbol{\vartheta}_0)\right\| \le C\max_{(k,l,m)\in\{1,\dots,K\}^3}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\mathscr{R}_T(\boldsymbol{\vartheta})\right\|\left\|\hat{\boldsymbol{\vartheta}}_T-\boldsymbol{\vartheta}_0\right\|$$

one of the most problematic terms in $\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\mathscr{R}_T(\boldsymbol{\vartheta})$ is $\left(\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right)'\left(\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)}\right)$ and notices

$$\frac{\partial}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)} = \mathbb{E}_{T-|j|}\left[\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)+\frac{\partial^2}{\partial\vartheta_m\partial\vartheta_l}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial}{\partial\vartheta_k}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]+ \mathbb{E}_{T-|j|}\left[\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_l}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial}{\partial\vartheta_m}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)+\frac{\partial}{\partial\vartheta_l}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_m}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]+ \mathbb{E}_{T-|j|}\left[\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_m}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial}{\partial\vartheta_l}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)+\frac{\partial}{\partial\vartheta_m}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial^2}{\partial\vartheta_k\partial\vartheta_l}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]+ \mathbb{E}_{T-|j|}\left[\frac{\partial}{\partial\vartheta_k}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial^2}{\partial\vartheta_m\partial\vartheta_l}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)+\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]$$

Thus

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left(\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right)'\left(\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)}\right)\right\| \le \left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^{9/4}\right\}^{4/9}\times \left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)}\right\|^{9/5}\right\}^{5/9}$$

and appliying Minkowski inequality

$$\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)}\right\|^{9/5}\right\}^{5/9} \leq \left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\mathbb{E}_{T-|j|}\left[\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]^{9/5}\right\|\right\}^{5/9} +$$

$$\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\mathbb{E}_{T-|j|}\left[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]^{9/5}\right\|\right\}^{5/9} +$$

$$C\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\mathbb{E}_{T-|j|}\left[\frac{\partial^2}{\partial\vartheta_m\partial\vartheta_l}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial}{\partial\vartheta_k}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]^{9/5}\right\|\right\}^{5/9} +$$

$$C\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\mathbb{E}_{T-|j|}\left[\frac{\partial}{\partial\vartheta_k}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\frac{\partial^2}{\partial\vartheta_m\partial\vartheta_l}\hat{z}_{t-|j|}(\boldsymbol{\vartheta},-\boldsymbol{\tau}_2)\right]^{9/5}\right\|\right\}^{5/9}$$

and each of these terms is bounded by a factor proportional to $C\|\boldsymbol{\tau}_2\|^4$ if and only if $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^{4.5}<\infty$, given that data is generated by the finite SVARMA model. For the rest of the terms, the previous condition is sufficient for their boundedness.

The other problematic term in $\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\mathscr{R}_T(\boldsymbol{\vartheta})$ is $2\mathrm{Re}\left\{\left(\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)\right)'\left(\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)}\right)\right\}$ where

$$\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)} = \frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)-\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)$$

$$\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2) = \left[\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\mathbb{E}_T\left(\varepsilon_{n,t}(\boldsymbol{\vartheta})\hat{z}_{n,t}(\boldsymbol{\vartheta},-\tau_{2,n})\right)\prod_{r\neq n}\hat{\varphi}_0^{\varepsilon_{r,t}(\boldsymbol{\vartheta})}(0,-\tau_{2,r})\right]_{n=1,\ldots,d}$$

Thus, by applying Holder inequality

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left(\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)\right)'\left(\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\overline{\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},.)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,.)}\right)\right\| \leq$$

$$\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right\|^{9/4}\right)^{4/9}\times$$

$$\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)-\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right\|^{9/5}\right)^{5/9}$$

by Minkowski inequality

$$\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right\|^{9/4}\right)^{4/9}\leq\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)\right\|^{9/4}\right)^{4/9}+\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right\|^{9/4}\right)^{4/9}$$

$$\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)-\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right\|^{9/5}\right)^{5/9}\leq\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\Delta^{(1,0)}\hat{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},-\boldsymbol{\tau}_2)\right\|^{9/5}\right)^{5/9}+\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right\|^{9/5}\right)^{5/9}$$

which are bounded as long as the condition $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^{4.5}<\infty$ holds.

Therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{\partial^3}{\partial\vartheta_k\partial\vartheta_m\partial\vartheta_l}\mathscr{R}_T(\boldsymbol{\vartheta})\right\|\leq C\frac{2}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\left(1-\frac{|j|}{T}\right)\int\|\boldsymbol{\tau}_2\|^{20/9}\,dW(\boldsymbol{\tau}_2)+C\frac{\pi}{3}\int\|\boldsymbol{\tau}_2\|^{20/9}\,dW(\boldsymbol{\tau}_2)\leq C$$

Since $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{R}}-\boldsymbol{\vartheta}_0=O_P(T^{-1/2})$, then

$$\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathbf{s}_T^{\mathscr{R}}(\hat{\boldsymbol{\vartheta}}_T)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}_0)\right\|=o_P(1)$$

5.8.2. Weak Convergence of $\sqrt{T}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}_0)$

$$
\begin{aligned}
T^{1/2}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}_0) &= T^{1/2}\frac{2}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\left(1-\frac{|j|}{T}\right)\int\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2 dW(\boldsymbol{\tau}_2)\\
&+T^{1/2}\frac{\pi}{3}\int\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right]^{MDIC}(0,\boldsymbol{\tau}_2)\right\|^2 dW(\boldsymbol{\tau}_2)\\
&=T^{1/2}\frac{4}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\left(1-\frac{|j|}{T}\right)\int\operatorname{Re}\left\{\left(\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},\boldsymbol{\tau}_2)\right)'\left(\frac{\partial}{\partial\boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2)\right)\right\}dW(\boldsymbol{\tau}_2)+\\
T^{1/2}\frac{2\pi}{3}&\int\operatorname{Re}\left\{\left(\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\tilde{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(0,\boldsymbol{\tau}_2)\right)'\left(\frac{\partial}{\partial\boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right)\right\}\\
&+o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2) &= i\,\mathbb{C}\mathrm{ov}\left(\boldsymbol{\varepsilon}_t,\frac{\partial}{\partial\boldsymbol{\vartheta}}z_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\right)+i\,\mathbb{C}\mathrm{ov}\left(\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)\boldsymbol{\varepsilon}_t,z_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\right)\\
&= i\,\mathbb{C}\mathrm{ov}\left(\boldsymbol{\varepsilon}_t,-i\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)\boldsymbol{\varepsilon}_{t-|j|}e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}\right)+i\,\mathbb{C}\mathrm{ov}\left(\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)\boldsymbol{\varepsilon}_t,e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}\right)\\
&= -i^2\,\mathbb{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}\right)\dot{\boldsymbol{\delta}}_{-|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2+i\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_{t-|j|}e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}\right)\\
&= \xi_1^{0,|j|}(\boldsymbol{\tau}_2)+\xi_2^{0,|j|}(\boldsymbol{\tau}_2)\\
\frac{\partial}{\partial\boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2) &= \xi_1^{0,0}(\boldsymbol{\tau}_2)+\xi_2^{0,0}(\boldsymbol{\tau}_2)\\
\frac{\partial}{\partial\boldsymbol{\vartheta}}\left[F^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right]^{MDIC}(0,-\boldsymbol{\tau}_2) &= i\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\{\mathbb{C}\mathrm{ov}\left(\varepsilon_{t,n}(\boldsymbol{\vartheta}_0),z_t(-\tau_{2,n},\boldsymbol{\vartheta}_0)\right)\prod_{m\neq n}\varphi^{\varepsilon(\boldsymbol{\vartheta}_0)}(-\tau_{2,m})\right\}\right]_{n=1,\dots,d}=\xi_3^{0,0}(\boldsymbol{\tau}_2)\\
\frac{\partial}{\partial\boldsymbol{\vartheta}}\mathbb{C}\mathrm{ov}\left(\varepsilon_{t,n}(\boldsymbol{\vartheta}_0),z_t(-\tau_{2,n},\boldsymbol{\vartheta}_0)\right) &= \boldsymbol{e}_n'\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0)\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{-i\tau_{2,n}\varepsilon_{t,n}}\right]-i\tau_{2,n}\boldsymbol{e}_n'\,\mathbb{E}\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'e^{-i\tau_{2,n}\varepsilon_{t,n}}\right]\dot{\boldsymbol{\delta}}_0'(\boldsymbol{\vartheta}_0)\boldsymbol{e}_n\\
\frac{\partial}{\partial\boldsymbol{\vartheta}}\left\{\prod_{m\neq n}\varphi^{\varepsilon(\boldsymbol{\vartheta}_0)}(-\tau_{2,m})\right\} &= \sum_{n_1\in\mathscr{I}_n}\boldsymbol{e}_{n_1}'\,\mathbb{E}\left[\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)\boldsymbol{\varepsilon}_te^{-i\tau_{2,n_1}\varepsilon_{t,n_1}}\right]\prod_{m\neq n_1}\varphi^{\varepsilon(\boldsymbol{\vartheta}_0)}(-\tau_{2,m})\\
&= \sum_{n_1\in\mathscr{I}_n}\boldsymbol{e}_{n_1}'\,\mathbb{E}\left[\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0)\boldsymbol{\varepsilon}_te^{-i\tau_{2,n_1}\varepsilon_{t,n_1}}\right]\prod_{m\neq n_1}\varphi^{\varepsilon(\boldsymbol{\vartheta}_0)}(-\tau_{2,m})=\xi_{3,n}^{0,0}(\boldsymbol{\tau}_2)
\end{aligned}
$$

with $\mathscr{I}_n=\{1,\dots,d\}/\{n\}$, the set of indexes from 1 to d excluding the n-th value. Thus:

$$
\begin{aligned}
\left(\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},\boldsymbol{\tau}_2)\right)'\left(\frac{\partial}{\partial\boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2)\right) &= \left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_tz_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\right)'\left(\xi_1^{0,|j|}(\boldsymbol{\tau}_2)+\xi_2^{0,|j|}(\boldsymbol{\tau}_2)\right)\\
&= \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{z_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\boldsymbol{\varepsilon}_t'\xi_1^{0,|j|}(\boldsymbol{\tau}_2)+z_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\boldsymbol{\varepsilon}_t'\xi_2^{0,|j|}(\boldsymbol{\tau}_2)\right\}
\end{aligned}
$$

and

$$\int \left(\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},\boldsymbol{\tau}_2)\right)' \left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2)\right) dW(\boldsymbol{\tau}_2) = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t' \int z_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\xi_1^{0,|j|}(\boldsymbol{\tau}_2) dW(\boldsymbol{\tau}_2)$$

$$+\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t' \int z_{t-|j|}(\boldsymbol{\vartheta}_0,-\boldsymbol{\tau}_2)\xi_2^{0,|j|}(\boldsymbol{\tau}_2) dW(\boldsymbol{\tau}_2)$$

$$=\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t' \left(E_{t-|j|}^0 + S_{t-|j|}^0\right)$$

and

$$\frac{4}{\pi}\sum_{j=1}^{T-1}\frac{1}{j^2}\left(1-\frac{|j|}{T}\right)\int \operatorname{Re}\left\{\left(\Delta^{(1,0)}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},\boldsymbol{\tau}_2)\right)' \left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2)\right)\right\} dW(\boldsymbol{\tau}_2) = \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{\varepsilon}_t'\left(E_{t-1}^0 + S_{t-1}^0\right)$$

$$E_{t-1}^0 = \frac{4}{\pi}\sum_{j=1}^{t-1}\frac{1}{j^2}E_{t-|j|}^0$$

$$S_{t-1}^0 = \frac{4}{\pi}\sum_{j=1}^{t-1}\frac{1}{j^2}S_{t-|j|}^0$$

On the other hand

$$\left(\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\tilde{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(0,\boldsymbol{\tau}_2)\right)' \left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2) - \frac{\partial}{\partial \boldsymbol{\vartheta}}\left[F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}^{(1,0)}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right) =$$

$$\left(\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\tilde{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(0,\boldsymbol{\tau}_2)\right)' \boldsymbol{E}^{0,0}(\boldsymbol{\tau}_2)$$

$$\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\varepsilon_{t,1}z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)\\ \vdots \\ \varepsilon_{t,d}z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2)\end{bmatrix}$$

$$z_t(\boldsymbol{\vartheta}_0,\boldsymbol{\tau}_2) = \prod_{m=1}^{d} e^{i\tau_{2,m}\varepsilon_{t,m}} - \prod_{m=1}^{d}\mathbb{E}\left[e^{i\tau_{2,m}\varepsilon_{t,m}}\right]$$

$$\left[\tilde{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(0,\boldsymbol{\tau}_2) = \frac{1}{T}\sum_{t=1}^{T}\left[\varepsilon_{t,n}z_{t,n}(\boldsymbol{\vartheta}_0,\tau_{2,n})\prod_{m\neq n}\mathbb{E}\left(e^{i\tau_{2,m}\varepsilon_{t,m}}\right)\right]_{n=1,\ldots,d}$$

$$\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\tilde{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(0,\boldsymbol{\tau}_2) = \frac{1}{T}\sum_{t=1}^{T}\left[\varepsilon_{t,n}e^{i\tau_{2,n}\varepsilon_{t,n}}\prod_{m\neq n}\mathbb{E}\left(e^{i\tau_{2,m}\varepsilon_{t,m}}\right)\right]_{n=1,\ldots,d}$$

thus

$$\frac{2\pi}{3}\int \operatorname{Re}\left\{\left(\Delta^{(1,0)}\tilde{\sigma}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2) - \left[\tilde{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(0,\boldsymbol{\tau}_2)\right)' \left(\frac{\partial}{\partial \boldsymbol{\vartheta}}\Delta^{(1,0)}\sigma_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\mathbf{0},-\boldsymbol{\tau}_2) - \frac{\partial}{\partial \boldsymbol{\vartheta}}\left[F_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}^{(1,0)}\right]^{MDIC}(0,-\boldsymbol{\tau}_2)\right)\right\} = \frac{1}{T}\sum_{t=1}^{T}X_t^0$$

where $X_t^0 = \frac{2\pi}{3}\left\{\sum_{n=1}^d \varepsilon_{t,n} \int \boldsymbol{E}_n^{\prime(0,0)}(\boldsymbol{\tau}_2) e^{i\tau_{2,n}\varepsilon_{t,n}} \prod_{m\neq n} \mathbb{E}\left(e^{i\tau_{2,m}\varepsilon_{t,m}}\right) dW(\boldsymbol{\tau}_2)\right\}'$ and $\boldsymbol{E}_n^{\prime(0,0)}(\boldsymbol{\tau}_2)$ represents the n-th column of $\left(\boldsymbol{E}^{0,0}(\boldsymbol{\tau}_2)\right)'$.

Therefore

$$T^{1/2}\mathbf{s}_T^{\mathscr{R}}(\boldsymbol{\vartheta}_0) = T^{1/2}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\frac{1}{T}\sum_{t=2}^T \boldsymbol{\varepsilon}_t'\left(E_{t-1}^0 + S_{t-1}^0\right) \\ \\ \frac{1}{T}\sum_{t=1}^T X_t^0\end{bmatrix} + o_P(1)$$

it is straightforward to notice that $\sum_{t=2}^T \boldsymbol{\varepsilon}_t'\left(E_{t-1}^0 + S_{t-1}^0\right)$ and $\sum_{t=1}^T X_t^0$ are m.d.s. under the Assumption (2.4).

5.9. Auxiliary Lemmas

5.9.1. Lemma 1

Since $\det(\boldsymbol{\Phi}(z,\boldsymbol{\vartheta})) \times \det(\boldsymbol{\Theta}(z,\boldsymbol{\vartheta})) \neq 0$ for all $z \in \mathbb{T}$, then their inverses

$$\boldsymbol{\Phi}^{-1}(z,\boldsymbol{\vartheta}) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Phi}_j^{(-1)}(\boldsymbol{\vartheta}) z^j$$
$$\boldsymbol{\Theta}^{-1}(z,\boldsymbol{\vartheta}) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Theta}_j^{(-1)}(\boldsymbol{\vartheta}) z^j$$

hold the following

$$\sup_{\theta\in\mathcal{V}} \left\|\boldsymbol{\Phi}^{-1}(z,\boldsymbol{\vartheta})\right\| \times \sup_{\theta\in\mathcal{V}} \left\|\boldsymbol{\Theta}^{-1}(z,\boldsymbol{\vartheta})\right\| \leq C < \infty$$

and we can assume

$$\sup_{\theta\in\mathcal{V}} \left\|\boldsymbol{\Phi}_j^{(-1)}(\boldsymbol{\vartheta})\right\| \leq C|j|^{-\mu_0}$$
$$\sup_{\theta\in\mathcal{V}} \left\|\boldsymbol{\Theta}_j^{(-1)}(\boldsymbol{\vartheta})\right\| \leq C|j|^{-\mu_0}$$

with $\mu_0 > 1$ and $|j| > 0$.

Now, without considering the matrix of contemporaneous effects, $\boldsymbol{\Psi}(z,\boldsymbol{\vartheta}) = \boldsymbol{\Phi}^{-1}(z,\boldsymbol{\vartheta})\boldsymbol{\Theta}(z,\boldsymbol{\vartheta})$

which is equivalent to

$$\boldsymbol{\Psi}(z,\boldsymbol{\vartheta})=\left(\sum_{j=-\infty}^{\infty}\boldsymbol{\Phi}_j^{(-1)}(\boldsymbol{\vartheta})z^j\right)\left(\sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\vartheta})z^j\right)$$
$$=\sum_{j_1=-\infty}^{\infty}\sum_{j_2=0}^{q}\boldsymbol{\Phi}_{j_1}^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{\Theta}_{j_2}(\boldsymbol{\vartheta})z^{j_1+j_2}=\sum_{j=-\infty}^{\infty}\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})z^j$$

where $\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})=\sum_{j_2=0}^{q}\boldsymbol{\Phi}_{j-j_2}^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{\Theta}_{j_2}(\boldsymbol{\vartheta})$.

From above, it is straightforward to notice

$$\|\boldsymbol{\Psi}(z,\boldsymbol{\vartheta})\|\leq\left\|\sum_{j=-\infty}^{\infty}\boldsymbol{\Phi}_j^{(-1)}(\boldsymbol{\vartheta})z^j\right\|\left\|\sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\vartheta})z^j\right\|\leq C\left\|\sum_{j=0}^{q}\boldsymbol{\Theta}_j(\boldsymbol{\vartheta})z^j\right\|<\infty$$

also note

$$\sup_{\theta\in\mathcal{V}}\left\|\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})\right\|=\sup_{\theta\in\mathcal{V}}\left\|\sum_{j_2=0}^{q}\boldsymbol{\Phi}_{j-j_2}^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{\Theta}_{j_2}(\boldsymbol{\vartheta})\right\|$$
$$\leq\sum_{j_2=0}^{q}\sup_{\theta\in\mathcal{V}}\left\|\boldsymbol{\Phi}_{j-j_2}^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{\Theta}_{j_2}(\boldsymbol{\vartheta})\right\|\leq C\sum_{j_2=0}^{q}|j-j_2|^{-\mu_0}|j_2|^{-\mu_0}$$
$$\leq C\begin{cases}2+\int_1^q|j_2|^{-2\mu_0}, & j=0\\ 1+\int_1^q|j_2|^{-\mu_0}, & j=1,\ldots,q\\ |q-1|^{-\mu_0}+\int_{q-1}^{j}|j_2|^{-\mu_0}, & j=q+1,q+2,\ldots\\ |j|^{-\mu_0}+\int_j^{j+q}|j_2|^{-\mu_0}, & j=\ldots,-2,-1\end{cases}$$

this implies that for $j=0$, $\sup_{\theta\in\mathcal{V}}\left\|\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})\right\|\leq C$ and for $|j|>0$, it holds $\sup_{\theta\in\mathcal{V}}\left\|\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})\right\|\leq C|j|^{1-\mu_0}$. Then, selecting $\mu_0>2$, we obtain $\sup_{\theta\in\mathcal{V}}\left\|\boldsymbol{\Psi}_j(\boldsymbol{\vartheta})\right\|\leq C|j|^{-\bar{\mu}_0}$ with $\bar{\mu}_0>1$.

A similar procedure can be applied to show absolute summability of $\boldsymbol{\Psi}^{-1}(z,\boldsymbol{\vartheta})$ and to find out that $\sup_{\theta\in\mathcal{V}}\left\|\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\right\|\leq C|j|^{-\bar{\mu}_0}$. □

5.9.2. Lemma 2

Under Assumptions (2.1)-(2.3)

$$\sup_{\boldsymbol{\vartheta}}\left|\sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\leq C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}|j|^{1-\mu_0},\ j\neq0 \tag{5.25}$$

Under Assumptions (2.1)-(2.3.(ii-iv)) and (2.4)

$$\sup_{\boldsymbol{\vartheta}} \left| \Delta^{(1,0)} \sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right| \leq C_{(\boldsymbol{\tau}_2)} |j|^{1/2-\mu_0},\ j \neq 0 \tag{5.26}$$

where $C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}$ is a constant term depending on $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$.

Proof.

For (5.25):

In case of causal and invertible representation, given the existence of second order moments, which provides the sufficient condition of $\mathbb{E}|\boldsymbol{\varepsilon}_t\|^{\delta} < \infty$ for $\delta > 1$, and the strong mixing behaviour of structural innovations, the process $z_t(\boldsymbol{\tau}, \boldsymbol{\vartheta})$ is a strong mixing process as well for any $\boldsymbol{\vartheta} \in \mathcal{V}$. Unfortunately, such a result is not guaranteed for possibly non-causal/non-invertible process. Doukhan and Louhichi (1999) provide useful tools for bounding covariances when one faces non-causal Bernoulli transformations of independent errors. Using the results in Dedecker et al. (2007); Doukhan and Louhichi (1999), we have that for each (point-wise case) $\boldsymbol{\vartheta} \in \mathcal{V}$, $\left| \sigma_j^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right| \leq C(\|\boldsymbol{\tau}_1\| + \|\boldsymbol{\tau}_2\|)|j|^{1-\mu_0}$ where $\mu_0 > 2$ is the decay rate of VARMA coefficients. For showing the uniform boundedness, we use the following:

$$z_t(\boldsymbol{\tau}; \boldsymbol{\vartheta}) = z_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta}) + \boldsymbol{\zeta}_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta})$$

where $z_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta})$ is defined similary to $z_t(\boldsymbol{\tau}; \boldsymbol{\vartheta})$ but using $\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}) = \boldsymbol{\delta}(L; \boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t \mathbf{1}\{|t| \leq m\}$. Notice that when $m \to \infty$, $\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}) \to \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})$. For convenience, we take $m = [j/2]$, where [x] denotes the nearest integer to x; and $\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta}) = z_t(\boldsymbol{\tau}; \boldsymbol{\vartheta}) - z_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta})$. Now, notice

$$\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta}) = \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})) - \mathbb{E}\left[\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}))\right]$$

thus

$$\begin{aligned}
\sup_{\boldsymbol{\vartheta}} \mathbb{E}\left|\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}; \boldsymbol{\vartheta})\right| &\leq \sup_{\boldsymbol{\vartheta}} \mathbb{E}\left|\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}))\right| + \sup_{\boldsymbol{\vartheta}} \left|\mathbb{E}\left[\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}))\right]\right| \\
&\leq 2 \sup_{\boldsymbol{\vartheta}} \mathbb{E}\left|\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}))\right| \\
&\leq 4 \sup_{\boldsymbol{\vartheta}} \mathbb{E}\left|\boldsymbol{\tau}'\left(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})\right)\right| \\
&\leq C(\|\boldsymbol{\tau}\|) \sum_{|j|>m} \sup_{\boldsymbol{\vartheta}} \left\|\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right\| = 2C(\|\boldsymbol{\tau}\|) \sum_{j>m} \sup_{\boldsymbol{\vartheta}} \left\|\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right\| \leq C(\|\boldsymbol{\tau}\|)|j|^{1-\mu_0}
\end{aligned}$$

Now, note that

$$
\begin{aligned}
\mathbb{E}\left[z_t(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})z_{t-j}(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\right] &= \mathbb{E}\left[z_t^m(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})z_{t-j}^m(\boldsymbol{\tau}_2,\boldsymbol{\vartheta})\right] + \mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})z_{t-j}^m(\boldsymbol{\tau}_2,\boldsymbol{\vartheta})\right] \\
&\quad + \mathbb{E}\left[\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})z_t^m(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\right] + \mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})\right] \\
\sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[z_t(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})z_{t-j}(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\right]\right| &\leq \sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[z_t^m(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})z_{t-j}^m(\boldsymbol{\tau}_2,\boldsymbol{\vartheta})\right]\right| + \sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})z_{t-j}^m(\boldsymbol{\tau}_2,\boldsymbol{\vartheta})\right]\right| \\
&\quad + \sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})z_t^m(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\right]\right| + \sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})\right]\right|
\end{aligned}
$$

when $m=[|j|/2]$, $\boldsymbol{\varepsilon}_t^m=\sum_{|l|\leq[|j|/2]}\boldsymbol{\delta}_l(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-l}$ and $\boldsymbol{\varepsilon}_{t-j}^m=\sum_{|l|\leq[|j|/2]}\boldsymbol{\delta}_l(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-j}$ are independent, then $\mathbb{E}\left[z_t^m(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})z_{t-j}^m(\boldsymbol{\tau}_2,\boldsymbol{\vartheta})\right]=0$. Besides, $|z_t^m(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})|\leq 2$, and $\mathbb{E}\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})=\mathbb{E}\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})$. Hence

$$
\begin{aligned}
\sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[z_t(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})z_{t-j}(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\right]\right| &\leq 2\sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})\right]\right| + 2\sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})\right]\right| \\
&\quad + \sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})\right]\right| \\
&\leq 2\sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})\right]\right| + 4\sup_{\boldsymbol{\vartheta}}\left|\mathbb{E}\left[\boldsymbol{\zeta}_{t-j}^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})\right]\right| \\
&\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)|j|^{1-\mu_0}
\end{aligned}
$$

where before last inequality uses the fact that $\left|\boldsymbol{\zeta}_t^m(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})\right|\leq 4$. Then, Equation (5.25) holds.

For (5.26):
Take $m=[j/2]$ and, as in the previous case, we use $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})=\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})+\xi_t^m(\boldsymbol{\vartheta})$ with $\xi_t^m(\boldsymbol{\vartheta})=\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})$. Besides

$$
\begin{aligned}
\sup_{\boldsymbol{\vartheta}}\mathbb{E}\left|\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right)-\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})\right)\right| &= \sup_{\boldsymbol{\vartheta}}\mathbb{E}\left|\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta},m)\right)-1\right| \\
&\leq 2\sup_{\boldsymbol{\vartheta}}\mathbb{E}\left\|\boldsymbol{\tau}'\sum_{|l|>m}\boldsymbol{\delta}_l(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-l}\right\| \\
&\leq C\|\boldsymbol{\tau}\|\sum_{|l|>m}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\delta}_l(\boldsymbol{\vartheta})\| \leq C\|\boldsymbol{\tau}\|\,|j|^{1-\mu_0}
\end{aligned}
$$

and

$$
\sup_{\boldsymbol{\vartheta}}\mathbb{E}\left|\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right)-\exp\left(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})\right)\right|^2 \leq 2\|\boldsymbol{\tau}\|^2\sup_{\boldsymbol{\vartheta}}\mathbb{E}\left\|\sum_{|l|>m}\boldsymbol{\delta}_l(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-l}\right\|^2 \leq C\|\boldsymbol{\tau}\|^2|j|^{1-2\mu_0}
$$

where we use

$$\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) \right\|^2 = \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \sum_{|j|>m} \boldsymbol{\delta}_j(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t-j} \right\|^2 \\ \leq C m^{1-2\mu_0}$$

thus

$$\begin{aligned} \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left| \boldsymbol{\zeta}_t^m(\boldsymbol{\tau}, \boldsymbol{\vartheta}) \right|^2 \leq & \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left| \exp\left(i \boldsymbol{\tau}' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right) - \exp\left(i \boldsymbol{\tau}' \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})\right) \right|^2 + 2 \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left| \exp\left(i \boldsymbol{\tau}' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right) - \exp\left(i \boldsymbol{\tau}' \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})\right) \right| \right)^2 \\ & + \sup_{\boldsymbol{\vartheta}} \left| \mathbb{E} \left[\exp\left(i \boldsymbol{\tau}' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right) - \exp\left(i \boldsymbol{\tau}' \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta})\right) \right] \right|^2 \\ \leq & C \left\| \boldsymbol{\tau} \right\|^2 |j|^{2-2\mu_0} \end{aligned}$$

Hence, since $\Delta^{(1,0)} \sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) = i\, \mathbb{E}\left(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) z_{t-|j|}(\boldsymbol{\vartheta}, \boldsymbol{\tau}_2)\right)$, we get

$$\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) z_{t-|j|}(\boldsymbol{\vartheta}, \boldsymbol{\tau}_2) \right\| = \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) z_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right\| + \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) \boldsymbol{\zeta}_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right\| + \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}) \boldsymbol{\zeta}_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right\|$$

by Horner inequality:

$$\begin{aligned} \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) z_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right\| &\leq \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) \right\|^2 \right)^{1/2} \times \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left| z_{t-|j|}^m(\boldsymbol{\tau}_2 \boldsymbol{\vartheta}) \right|^2 \right)^{1/2} \\ &\leq C |j|^{1/2-\mu_0} \\ \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) \boldsymbol{\zeta}_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right\| &\leq \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left| \boldsymbol{\zeta}_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right|^2 \right)^{1/2} \times \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\xi}_t^m(\boldsymbol{\vartheta}) \right\|^2 \right)^{1/2} \\ &\leq C \left\| \boldsymbol{\tau}_2 \right\| |j|^{3/2-2\mu_0} \\ \sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}) \boldsymbol{\zeta}_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right\| &\leq \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}) \right\|^2 \right)^{1/2} \times \left(\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left| \boldsymbol{\zeta}_{t-|j|}^m(\boldsymbol{\tau}_2, \boldsymbol{\vartheta}) \right|^2 \right)^{1/2} \\ &\leq C \left\| \boldsymbol{\tau}_2 \right\| |j|^{1-\mu_0} \end{aligned}$$

where we use the fact that $\sup_{\boldsymbol{\vartheta}} \mathbb{E} \left\| \boldsymbol{\varepsilon}_t^m(\boldsymbol{\vartheta}) \right\|^2 < \infty$. □

5.9.3. Lemma 3

Under Assumptions (2.1)-(2.3), it holds

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| = C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}(T-|j|)^{-1} \tag{5.27}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 = C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}(T-|j|)^{-2} \tag{5.28}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| = C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}T^{-1} \tag{5.29}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 = C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}T^{-2} \tag{5.30}$$

where $\hat{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \prod_{m=1}^{d}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^{d}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},0)\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2m})$ and $\overline{\sigma}_{0,ICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \prod_{m=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^{d}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},0)\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2m})$.

Proof.

For (5.27):

Notice

$$\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \left(\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right)$$

$$\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| \leq \left|\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| + \left|\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right|$$

and

$$\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\exp\left[i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})+i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right]-\exp\left[i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right]\right\}$$

By Triangle Inequality and the result stated in Billingsley (2008, Section 26-pp.343):

$$\begin{aligned}
\left|\hat{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| &\leq \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left|\boldsymbol{\tau}_1'(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))+\boldsymbol{\tau}_2'\left(\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right)\right| \\
&= \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left|\boldsymbol{\tau}_1'\left(\sum_{l=-\infty}^{t-2}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}+\sum_{l=t-T+1}^{\infty}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}\right)\right. \\
&\quad \left.+\boldsymbol{\tau}_2'\left(\sum_{l=-\infty}^{t-2}\boldsymbol{\Psi}_l^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-j-l}+\sum_{l=t-T+1}^{\infty}\boldsymbol{\Psi}_l^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-j-l}\right)\right|
\end{aligned}$$

hence

$$
\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| &\leq C\left\|\boldsymbol{\tau}_1\right\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\sum_{l=-\infty}^{t-T-1}\sup_{\boldsymbol{\vartheta}}\left\|\boldsymbol{\Psi}^{(-1)}_{l}(\boldsymbol{\vartheta})\right\|+\sum_{l=t}^{\infty}\sup_{\boldsymbol{\vartheta}}\left\|\boldsymbol{\Psi}^{(-1)}_{l}(\boldsymbol{\vartheta})\right\|\right\}\\
&\quad+C\left\|\boldsymbol{\tau}_2\right\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\sum_{l=-\infty}^{t-|j|-T-1}\sup_{\boldsymbol{\vartheta}}\left\|\boldsymbol{\Psi}^{(-1)}_{l}(\boldsymbol{\vartheta})\right\|+\sum_{l=t-|j|}^{\infty}\sup_{\boldsymbol{\vartheta}}\left\|\boldsymbol{\Psi}^{(-1)}_{l}(\boldsymbol{\vartheta})\right\|\right\}\\
&\leq\frac{C\left(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|\right)}{T-|j|}\sum_{t=1+|j|}^{T}\left\{(T+1-t)^{1-\mu_0}+t^{1-\mu_0}+(t-|j|)^{1-\mu_0}+(T+|j|+1-t)^{1-\mu_0}\right.\\
&\leq\frac{2C\left(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|\right)}{T-|j|}\left(\sum_{t=1}^{T-|j|}t^{1-\mu_0}+\sum_{t=1+|j|}^{T}t^{1-\mu_0}\right)\leq\frac{C\left(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|\right)}{T-|j|},\ \mu_0>2
\end{aligned}
$$

where second inequality exploits the fact that $\mathbb{E}\|\boldsymbol{Y}_t\|<\infty$ for all $t\in\mathbb{Z}$. Besides, by Triangle inequality and the fact that $|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})|\leq 1$ and $|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})|\leq 1$:

$$
\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|\leq\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\right|+\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|
$$

hence, using the previous result twice, we get:

$$
\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|\leq\frac{C\left(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|\right)}{T-|j|}
$$

and

$$
\therefore\quad\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\leq\frac{C\left(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|\right)}{T-|j|}=C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}(T-|j|)^{-1}
$$

where $C_{(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)}=C\left(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|\right)$.

For (5.28):

Note that

$$
\begin{aligned}
\left|\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 &\leq\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2+\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|^2\\
&\quad+2\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\cdot\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|
\end{aligned}
$$

First, analyzing the term $\left|\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2$:

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 &= \mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} e^{i\left(\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right)}-e^{i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)}\right|^2 \\
&= \mathbb{E}\sup_{\boldsymbol{\vartheta}}\frac{1}{(T-|j|)^2}\sum_{t=1+|j|}^{T}\sum_{t'=1+|j|}^{T}\left\{e^{i\left(\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right)}-e^{i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})\right)}\right\}\times \\
&\qquad \left\{e^{i\left(\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_{t'-|j|}(\boldsymbol{\vartheta})\right)}-e^{i\left(\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})+\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t'-|j|}(\boldsymbol{\vartheta})\right)}\right\} \\
&\le C\frac{1}{(T-|j|)^2}\sum_{t=1+|j|}^{T}\sum_{t'=1+|j|}^{T}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\{\left|\boldsymbol{\tau}_1'(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))+\boldsymbol{\tau}_2'\left(\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right)\right|\times\right. \\
&\qquad \left.\left|\boldsymbol{\tau}_1'(\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta}))+\boldsymbol{\tau}_2'\left(\boldsymbol{\varepsilon}_{t'-|j|}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t'-|j|}(\boldsymbol{\vartheta})\right)\right|\right\} \\
&\le C\frac{1}{(T-|j|)^2}\sum_{t=1+|j|}^{T}\sum_{t'=1+|j|}^{T}\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\boldsymbol{\tau}_1'(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))+\boldsymbol{\tau}_2'\left(\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right)\right|^2\right\}^{1/2}\times \\
&\qquad \left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\boldsymbol{\tau}_1'(\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta}))+\boldsymbol{\tau}_2'\left(\boldsymbol{\varepsilon}_{t'-|j|}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t'-|j|}(\boldsymbol{\vartheta})\right)\right|^2\right\}^{1/2} \\
&\le C\frac{1}{(T-|j|)^2}\left(\sum_{t=1+|j|}^{T}\left\{\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\boldsymbol{\tau}_1'(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))+\boldsymbol{\tau}_2'\left(\boldsymbol{\varepsilon}_{t-|j|}(\boldsymbol{\vartheta})-\hat{\boldsymbol{\varepsilon}}_{t-|j|}(\boldsymbol{\vartheta})\right)\right|^2\right\}^{1/2}\right)^2
\end{aligned}$$

and using the results for equation (5.27), we get

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 &\le C\frac{\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)}{(T-|j|)^2}\left(\sum_{t=1+|j|}^{T}\left\{(T+1-t)^{1-\mu_0}+t^{1-\mu_0}+\right.\right. \\
&\qquad \left.\left.(t-|j|)^{1-\mu_0}+(T+|j|+1-t)^{1-\mu_0}\right\}\right)^2 \\
&\le 4C\frac{\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)}{(T-|j|)^2}\left(\sum_{t=1+|j|}^{T}t^{1-\mu_0}+\sum_{t=1}^{T-|j|}t^{1-\mu_0}\right)^2 \\
&\le C\frac{\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)}{(T-|j|)^2}\left(\sum_{t=1}^{\infty}t^{1-\mu_0}\right)^2 \\
&\le C\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)(T-|j|)^{-2}, \quad \text{for } \mu_0>2
\end{aligned}$$

Second, for $\left|\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right|^2$: we apply the same procedure and get

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right|^2 \le C\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)(T-|j|)^{-2}, \quad \text{for } \mu_0>2$$

For the last term, we use Cauchy-Schwarz and get:

$$
\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|\cdot\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|\\
&\qquad\le\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2\right)^{1/2}\times\\
&\qquad\quad\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right|^2\right)^{1/2}\\
&\qquad\le C\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)(T-|j|)^{-2},\quad\text{for }\mu_0>2
\end{aligned}
$$

For (5.29):

Note that

$$
\begin{aligned}
\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{0,ICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)&=\prod_{m=1}^{d}\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}_0(\tau_{1m},\tau_{2m})-\prod_{m=1}^{d}\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(\tau_{1m},\tau_{2m})\\
&\quad-\left\{\left(\prod_{m=1}^{d}\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(\tau_{1m},0)-\prod_{m=1}^{d}\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(\tau_{1m},0)\right)\prod_{m=1}^{d}\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(0,\tau_{2m})\right\}\\
&\quad-\left\{\prod_{m=1}^{d}\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(\tau_{1m},0)\left(\prod_{i=1}^{d}\hat{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(0,\tau_{2m})-\prod_{m=1}^{d}\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(0,\tau_{2m})\right)\right\}
\end{aligned}
$$

after some laborious algebra, it can be shown:

$$
\prod_{m=1}^{d}\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}_0(\tau_{1m},\tau_{2m})=\sum_{m=0}^{d}\sum_{c=1}^{n_m}\left\{\prod_{j\in\mathscr{J}_{m,c}}\left(\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}_0(\tau_{1m},\tau_{2m})-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(\tau_{1m},\tau_{2m})\right)\prod_{j\in\mathscr{J}^{\complement}_{m,c}}\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_0(\tau_{1m},\tau_{2m})\right\}
$$

where $\mathscr{J}_{m,c}\subset\{1,2,\ldots,d\}$, $\mathscr{J}_m:=\{\mathscr{J}_{m,n},1\le j\le n_m\}$ with $n_m=\binom{d}{d-m}$ such that $\mathscr{J}_{m,n}$ excludes m elements from index set $\{1,\ldots,d\}$. For instance, if $d=4$ and $m=2$, then $n_m=\binom{4}{2}=\frac{4!}{2!\times 2!}=6$ and $\mathscr{J}_2=\left\{\mathscr{J}_{2,1},\mathscr{J}_{2,2},\ldots,\mathscr{J}_{2,6}\right\}$ with $\mathscr{J}_{2,1}=\{1,2\}$, $\mathscr{J}_{2,2}=\{1,3\}$, $\mathscr{J}_{2,3}=\{1,4\}$, $\mathscr{J}_{2,4}=\{2,3\}$, $\mathscr{J}_{2,5}=\{2,4\}$ and $\mathscr{J}_{2,6}=\{3,4\}$.

Thus

$$\prod_{m=1}^{d} \hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^{d} \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) = \sum_{m=0}^{d-1}\sum_{c=1}^{n_m}\left\{\prod_{j\in\mathscr{J}_{m,c}}\left(\hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right)\prod_{j\in\mathscr{J}_{m,c}^{\complement}}\overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right\}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\prod_{m=1}^{d} \hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^{d} \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right| \le \sum_{m=0}^{d-1}\sum_{c=1}^{n_m}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\prod_{j\in\mathscr{J}_{m,c}}\left(\hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right)\prod_{j\in\mathscr{J}_{m,c}^{\complement}}\overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right|$$

and given that $\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right| \le C(|\tau_{1m}|+|\tau_{2m}|)T^{-1}$ and $\sup_{\boldsymbol{\vartheta}}\left|\overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right| \le 1$, then

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\prod_{m=1}^{d} \hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m}) - \prod_{m=1}^{d} \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},\tau_{2m})\right| \le C\left(\sum_{m=1}^{d-1}\{|\tau_{1m}|+|\tau_{2m}|\}\right)T^{-1}$$

applying this result to the terms $\left\{\left(\prod_{m=1}^d \hat{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},0) - \prod_{m=1}^d \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},0)\right)\prod_{m=1}^d \hat{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(0,\tau_{2m})\right\}$ and $\left\{\prod_{m=1}^d \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1m},0)\left(\prod_{i=1}^d \hat{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(0,\tau_{2m}) - \prod_{m=1}^d \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(0,\tau_{2m})\right)\right\}$, and using triangle inequality, we get:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}_{0,ICA}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \overline{\sigma}_{0,ICA}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| \le C\left(\sum_{m=1}^{d-1}\{|\tau_{1m}|+|\tau_{2m}|\}\right)T^{-1} \le C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)T^{-1}$$

For (5.30):

Using a similar procedure than in parts (5.27)-(5.29), we obtain $\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\varphi}_0^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\tau_{1i},\tau_{2i}) - \overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(\tau_{1i},\tau_{2i})\right|^2 \le C(|\tau_{1i}|^2+|\tau_{2i}|^2)T^{-2}$.

Thus

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\hat{\sigma}_{0,ICA}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \overline{\sigma}_{0,ICA}^{\varepsilon(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 &\le C\left(\sum_{m=1}^{d-1}\left\{|\tau_{1m}|^2+|\tau_{2m}|^2\right\}\right)T^{-2} \\ &\le C\left(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2\right)T^{-2}\end{aligned}$$

□

5.9.4. Lemma 4

Under Assumptions (2.1)-(2.3)

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\bar{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| \le C(\|\boldsymbol{\tau}_1\|)(T-|j|)^{-1/2} \tag{5.31}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\bar{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 \le C(\|\boldsymbol{\tau}_1\|)(T-|j|)^{-1/2} \tag{5.32}$$

Proof.

Part (5.31):

Since

$$\begin{aligned}\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) &= \frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\tau}_2,\boldsymbol{\vartheta})z_{t-|j|}(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\\ &= \overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2)-\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)+\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2)\end{aligned}$$

thus

$$\bar{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = -\left\{\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\right\}\left\{\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2)\right\}$$

applying Hörner inequality

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\bar{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right| &\le \left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\right|^2\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2)\right|^2\right)^{1/2}\\ &\le \mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\right|^2 = \mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\tau}_1,\boldsymbol{\vartheta})\right|^2\\ &\le C\|\boldsymbol{\tau}_1\|(T-|j|)^{-\bar{\mu}} \le C\|\boldsymbol{\tau}_1\|(T-|j|)^{-1/2}\end{aligned}$$

last inequality uses Lemma (5.9.5) with $\mu_0 > 2$ and $\mu_2 = \mu_1 \ge 1$.

Part (5.32): Notice

$$\begin{aligned}\left|\bar{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 &= \left|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\right|^2\left|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_2)\right|^2\\ &\le 4\left|\overline{\varphi}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})-\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1)\right|^2\end{aligned}$$

as a consequence, we have:

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\bar{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right|^2 \le C(\|\boldsymbol{\tau}_1\|)(T-|j|)^{-1/2}$$

$\square$

5.9.5. Lemma 5

Under assumptions (2.1)-(2.3), it holds

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\tau},\boldsymbol{\vartheta})\right| \leq C(\|\boldsymbol{\tau}\|)(T-|j|)^{-\mu_3}, \quad \mu_3>0 \tag{5.33}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\tau},\boldsymbol{\vartheta})\right|^2 \leq C(\|\boldsymbol{\tau}\|)(T-|j|)^{-\bar{\mu}}, \quad \bar{\mu}>0 \tag{5.34}$$

Proof.

For (5.33):

We proceed in three steps.

1. First, we take $\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \delta_0(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t$. The objective is to prove the uniform law of large numbers. Since $\boldsymbol{\varepsilon}_t$ is an i.i.d sequence, then $\{\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\}$ inherits that feature. The function $z(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}),\boldsymbol{\tau})$, by definition, is bounded and continuous. Besides, following Hoffmann-Jørgensen (1985) or Peskir and Weber (1994), and their definition of eventually totally bounded, the function $z(\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}),\boldsymbol{\tau})$ satisfies

$$\inf_{T\geq 1}\frac{1}{T}\mathbb{E}\sup_{\boldsymbol{\vartheta},\tilde{\boldsymbol{\vartheta}}\in U_\epsilon}\left|\sum_{t=1}^{T}(z_t(\boldsymbol{\vartheta})-z_t(\tilde{\boldsymbol{\vartheta}}))\right| < \epsilon,\ \forall\, \epsilon>0$$

where $U_\epsilon \in U(\mathcal{V})$ and $U(\mathcal{V})$ the family of finite covers of $\mathcal{V}$. Then, following Hoffmann-Jørgensen (1985); Peskir and Weber (1994), we get

$$\sup_{\theta\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T} z_t(\boldsymbol{\vartheta})\right| \xrightarrow{p} 0$$

and

$$\mathbb{E}\sup_{\theta\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T} z_t(\boldsymbol{\vartheta})\right| \to 0$$

That is, the uniform law of large numbers holds for stationary sequence of i.i.d random vector $\{\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\}$.

2. Now, let consider the case $\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) = \bar{\delta}(\boldsymbol{\vartheta};L) = \sum_{j=-m}^{m}\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t-j}$, where $\boldsymbol{\delta}_j(\boldsymbol{\vartheta})$ is coefficient of the infinite possibly two sided polynomial $\delta(\boldsymbol{\vartheta};L)$ for $i=0,\pm1,\dots,\pm m$ with $m<\infty$. Given the finiteness of m and the fact that $\boldsymbol{\varepsilon}_t$ is and i.i.d process satisfying the strong mixing property, it is straightforward to see that the process $\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})$ has

an $\alpha(k)$-strong mixing coefficient that converges to zero as long as k goes to infinity. We assume that $\alpha(k) = O(|k|^{-\mu_1})$ for some $\mu_1 > 0$. Under this condition, Nobel and Dembo (1993) showed that:

$$\sup_{A} \left|\mathbb{P}(A_1) - \mathbb{P}^{\star}(A_1)\right| \leq l\alpha_j,\ A \in \sigma(\{\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}), \bar{\boldsymbol{\varepsilon}}_{t+j}(\boldsymbol{\vartheta}), \ldots, \bar{\boldsymbol{\varepsilon}}_{t+lj}(\boldsymbol{\vartheta})\})$$

where $\mathbb{P}^{\star}$ is probability measure under independence.

Now, taking a grid of size $l_T = \left\lfloor \frac{T}{n_T} \right\rfloor$, and noticing that $T = l_T n_T + R_T$ with $R_T \in \{0, \ldots, n_T - 1\}$. Thus, the set of integers $\{1, \ldots, T\} = \bigcup_{m=1}^{n_T} \mathscr{I}_m$, with

$$\begin{aligned} \mathscr{I}_m &= \{m, m+n_T, \ldots, m+n_T l_T\}, & 1 \leq m \leq R_T \\ \mathscr{I}_m &= \{m, m+n_T, \ldots, m+n_T(l_T-1)\}, & R_T + 1 \leq m \leq n_T \end{aligned}$$

then, for the measurable function $\bar{z}_t(\boldsymbol{\vartheta}; \boldsymbol{\tau}) = \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})) - \mathbb{E}[\exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))]$, we have:

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \bar{z}_t(\boldsymbol{\vartheta}, \boldsymbol{\tau}) &= \frac{1}{T}\left[\sum_{m=1}^{R_T}\sum_{j \in \mathscr{I}_m} \bar{z}_j(\boldsymbol{\vartheta}, \boldsymbol{\tau}) + \sum_{m=R_T+1}^{n_T}\sum_{j \in \mathscr{I}_m} \bar{z}_j(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right] \\ &= \frac{1}{T}\left[(l_T+1)\sum_{m=1}^{R_T}\frac{1}{l_T+1}\sum_{j \in \mathscr{I}_m} \bar{z}_j(\boldsymbol{\vartheta}, \boldsymbol{\tau}) + l_T \sum_{m=R_T+1}^{n_T}\frac{1}{l_T}\sum_{j \in \mathscr{I}_m} \bar{z}_j(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right] \end{aligned}$$

calling $\bar{\zeta}_T(\boldsymbol{\vartheta}, \boldsymbol{\tau}) = \frac{1}{T}\sum_{t=1}^{T} \bar{z}_t(\boldsymbol{\vartheta}, \boldsymbol{\tau})$ and $\bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau}) = \frac{1}{\#(\mathscr{I}_m)}\sum_{t \in \mathscr{I}_m} \bar{z}_t(\boldsymbol{\vartheta}, \boldsymbol{\tau})$, then:

$$\left|\bar{\zeta}_T(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| \leq \frac{l_T+1}{T}\left|\sum_{m=1}^{R_T} \bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| + \frac{l_T}{T}\left|\sum_{m=R_T+1}^{n_T} \bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| \leq \frac{l_T+1}{T}\left|\sum_{m=1}^{n_T} \bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right|$$

thus

$$\mathbb{E}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| \leq \frac{l_T+1}{T}\sum_{m=1}^{n_T} \mathbb{E}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right|$$

Given the mixing condition, and the fact that uniform convergence in mean under independency holds, Nobel and Dembo (1993) showed

$$\left|\mathbb{E}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| - \mathbb{E}_{\mathbb{P}^{\star}}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right|\right| \leq 2\alpha_{n_T}.$$

Besides, since $\mathbb{E}_{\mathbb{P}^{\star}}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T^i(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| \to 0$, without loss of generality, we can assume $\mathbb{E}_{\mathbb{P}^{\star}}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T^m(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| = O(T^{-\mu_2})$ with $\mu_2 > 0$. Consequently

$$\mathbb{E}\sup_{\boldsymbol{\vartheta} \in \mathscr{V}} \left|\bar{\zeta}_T(\boldsymbol{\vartheta}, \boldsymbol{\tau})\right| \leq \frac{(l_T+1)n_T}{T}\left[O\left((l_T+1)^{-\mu_2}\right) + O\left(n_T^{-\mu_1}\right)\right] = O(T^{-\min\{\mu_1, \mu_2\}/2}) \tag{5.35}$$

it is clear that $\frac{(l_T+1)n_T}{T} = O(1)$, and assuming that l_T and n_T are $O(T^{1/2})$, then last equality arises.

3. Now, for extending the result to the general case of a Bernoulli shift we use the previous step and make the following decomposition

$$z_t(\boldsymbol{\vartheta};\boldsymbol{\tau}) = [z_t(\boldsymbol{\vartheta};\boldsymbol{\tau}) - \bar{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau})] + \bar{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau})$$

and

$$z_t(\boldsymbol{\vartheta};\boldsymbol{\tau}) - \bar{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau}) = \exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})) - \mathbb{E}\left[\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right]$$

where $\bar{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau})$ was defined in the previous step.

Thus

$$\frac{1}{T}\sum_{t=1}^{T}(z_t(\boldsymbol{\vartheta};\boldsymbol{\tau}) - \bar{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau})) = \frac{1}{T}\sum_{t=1}^{T}\left(\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right) - \mathbb{E}\left[\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right]$$

and

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}(z_t(\boldsymbol{\vartheta};\boldsymbol{\tau}) - \bar{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau}))\right| \le \mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right)\right| + \sup_{\theta\in\mathcal{V}}\left|\mathbb{E}\left[\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}\right.\right.$$

Now, using the fact that $\mathbb{E}\left\|\boldsymbol{\varepsilon}_t\right\| < \infty$, we get

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right)\right| &\le \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\sup_{\theta\in\mathcal{V}}\left|\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right| \\
&\le \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\sup_{\theta\in\mathcal{V}}\left|\boldsymbol{\tau}'[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})]\right| \le \frac{\|\boldsymbol{\tau}\|\,\mathbb{E}\left\|\boldsymbol{\varepsilon}_t\right\|}{T}\sum_{t=1}^{T}\mathbb{E}\sup_{\theta\in\mathcal{V}}\left\|\sum_{|j|>m}\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right\| \\
&\le \|\boldsymbol{\tau}\|\,\mathbb{E}\left\|\boldsymbol{\varepsilon}_t\right\|\left[\sum_{|j|>m}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left\|\boldsymbol{\delta}_j(\boldsymbol{\vartheta})\right\|\right] = C\,\|\boldsymbol{\tau}\|\,O(m^{-\mu_0})
\end{aligned}$$

besides

$$\sup_{\theta\in\mathcal{V}}\left|\mathbb{E}\left[\exp(i\boldsymbol{\tau}'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})) - \exp(i\boldsymbol{\tau}'\bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}))\right]\right| \le \sup_{\theta\in\mathcal{V}}\mathbb{E}\left|\boldsymbol{\tau}'[\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \bar{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})]\right| = C\,\|\boldsymbol{\tau}\|\,O(m^{-\mu_0})$$

Hence

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right| \le \mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}(z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})-\tilde{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau}))\right| + \mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right|$$
$$\le C\,\|\boldsymbol{\tau}\|\,O(m^{-\mu_0}) + \mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{z}_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right|$$

For $m=\sqrt{T}$, we get

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right| \le C\|\boldsymbol{\tau}\|O(T^{-\mu_3}) \tag{5.36}$$

with $\mu_3=\frac{1}{2}\min\{\mu_0,\,\min\{\mu_1,\,\mu_2\}\}$.

For (5.34):

Given any $\mathbf{x}\in\mathbb{C}$, $|\mathbf{x}|^2=\mathbf{x}\cdot\mathbf{x}^\star$ where $\mathbf{x}^\star$ is the conjugate of $\mathbf{x}$. Thus

$$\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right|^2 = \left(\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right)\left(\frac{1}{T}\sum_{t=1}^{T}z_t^\star(\boldsymbol{\vartheta};\boldsymbol{\tau})\right) = \left(\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right)\left(\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};-\boldsymbol{\tau})\right) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})z_{t'}(\boldsymbol{\vartheta};-\boldsymbol{\tau})$$
$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\left\{z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})z_{t'}(\boldsymbol{\vartheta};-\boldsymbol{\tau})-\sigma_{|t-t'|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})\right\} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sigma_{|t-t'|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})$$

Now, notice that $|t-t'|\in\{0,1,\ldots,T-1\}$, then

$$\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right|^2 = \frac{1}{T^2}\sum_{j=0}^{T-1}\sum_{\mathcal{T}_j}\left\{z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})z_{t'}(\boldsymbol{\vartheta};-\boldsymbol{\tau})-\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})\right\} + \frac{1}{T^2}\sum_{j=0}^{T-1}\sum_{\mathcal{T}_j}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})$$

with $\mathcal{T}_j=\{(t,t'):|t-t'|=j\}$. Consequently

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right|^2 \le \frac{1}{T^2}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\sum_{j=0}^{T-1}\sum_{\mathcal{T}_j}\left\{z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})z_{t'}(\boldsymbol{\vartheta};-\boldsymbol{\tau})-\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})\right\}\right| + \frac{1}{T^2}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\sum_{j=0}^{T-1}\sum_{\mathcal{T}_j}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})\right|$$

and

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\frac{1}{\#(\mathcal{T}_j)}\sum_{\mathcal{T}_j}\left\{z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})z_{t'}(\boldsymbol{\vartheta};-\boldsymbol{\tau})-\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})\right\}\right| \le C\|\boldsymbol{\tau}\|(T-|j|)^{-\mu_3}$$

From Lemma (5.9.2) and given that $\#(\mathcal{T}_j)=O(T-|j|)$,

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{V}}\left|\sum_{\mathcal{T}_j}\sigma_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau},-\boldsymbol{\tau})\right| \le C\|\boldsymbol{\tau}\|(T-|j|)|j|^{1-\mu_0}$$

Accordingly to these results, we get

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\vartheta}\in\bar{\mathcal{V}}}\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\boldsymbol{\vartheta};\boldsymbol{\tau})\right|^2 &\leq C\|\boldsymbol{\tau}\|\left(\frac{1}{T^2}\sum_{j=0}^{T-1}(T-|j|)^{1-\mu_3}+\frac{1}{T^2}\left[T+\sum_{j=1}^{T-1}(T-|j|)|j|^{1-\mu_0}\right]\right)\\
&\leq C\|\boldsymbol{\tau}\|\left(\frac{1}{T^2}\sum_{j=1}^{T}|j|^{1-\mu_3}+T^{-1}+\frac{1}{T}\sum_{j=1}^{T-1}|j|^{1-\mu_0}\right)\\
&\leq C\|\boldsymbol{\tau}\|\left[T^{-\mu_3}+T^{-1}+T^{1-\mu_0}\right]\leq C\|\boldsymbol{\tau}\|T^{-\bar{\mu}}
\end{aligned}$$

where $\bar{\mu}=\min\{1,\mu_3,\mu_0-1\}>0$. Selecting $\mu_3=1$ and $\mu_0>2$, then $\bar{\mu}=1$

□

5.9.6. Lemma 6

Under assumptions (2.1)-(2.3), it holds

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}}\right)\right\|\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)(T-|j|)^{-1} \tag{5.37}$$

$$\mathbb{E}\left|\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}-\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}\right|^2\leq C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)(T-|j|)^{-2} \tag{5.38}$$

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}}\right)\right\|^2\leq C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2) \tag{5.39}$$

Proof.
For (5.37): notice that

$$\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}}\right)-\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}}\right)=\partial\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)-\partial\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)$$

and

$$\begin{aligned}
\partial\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\frac{\partial}{\partial\boldsymbol{\vartheta}}\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)&=\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\partial z_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)z_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)+z_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)\partial z_{t-|j|}(\boldsymbol{\vartheta},\boldsymbol{\tau}_2)\right]\\
\partial z_t(\boldsymbol{\vartheta},\boldsymbol{\tau}_1)&=e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}(i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}))-\mathbb{E}[e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}(i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}))]
\end{aligned}$$

hence (without including the terms $(\boldsymbol{\tau_1},\boldsymbol{\tau_2})$)

$$\partial\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\partial\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\left(\partial\hat{z}_t\hat{z}_{t-|j|}-\partial z_tz_{t-|j|}\right)+\left(\hat{z}_t\partial\hat{z}_{t-|j|}-z_t\partial z_{t-|j|}\right)\right\}$$

with

$$\partial\hat{z}_t\hat{z}_{t-|j|} - \partial z_t z_{t-|j|} = \left(\partial\varphi(\boldsymbol{\tau}_1) - \partial\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0})\right)\hat{z}_{t-|j|}(\boldsymbol{\tau}_2;\boldsymbol{\vartheta}) + (\varphi(\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2))\partial z_t(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})$$

$$\hat{z}_t\partial\hat{z}_{t-|j|} - z_t\partial z_{t-|j|} = \left(\partial\varphi(\boldsymbol{\tau}_2) - \partial\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right)\hat{z}_t(\boldsymbol{\tau}_1;\boldsymbol{\vartheta}) + (\varphi(\boldsymbol{\tau}_1) - \hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0}))\partial z_{t-|j|}(\boldsymbol{\tau}_2;\boldsymbol{\vartheta})$$

replacing these terms into the equation above, we get

$$\begin{aligned}
\partial\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|} - \partial\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}_{|j|} &= (\varphi(\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2))\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\partial z_t(\boldsymbol{\tau}_1;\boldsymbol{\vartheta}) + (\varphi(\boldsymbol{\tau}_1) - \hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0}))\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\partial z_{t-|j|}(\boldsymbol{\tau}_2;\boldsymbol{\vartheta}) \\
&= (\varphi(\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2))(\partial\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0}) - \partial\varphi(\boldsymbol{\tau}_1)) + (\varphi(\boldsymbol{\tau}_1) - \hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0}))(\partial\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2) - \partial\varphi(\boldsymbol{\tau}_2))
\end{aligned}$$

where last equality comes from the fact that

$$\partial\hat{\varphi}_{|j|}(\boldsymbol{\tau}_1,\mathbf{0}) - \partial\varphi(\boldsymbol{\tau}_1) = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) - \mathbb{E}\left[e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]\right\} = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\partial z_t(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})$$

It is clear that $\mathbb{E}[\partial z_t(\boldsymbol{\tau}_1;\boldsymbol{\vartheta})] = \mathbf{0}$ whenever the term $\mathbb{E}\left[e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]$ is bounded. By Cauchy-Schwarz:

$$\left|\mathbb{E}\left[e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]\right| \le \left\{\mathbb{E}\left|e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right|^2\mathbb{E}|\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})|^2\right\}^{1/2}$$

then, we need to show boundedness of $\mathbb{E}|\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})|^2$. Using Frobenius norm, we have

$$\begin{aligned}
\mathbb{E}|\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})|^2 &= \mathbb{E}\,\mathrm{Tr}\left(\sum_{i=1}^{d}\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\partial\boldsymbol{\delta}^{(i)\prime}(L,\boldsymbol{\vartheta})\right) \\
&= \mathrm{Tr}\sum_{i=1}^{d}\mathbb{E}\left[\partial\boldsymbol{\delta}^{(i)}(L,\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\partial\boldsymbol{\delta}^{(i)\prime}(L,\boldsymbol{\vartheta})\right] \\
&\le \mathbb{E}\,\|\boldsymbol{\varepsilon}_t\|^2\mathrm{Tr}\sum_{i=1}^{d}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{l=-\infty}^{\infty}\partial\boldsymbol{\delta}^{(i)}_l(\boldsymbol{\vartheta})\right\|^2
\end{aligned}$$

so, it is bounded because of $\partial\boldsymbol{\delta}^{(i)}(\boldsymbol{\vartheta})$ is the multiplication of absolutely summable filters and the Assumption (2.3.(iv)) about finiteness of $\mathbb{E}\,\|\boldsymbol{\varepsilon}_t\|^2$.

$$\begin{aligned}
\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\hat{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}}\right) - \frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{|j|}}\right)\right\| &\le C(\|\boldsymbol{\tau}_1\| + \|\boldsymbol{\tau}_2\|)\mathbb{E}\left\|(\varphi(-\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}(\mathbf{0},-\boldsymbol{\tau}_2))(\partial\hat{\varphi}_{|j|}(-\boldsymbol{\tau}_1,\mathbf{0}) - \partial\varphi(-\boldsymbol{\tau}_1))\right\| \\
&\le C(\|\boldsymbol{\tau}_1\| + \|\boldsymbol{\tau}_2\|)\left\{\mathbb{E}\left|\varphi(-\boldsymbol{\tau}_2) - \hat{\varphi}_{|j|}(\mathbf{0},-\boldsymbol{\tau}_2)\right|^2\mathbb{E}\left|\partial\hat{\varphi}_{|j|}(-\boldsymbol{\tau}_1,\mathbf{0}) - \partial\varphi(-\boldsymbol{\tau}_1)\right|^2\right\}^{1/2} \\
&= C(\|\boldsymbol{\tau}_1\| + \|\boldsymbol{\tau}_2\|)O((T-j)^{-1})
\end{aligned}$$

For (5.38): note that

$$\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)} - \tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)} = \left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1)\right)\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_{t-|j|}(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_2)\right)$$

hence, given that $z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1)$ and $z_{t-|j|}(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_2)$ are independent:

$$\begin{aligned}
\mathbb{E}\left|\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)} - \tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}\right|^2 &= \mathbb{E}\left|\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1)\right)\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_{t-|j|}(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_2)\right)\right|^2 \\
&= \mathbb{E}\left|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1)\right|^2 \mathbb{E}\left|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T} z_{t-|j|}(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_2)\right|^2 \\
&= O((T-j)^{-2})
\end{aligned}$$

For (5.39): note that

$$\left\|\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\| \leq \left\|\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right) - \mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\| + \left\|\mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|$$

then

$$\begin{aligned}
\mathbb{E}\left\|\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|^2 \leq{}& \mathbb{E}\left\|\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right) - \mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|^2 + \left\|\mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|^2 \\
&+ 2\mathbb{E}\left\|\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right) - \mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\| \cdot \left\|\mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\|
\end{aligned}$$

with

$$\frac{\partial}{\partial \boldsymbol{\vartheta}}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\partial z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1) z_{t-|j|}(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_2) + z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1)\partial z_{t-|j|}(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_2)\right]$$

$$\partial z_t^0(\boldsymbol{\tau}_1) = \partial z_t(\boldsymbol{\vartheta}_0, \boldsymbol{\tau}_1) = e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}(i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0)) - \mathbb{E}[e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}(i\boldsymbol{\tau}_1'\partial\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}_0))]$$

$$\mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \left(\varphi(\boldsymbol{\tau}_1)\mathbb{E}\left[i\boldsymbol{\tau}_1' e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}\partial\boldsymbol{\delta}_{|j|}^0(\boldsymbol{I}_{\bar{d}} \otimes \boldsymbol{\varepsilon}_{t-|j|})\right] + \varphi(\boldsymbol{\tau}_2)\mathbb{E}\left[i\boldsymbol{\tau}_2' e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\partial\boldsymbol{\delta}_{-|j|}^0(\boldsymbol{I}_{\bar{d}} \otimes \boldsymbol{\varepsilon}_t)\right]\right)$$

note that, under boundedness assumption of $\partial\varphi_{|j|}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$, $\mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ is bounded by some constant that depends on the size of $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$.

Besides

$$\mathbb{E}\left\|\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right) - \mathbb{E}\frac{\partial}{\partial \boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}}\right)\right\| = o(1)$$

Now, by Minkowski inequality:

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\vartheta_0)}}\right)-\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\vartheta_0)}}\right)\right\|^2\leq\left[\left(\mathbb{E}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\partial z_t^0(\boldsymbol{\tau}_1)z_{t-|j|}^0(\boldsymbol{\tau}_2)-\mathbb{E}\,\partial z_t^0(\boldsymbol{\tau}_1)z_{t-|j|}^0(\boldsymbol{\tau}_2)\right\}\right\|^2\right)^{1/2}\right.$$
$$\left.+\left(\mathbb{E}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{z_t^0(\boldsymbol{\tau}_1)\partial z_{t-|j|}^0(\boldsymbol{\tau}_2)-\mathbb{E}\,z_t^0(\boldsymbol{\tau}_1)\partial z_{t-|j|}^0(\boldsymbol{\tau}_2)\right\}\right\|^2\right)^{1/2}\right]^2$$

since

$$\mathbb{E}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left\{\partial z_t^0(\boldsymbol{\tau}_1)z_{t-|j|}^0(\boldsymbol{\tau}_2)-\mathbb{E}\,\partial z_t^0(\boldsymbol{\tau}_1)z_{t-|j|}^0(\boldsymbol{\tau}_2)\right\}\right\|^2\leq C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)(T-|j|)^{-1}$$

therefore

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\vartheta_0)}}\right)-\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\vartheta_0)}}\right)\right\|^2\leq C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)(T-|j|)^{-1}\tag{5.40}$$

Furthermore

$$\begin{aligned}\left\|\mathbb{E}\frac{\partial}{\partial\boldsymbol{\vartheta}}\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\vartheta_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right\|&\leq\left\|\varphi(\boldsymbol{\tau}_1)\mathbb{E}\left[i\boldsymbol{\tau}_1'e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}\partial\boldsymbol{\delta}_{|j|}^0(\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_{t-|j|})\right]\right\|+\left\|\varphi(\boldsymbol{\tau}_2)\mathbb{E}\left[i\boldsymbol{\tau}_2'e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\partial\boldsymbol{\delta}_{-|j|}^0(\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_t)\right]\right\|\\&\leq C\|\boldsymbol{\tau}_1\|\left\|\mathbb{E}\left[e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t-|j|}}(\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_{t-|j|})\right]\right\|+C\|\boldsymbol{\tau}_2\|\left\|\mathbb{E}\left[e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}(\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_t)\right]\right\|\\&\leq C\|\boldsymbol{\tau}_1\|\,\mathbb{E}\|\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_{t-|j|}\|^2+C\|\boldsymbol{\tau}_2\|\,\mathbb{E}\|\boldsymbol{I}_{\bar{d}}\otimes\boldsymbol{\varepsilon}_t\|^2\leq C(\|\boldsymbol{\tau}_1\|+\|\boldsymbol{\tau}_2\|)\end{aligned}$$

hence

$$\mathbb{E}\left\|\frac{\partial}{\partial\boldsymbol{\vartheta}}\left(\overline{\tilde{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\vartheta_0)}}\right)\right\|^2=C(\|\boldsymbol{\tau}_1\|^2+\|\boldsymbol{\tau}_2\|^2)\tag{5.41}$$

□

5.9.7. Lemma 7

Under assumptions (2.1), (2.2), and (2.4)

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\boldsymbol{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{0},\boldsymbol{\tau}_2)\right\|\leq C\|\boldsymbol{\tau}_2\|(T-|j|)^{-1}\tag{5.42}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\boldsymbol{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\boldsymbol{0},\boldsymbol{\tau}_2)\right\|^2\leq C(T-|j|)^{-1}\tag{5.43}$$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left[\hat{F}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(\boldsymbol{0},\boldsymbol{\tau}_2)-\left[\hat{F}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}^{(1,0)}\right]^{MDIC}(\boldsymbol{0},\boldsymbol{\tau}_2)\right\|^2\leq C(T-|j|)^{-1}\tag{5.44}$$

Proof.

Part 5.42:

Notice

$$\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)=i\left\{\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right\}$$
$$-i\left\{\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\overline{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right\}$$

Analyzing the first term in the RHS

$$\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]=\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}+$$
$$\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\left[e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]$$

with

$$\sup_{\theta\in\mathcal{V}}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}\right\|\leq\frac{C}{T-|j|}\sum_{t=1+|j|}^{T}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{l=-\infty}^{t-2}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}+\sum_{l=t-T+1}^{\infty}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}\right\|$$
$$\sup_{\theta\in\mathcal{V}}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\left[e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]\right\|\leq\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\sup_{\boldsymbol{\vartheta}}\left\|\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\left[e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]\right\|$$

and

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{l=-\infty}^{t-2}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}+\sum_{l=t-T+1}^{\infty}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}\right\|\leq C\left\{(T+1-t)^{1-\mu_0}+t^{1-\mu_0}+(t-|j|)^{1-\mu_0}+(T+|j|+1-t)^{1-\mu_0}\right\}$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\left[e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]\right\|\leq\left(\mathbb{E}\sup_{\boldsymbol{\vartheta}}\|\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\|^2\,\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\sum_{l=-\infty}^{t-2}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-l}+\sum_{l=t-T+1}^{\infty}\boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{Y}_{t-}\right.\right.$$
$$\leq C\,\|\boldsymbol{\tau}_2\|\left\{(T+1-t)^{1-\mu_0}+t^{1-\mu_0}+(t-|j|)^{1-\mu_0}+(T+|j|+1-t)^{1-}\right.$$

these results exploit the finitenes of $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^2$ and the summability of filters.

Hence

$$\mathbb{E}\sup_{\theta\in\mathcal{V}}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\right]e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}\right\|\leq\frac{C}{T-|j|}\left\{\sum_{t=1}^{T-|j|}t^{1-\mu_0}+\sum_{t=1+|j|}^{T}t^{1-\mu_0}\right\}\leq C(T-|j|)^{1-\mu_0}$$
$$\mathbb{E}\sup_{\theta\in\mathcal{V}}\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\left[e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]\right\|\leq C\,\|\boldsymbol{\tau}_2\|\,(T-|j|)^{1-\mu_0}$$

and by applying triangle inequality and selecting $\mu_0 > 2$

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right] \right\| \leq C \left\| \boldsymbol{\tau}_2 \right\| (T-|j|)^{-1}$$

Now, the second term in the RHS

$$\frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left\{ \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) \hat{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \overline{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) \right\} = \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left\{ \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \right\} \hat{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) + \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \left\{ \hat{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) - \overline{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) \right\}$$

hence

$$\mathbb{E}\sup_{\theta \in \mathcal{V}} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left\{ \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \right\} \hat{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) \right\| \leq C (T-|j|)^{1-\mu_0}$$

$$\mathbb{E}\sup_{\theta \in \mathcal{V}} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \left\{ \hat{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) - \overline{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) \right\} \right\| \leq C \left\| \boldsymbol{\tau}_2 \right\| (T-|j|)^{-\mu_0}$$

Therefore, ensambling these two previous results and selecting $\mu_0 > 2$, we get

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}} \left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\| \leq C \|\boldsymbol{\tau}_2\| (T-|j|)^{-1}$$

Part 5.43:

Note

$$\left\| \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) - \Delta^{(1,0)} \hat{\sigma}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2 \leq C \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right\|^2 + C \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) \hat{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \overline{\varphi}_{|j|}(\mathbf{0}, \boldsymbol{\tau}_2) \right\|^2$$

The first term in the RHS

$$\left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right] \right\|^2 \leq C \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \left[\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \right] e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} \right\|^2 + C \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \left[e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right] \right\|^2$$

note

$$\left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} [\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})] e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} \right\|^2 = \frac{1}{(T-|j|)^2} \sum_{t=1+|j|}^{T} \sum_{t'=1+|j|}^{T} \nabla \hat{\boldsymbol{\varepsilon}}_t'(\boldsymbol{\vartheta}) \nabla \hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta}) e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} \overline{e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta})}}$$

where $\nabla \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) = \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) = \sum_{l=-\infty}^{t-2} \boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t-l} + \sum_{l=t-T+1}^{\infty} \boldsymbol{\Psi}_j^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t-l}$. And

$$\begin{aligned}
\nabla \hat{\boldsymbol{\varepsilon}}_t'(\boldsymbol{\vartheta}) \nabla \hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta}) &= \left(\sum_{l=-\infty}^{t-2} \boldsymbol{\Psi}_l^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t-l} + \sum_{l=t-T+1}^{\infty} \boldsymbol{\Psi}_l^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t-l} \right)' \left(\sum_{l=-\infty}^{t'-2} \boldsymbol{\Psi}_l^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t'-l} + \sum_{l=t'-T+1}^{\infty} \boldsymbol{\Psi}_l^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t'-l} \right) \\
&= \sum_{l=-\infty}^{t-2} \sum_{l'=-\infty}^{t'-2} \boldsymbol{Y}_{t-l}' \boldsymbol{\Psi}_l^{(-1)\prime}(\boldsymbol{\vartheta}) \boldsymbol{\Psi}_{l'}^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t'-l'} + \sum_{l=t-T+1}^{\infty} \sum_{l'=t'-T+1}^{\infty} \boldsymbol{Y}_{t-l}' \boldsymbol{\Psi}_l^{(-1)\prime}(\boldsymbol{\vartheta}) \boldsymbol{\Psi}_{l'}^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t'-l'} \\
&\quad + \sum_{l=-\infty}^{t-2} \sum_{l'=t'-T+1}^{\infty} \boldsymbol{Y}_{t-l}' \boldsymbol{\Psi}_l^{(-1)\prime}(\boldsymbol{\vartheta}) \boldsymbol{\Psi}_{l'}^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t'-l'} + \sum_{l=t-T+1}^{\infty} \sum_{l'=-\infty}^{t'-2} \boldsymbol{Y}_{t-l}' \boldsymbol{\Psi}_l^{(-1)\prime}(\boldsymbol{\vartheta}) \boldsymbol{\Psi}_{l'}^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{Y}_{t'-l'}
\end{aligned}$$

which implies

$$\begin{aligned}
\mathbb{E} \sup_{\boldsymbol{\vartheta}} \left\| \nabla \hat{\boldsymbol{\varepsilon}}_t'(\boldsymbol{\vartheta}) \nabla \hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta}) \right\| &\leq C \sum_{l=-\infty}^{t-2} \sum_{l'=-\infty}^{t'-2} |l|^{1-\mu_0} \left| l' \right|^{1-\mu_0} + C \sum_{l=t-T+1}^{\infty} \sum_{l'=t'-T+1}^{\infty} |l|^{1-\mu_0} \left| l' \right|^{1-\mu_0} \\
&\quad + C \sum_{l=-\infty}^{t-2} \sum_{l'=t'-T+1}^{\infty} |l|^{1-\mu_0} \left| l' \right|^{1-\mu_0} + C \sum_{l=t-T+1}^{\infty} \sum_{l'=-\infty}^{t'-2} |l|^{1-\mu_0} \left| l' \right|^{1-\mu_0} \\
&\leq C \left| t-2 \right|^{2-\mu_0} \left| t'-2 \right|^{2-\mu_0} - C \left| t-T+1 \right|^{2-\mu_0} \left| t'-T+1 \right|^{2-\mu_0}
\end{aligned}$$

And

$$\sup_{\theta \in \mathcal{V}} \left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} [\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta}) - \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})] e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} \right\|^2 \leq \frac{C}{(T-|j|)^2} \left\{ (T-|j|)^{6-2\mu_0} \right\} \leq C(T-|j|)^{-2}, \ \mu_0 \geq 3$$

Now, the term

$$\begin{aligned}
&\left\| \frac{1}{T-|j|} \sum_{t=1+|j|}^{T} \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta}) \left[e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right] \right\|^2 = \\
&\qquad \frac{1}{(T-|j|)^{-2}} \sum_{t=1+|j|}^{T} \sum_{t'=1+|j|}^{T} \boldsymbol{\varepsilon}_t'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}) \left[e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right] \left[\overline{e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta})} - e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})}} \right]
\end{aligned}$$

and

$$\mathbb{E} \sup_{\theta \in \mathcal{V}} \boldsymbol{\varepsilon}_t'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}) \left[e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})} - e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})} \right] \left[\overline{e^{i\boldsymbol{\tau}_2' \hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta})} - e^{i\boldsymbol{\tau}_2' \boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})}} \right] \leq C \, \mathbb{E} \sup_{\boldsymbol{\vartheta}} \boldsymbol{\varepsilon}_t'(\boldsymbol{\vartheta}) \boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta}) \leq \begin{cases} C \left| t-t' \right|^{1-\mu_0}, & \left| t-t' \right| > 0 \\ C, & \left| t-t' \right| = 0 \end{cases}$$

thus

$$\mathbb{E}\sup_{\theta\in\mathcal{V}}\boldsymbol{\varepsilon}_t'(\boldsymbol{\vartheta})\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})\left[e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})}\right]\left[\overline{e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_{t'}(\boldsymbol{\vartheta})}-e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_{t'}(\boldsymbol{\vartheta})}}\right]\leq$$

$$\frac{1}{(T-|j|)^2}\left\{C(T-|j|)+C\sum_{|j|=1}^{T-|j|}|j|^{1-\mu_0}\right\}\leq C(T-|j|)^{-1},\ \mu_0\geq 3$$

Now, the second term in the RHS

$$\left\|\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\boldsymbol{\varepsilon}_t(\boldsymbol{\vartheta})\overline{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\leq C\left\|\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\nabla\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})\right)\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2$$

$$+C\left\|\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\vartheta)\right)\left[\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right]\right\|^2$$

thus

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\nabla\hat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\vartheta})\right)\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\leq C(T-|j|)^{-2},\ \mu_0\geq 3$$

and

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left(\frac{1}{T-|j|}\sum_{t=1+|j|}^{T}\boldsymbol{\varepsilon}_t(\vartheta)\right)\left[\hat{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)-\overline{\varphi}_{|j|}(\mathbf{0},\boldsymbol{\tau}_2)\right]\right\|^2\leq C(T-|j|)^{-1},\ \mu_0\geq 3$$

Therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\hat{\varepsilon}(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)-\Delta^{(1,0)}\hat{\sigma}_{|j|}^{\varepsilon(\boldsymbol{\vartheta})}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\leq C(T-|j|)^{-1}$$

Part (5.44):

Notice that

$$\left\|\left[\hat{F}^{(1,0)}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2=$$

$$\sum_{m=1}^{d}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{t,m}\hat{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\hat{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(0,\tau_{2n})-\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t,m}\tilde{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\overline{\varphi}_0^{\varepsilon(\boldsymbol{\vartheta})}(0,\tau_{2n})\right\|^2$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{t,m}\hat{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})-\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t,m}\tilde{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})=$$
$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{t,m}-\varepsilon_{t,m}\right)\hat{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})$$
$$+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t,m}\left\{\hat{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})-\tilde{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})\right.$$

applying the similar procedure than above

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{t,m}-\varepsilon_{t,m}\right)\hat{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})\right\|^2\leq CT^{-1}$$
$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t,m}\left\{\hat{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\hat{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})-\tilde{z}_t(\tau_{2m},\boldsymbol{\vartheta})\prod_{n\neq m}\overline{\varphi}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}(0,\tau_{2n})\right\}\right\|^2\leq CT^{-1}$$

therefore

$$\mathbb{E}\sup_{\boldsymbol{\vartheta}}\left\|\left[\hat{F}^{(1,0)}_{\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)-\left[\hat{F}^{(1,0)}_{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta})}\right]^{MDIC}(\mathbf{0},\boldsymbol{\tau}_2)\right\|^2\leq dCT^{-1}$$

□

5.10. Closed Form Solutions

In this section we detail the closed form solution when $dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=dW(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_2)$ and $dW(\boldsymbol{\tau})=(2\pi)^{-d/2}\exp\left(-\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\tau}\right)d\boldsymbol{\tau}$.

5.10.1. Estimation with i.i.d innovations

$$\int\left|\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2dW(\tau_1,\tau_2)$$

Since

$$\left|\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2=\left(\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)\right)\left(\overline{\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)}\right)$$
$$=\hat{\sigma}_0(\tau_1,\tau_2)\left(\overline{\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)}\right)-\hat{s}_0^{IICA}(\tau_1,\tau_2)\left(\overline{\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)}\right)$$

with

$$\hat{\sigma}_0(\tau_1,\tau_2)\left(\overline{\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)}\right) = \hat{\sigma}_0(\tau_1,\tau_2)\overline{\hat{\sigma}_0(\tau_1,\tau_2)} - \hat{\sigma}_0(\tau_1,\tau_2)\overline{\hat{s}_0^{IICA}(\tau_1,\tau_2)}$$
$$= \hat{\sigma}_0(\tau_1,\tau_2)\hat{\sigma}_0(-\tau_1,-\tau_2) - \hat{\sigma}_0(\tau_1,\tau_2)\hat{s}_0^{IICA}(-\tau_1,-\tau_2)$$

$$\hat{s}_0^{IICA}(\tau_1,\tau_2)\left(\overline{\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)}\right) = \hat{s}_0^{IICA}(\tau_1,\tau_2)\overline{\hat{\sigma}_0(\tau_1,\tau_2)} - \hat{s}_0^{IICA}(\tau_1,\tau_2)\overline{\hat{s}_0^{IICA}(\tau_1,\tau_2)}$$
$$= \hat{s}_0^{IICA}(\tau_1,\tau_2)\hat{\sigma}_0(-\tau_1,-\tau_2) - \hat{s}_0^{IICA}(\tau_1,\tau_2)\hat{s}_0^{IICA}(-\tau_1,-\tau_2)$$

then

$$\int \left|\hat{\sigma}_0(\tau_1,\tau_2)-\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2 dW(\tau_1,\tau_2) = \int |\hat{\sigma}_0(\tau_1,\tau_2)|^2 dW(\tau_1,\tau_2) + \int \left|\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2 dW(\tau_1,\tau_2)$$
$$-2\int \hat{\sigma}_0(\tau_1,\tau_2)\overline{\hat{s}_0^{IICA}(\tau_1,\tau_2)}dW(\tau_1,\tau_2)$$

5.10.2. Computing $\int |\hat{\sigma}_0(\tau_1,\tau_2)|^2 dW(\tau_1,\tau_2)$

$$\begin{aligned}
|\hat{\sigma}_0(\tau_1,\tau_2)|^2 &= \hat{\sigma}_0(\tau_1,\tau_2)\hat{\sigma}_0(-\tau_1,-\tau_2)\\
&= \left\{\hat{\psi}_0(\tau_1,\tau_2)-\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\right\}\left\{\hat{\psi}_0(-\tau_1,-\tau_2)-\hat{\psi}_0(-\tau_1,0)\hat{\psi}_0(0,-\tau_2)\right\}\\
&= \hat{\psi}_0(\tau_1,\tau_2)\left\{\hat{\psi}_0(-\tau_1,-\tau_2)-\hat{\psi}_0(-\tau_1,0)\hat{\psi}_0(0,-\tau_2)\right\}\\
&\quad -\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\left\{\hat{\psi}_0(-\tau_1,-\tau_2)-\hat{\psi}_0(-\tau_1,0)\hat{\psi}_0(0,-\tau_2)\right\}\\
&= \left|\hat{\psi}_0(\tau_1,\tau_2)\right|^2+\left|\hat{\psi}_0(\tau_1,0)\right|^2\left|\hat{\psi}_0(0,\tau_2)\right|^2\\
&\quad -\hat{\psi}_0(\tau_1,\tau_2)\hat{\psi}_0(-\tau_1,0)\hat{\psi}_0(0,-\tau_2)-\hat{\psi}_0(-\tau_1,-\tau_2)\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)
\end{aligned}$$

Now, for $W(\tau_1,\tau_2)=W(\tau_1)W(\tau_2)$ multiplication of standard Gaussian weight functions, with $W(\tau)=(2\pi)^{-d/2}\exp\left(-\frac{1}{2}\tau'\tau\right)$.

$$\int \left|\hat{\psi}_0(\tau_1,\tau_2)\right|^2 dW(\tau_1,\tau_2) = \int\left(\frac{1}{T}\sum_{t=1}^{T}\exp\left\{i(\tau_1'\epsilon_t(\vartheta)+\tau_2'\epsilon_t(\vartheta))\right\}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\exp\left\{-i(\tau_1'\epsilon_t(\vartheta)+\tau_2'\epsilon_t(\vartheta))\right\}\right)dW(\tau_1,\tau_2)$$
$$=\frac{(2\pi)^{-d}}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta)\right\|^2\right)\exp\left(-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta)\right\|^2\right)$$

$$\int \left|\hat{\psi}_0(\tau_1,0)\right|^2 \left|\hat{\psi}_0(0,\tau_2)\right|^2 dW(\tau_1,\tau_2) = \int \left|\hat{\psi}_0(\tau_1,0)\right|^2 dW(\tau_1) \int \left|\hat{\psi}_0(0,\tau_2)\right|^2 dW(\tau_2)$$
$$= \left\{\int \left|\hat{\psi}_0(\tau_1,0)\right|^2 dW(\tau_1)\right\}^2$$
$$\int \left|\hat{\psi}_0(\tau_1,0)\right|^2 dW(\tau_1) = \frac{(2\pi)^{-d/2}}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta)\right\|^2\right)$$

$$\int \hat{\psi}_0(\tau_1,\tau_2)\hat{\psi}_0(-\tau_1,0)\hat{\psi}_0(0,-\tau_2)dW(\tau_1,\tau_2) = \int \left(\frac{1}{T}\sum_{t=1}^{T}\exp\left\{i(\tau_1'\epsilon_t(\vartheta)+\tau_2'\epsilon_t(\vartheta))\right\}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\exp\left\{-i\tau_1'\epsilon_t(\vartheta)\right\}\right)\times$$
$$\left(\frac{1}{T}\sum_{t=1}^{T}\exp\left\{-i\tau_2'\epsilon_t(\vartheta)\right\}\right)dW(\tau_1,\tau_2)$$
$$= \frac{(2\pi)^{-d}}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\exp\left(-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta)\right\|^2\right)\exp\left(-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_3}(\vartheta)\right\|\right.$$
$$= \int \hat{\psi}_0(-\tau_1,-\tau_2)\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)dW(\tau_1,\tau_2)$$

Then

$$\int |\hat{\sigma}_0(\tau_1,\tau_2)|^2 dW(\tau_1,\tau_2) = \int \left|\hat{\psi}_0(\tau_1,\tau_2)\right|^2 dW(\tau_1,\tau_2) + \int \left|\hat{\psi}_0(\tau_1,0)\right|^2 \left|\hat{\psi}_0(0,\tau_2)\right|^2 dW(\tau_1,\tau_2)$$
$$-2\int \hat{\psi}_0(\tau_1,\tau_2)\hat{\psi}_0(-\tau_1,0)\hat{\psi}_0(0,-\tau_2)dW(\tau_1,\tau_2)$$

5.10.3. Computing $\int \left|\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2 dW(\tau_1,\tau_2)$

$$\left|\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2 = \hat{s}_0^{IICA}(\tau_1,\tau_2)\overline{\hat{s}_0^{IICA}(\tau_1,\tau_2)} = \hat{s}_0^{IICA}(\tau_1,\tau_2)\hat{s}_0^{IICA}(-\tau_1,-\tau_2)$$
$$\hat{s}_0^{IICA}(\tau_1,\tau_2) = \prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},\tau_{j2}) - \prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},0)\hat{\psi}_0(0,\tau_{j2})$$

thus

$$\left|\hat{s}_0^{IICA}(\tau_1,\tau_2)\right|^2 = \left(\prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},\tau_{j2}) - \prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},0)\hat{\psi}_0(0,\tau_{j2})\right)\left(\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) - \prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})\right)$$
$$= \prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},\tau_{j2})\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) + \prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},0)\hat{\psi}_0(0,\tau_{j2})\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})$$
$$-\prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},\tau_{j2})\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) - \prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2})\prod_{j=1}^{d}\hat{\psi}_0(\tau_{j1},0)\hat{\psi}_0(0,\tau_{j2})$$

and

$$\prod_{j=1}^{d} \hat{\psi}_0(\tau_{j1},\tau_{j2}) \prod_{j=1}^{d} \hat{\psi}_0(-\tau_{j1},-\tau_{j2}) = \prod_{j=1}^{d} \left|\hat{\psi}_0(\tau_{j1},\tau_{j2})\right|^2$$

$$\prod_{j=1}^{d} \hat{\psi}_0(\tau_{j1},0)\hat{\psi}_0(0,\tau_{j2}) \prod_{j=1}^{d} \hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) = \prod_{j=1}^{d} \left|\hat{\psi}_0(\tau_{j1},0)\right|^2 \left|\hat{\psi}_0(0,\tau_{j2})\right|^2$$

Since $W(\tau_1)$ and $W(\tau_2)$ are standard Gaussian weight functions, then:

$$\begin{aligned}\int \prod_{j=1}^{d} \left|\hat{\psi}_0(\tau_{j1},\tau_{j2})\right|^2 dW(\tau_1,\tau_2) &= \prod_{j=1}^{d} \int \left|\hat{\psi}_0(\tau_{j1},\tau_{j2})\right|^2 dW(\tau_{j1},\tau_{j2}) \\ &= \prod_{j=1}^{d} \left(\frac{(2\pi)^{-1}}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \exp(-\left|\epsilon_{j,t_1}(\vartheta) - \epsilon_{j,t_2}(\vartheta)\right|^2) \right)\end{aligned}$$

$$\begin{aligned}\int \prod_{j=1}^{d} \left|\hat{\psi}_0(\tau_{j1},0)\right|^2 \left|\hat{\psi}_0(0,\tau_{j2})\right|^2 dW(\tau_1,\tau_2) &= \prod_{j=1}^{d} \left\{ \int \left|\hat{\psi}_0(\tau_{j1},0)\right|^2 dW(\tau_{j1}) \right\} \left\{ \int \left|\hat{\psi}_0(0,\tau_{j2})\right|^2 dW(\tau_{j2}) \right\} \\ &= \left\{ \prod_{j=1}^{d} \left\{ \int \left|\hat{\psi}_0(\tau_{j1},0)\right|^2 dW(\tau_{j1}) \right\} \right\}^2 \\ &= \left[\prod_{j=1}^{d} \left\{ \frac{(2\pi)^{-1/2}}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \exp\left(-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta) - \epsilon_{j,t_2}(\vartheta)\right|^2\right) \right\} \right]^2\end{aligned}$$

$$\begin{aligned}\int \prod_{j=1}^{d} \hat{\psi}_0(\tau_{j1},\tau_{j2}) \prod_{j=1}^{d} \hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) dW(\tau_1,\tau_2) &= \int \prod_{j=1}^{d} \hat{\psi}_0(\tau_{j1},\tau_{j2})\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) dW(\tau_1,\tau_2) \\ &= \prod_{j=1}^{d} \int \hat{\psi}_0(\tau_{j1},\tau_{j2})\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) dW(\tau_{j1},\tau_{j2})\end{aligned}$$

$$\begin{aligned}\int \hat{\psi}_0(\tau_{j1},\tau_{j2})\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) dW(\tau_{j1},\tau_{j2}) = \frac{(2\pi)^{-1}}{T^3} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \Bigg\{ &\exp\left(-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta) - \epsilon_{j,t_2}(\vartheta)\right|^2\right) \\ &\exp\left(-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta) - \epsilon_{j,t_3}(\vartheta)\right|^2\right) \Bigg\}\end{aligned}$$

5.10.4. Computing $\int \hat{\sigma}_0(\tau_1,\tau_2)\hat{s}_0^{IICA}(-\tau_1,-\tau_2)dW(\tau_1,\tau_2)$

Now

$$\begin{aligned}
\hat{\sigma}_0(\tau_1,\tau_2)\hat{s}_0^{IICA}(-\tau_1,-\tau_2) &= \{\hat{\psi}_0(\tau_1,\tau_2)-\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\}\left\{\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2})-\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})\right\} \\
&= \hat{\psi}_0(\tau_1,\tau_2)\left\{\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2})-\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})\right\} \\
&\quad -\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\left\{\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2})-\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})\right\}
\end{aligned}$$

with

$$\begin{aligned}
\hat{\psi}_0(\tau_1,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) &= \left(\frac{1}{T}\sum_{t_1=1}^{T}\exp\left(i\tau_1'\epsilon_{t_1}(\vartheta)+i\tau_2'\epsilon_{t_1}(\vartheta)\right)\right)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \left(\frac{1}{T}\sum_{t_1=1}^{T}\exp\left(i\tau_1'\epsilon_{t_1}(\vartheta)\right)\exp\left(i\tau_2'\epsilon_{t_1}(\vartheta)\right)\right)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \left(\frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\exp(i\tau_{j2}\epsilon_{j,t_1})\right)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\exp(i\tau_{j2}\epsilon_{j,t_1})\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\frac{1}{T}\sum_{t_2=1}^{T}\{\exp(-i\tau_{j1}\epsilon_{j,t_2}-i\tau_{j2}\epsilon_{j,t_2})\exp(i\tau_{j1}\epsilon_{j,t_1})\exp(i\tau_{j2}\epsilon_{j,t_1})\} \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\left[\frac{1}{T}\sum_{t_2=1}^{T}\{\exp(i\tau_{j1}(\epsilon_{j,t_1}-\epsilon_{j,t_2}))\exp(i\tau_{j2}(\epsilon_{j,t_1}-\epsilon_{j,t_2}))\}\right]
\end{aligned}$$

$$\begin{aligned}
\hat{\psi}_0(\tau_1,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) &= \left(\frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\exp(i\tau_{j2}\epsilon_{j,t_1})\right)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\hat{\psi}_0(-\tau_{j1},0)\exp(i\tau_{j2}\epsilon_{j,t_1})\hat{\psi}_0(0,-\tau_{j2}) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\left\{\frac{1}{T}\sum_{t_2=1}^{T}\exp(i\tau_{j1}(\epsilon_{j,t_1}-\epsilon_{j,t_2}))\right\}\left\{\frac{1}{T}\sum_{t_3=1}^{T}\exp(i\tau_{j2}(\epsilon_{j,t_1}-\epsilon_{j,t_3}))\right\} \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\exp(i\tau_{j1}(\epsilon_{j,t_1}-\epsilon_{j,t_2}))\exp(i\tau_{j2}(\epsilon_{j,t_1}-\epsilon_{j,t_3}))\right]
\end{aligned}$$

$$
\begin{aligned}
\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) &= \left(\frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\right)\left(\frac{1}{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j2}\epsilon_{j,t_2})\right)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\prod_{j=1}^{d}\exp(i\tau_{j2}\epsilon_{j,t_2})\right]\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\exp(i\tau_{j1}\epsilon_{j,t_1})\exp(i\tau_{j2}\epsilon_{j,t_2})\hat{\psi}_0(-\tau_{j1},-\tau_{j2}) \\
&= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\frac{1}{T}\sum_{t_3=1}^{T}\exp(i\tau_{j1}(\epsilon_{j,t_1}-\epsilon_{j,t_3}))\exp(i\tau_{j2}(\epsilon_{j,t_2}-\epsilon_{j,t_3}))
\end{aligned}
$$

$$
\begin{aligned}
\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) &= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\exp(i\tau_1'\epsilon_{t_1})\exp(i\tau_2'\epsilon_{t_2})\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2}) \\
&= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\left\{\frac{1}{T}\sum_{t_3=1}^{T}\exp(i\tau_{j1}(\epsilon_{j,t_1}-\epsilon_{j,t_3}))\right\}\left\{\frac{1}{T}\sum_{t_4=1}^{T}\exp(i\tau_{j2}(\epsilon_{j,t_2}-\epsilon_{j,
\end{aligned}
$$

$$
= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\exp(i\tau_{j1}(\epsilon_{j,t_1}-\epsilon_{j,t_3}))\exp(i\tau_{j2}(\epsilon_{j,t_2}-\epsilon_{j,t_4}))\right]
$$

thus

$$
\begin{aligned}
\int\hat{\psi}_0(\tau_1,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2})dW(\tau_1,\tau_2) &= \frac{1}{T^2}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\sum_{t_2=1}^{T}\left\{\exp(-\frac{1}{2}\left|\epsilon_{j,t_1}-\epsilon_{j,t_2}\right|^2)\exp(-\frac{1}{2}\left|\epsilon_{j,t_1}-\epsilon_{j,t_2}\right|^2)\right\} \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\left[\frac{1}{T}\sum_{t_2=1}^{T}\left\{\exp(-\left|\epsilon_{j,t_1}-\epsilon_{j,t_2}\right|^2)\right\}\right]
\end{aligned}
$$

$$
\int\hat{\psi}_0(\tau_1,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})dW(\tau_1,\tau_2) = \frac{1}{T}\sum_{t_1=1}^{T}\prod_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\exp(-\frac{1}{2}\left|\epsilon_{j,t_1}-\epsilon_{j,t_2}\right|^2)\exp(-\frac{1}{2}\left|\epsilon_{j,t_1}-\epsilon_{j,t_3}\right|^2
$$

$$
\int\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},-\tau_{j2})dW(\tau_1,\tau_2) = \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\frac{1}{T}\sum_{t_3=1}^{T}\exp(-\frac{1}{2}\left|\epsilon_{j,t_1}-\epsilon_{j,t_3}\right|^2)\exp(-\frac{1}{2}\left|\epsilon_{j,t_2}-\epsilon_{j,t_3}\right|^2)
$$

$$
\int\hat{\psi}_0(\tau_1,0)\hat{\psi}_0(0,\tau_2)\prod_{j=1}^{d}\hat{\psi}_0(-\tau_{j1},0)\hat{\psi}_0(0,-\tau_{j2})dW(\tau_1,\tau_2) =
$$

$$
\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\prod_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\exp(-\frac{1}{2}\left|\epsilon_{j,t_1}-\epsilon_{j,t_3}\right|^2)\exp(-\frac{1}{2}\left|\epsilon_{j,t_2}-\epsilon_{j,t_4}\right.
$$

5.10.5. Asymptotic Variance - i.i.d innovations

For estimating the asymptotic variance of $\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}}$, we use $\hat{\boldsymbol{\varepsilon}}_t \equiv \hat{\boldsymbol{\varepsilon}}_t(\hat{\boldsymbol{\vartheta}}_T^{\mathscr{L}})$ and $\hat{z}_t^0(\boldsymbol{\tau}) = e^{i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t} - \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau})$ with $\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}) = \frac{1}{T}\sum_{s=1}^{T-1} e^{i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_s}$.

Estimation of x_t^0

According to previous definitions, we have

$$
\begin{aligned}
\hat{x}_t^0 &= \int \hat{z}_t^0(\boldsymbol{\tau})\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau})dW(\boldsymbol{\tau}) \\
&= \int e^{i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau})dW(\boldsymbol{\tau}) - \int \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau})\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau})dW(\boldsymbol{\tau}) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\int \exp\left(i\boldsymbol{\tau}'[\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}]\right)dW(\boldsymbol{\tau}) - \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\int \exp\left(i\boldsymbol{\tau}'[\hat{\boldsymbol{\varepsilon}}_{t_1} - \hat{\boldsymbol{\varepsilon}}_{t_2}]\right)dW(\boldsymbol{\tau}) \\
&= \frac{1}{T}\sum_{t_1=1}^{T}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) - \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_1} - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right)
\end{aligned} \tag{5.45}
$$

Estimation of e_t^0

Now, for finding $\hat{e}_t^0$, let notice first:

$$
\begin{aligned}
\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}]'\right) \otimes \boldsymbol{\tau}_1' &= \begin{bmatrix} \mathbb{E}_T[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}]' \otimes \boldsymbol{\tau}_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbb{E}_T[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}]' \otimes \boldsymbol{\tau}_1' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{E}_T[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}]' \otimes \boldsymbol{\tau}_1' \end{bmatrix} \\
&= \frac{i}{T}\sum_{t_1=1}^{T}\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \left\{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\}' \\
&\quad - \frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \left\{\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\}'
\end{aligned}
$$

where $\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right] = \frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}}$.

Besides

$$\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right] = \frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}} = \begin{bmatrix} \frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}} \\ \frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,2} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}} \\ \vdots \\ \frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,d} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}} \end{bmatrix} = \begin{bmatrix} \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right) \\ \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,2} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right) \\ \vdots \\ \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,d} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right) \end{bmatrix}$$

$$\begin{aligned}
\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right]' \otimes \boldsymbol{\tau}' &= \left(\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right)\boldsymbol{\tau}' \quad \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,2} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right)\boldsymbol{\tau}' \quad \cdots \quad \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,d} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right)\boldsymbol{\tau}'\right) \\
&= \frac{i}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \left\{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\}' \\
&\quad - \frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \left\{\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\}'
\end{aligned}$$

and

$$\begin{aligned}
\int \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right)\boldsymbol{\tau}\hat{z}_t^0 dW(\boldsymbol{\tau}) &= \left[\int \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right)\tau_j\hat{z}_t^0 dW(\boldsymbol{\tau})\right] \\
&= \frac{i}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\left\{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\} \\
&\quad - \frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\left\{\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\} \\
\int \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}\right)\tau_j\hat{z}_t^0 dW(\boldsymbol{\tau}) &= \int\left\{\frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right\}\tau_j\hat{z}_t^0 dW(\boldsymbol{\tau}) \\
&= \int\left\{\frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right\}\tau_j e^{i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t} dW(\boldsymbol{\tau}) \\
&\quad - \int\left\{\frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right\}\tau_j\hat{\varphi}_0^{\hat{\varepsilon}}(\boldsymbol{\tau}) dW(\boldsymbol{\tau})
\end{aligned}$$

with

$$\begin{aligned}
\int\left\{\frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right\}\tau_j e^{i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t} dW(\boldsymbol{\tau}) &= \frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1}\left\{\int \tau_j e^{i\boldsymbol{\tau}'\left(\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\right)} dW(\boldsymbol{\tau})\right\} \\
&= \frac{i}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1}\left(\hat{\boldsymbol{\varepsilon}}_{t,j} - \hat{\boldsymbol{\varepsilon}}_{t_1,j}\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \\
\int\left\{\frac{1}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1} e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right\}\tau_j\hat{\varphi}_0^{\hat{\varepsilon}}(\boldsymbol{\tau}) dW(\boldsymbol{\tau}) &= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1}\left\{\int \tau_j e^{i\boldsymbol{\tau}'\left(\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\right)} dW(\boldsymbol{\tau})\right\} \\
&= \frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1,1}\left(\hat{\boldsymbol{\varepsilon}}_{t_2,j} - \hat{\boldsymbol{\varepsilon}}_{t_1,j}\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)
\end{aligned}$$

Therefore, $\hat{e}_t^0$ is given by

$$
\begin{aligned}
i\hat{e}_t^0 &= \int \mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T[\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}'\hat{\boldsymbol{\varepsilon}}_t}]\right) \otimes \boldsymbol{\tau}_1'\right] \hat{z}_t^0(\boldsymbol{\tau}) dW(\boldsymbol{\tau}) \\
&= \frac{i}{T}\sum_{t_1=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \left\{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\}'\right) \\
&\quad - \frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \left\{\hat{\boldsymbol{\varepsilon}}_{t_2} - \hat{\boldsymbol{\varepsilon}}_{t_1}\right\}'\right) \qquad (5.46)
\end{aligned}
$$

Estimation of $\tilde{e}_t^{0,1}$

We know that

$$
\tilde{e}_t^{0,1} = \mathrm{Re}\left(i\int\int z_t^0(\boldsymbol{\tau}_1) z_t^0(\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]'\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\right] dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)
$$

Now, note that

$$
\begin{aligned}
&\int\int \left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}\right]\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right] \hat{z}_t^0(\boldsymbol{\tau}_1)\hat{z}_t^0(\boldsymbol{\tau}_2) dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \\
&\qquad \int\int \left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right] e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t} dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) \\
&\qquad - \int\int \left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right] e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2) dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) \\
&\qquad - \int\int \left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right] e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1) dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) \\
&\qquad + \int\int \left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right) \otimes (\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right] \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2) dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)
\end{aligned}
$$

where

$$\int\int\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\boldsymbol{I}_{\bar{d}}\otimes\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

$$\int\int\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\boldsymbol{I}_{\bar{d}}\otimes\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

$$\int\int\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\boldsymbol{I}_{\bar{d}}\otimes\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

$$\int\int\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\boldsymbol{I}_{\bar{d}}\otimes\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

and

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_1\right]e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$
$$+\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_2\right]e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_1\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$
$$+\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_2\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$
$$+\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_2\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

notice that

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_1\right]e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\frac{i}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_1\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_2\right]e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

$$\int\int\left[\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right]\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\frac{i}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

Therefore

$$\begin{aligned}
\widehat{e}_t^{0,1}=i^2 2\Bigg\{&\frac{1}{T}\sum_{t_1=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}'\exp\left(-\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\left\{\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\}'\right)\\
&-\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}'\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\}'\right)\\
&-\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}'\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\}'\right)\\
&+\frac{1}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}'\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\}\prime\right)\Bigg\}
\end{aligned}\tag{5.47}$$

Estimation of $\tilde{e}_t^{0,2}$

We know that

$$\tilde{e}_t^{0,2}=\mathrm{Re}\left(-i\int\int z_t^0(\boldsymbol{\tau}_1)z_t^0(\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]'\right)\otimes\boldsymbol{\tau}_1'\right]\right)$$

Let start by

$$\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)\hat{z}_t^0(\boldsymbol{\tau}_1)\hat{z}_t^0(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\begin{aligned}&\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\\
-&\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\\
-&\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\\
+&\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\end{aligned}$$

and

$$\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)e^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

$$\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\frac{i}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

$$\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\frac{i}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t-\hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

$$\int\int\left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1}e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1)\hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\frac{i}{T^4}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right)\exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_4}-\hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right)\otimes\{\hat{\boldsymbol{\varepsilon}}_{t_3}-\hat{\boldsymbol{\varepsilon}}_{t_1}\}$$

thus

$$\widehat{\tilde{e}}_t^{0,2} = -i^2 \left\{ \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right) \right.$$
$$- \frac{1}{T^3} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right)$$
$$- \frac{1}{T^3} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right)$$
$$\left. + \frac{1}{T^4} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_4} - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right) \right\} \tag{5.48}$$

Estimation of $\tilde{e}_t^{0,3}$

We know that

$$\tilde{e}_t^{0,3} = \mathrm{Re}\left(-i \int\int z_t^0(\boldsymbol{\tau}_1) z_t^0(\boldsymbol{\tau}_2) \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1) \mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t} \right]' \right) \otimes \boldsymbol{\tau}_2' \right] \right)$$

Similar to previous part

$$\widehat{\tilde{e}}_t^{0,3} = -i^2 \left\{ \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right) \right.$$
$$- \frac{1}{T^3} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right)$$
$$- \frac{1}{T^3} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right)$$
$$\left. + \frac{1}{T^4} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}_{t_1}' \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\|^2\right) \exp\left(-\frac{1}{2}\|\hat{\boldsymbol{\varepsilon}}_{t_4} - \hat{\boldsymbol{\varepsilon}}_{t_2}\|^2\right) \otimes \{\hat{\boldsymbol{\varepsilon}}_{t_3} - \hat{\boldsymbol{\varepsilon}}_{t_1}\}' \right) \right\} \tag{5.49}$$

Estimation of $\breve{e}^{0,1}$

We know that

$$\breve{e}^{0,1} = \mathrm{Re}\left(i \int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t} \right]' \right) \otimes (\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2)' \right] dW(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \right)$$

Let start by

$$\int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)\right]\hat{\sigma}^{\hat{\boldsymbol{\varepsilon}}}_{0,IICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\sigma}^{\hat{\boldsymbol{\varepsilon}}}_{0,IICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) + \int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_2\right]\hat{\sigma}^{\hat{\boldsymbol{\varepsilon}}}_{0,IICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

and

$$\int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\hat{\sigma}^{\hat{\boldsymbol{\varepsilon}}}_{0,IICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\prod_{l=1}^{d}\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}}_0(\tau_{1l},\tau_{2l})dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) - \int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\prod_{l=1}^{d}\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}}_0(\tau_{1l})\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}}_0(\tau_{2l})dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$

thus

$$\int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\prod_{l=1}^{d}\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}}_0(\tau_{1l},\tau_{2l})dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \frac{i}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\otimes\hat{\boldsymbol{\eta}}_{1,t_1}$$

where $\hat{\boldsymbol{\eta}}_{1,t_1} = \begin{bmatrix} \left\{\prod_{l\neq 1}\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-|\hat{\varepsilon}_{t_2,l}-\hat{\varepsilon}_{t_1,l}|^2\right)\right\}\left(\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-|\hat{\varepsilon}_{t_2,1}-\hat{\varepsilon}_{t_1,1}|^2\right)\{\hat{\varepsilon}_{t_2,1}-\hat{\varepsilon}_{t_1,1}\}\right) \\ \vdots \\ \left\{\prod_{l\neq d}\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-|\hat{\varepsilon}_{t_2,l}-\hat{\varepsilon}_{t_1,l}|^2\right)\right\}\left(\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-|\hat{\varepsilon}_{t_2,d}-\hat{\varepsilon}_{t_1,d}|^2\right)\{\hat{\varepsilon}_{t_2,d}-\hat{\varepsilon}_{t_1,d}\}\right)\end{bmatrix}$.

Also

$$\int\int \left[\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_{t_1}}\right)\otimes\boldsymbol{\tau}_1\right]\prod_{l=1}^{d}\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}}_0(\tau_{1l})\hat{\varphi}^{\hat{\boldsymbol{\varepsilon}}}_0(\tau_{2l})dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) = \frac{i}{T}\sum_{t_1=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}\left(\prod_{l=1}^{d}\frac{1}{T}\sum_{t_3=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_3,l}-\hat{\varepsilon}_{t_1,l}|^2\right)\right)\otimes\hat{\boldsymbol{\eta}}_{2,t_1}$$

where $\hat{\boldsymbol{\eta}}_{2,t_1} = \begin{bmatrix} \left\{\prod_{l\neq 1}\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_2,l}-\hat{\varepsilon}_{t_1,l}|^2\right)\right\}\left(\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_2,1}-\hat{\varepsilon}_{t_1,1}|^2\right)\{\hat{\varepsilon}_{t_2,1}-\hat{\varepsilon}_{t_1,1}\}\right) \\ \vdots \\ \left\{\prod_{l\neq d}\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_2,l}-\hat{\varepsilon}_{t_1,l}|^2\right)\right\}\left(\frac{1}{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_2,d}-\hat{\varepsilon}_{t_1,d}|^2\right)\{\hat{\varepsilon}_{t_2,d}-\hat{\varepsilon}_{t_1,d}\}\right)\end{bmatrix}$.

Therefore

$$\breve{\hat{e}}_t^{0,1} = 2i^2 \left\{ \frac{1}{T} \sum_{t_1=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}'_{t_1} \otimes \hat{\boldsymbol{\eta}}'_{1,t_1} \right) - \frac{1}{T} \sum_{t_1=1}^{T} \mathrm{vec}\left(\boldsymbol{I}_{\bar{d}} \otimes \hat{\boldsymbol{\varepsilon}}'_{t_1} \left(\prod_{l=1}^{d} \frac{1}{T} \sum_{t_3=1}^{T} \exp\left(-\frac{1}{2} |\hat{\varepsilon}_{t_3,l} - \hat{\varepsilon}_{t_1,l}|^2 \right) \right) \otimes \hat{\boldsymbol{\eta}}'_{2,t_1} \right) \right\} \tag{5.50}$$

Estimation of $\breve{e}^{0,2}$

We know that

$$\breve{e}^{0,2} = \mathrm{Re}\left(-i \int\int \sigma_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2) \mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}'_1 \boldsymbol{\varepsilon}_t} \right]' \right) \otimes \boldsymbol{\tau}'_1 \right] \right)$$

Let start by

$$\int\int \left[\mathbb{E}_T \left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'_1 \hat{\boldsymbol{\varepsilon}}_{t_1}} \right) \otimes \boldsymbol{\tau}_1 \right] \hat{\sigma}_{0,IICA}^{\hat{\boldsymbol{\varepsilon}}}(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2) dW(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) =$$
$$\int\int \left[\mathbb{E}_T \left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'_1 \hat{\boldsymbol{\varepsilon}}_{t_1}} \right) \otimes \boldsymbol{\tau}_1 \right] \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2) \prod_{l=1}^{d} \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\tau_{1l}, \tau_{2l}) dW(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$$
$$- \int\int \left[\mathbb{E}_T \left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'_1 \hat{\boldsymbol{\varepsilon}}_{t_1}} \right) \otimes \boldsymbol{\tau}_1 \right] \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2) \prod_{l=1}^{d} \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\tau_{1l}) \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\tau_{2l}) dW(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$$

and

$$\int\int \left[\mathbb{E}_T \left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'_1 \hat{\boldsymbol{\varepsilon}}_{t_1}} \right) \otimes \boldsymbol{\tau}_1 \right] \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2) \prod_{l=1}^{d} \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\tau_{1l}, \tau_{2l}) dW(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \frac{i}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \hat{\boldsymbol{\varepsilon}}_{t_1} \otimes \hat{\boldsymbol{\Upsilon}}_{1,t_1,t_2}$$

$$\int\int \left[\mathbb{E}_T \left(\hat{\boldsymbol{\varepsilon}}_{t_1} e^{-i\boldsymbol{\tau}'_1 \hat{\boldsymbol{\varepsilon}}_{t_1}} \right) \otimes \boldsymbol{\tau}_1 \right] \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(-\boldsymbol{\tau}_2) \prod_{l=1}^{d} \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\tau_{1l}) \hat{\varphi}_0^{\hat{\boldsymbol{\varepsilon}}}(\tau_{2l}) dW(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) =$$
$$\frac{i}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \hat{\boldsymbol{\varepsilon}}_{t_1} \left(\prod_{l=1}^{d} \frac{1}{T} \sum_{t_3=1}^{T} \exp\left(-\frac{1}{2} |\hat{\varepsilon}_{t_3,l} - \hat{\varepsilon}_{t_2,l}|^2 \right) \right) \otimes \hat{\boldsymbol{\eta}}_{2,t_1}$$

where

$$\hat{\boldsymbol{\Upsilon}}_{1,t_1,t_2} = \begin{bmatrix} \left\{ \prod_{l \neq 1} \frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,l} - \hat{\varepsilon}_{t_1,l}|^2} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,l} - \hat{\varepsilon}_{t_2,l}|^2} \right\} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,1} - \hat{\varepsilon}_{t_1,1}|^2} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,1} - \hat{\varepsilon}_{t_2,1}|^2} \{\hat{\varepsilon}_{t_3,1} - \hat{\varepsilon}_{t_1,1}\} \right) \\ \vdots \\ \left\{ \prod_{l \neq d} \frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,l} - \hat{\varepsilon}_{t_1,l}|^2} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,l} - \hat{\varepsilon}_{t_2,l}|^2} \right\} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,d} - \hat{\varepsilon}_{t_1,d}|^2} e^{-\frac{1}{2}|\hat{\varepsilon}_{t_3,d} - \hat{\varepsilon}_{t_2,d}|^2} \{\hat{\varepsilon}_{t_3,d} - \hat{\varepsilon}_{t_1,d}\} \right) \end{bmatrix}$$

therefore

$$\widehat{\breve{e}}^{0,2} = -i^2\left\{\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}'_{t_1}\otimes\hat{\boldsymbol{\gamma}}'_{1,t_1,t_2}\right)-\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}'_{t_1}\left(\prod_{l=1}^{d}\frac{1}{T}\sum_{t_3=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_3,l}-\hat{\varepsilon}_{t_2,l}|^2\right)\right)\otimes\hat{\boldsymbol{\eta}}'_{2,t_1}\right)\right\} \tag{5.51}$$

Estimation $\breve{e}^{0,3}$

We know

$$\breve{e}^{0,3} = \mathrm{Re}\left(-i\int\int\sigma^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}_{0,IICA}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\varphi^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\mathrm{vec}\left[\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}'_2\boldsymbol{\varepsilon}_t}\right]'\right)\otimes\boldsymbol{\tau}'_2\right]dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)$$

which is equivalent to $\breve{e}^{0,2}$. Thus

$$\widehat{\breve{e}}^{0,3} = -i^2\left\{\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}'_{t_1}\otimes\hat{\boldsymbol{\gamma}}'_{1,t_1,t_2}\right)-\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{vec}\left(\boldsymbol{I}_{\bar{d}}\otimes\hat{\boldsymbol{\varepsilon}}'_{t_1}\left(\prod_{l=1}^{d}\frac{1}{T}\sum_{t_3=1}^{T}\exp\left(-\frac{1}{2}|\hat{\varepsilon}_{t_3,l}-\hat{\varepsilon}_{t_2,l}|^2\right)\right)\otimes\hat{\boldsymbol{\eta}}'_{2,t_1}\right)\right\} \tag{5.52}$$

Estimation of $\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)$

From the definition of residuals we have that $\boldsymbol{\delta}(L,\boldsymbol{\vartheta}) = \boldsymbol{\Psi}^{-1}(L,\boldsymbol{\vartheta})\boldsymbol{\Psi}(L,\boldsymbol{\vartheta}_0)$. Besides, from asymptotic distribution part, we know

$$\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}) = \left[\tfrac{\partial}{\partial\vartheta_1}\boldsymbol{\delta}(L,\boldsymbol{\vartheta}) \quad \tfrac{\partial}{\partial\vartheta_2}\boldsymbol{\delta}(L,\boldsymbol{\vartheta}) \quad \dots \quad \tfrac{\partial}{\partial\vartheta_K}\boldsymbol{\delta}(L,\boldsymbol{\vartheta})\right]$$

and we take $K = \bar{d}$. Moreover, from the structure of parameter vector, we have: d^2p_+ non-causal parameters, d^2p_- causal parameters, d^2q_+ non-invertible parameters, d^2q_- invertible parameters and d^2 static parameters. We can see $\boldsymbol{\vartheta} = [\boldsymbol{\phi}'_+,\ \boldsymbol{\phi}'_-,\ \boldsymbol{\theta}'_+,\ \boldsymbol{\theta}'_-,\ \boldsymbol{b}]$. The column vector $\boldsymbol{\phi}_+$ has dimension d^2p_+; the column vector $\boldsymbol{\phi}_-$ has dimension d^2p_-; the column vector $\boldsymbol{\theta}_+$ has dimension d^2q_+; the column vector $\boldsymbol{\theta}_-$ has dimension d^2q_-; and $\boldsymbol{b}$ has dimension d^2.

When $p_+ > 0$, the derivative of $\boldsymbol{\delta}(L,\boldsymbol{\vartheta})$ with respect to i-th element of $\boldsymbol{\phi}_+$:

$$\begin{aligned}
\frac{\partial}{\partial\phi_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}) &= \frac{\partial}{\partial\phi_{+,i}}\left(\boldsymbol{\Psi}^{-1}(L;\boldsymbol{\vartheta})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)\right) \\
&= \left(\frac{\partial}{\partial\phi_{+,i}}\boldsymbol{\Psi}^{-1}(L;\boldsymbol{\vartheta})\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\frac{\partial}{\partial\phi_{+,i}}\boldsymbol{\Phi}(L;\boldsymbol{\phi})\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\frac{\partial}{\partial\phi_{+,i}}\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}L^{\lceil i/d^2\rceil}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_-(L;\boldsymbol{\phi}_-)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil} \qquad (5.53)
\end{aligned}$$

where $l(i) = j(i) - \left\lfloor \frac{j(i)-1}{d} \right\rfloor d$, $m(i) = 1 + \left\lfloor \frac{j(i)-1}{d} \right\rfloor$ and $j(i) = i - \left\lfloor \frac{i-1}{d^2} \right\rfloor d^2$. $\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}$ is a matrix with 1 at $(l(i), m(i))$ position, 0 otherwise.

When $p_- > 0$, the derivative of $\boldsymbol{\delta}(L,\boldsymbol{\vartheta})$ with respect to i-th element of $\boldsymbol{\phi}_-$:

$$\frac{\partial}{\partial\phi_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}) = \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil} \qquad (5.54)$$

When $q_+ > 0$, the derivative of $\boldsymbol{\delta}(L,\boldsymbol{\vartheta})$ with respect to i-th element of $\boldsymbol{\theta}_+$:

$$\begin{aligned}
\frac{\partial}{\partial\theta_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}) &= \boldsymbol{B}^{-1}(\boldsymbol{b})\left(\frac{\partial}{\partial\theta+,i}\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta})\right)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= \boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\frac{\partial}{\partial\theta+,i}\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\right)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\frac{\partial}{\partial\theta+,i}\boldsymbol{\Theta}_+(L,\boldsymbol{\theta}_+)\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \\
&= -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil}
\end{aligned} \qquad (5.55)$$

When $q_- > 0$, the derivative of $\boldsymbol{\delta}(L,\boldsymbol{\vartheta})$ with respect to i-th element of $\boldsymbol{\theta}_-$:

$$\frac{\partial}{\partial\theta_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}) = -\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_-)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_+)\boldsymbol{\Phi}(L;\boldsymbol{\phi})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil} \qquad (5.56)$$

For elements in contemporaneous effect matrix

$$\begin{aligned}\frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}) &= -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\frac{\partial}{\partial b_i}\boldsymbol{B}(\boldsymbol{b})\right)\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L;\boldsymbol{\vartheta})\boldsymbol{\Phi}(L;\boldsymbol{\vartheta})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)\\ &= -\boldsymbol{B}^{-1}(\boldsymbol{b})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}^{-1}(\boldsymbol{b})\boldsymbol{\Theta}^{-1}(L;\boldsymbol{\vartheta})\boldsymbol{\Phi}(L;\boldsymbol{\vartheta})\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0) \end{aligned} \tag{5.57}$$

Evaluating these derivatives at $\boldsymbol{\vartheta}_0$, we obtain:

$$\begin{aligned}
\frac{\partial}{\partial\phi_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= \boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_+^{-1}(L;\boldsymbol{\phi}_{0,+})\boldsymbol{\Theta}(L;\boldsymbol{\theta}_0)\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\
\frac{\partial}{\partial\phi_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= \boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\theta}_0)\boldsymbol{\Phi}_+(L;\boldsymbol{\phi}_{0,+})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Psi}(L;\boldsymbol{\vartheta}_0)L^{\lceil i/d^2\rceil}\\
\frac{\partial}{\partial\theta_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_{0,-})\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_{0,+})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Theta}_-(L,\boldsymbol{\theta}_{0,-})\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\
\frac{\partial}{\partial\theta_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_{0,-})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\
\frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)
\end{aligned}$$

In general, these terms are infinite polynomials, with the exception of derivatives respect to coefficients of static component. In particular cases, they simplify to.

1. **Causal SVARMA**$(p,0)$**:** In this case $q=0$ and $p_+=0$. Thus, we have:

$$\begin{aligned}
\frac{\partial}{\partial\phi_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= \boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_-^{-1}(L;\boldsymbol{\phi}_{0,-})\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\
&= \sum_{l=0}^{\infty}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_{-,l}^{(-1)}(\boldsymbol{\phi}_{0,-})\boldsymbol{B}(\boldsymbol{b}_0)L^{l+\lceil i/d^2\rceil}\\
\frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)
\end{aligned}$$

then, the coefficient at j-th lag of $\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)$ is

$$\dot{\boldsymbol{\delta}}_j(\boldsymbol{\vartheta}_0) = \begin{cases} \left[\mathbf{1}_{1\times d^2p}\otimes\mathbf{0}_d \,\vdots\, -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_1 \,\vdots\, -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_2 \,\vdots\, \cdots\, -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_d\right],\ j=0\\[2em] \left[\left\{\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{e}_{l(1)}\boldsymbol{e}'_{m(1)}\boldsymbol{\Phi}^{(-1)}_{-,j-\lceil i/d^2\rceil}(\boldsymbol{\phi}_0)\boldsymbol{B}(\boldsymbol{b}_0)\right\}_{i=1}^{d^2p} \,\vdots\, \mathbf{1}_{1\times d^2}\otimes\mathbf{0}_d\right],\ j>0\\[2em] \mathbf{1}_{d^2(p+1)}\otimes\mathbf{0}_{d^2},\ j<0 \end{cases}$$

2. **Non-Causal SVARMA$(p,0)$:** In this case $q=0$ and $p_-=0$. Thus, we have:

$$\begin{aligned}\frac{\partial}{\partial\phi_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= \boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{\Phi}_+^{-1}(L;\boldsymbol{\phi}_{0,+})\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\ &= \sum_{l=p_+}^{\infty}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}^{(-1)}_{+,-l}(\boldsymbol{\phi}_{0,+})\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil-l}\\ \frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\end{aligned}$$

then, the coefficient at j-th lag of $\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)$ is

$$\dot{\boldsymbol{\delta}}_j(\boldsymbol{\vartheta}_0)=\begin{cases}\left[\mathbf{1}_{1\times d^2(p-1)}\otimes\mathbf{0}_d\ \vdots\ \left\{\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}^{(-1)}_{+,-p}(\boldsymbol{\phi}_0)\boldsymbol{B}(\boldsymbol{b}_0)\right\}_{i=1+d^2(p-1)}^{d^2p}\ \vdots\ \left\{-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_i\right\}_{i=1}^{d}\right],\ j=0\\ \left[\left\{\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}^{(-1)}_{+,\lceil i/d^2\rceil-j}(\boldsymbol{\phi}_0)\boldsymbol{B}(\boldsymbol{b}_0)\right\}_{i=1}^{d^2p}\ \vdots\ \mathbf{1}_{1\times d^2}\otimes\mathbf{0}_d\right],\ j<0\\ \mathbf{1}_{d^2(p+1)}\otimes\mathbf{0}_{d^2},\ j>0\end{cases}$$

3. **Invertible SVARMA$(0,q)$:** In this case $p=0$ and $q_+=0$. Thus, we have:

$$\begin{aligned}\frac{\partial}{\partial\theta_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_-^{-1}(L,\boldsymbol{\theta}_{0,-})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\ &= \sum_{l=0}^{\infty}-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}^{(-1)}_{-,l}(\boldsymbol{\theta}_{0,-})\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{B}(\boldsymbol{b}_0)L^{l+\lceil i/d^2\rceil}\\ \frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\end{aligned}$$

then, the coefficient at j-th lag of $\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)$ is

$$\dot{\boldsymbol{\delta}}_j(\boldsymbol{\vartheta}_0)=\begin{cases}\left[\mathbf{1}_{1\times d^2q}\otimes\mathbf{0}_d\ \vdots\ -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_1\ \vdots\ -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_2\ \vdots\ \cdots\ -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_d\right],\ j=0\\ \left[\left\{-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}^{(-1)}_{-,j-\lceil i/d^2\rceil}(\boldsymbol{\theta}_{0,-})\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{B}(\boldsymbol{b}_0)\right\}_{i=1}^{d^2q}\ \vdots\ \mathbf{1}_{1\times d^2}\otimes\mathbf{0}_d\right],\ j>0\\ \mathbf{1}_{d^2(q+1)}\otimes\mathbf{0}_{d^2},\ j<0\end{cases}$$

4. **Non-Invertible SVARMA$(0,q)$:** In this case $p=0$ and $q_-=0$. Thus, we have:

$$\begin{aligned}\frac{\partial}{\partial\theta_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_+^{-1}(L,\boldsymbol{\theta}_{0,+})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}(\boldsymbol{b}_0)L^{\lceil i/d^2\rceil}\\ &= \sum_{l=q_+}^{\infty} -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}^{(-1)}_{+,-l}(\boldsymbol{\theta}_{0,+})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}(\boldsymbol{b}_0)L^{-l+\lceil i/d^2\rceil}\\ \frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\end{aligned}$$

then, the coefficient at j-th lag of $\dot{\boldsymbol{\delta}}(L,\boldsymbol{\vartheta}_0)$ is

$$\dot{\boldsymbol{\delta}}_j(\boldsymbol{\vartheta}_0)=\begin{cases}\left[\mathbf{1}_{1\times d^2(q-1)}\otimes\mathbf{0}_d \,\vdots\, \left\{\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}^{(-1)}_{+,-q}(\boldsymbol{\phi}_0)\boldsymbol{B}(\boldsymbol{b}_0)\right\}_{i=1+d^2(q-1)}^{d^2q} \,\vdots\, \left\{-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\otimes\boldsymbol{e}'_i\right\}_{i=1}^{d}\right],\ j=0\\[2ex] \left[\left\{-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}^{(-1)}_{+,-j+\lceil i/d^2\rceil}(\boldsymbol{\theta}_{0,+})\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)\boldsymbol{B}(\boldsymbol{b}_0)\right\}_{i=1}^{d^2q} \,\vdots\, \mathbf{1}_{1\times d^2}\otimes\mathbf{0}_d\right],\ j<0\\[2ex] \mathbf{1}_{d^2(q+1)}\otimes\mathbf{0}_{d^2},\ j>0\end{cases}$$

5. **General SVARMA(p,q):** Now, consider $0<p_+<p$ and $0<q_+<q$. Then, in this case

$$\begin{aligned}\boldsymbol{\Phi}^{-1}(L,\boldsymbol{\vartheta}_0)=\left(\sum_{l=-\infty}^{\infty}\boldsymbol{\Phi}^{(-1)}_{-,l}L^l\right)\left(\sum_{l=-\infty}^{\infty}\boldsymbol{\Phi}^{(-1)}_{+,l}L^l\right) &= \sum_{l_c=-\infty}^{\infty}\sum_{l_n=-\infty}^{\infty}\boldsymbol{\Phi}^{(-1)}_{-,l_c}\boldsymbol{\Phi}^{(-1)}_{+,l_n}L^{l_c+l_n}\\ &= \sum_{l=-\infty}^{\infty}\boldsymbol{\Phi}^{(-1)}_{l}L^j\end{aligned}$$

with $\boldsymbol{\Phi}^{(-1)}_{-,l}=\mathbf{0}$ for $l<0$, $\boldsymbol{\Phi}^{(-1)}_{+,l}=\mathbf{0}$ for $l>-p_++1$; and $\boldsymbol{\Phi}^{(-1)}_{l}=\sum_{l_c=-\infty}^{\infty}\boldsymbol{\Phi}^{(-1)}_{-,l_c}\boldsymbol{\Phi}^{(-1)}_{+,l-l_c}$.

And

$$\begin{aligned}\boldsymbol{\Theta}^{-1}(L,\boldsymbol{\vartheta}_0)=\left(\sum_{l=-\infty}^{\infty}\boldsymbol{\Theta}^{(-1)}_{-,l}L^l\right)\left(\sum_{l=-\infty}^{\infty}\boldsymbol{\Theta}^{(-1)}_{+,l}L^l\right) &= \sum_{l_\iota=-\infty}^{\infty}\sum_{l_{n\iota}=-\infty}^{\infty}\boldsymbol{\Theta}^{(-1)}_{-,l_\iota}\boldsymbol{\Theta}^{(-1)}_{+,l_{n\iota}}L^{l_\iota+l_{n\iota}}\\ &= \sum_{l=-\infty}^{\infty}\boldsymbol{\Theta}^{(-1)}_{l}L^j\end{aligned}$$

with $\boldsymbol{\Theta}^{(-1)}_{-,l}=\mathbf{0}$ for $l<0$, $\boldsymbol{\Theta}^{(-1)}_{+,l}=\mathbf{0}$ for $l>-q_++1$; and $\boldsymbol{\Theta}^{(-1)}_{l}=\sum_{l_\iota=-\infty}^{\infty}\boldsymbol{\Theta}^{(-1)}_{-,l_\iota}\boldsymbol{\Theta}^{(-1)}_{+,l-l_\iota}$.

Hence

$$\frac{\partial}{\partial \phi_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) = \sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}\sum_{l_3=0}^{q}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{l_1}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}_{+,l_2}^{(-1)}\boldsymbol{\Theta}_{l_3}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+l_2+l_3+\lceil i/d^2\rceil}$$

$$\frac{\partial}{\partial \phi_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) = \sum_{l_1=-\infty}^{\infty}\sum_{l_2=0}^{p_+}\sum_{l_3=-\infty}^{\infty}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{l_1}^{(-1)}\boldsymbol{\Phi}_{+,l_2}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Psi}_{l_3}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+l_2+l_3+\lceil i/d^2\rceil}$$

$$\frac{\partial}{\partial \theta_{+,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) = -\sum_{l_1=-\infty}^{\infty}\sum_{l_2=0}^{q_-}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{l_1}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Theta}_{-,l_2}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+l_2+\lceil i/d^2\rceil}$$

$$\frac{\partial}{\partial \theta_{-,i}}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) = -\sum_{l_1=-\infty}^{\infty}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{-,l_1}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+\lceil i/d^2\rceil}$$

$$\frac{\partial}{\partial b_i}\boldsymbol{\delta}(L;\boldsymbol{\vartheta}_0) = -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)$$

However, dealing with an infinite, two-sided polynomial is problematic. For that reason, we truncate them. Besides, since $\boldsymbol{\Phi}_-(L,\boldsymbol{\vartheta}_0)$ and $\boldsymbol{\Phi}_+(L,\boldsymbol{\vartheta}_0)$ are analytic outside and inside unit circle, respectively; we get:

$$\bar{\boldsymbol{\Phi}}^{-1}(L,\boldsymbol{\vartheta}_0) = \left(\sum_{l_c=0}^{\bar{l}}\boldsymbol{\Phi}_{-,l_c}^{(-1)}L^{l_c}\right)\left(\sum_{l_n=-\bar{l}-p_++1}^{-p_+}\boldsymbol{\Phi}_{+,l_n}^{(-1)}L^{l_n}\right) = \sum_{l=-\bar{l}-p_++1}^{\bar{l}-p_+}\bar{\boldsymbol{\Phi}}_l^{(-1)}L^l$$

with $\bar{\boldsymbol{\Phi}}_l^{(-1)} = \sum_{l_c=0}^{\bar{l}}\boldsymbol{\Phi}_{-,l_c}^{(-1)}\boldsymbol{\Phi}_{+,l-l_c}^{(-1)}$.

Similar to AR polynomial, for MA polynomial we truncate and obtain:

$$\bar{\boldsymbol{\Theta}}^{-1}(L,\boldsymbol{\vartheta}_0) = \left(\sum_{l_\iota=0}^{\bar{l}}\boldsymbol{\Theta}_{-,l_\iota}^{(-1)}L^{l_\iota}\right)\left(\sum_{l_{n\iota}=-\bar{l}-q_++1}^{-q_+}\boldsymbol{\Theta}_{+,l_{n\iota}}^{(-1)}L^{l_{n\iota}}\right) = \sum_{l=-\bar{l}-q_++1}^{\bar{l}-q_+}\bar{\boldsymbol{\Theta}}_l^{(-1)}L^l$$

with $\bar{\boldsymbol{\Theta}}_l^{(-1)} = \sum_{l_\iota=0}^{\bar{l}}\boldsymbol{\Theta}_{-,l_\iota}^{(-1)}\boldsymbol{\Theta}_{+,l-l_\iota}^{(-1)}$.

Thus, the truncated derivative polynomials

$$
\begin{aligned}
\frac{\partial}{\partial \phi_{+,i}}\bar{\boldsymbol{\delta}}(L;\boldsymbol{\vartheta}_0) &= \sum_{l_1=-\bar{l}-q_++1}^{\bar{l}-q_+}\sum_{l_2=-\bar{l}-p_++1}^{-p_+}\sum_{l_3=0}^{q}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{l_1}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}_{+,l_2}^{(-1)}\boldsymbol{\Theta}_{l_3}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+l_2+l_3+\lceil i/d^2\rceil}\\
\frac{\partial}{\partial \phi_{-,i}}\bar{\boldsymbol{\delta}}(L;\boldsymbol{\vartheta}_0) &= \sum_{l_1=-\bar{l}-p_++1}^{\bar{l}-p_+}\sum_{l_2=0}^{p_+}\sum_{l_3=-2\bar{l}-p_+-q_++2}^{2\bar{l}-p_+-q_+}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{l_1}^{(-1)}\boldsymbol{\Phi}_{+,l_2}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Psi}_{l_3}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+l_2+l_3+\lceil i/d^2\rceil}\\
\frac{\partial}{\partial \theta_{+,i}}\bar{\boldsymbol{\delta}}(L;\boldsymbol{\vartheta}_0) &= -\sum_{l_1=-\bar{l}-q_++1}^{\bar{l}-q_+}\sum_{l_2=0}^{q_-}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{l_1}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Theta}_{-,l_2}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+l_2+\lceil i/d^2\rceil}\\
\frac{\partial}{\partial \theta_{-,i}}\bar{\boldsymbol{\delta}}(L;\boldsymbol{\vartheta}_0) &= -\sum_{l_1=0}^{\bar{l}}\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{-,l_1}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{B}(\boldsymbol{b}_0)L^{l_1+\lceil i/d^2\rceil}\\
\frac{\partial}{\partial b_i}\bar{\boldsymbol{\delta}}(L;\boldsymbol{\vartheta}_0) &= -\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right)
\end{aligned}
$$

and the j-th coefficient of each polynomial is

$$
\dot{\boldsymbol{\delta}}_j^{(i)}(\boldsymbol{\vartheta}_0) = \begin{cases}
\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\sum_{l_3=0}^{q}\sum_{l_2=-\bar{l}-p_++1}^{-p_+}\boldsymbol{\Theta}_{j-\lceil i/d^2\rceil-l_3-l_2}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Phi}_{+,l_2}^{(-1)}\boldsymbol{\Theta}_{l_3}\boldsymbol{B}(\boldsymbol{b}_0), & j\in[-2\bar{l}-(p_++q_+-2),\bar{l}+q-q_+ \\
 & \vartheta_i\in\boldsymbol{\phi}_+ \\
\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\sum_{l_2=0}^{p_+}\sum_{l_3=-2\bar{l}-p_+-q_++2}^{2\bar{l}-p_+-q_+}\boldsymbol{\Theta}_{j-\lceil i/d^2\rceil-l_2-l_3}^{(-1)}\boldsymbol{\Phi}_{+,l_2}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Psi}_{l_3}\boldsymbol{B}(\boldsymbol{b}_0), & j\in[-3\bar{l}-2p_+-q_++3,3\bar{l}-p_+- \\
 & \vartheta_i\in\boldsymbol{\phi}_- \\
-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\sum_{l_2=0}^{q_-}\boldsymbol{\Theta}_{j-\lceil i/d^2\rceil-l_2}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{\Theta}_{-,l_2}\boldsymbol{B}(\boldsymbol{b}_0), & j\in[-\bar{l}-q_++1,\bar{l}+q-q_+] \\
 & \vartheta_i\in\boldsymbol{\theta}_+ \\
-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\boldsymbol{\Theta}_{-,j}^{(-1)}\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\boldsymbol{B}(\boldsymbol{b}_0), & j\in[0,\bar{l}] \\
 & \vartheta_i\in\boldsymbol{\theta}_- \\
-\boldsymbol{B}^{-1}(\boldsymbol{b}_0)\left(\boldsymbol{e}_{l(i)}\boldsymbol{e}'_{m(i)}\right), & j=0,\ \vartheta_i\in\boldsymbol{b}
\end{cases}
$$

5.10.6. Estimation of $\mathbb{E}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{L}_T(\boldsymbol{\vartheta}_0)$

From Theorem (2.5) we know that

$$
\begin{aligned}
\mathbb{E}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{L}_T(\boldsymbol{\vartheta}_0) &= \frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}\mathscr{L}_0(\boldsymbol{\vartheta}_0)\\
&= \frac{4}{\pi}\sum_{j=1}^{\infty}\frac{1}{j^2}\int\left(\mathbb{E}\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\right)'\left(\mathbb{E}\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)+\\
&\quad \frac{4\pi}{3}\int\left(\mathbb{E}\dot{\tilde{\sigma}}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)-\mathbb{E}\dot{\tilde{\sigma}}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\right)'\left(\mathbb{E}\dot{\tilde{\sigma}}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\mathbb{E}\dot{\tilde{\sigma}}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)
\end{aligned}
$$

From what we obtained in the proof of Theorem (2.5), we get:

$$\left(\mathbb{E}\,\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\right)'\left(\mathbb{E}\,\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)=$$
$$\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\right\}\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right\}$$
$$+\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\right\}\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right\}$$
$$+\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\right\}\dot{\boldsymbol{\delta}}_{-|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right\}$$
$$+\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\right\}\dot{\boldsymbol{\delta}}_{-|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right\}$$

since

$$\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)=\begin{bmatrix}\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2) & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2) & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\end{bmatrix}$$
$$\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)=\begin{bmatrix}\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(1)}(\boldsymbol{\vartheta}_0) & \boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(2)}(\boldsymbol{\vartheta}_0) & \boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(3)}(\boldsymbol{\vartheta}_0) & \cdots & \boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})}(\boldsymbol{\vartheta}_0)\end{bmatrix}$$
$$\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)=\begin{bmatrix}\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(1)}(\boldsymbol{\vartheta}_0) & \dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(2)}(\boldsymbol{\vartheta}_0) & \cdots & \dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})}(\boldsymbol{\vartheta}_0)\\ \dot{\boldsymbol{\delta}}_{|j|}^{(2)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(1)}(\boldsymbol{\vartheta}_0) & \dot{\boldsymbol{\delta}}_{|j|}^{(2)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(2)}(\boldsymbol{\vartheta}_0) & \cdots & \dot{\boldsymbol{\delta}}_{|j|}^{(2)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})}(\boldsymbol{\vartheta}_0)\\ \vdots & \vdots & \ddots & \vdots\\ \dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(1)}(\boldsymbol{\vartheta}_0) & \dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(2)}(\boldsymbol{\vartheta}_0) & \cdots & \dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})}(\boldsymbol{\vartheta}_0)\end{bmatrix}$$

thus

$$\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\right\}\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right\}=$$
$$\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right]_{l_1,l_2=1}^{\bar{d}}$$
$$\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\right\}\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right\}=$$
$$\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right]_{l_1,l_2=1}^{\bar{d}}$$
$$\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\right\}\dot{\boldsymbol{\delta}}_{-|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right\}=$$
$$\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\dot{\boldsymbol{\delta}}_{-|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right]_{l_1,l_2=1}^{\bar{d}}$$
$$\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\right\}\dot{\boldsymbol{\delta}}_{-|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right\}=$$
$$\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\dot{\boldsymbol{\delta}}_{-|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_2\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right]_{l_1,l_2=1}^{\bar{d}}$$

and

$$\int \mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$
$$=\int\left|\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right|^2 dW(\boldsymbol{\tau}_2)\int \mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)dW(\boldsymbol{\tau}_1)$$
$$=\int\left|\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right|^2 dW(\boldsymbol{\tau}_2)\int \operatorname{vec}\left(\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\right)'\operatorname{vec}\left(\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\right)dW(\boldsymbol{\tau}_1)$$
$$=\int\left|\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right|^2 dW(\boldsymbol{\tau}_2)\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\right)'\int \boldsymbol{\tau}_1\boldsymbol{\tau}_1'\otimes\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)dW(\boldsymbol{\tau}_1)\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)\prime}(\boldsymbol{\vartheta}_0)\right)$$

$$\int \mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$
$$=\int \mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$$
$$=\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(l_1)\prime}(\boldsymbol{\vartheta}_0)\right)'\int \boldsymbol{\tau}_1\otimes\mathbb{E}\left(\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\int \mathbb{E}\left(\boldsymbol{\varepsilon}_t' e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\otimes\boldsymbol{\tau}_2'dW(\boldsymbol{\tau}_2)\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(l_2)}(\boldsymbol{\vartheta}_0)\right)$$

as consequence, we get:

$$\int \boldsymbol{\tau}_1\boldsymbol{\tau}_1'\otimes\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right)\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_t' e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right)dW(\boldsymbol{\tau}_1)=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left\{\boldsymbol{I}_d-(\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1})(\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1})'\right\}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\hat{\boldsymbol{\varepsilon}}_{t_2}'\exp\left(-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\|^2\right)$$
$$\int \boldsymbol{\tau}_1\otimes\mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_t e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right)\hat{\varphi}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1)=\frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}(\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1})\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\|^2\right)$$
$$\int \mathbb{E}_T\left(\hat{\boldsymbol{\varepsilon}}_t' e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\right)\hat{\varphi}(-\boldsymbol{\tau}_2)\otimes\boldsymbol{\tau}_2'dW(\boldsymbol{\tau}_2)=\frac{i}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\hat{\boldsymbol{\varepsilon}}_{t_1}'\exp\left(-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_1}-\hat{\boldsymbol{\varepsilon}}_{t_2}\right\|^2\right)\otimes(\hat{\boldsymbol{\varepsilon}}_{t_1}-\hat{\boldsymbol{\varepsilon}}_{t_2})'$$

therefore

$$\int\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\right\}\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_1'\dot{\boldsymbol{\delta}}_{|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)\right\}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$\hat{\alpha}\begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)'\\ \vdots\\ \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left\{\boldsymbol{I}_d-\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}'\right\}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\hat{\boldsymbol{\varepsilon}}_{t_2}'e^{-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\|^2}\right]\begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) & \cdots & \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\end{bmatrix}$$
$$\int\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t' e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\right\}\dot{\boldsymbol{\delta}}_{|j|}'(\boldsymbol{\vartheta}_0)\boldsymbol{\tau}_1\boldsymbol{\tau}_2'\dot{\boldsymbol{\delta}}_{-|j|}(\boldsymbol{\vartheta}_0)\left\{\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)\right\}dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$
$$-\begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)'\\ \vdots\\ \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\star}\boldsymbol{\Sigma}_T^{\star\prime}\begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) & \cdots & \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\end{bmatrix}$$

where $\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}=\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}$ and $\hat{\alpha}=\int\left|\hat{\varphi}(\boldsymbol{\tau}_2)\right|^2 dW(\boldsymbol{\tau}_2)=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\exp\left(-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_1}-\hat{\boldsymbol{\varepsilon}}_{t_2}\right\|^2\right)$, $\boldsymbol{\Sigma}_T^{\star}=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}(\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1})\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\exp\left(-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\|^2\right)$.

Hence, the sample estimate of $\int\left(\mathbb{E}\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\right)'\left(\mathbb{E}\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)$ is

$$
\begin{aligned}
\widehat{\int\left(\mathbb{E}\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\right)'\left(\mathbb{E}\dot{\tilde{\sigma}}_{|j|}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)dW} = \hat{\alpha}\Bigg\{ & \begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)' \\ \vdots \\ \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\ddagger}\left[\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) \quad \cdots \quad \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right] \\
& + \begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)' \\ \vdots \\ \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\ddagger}\left[\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) \quad \cdots \quad \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]\Bigg\} \\
& - \Bigg\{\begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)' \\ \vdots \\ \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\star}\boldsymbol{\Sigma}_T^{\star\prime}\left[\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) \quad \cdots \quad \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right] \\
& + \begin{bmatrix}\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)' \\ \vdots \\ \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{-|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\star}\boldsymbol{\Sigma}_T^{\star\prime}\left[\operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) \quad \cdots \quad \operatorname{vec}\left(\dot{\boldsymbol{\delta}}_{|j|}^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]\Bigg\}
\end{aligned}
$$

where $\boldsymbol{\Sigma}_T^{\ddagger} = \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left\{\boldsymbol{I}_d - \Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}'\right\}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\hat{\boldsymbol{\varepsilon}}_{t_2}'e^{-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\|^2}$.

Besides, since

$$
\begin{aligned}
\mathbb{E}\dot{\tilde{\sigma}}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2) - \mathbb{E}\dot{\tilde{\sigma}}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2) = & -i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\left(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0)\right)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right) \\
& + i(\boldsymbol{\tau}_1)'\left(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0)\right)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2) \\
& + i(\boldsymbol{\tau}_2)'\left(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0)\right)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)
\end{aligned}
$$

thus

$$\left(\mathbb{E}\dot{\tilde{\sigma}}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)-\mathbb{E}\dot{\tilde{\sigma}}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1,-\boldsymbol{\tau}_2)\right)'\left(\mathbb{E}\dot{\tilde{\sigma}}_0^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)-\mathbb{E}\dot{\tilde{\sigma}}_{0,IICA}^{\boldsymbol{\varepsilon}(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)\right)=$$

$$\begin{aligned}
&\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)+\\
&\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)(\boldsymbol{\tau}_1)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)+\\
&\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)(\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)+\\
&\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_1)(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)+\\
&\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_1)(\boldsymbol{\tau}_1)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)+\\
&\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_2)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_1)(\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)+\\
&\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_2)(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\boldsymbol{\varepsilon}_t}\right]\right)+\\
&\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_2)(\boldsymbol{\tau}_1)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i\boldsymbol{\tau}_1'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_2)+\\
&\varphi^{(\boldsymbol{\vartheta}_0)}(-\boldsymbol{\tau}_1)\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_t'e^{-i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))'(\boldsymbol{\tau}_2)(\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\boldsymbol{\vartheta}_0))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}\left[\boldsymbol{\varepsilon}_te^{i\boldsymbol{\tau}_2'\boldsymbol{\varepsilon}_t}\right]\right)\varphi^{(\boldsymbol{\vartheta}_0)}(\boldsymbol{\tau}_1)
\end{aligned}$$

Then, the sample counterpart of each term of the right hand side is:

$$\int\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t'e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_T))'(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_T))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_te^{i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\begin{bmatrix}\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)'\\ \vdots\\ \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\dagger,(0)}\left[\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)\quad\cdots\quad\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]$$

$$\int\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t'e^{-i(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_0))'(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2)(\boldsymbol{\tau}_1)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_T))\left(\boldsymbol{I}_{\bar{d}}\otimes\mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_te^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)\hat{\varphi}^{(\hat{\boldsymbol{\vartheta}}_T)}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2)=$$

$$\begin{bmatrix}\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)'\\ \vdots\\ \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\dagger,(1)}\left[\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)\quad\cdots\quad\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]$$

$$-\begin{bmatrix}\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)'\\ \vdots\\ \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)'\end{bmatrix}\boldsymbol{\Sigma}_T^{\dagger,(2)}\left[\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)\quad\cdots\quad\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]$$

$$\int \hat{\varphi}^{(\hat{\boldsymbol{\vartheta}}_T)}(-\boldsymbol{\tau}_2)\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t' e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_0))'(\boldsymbol{\tau}_1)(\boldsymbol{\tau}_1)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_0))\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t e^{i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)\hat{\varphi}^{(\hat{\boldsymbol{\vartheta}}_T)}(\boldsymbol{\tau}_2)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) =$$

$$\hat{\alpha}\begin{bmatrix} \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)' \\ \vdots \\ \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)' \end{bmatrix} \boldsymbol{\Sigma}_T^{\ddagger}\left[\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) \quad \cdots \quad \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]$$

$$\int \hat{\varphi}^{(\hat{\boldsymbol{\vartheta}}_T)}(-\boldsymbol{\tau}_2)\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t' e^{-i\boldsymbol{\tau}_1'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_T))'(\boldsymbol{\tau}_1)(\boldsymbol{\tau}_2)'(\boldsymbol{S}[\circ]\dot{\boldsymbol{\delta}}_0(\hat{\boldsymbol{\vartheta}}_0))\left(\boldsymbol{I}_{\bar{d}} \otimes \mathbb{E}_T\left[\hat{\boldsymbol{\varepsilon}}_t e^{i\boldsymbol{\tau}_2'\hat{\boldsymbol{\varepsilon}}_t}\right]\right)\hat{\varphi}^{(\hat{\boldsymbol{\vartheta}}_T)}(\boldsymbol{\tau}_1)dW(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) =$$

$$-\begin{bmatrix} \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right)' \\ \vdots \\ \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)' \end{bmatrix} \boldsymbol{\Sigma}_T^{\star}\boldsymbol{\Sigma}_T^{\star\prime}\left[\operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(1)\prime}(\boldsymbol{\vartheta}_0)\right) \quad \cdots \quad \operatorname{vec}\left(\boldsymbol{S}\circ\dot{\boldsymbol{\delta}}_0^{(\bar{d})\prime}(\boldsymbol{\vartheta}_0)\right)\right]$$

where $\boldsymbol{\Sigma}_T^{\dagger,(0)} = \frac{2}{T^2}\sum_{t_1=1}^T\sum_{t_2=1}^T\left\{\boldsymbol{I}_d - 2\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_1}'\right\}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\hat{\boldsymbol{\varepsilon}}_{t_2}' e^{-\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_1}\right\|^2}$,
$\boldsymbol{\Sigma}_T^{\dagger,(1)} = \frac{1}{T^3}\sum_{t_1=1}^T\sum_{t_2=1}^T\sum_{t_3=1}^T\left\{\boldsymbol{I}_d - \Delta\hat{\boldsymbol{\varepsilon}}_{t_1,t_3}\Delta\hat{\boldsymbol{\varepsilon}}_{t_1,t_3}'\right\}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\hat{\boldsymbol{\varepsilon}}_{t_3}' e^{-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_1}-\hat{\boldsymbol{\varepsilon}}_{t_3}\right\|^2} e^{-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_3}\right\|^2}$,
$\boldsymbol{\Sigma}_T^{\dagger,(2)} = \frac{1}{T^3}\sum_{t_1=1}^T\sum_{t_2=1}^T\sum_{t_3=1}^T\left\{\Delta\hat{\boldsymbol{\varepsilon}}_{t_2,t_3}\Delta\hat{\boldsymbol{\varepsilon}}_{t_1,t_3}'\right\}\otimes\hat{\boldsymbol{\varepsilon}}_{t_1}\hat{\boldsymbol{\varepsilon}}_{t_3}' e^{-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_2}-\hat{\boldsymbol{\varepsilon}}_{t_3}\right\|^2} e^{-\frac{1}{2}\left\|\hat{\boldsymbol{\varepsilon}}_{t_1}-\hat{\boldsymbol{\varepsilon}}_{t_3}\right\|^2}$.

5.10.7. Estimation with m.d.s. innovations

$$\int\left\|\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2) - \left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right\|^2 dW(\boldsymbol{\tau}_1).$$

Notice that

$$\begin{aligned}
&\left\|\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2) - \left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right\|^2 \\
&= \left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2) - \left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right)'\left(\overline{\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2) - \left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right) \\
&= \left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right)'\left(\overline{\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)}\right) - \left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right)'\left(\overline{\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right) \\
&- \left(\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right)'\left(\overline{\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)}\right) \\
&+ \left(\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right)'\left(\overline{\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right)
\end{aligned}$$

Part I: Analyzing $\left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right)'\left(\overline{\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)}\right)$.

$$\left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right)'\left(\overline{\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)}\right) = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t(\boldsymbol{\vartheta})\hat{z}_t(\boldsymbol{\tau}_2)\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t(\boldsymbol{\vartheta})\hat{z}_t(-\boldsymbol{\tau}_2)\right)$$
$$= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{\epsilon}'_{t_1}(\boldsymbol{\vartheta})\boldsymbol{\epsilon}_{t_2}(\boldsymbol{\vartheta})\hat{z}_{t_1}(\boldsymbol{\tau}_2)\hat{z}_{t_2}(-\boldsymbol{\tau}_2)$$

and

$$\hat{z}_{t_1}(\boldsymbol{\tau}_2)\hat{z}_{t_2}(-\boldsymbol{\tau}_2) = \left(e^{i\tau_2'\epsilon_{t_1}(\vartheta)} - \hat{\psi}_0(0,\tau_2)\right)\left(e^{-i\tau_2'\epsilon_{t_2}(\vartheta)} - \hat{\psi}_0(0,-\tau_2)\right)$$
$$= e^{i\tau_2'(\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta))} - e^{i\tau_2'\epsilon_{t_1}(\vartheta)}\hat{\psi}_0(0,-\tau_2) - e^{-i\tau_2'\epsilon_{t_2}(\vartheta)}\hat{\psi}_0(0,\tau_2) + \hat{\psi}_0(0,\tau_2)\hat{\psi}_0(0,-\tau_2)$$

thus

$$\int e^{i\tau_2'(\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta))}dW(\tau_2) = e^{-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta)\right\|^2}$$
$$\int e^{i\tau_2'\epsilon_{t_1}(\vartheta)}\hat{\psi}_0(0,-\tau_2)dW(\tau_2) = \frac{1}{T}\sum_{t_3=1}^{T}e^{-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_3}(\vartheta)\right\|^2}$$
$$\int e^{-i\tau_2'\epsilon_{t_2}(\vartheta)}\hat{\psi}_0(0,\tau_2)dW(\tau_2) = \frac{1}{T}\sum_{t_3=1}^{T}e^{-\frac{1}{2}\left\|\epsilon_{t_2}(\vartheta)-\epsilon_{t_3}(\vartheta)\right\|^2}$$
$$\int \hat{\psi}_0(0,\tau_2)\hat{\psi}_0(0,-\tau_2)dW(\tau_2) = \frac{1}{T^2}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\exp\left(-\frac{1}{2}\left\|\epsilon_{t_3}(\vartheta)-\epsilon_{t_4}(\vartheta)\right\|^2\right)$$

therefore

$$\int\left\|\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right\|^2 dW(\boldsymbol{\tau}_2) = \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{\epsilon}'_{t_1}(\boldsymbol{\vartheta})\boldsymbol{\epsilon}_{t_2}(\boldsymbol{\vartheta})e^{-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_2}(\vartheta)\right\|^2}$$
$$-\frac{1}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{\epsilon}'_{t_1}(\boldsymbol{\vartheta})\boldsymbol{\epsilon}_{t_2}(\boldsymbol{\vartheta})\sum_{t_3=1}^{T}e^{-\frac{1}{2}\left\|\epsilon_{t_1}(\vartheta)-\epsilon_{t_3}(\vartheta)\right\|^2}$$
$$-\frac{1}{T^3}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{\epsilon}'_{t_1}(\boldsymbol{\vartheta})\boldsymbol{\epsilon}_{t_2}(\boldsymbol{\vartheta})\sum_{t_3=1}^{T}e^{-\frac{1}{2}\left\|\epsilon_{t_2}(\vartheta)-\epsilon_{t_3}(\vartheta)\right\|^2}$$
$$+\frac{1}{T^4}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\boldsymbol{\epsilon}'_{t_1}(\boldsymbol{\vartheta})\boldsymbol{\epsilon}_{t_2}(\boldsymbol{\vartheta})\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}e^{-\frac{1}{2}\left\|\epsilon_{t_3}(\vartheta)-\epsilon_{t_4}(\vartheta)\right\|^2}$$

Part II: Analyzing $\left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right)'\left(\overline{\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right)$.

$$\begin{aligned}
\left(\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right)'\left(\overline{\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right) &= \left(\frac{1}{T}\begin{bmatrix}\sum_{t=1}^T \epsilon_{1,t}\hat{z}_t(\tau_2)\\ \vdots \\ \sum_{t=1}^T \epsilon_{d,t}\hat{z}_t(\tau_2)\end{bmatrix}\right)'\left(\frac{1}{T}\begin{bmatrix}\sum_{t=1}^T \epsilon_{1,t}\hat{z}_{1,t}(-\tau_{2,1})\prod_{m\neq 1}\hat{\psi}_{0,m}(-\tau_{2,m})\\ \vdots \\ \sum_{t=1}^T \epsilon_{d,t}\hat{z}_{d,t}(-\tau_{2,d})\prod_{m\neq d}\hat{\psi}_{0,m}(-\tau_{2,m})\end{bmatrix}\right)\\
&= \sum_{j=1}^d \left\{\frac{1}{T^2}\sum_{t_1=1}^T\sum_{t_2=1}^T \epsilon_{j,t_1}\epsilon_{j,t_2}\hat{z}_{t_1}(\tau_2)\hat{z}_{j,t_2}(-\tau_{2,1})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\right\}
\end{aligned}$$

and

$$\begin{aligned}
\hat{z}_{t_1}(\tau_2)\hat{z}_{j,t_2}(-\tau_{2,1})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m}) &= \left(e^{i\tau_2'\epsilon_{t_1}(\vartheta)}-\hat{\psi}_0(0,\tau_2)\right)\left(e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}-\hat{\psi}_{0,j}(0,-\tau_{2,j})\right)\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\\
&= \left(e^{i\tau_2'\epsilon_{t_1}(\vartheta)}e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}-e^{i\tau_2'\epsilon_{t_1}(\vartheta)}\hat{\psi}_{0,j}(0,-\tau_{2,j})\right)\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\\
&\quad -\left(\hat{\psi}_0(0,\tau_2)e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}-\hat{\psi}_0(0,\tau_2)\hat{\psi}_{0,j}(0,-\tau_{2,j})\right)\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\\
e^{i\tau_2'\epsilon_{t_1}(\vartheta)}e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m}) &= e^{i\sum_{m=1}^d \tau_{2,m}\epsilon_{m,t_1}(\vartheta)}e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\\
&= e^{i\tau_{2,j}(\epsilon_{j,t_1}(\vartheta)-\epsilon_{j,t_2}(\vartheta))}\prod_{m\neq j}\left(\frac{1}{T}\sum_{t_3=1}^T e^{i\tau_{2,m}(\epsilon_{m,t_1}(\vartheta)-\epsilon_{m,t_3}(\vartheta))}\right)\\
e^{i\tau_2'\epsilon_{t_1}(\vartheta)}\hat{\psi}_{0,j}(0,-\tau_{2,j})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m}) &= \prod_{m=1}^d\left(\frac{1}{T}\sum_{t=1}^T e^{i\tau_{2,m}(\epsilon_{m,t_1}(\vartheta)-\epsilon_{m,t}(\vartheta))}\right)\\
\hat{\psi}_0(0,\tau_2)e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m}) &= \left(\frac{1}{T}\sum_{t=1}^T e^{i\sum_{m=1}^d \tau_{2,m}\epsilon_{m,t}(\vartheta)}e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)}\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\right)\\
&= \frac{1}{T}\sum_{t=1}^T\left\{e^{i\tau_{2,j}(\epsilon_{j,t}(\vartheta)-\epsilon_{j,t_2}(\vartheta))}\prod_{m\neq j}\left(\frac{1}{T}\sum_{t_3=1}^T e^{i\tau_{2,m}(\epsilon_{m,t}(\vartheta)-\epsilon_{m,t_3}(\vartheta))}\right)\right\}\\
\hat{\psi}_0(0,\tau_2)\hat{\psi}_{0,j}(0,-\tau_{2,j})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m}) &= \frac{1}{T}\sum_{t=1}^T\left\{\prod_{m=1}^d\left(\frac{1}{T}\sum_{t_3=1}^T e^{i\tau_{2,m}(\epsilon_{m,t}(\vartheta)-\epsilon_{m,t_3}(\vartheta))}\right)\right\}
\end{aligned}$$

thus

$$\int \left(e^{i\tau_2' \epsilon_{t_1}(\vartheta)} e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)} \prod_{m\neq j} \hat{\psi}_{0,m}(-\tau_{2,m}) \right) dW(\tau_2) = e^{-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta)-\epsilon_{j,t_2}(\vartheta)\right|^2} \prod_{m\neq j} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t_1}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right)$$

$$\int \left(e^{i\tau_2' \epsilon_{t_1}(\vartheta)} \hat{\psi}_{0,j}(0,-\tau_{2,j}) \prod_{m\neq j} \hat{\psi}_{0,m}(-\tau_{2,m}) \right) dW(\tau_2) = \prod_{m=1}^{d} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t_1}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right)$$

$$\int \left(\hat{\psi}_0(0,\tau_2) e^{-i\tau_{2_j}\epsilon_{j,t_2}(\vartheta)} \prod_{m\neq j} \hat{\psi}_{0,m}(-\tau_{2,m}) \right) dW(\tau_2) = \frac{1}{T} \sum_{t_4=1}^{T} \left\{ e^{-\frac{1}{2}\left|\epsilon_{j,t}(\vartheta)-\epsilon_{j,t_2}(\vartheta)\right|^2} \prod_{m\neq j} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t_4}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right) \right\}$$

$$\int \left(\hat{\psi}_0(0,\tau_2) \hat{\psi}_{0,j}(0,-\tau_{2,j}) \prod_{m\neq j} \hat{\psi}_{0,m}(-\tau_{2,m}) \right) dW(\tau_2) = \frac{1}{T} \sum_{t_4=1}^{T} \left\{ \prod_{m=1}^{d} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right) \right\}$$

therefore

$$\begin{aligned}
\int \hat{z}_{t_1}(\tau_2) \hat{z}_{j,t_2}(-\tau_{2,1}) \prod_{m\neq j} \hat{\psi}_{0,m}(-\tau_{2,m}) dW(\tau_2) = {} & e^{-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta)-\epsilon_{j,t_2}(\vartheta)\right|^2} \prod_{m\neq j} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t_1}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right) \\
& - \prod_{m=1}^{d} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t_1}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right) \\
& - \frac{1}{T} \sum_{t_4=1}^{T} \left\{ e^{-\frac{1}{2}\left|\epsilon_{j,t}(\vartheta)-\epsilon_{j,t_2}(\vartheta)\right|^2} \prod_{m\neq j} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t_4}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right) \right\} \\
& + \frac{1}{T} \sum_{t_4=1}^{T} \left\{ \prod_{m=1}^{d} \left(\frac{1}{T} \sum_{t_3=1}^{T} e^{-\frac{1}{2}\left|\epsilon_{m,t}(\vartheta)-\epsilon_{m,t_3}(\vartheta)\right|^2} \right) \right\}
\end{aligned}$$

and

$$\begin{aligned}
\int \left(\Delta^{(1,0)} \hat{\sigma}_0(0,\boldsymbol{\tau}_2) \right)' \left(\overline{\left[\Delta^{(1,0)} \hat{\sigma}_0(0,\boldsymbol{\tau}_2) \right]^{MDIC}} \right) dW(\tau_2) = & \\
\sum_{j=1}^{d} \left\{ \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \epsilon_{j,t_1} \epsilon_{j,t_2} \int \left\{ \hat{z}_{t_1}(\tau_2) \hat{z}_{j,t_2}(-\tau_{2,1}) \prod_{m\neq j} \hat{\psi}_{0,m}(-\tau_{2,m}) \right\} dW(\tau_2) \right\} &
\end{aligned}$$

Part III: Analyzing $\left(\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right)'\left(\overline{\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right)$.

$$
\begin{aligned}
&\left(\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}\right)'\left(\overline{\left[\Delta^{(1,0)}\hat{\sigma}_0(0,\boldsymbol{\tau}_2)\right]^{MDIC}}\right)\\
&\quad=\left(\frac{1}{T}\begin{bmatrix}\sum_{t=1}^{T}\epsilon_{1,t}\hat{z}_{1,t}(\tau_{2,1})\prod_{m\neq 1}\hat{\psi}_{0,m}(\tau_{2,m})\\ \vdots\\ \sum_{t=1}^{T}\epsilon_{d,t}\hat{z}_{d,t}(\tau_{2,d})\prod_{m\neq d}\hat{\psi}_{0,m}(\tau_{2,m})\end{bmatrix}\right)'\left(\frac{1}{T}\begin{bmatrix}\sum_{t=1}^{T}\epsilon_{1,t}\hat{z}_{1,t}(-\tau_{2,1})\prod_{m\neq 1}\hat{\psi}_{0,m}(-\tau_{2,m})\\ \vdots\\ \sum_{t=1}^{T}\epsilon_{d,t}\hat{z}_{d,t}(-\tau_{2,d})\prod_{m\neq d}\hat{\psi}_{0,m}(-\tau_{2,m})\end{bmatrix}\right)\\
&\quad=\frac{1}{T^2}\sum_{j=1}^{d}\left(\sum_{t=1}^{T}\epsilon_{j,t}\hat{z}_{j,t}(\tau_{2,j})\prod_{m\neq j}\hat{\psi}_{0,m}(\tau_{2,m})\right)\left(\sum_{t=1}^{T}\epsilon_{j,t}\hat{z}_{j,t}(-\tau_{2,j})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\right)\\
&\quad=\frac{1}{T^2}\sum_{j=1}^{d}\left\{\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\epsilon_{j,t_1}\epsilon_{j,t_2}\hat{z}_{j,t_1}(\tau_{2,j})\hat{z}_{j,t_2}(-\tau_{2,j})\prod_{m\neq j}\hat{\psi}_{0,m}(\tau_{2,m})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})\right\}
\end{aligned}
$$

From above

$$
\begin{aligned}
\int\hat{z}_{j,t_1}(\tau_{2,j})\hat{z}_{j,t_2}(-\tau_{2,j})dW(\tau_{2,j})&=e^{-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta)-\epsilon_{j,t_2}(\vartheta)\right|^2}-\frac{1}{T}\sum_{t_3=1}^{T}e^{-\frac{1}{2}\left|\epsilon_{j,t_1}(\vartheta)-\epsilon_{j,t_3}(\vartheta)\right|^2}\\
&\quad-\frac{1}{T}\sum_{t_3=1}^{T}e^{-\frac{1}{2}\left|\epsilon_{j,t_2}(\vartheta)-\epsilon_{j,t_3}(\vartheta)\right|^2}+\frac{1}{T^2}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}e^{-\frac{1}{2}\left|\epsilon_{j,t_3}(\vartheta)-\epsilon_{j,t_4}(\vartheta)\right|^2}
\end{aligned}
$$

$$
\begin{aligned}
\int\prod_{m\neq j}\hat{\psi}_{0,m}(\tau_{2,m})\prod_{m\neq j}\hat{\psi}_{0,m}(-\tau_{2,m})dW\left(\prod_{m\neq j}\tau_{2,m}\right)&=\prod_{m\neq j}\int\hat{\psi}_{0,m}(\tau_{2,m})\hat{\psi}_{0,m}(-\tau_{2,m})dW(\tau_{2,m})\\
&=\prod_{m\neq j}\left(\frac{1}{T^2}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}e^{-\frac{1}{2}\left|\epsilon_{m,t_3}(\vartheta)-\epsilon_{m,t_4}(\vartheta)\right|^2}\right)
\end{aligned}
$$

APPENDIX - CHAPTER II

5.11. Proof of Proposition 3.1

Given $G_k^2(\lambda) = \boldsymbol{\delta}\left(e^{-i\lambda};\boldsymbol{\vartheta}_f\right)\left(\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime}\boldsymbol{\Omega}^{\otimes(k-1)}\mathbf{v}_k^{\boldsymbol{\varepsilon}}\right)\boldsymbol{\delta}'\left(e^{i\lambda};\boldsymbol{\vartheta}_f\right)$ and the fact that either the SVARMA and RF-VARMA do not present roots on the unit circle (by condition (2.2) in the paper), then $\det(\boldsymbol{\delta}(z;\boldsymbol{\vartheta}_f)) = \det\left(\boldsymbol{\Phi}^{-1}(z;\boldsymbol{\vartheta}_f)\boldsymbol{\Phi}(z;\boldsymbol{\vartheta}_{0,1})\right) \neq 0$ for all $|z| = 1$. This implies that $\boldsymbol{\delta}(z;\boldsymbol{\vartheta}_f)$ is full rank for any z on the unit circle. Thus

$$\begin{aligned}
\operatorname{rank}\left(G_k^2(\lambda)\right) &= \operatorname{rank}\left[\boldsymbol{\delta}\left(e^{-i\lambda};\boldsymbol{\vartheta}_f\right)\left(\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime}\boldsymbol{\Omega}^{\otimes(k-1)}\mathbf{v}_k^{\boldsymbol{\varepsilon}}\right)\boldsymbol{\delta}'\left(e^{i\lambda};\boldsymbol{\vartheta}_f\right)\right] \\
&= \operatorname{rank}\left[\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime}\boldsymbol{\Omega}^{\otimes(k-1)}\mathbf{v}_k^{\boldsymbol{\varepsilon}}\right] \\
&= \operatorname{rank}\left[\mathbf{v}_k^{\boldsymbol{\varepsilon}\prime}\mathbf{v}_k^{\boldsymbol{\varepsilon}}\right],
\end{aligned}$$

where last equality comes from the fact that $\boldsymbol{\Omega}$ is squared, full rank matrix, implying that $\boldsymbol{\Omega}^{\otimes(k-1)}$ is also squared and full rank matrix. Thus, this implies that the dimension of null space spanned by $\boldsymbol{\Omega}^{\otimes(k-1)}\mathbf{v}_k^{\boldsymbol{\varepsilon}}$ is equal to the dimension of null space generated by $\mathbf{v}_k^{\boldsymbol{\varepsilon}}$. This conclusion holds because of

$$\begin{aligned}
\boldsymbol{\Omega}^{\otimes(k-1)}\mathbf{v}_k^{\boldsymbol{\varepsilon}} &= \boldsymbol{\Omega}^{\otimes(k-1)}\begin{bmatrix}\kappa_{k,1}\boldsymbol{e}_1^{\otimes(k-1)} & \cdots & \kappa_{k,d}\boldsymbol{e}_d^{\otimes(k-1)}\end{bmatrix} \\
&= \begin{bmatrix}\kappa_{k,1}\boldsymbol{\sigma}_1^{\otimes(k-1)} & \cdots & \kappa_{k,d}\boldsymbol{\sigma}_d^{\otimes(k-1)}\end{bmatrix}
\end{aligned}$$

Therefore

$$\operatorname{rank}\left(G_k^2(\lambda)\right) = d_k,\ k = 3,4$$

where d_k is the number of non-zero marginal cumulants $\{\kappa_{k,j}\}_{j=1}^d$

Now, the rectangular array $G_{34}^2(\lambda) = \begin{bmatrix}G_3^2(\lambda)\\ G_4^2(\lambda)\end{bmatrix}$ is

$$G_{34}^2(\lambda) = \begin{bmatrix}\boldsymbol{\delta}\left(e^{-i\lambda};\boldsymbol{\vartheta}_f\right) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\delta}\left(e^{-i\lambda};\boldsymbol{\vartheta}_f\right)\end{bmatrix}\begin{bmatrix}\mathbf{v}_3^{\boldsymbol{\varepsilon}\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\boldsymbol{\varepsilon}}\\ \mathbf{v}_4^{\boldsymbol{\varepsilon}\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\boldsymbol{\varepsilon}}\end{bmatrix}\boldsymbol{\delta}'\left(e^{i\lambda};\boldsymbol{\vartheta}_f\right),$$

where

$$\mathbf{v}_k^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes(k-1)}\mathbf{v}_k^{\varepsilon} = \begin{bmatrix} \kappa_{k,1}\boldsymbol{e}_1^{\otimes(k-1)\prime} \\ \vdots \\ \kappa_{k,d}\boldsymbol{e}_d^{\otimes(k-1)\prime} \end{bmatrix} \boldsymbol{\Omega}^{\otimes(k-1)} \begin{bmatrix} \kappa_{k,1}\boldsymbol{e}_1^{\otimes(k-1)} & \cdots & \kappa_{k,d}\boldsymbol{e}_d^{\otimes(k-1)} \end{bmatrix}$$

$$= \begin{bmatrix} \kappa_{k,1}\boldsymbol{e}_1^{\otimes(k-1)\prime} \\ \vdots \\ \kappa_{k,d}\boldsymbol{e}_d^{\otimes(k-1)\prime} \end{bmatrix} \begin{bmatrix} \kappa_{k,1}\boldsymbol{\sigma}_1^{\otimes(k-1)} & \cdots & \kappa_{k,d}\boldsymbol{\sigma}_d^{\otimes(k-1)} \end{bmatrix}$$

$$= \begin{bmatrix} \kappa_{k,1}^2\sigma_{11}^{k-1} & \kappa_{k,1}\kappa_{k,2}\sigma_{12}^{k-1} & \kappa_{k,1}\kappa_{k,3}\sigma_{13}^{k-1} & \cdots & \kappa_{k,1}\kappa_{k,d}\sigma_{1d}^{k-1} \\ \kappa_{k,1}\kappa_{k,2}\sigma_{12}^{k-1} & \kappa_{k,2}^2\sigma_{22}^{k-1} & \kappa_{k,2}\kappa_{k,3}\sigma_{23}^{k-1} & \cdots & \kappa_{k,2}\kappa_{k,d}\sigma_{2d}^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{k,1}\kappa_{k,d}\sigma_{1d}^{k-1} & \kappa_{k,2}\kappa_{k,d}\sigma_{2d}^{k-1} & \kappa_{k,3}\kappa_{k,d}\sigma_{3d}^{k-1} & \cdots & \kappa_{k,d}^2\sigma_{dd}^{k-1} \end{bmatrix},$$

thus, it is evident that

$$\operatorname{rank}\left(G_{34}^2(\lambda)\right) = \operatorname{rank}\left(\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}\right)$$

with

$$\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix} = \begin{bmatrix} \kappa_{3,1}^2\sigma_{11}^2 & \kappa_{3,1}\kappa_{3,2}\sigma_{12}^2 & \kappa_{3,1}\kappa_{3,3}\sigma_{13}^2 & \cdots & \kappa_{3,1}\kappa_{3,d}\sigma_{1d}^2 \\ \kappa_{3,1}\kappa_{3,2}\sigma_{12}^2 & \kappa_{3,2}^2\sigma_{22}^2 & \kappa_{3,2}\kappa_{3,3}\sigma_{23}^2 & \cdots & \kappa_{3,2}\kappa_{3,d}\sigma_{2d}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{3,1}\kappa_{3,d}\sigma_{1d}^2 & \kappa_{3,2}\kappa_{3,d}\sigma_{2d}^2 & \kappa_{3,3}\kappa_{3,d}\sigma_{3d}^2 & \cdots & \kappa_{3,d}^2\sigma_{dd}^2 \\ \kappa_{4,1}^2\sigma_{11}^3 & \kappa_{4,1}\kappa_{4,2}\sigma_{12}^3 & \kappa_{4,1}\kappa_{4,3}\sigma_{13}^3 & \cdots & \kappa_{4,1}\kappa_{4,d}\sigma_{1d}^3 \\ \kappa_{4,1}\kappa_{4,2}\sigma_{12}^3 & \kappa_{4,2}^2\sigma_{22}^3 & \kappa_{4,2}\kappa_{4,3}\sigma_{23}^3 & \cdots & \kappa_{4,2}\kappa_{4,d}\sigma_{2d}^3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \kappa_{4,1}\kappa_{4,d}\sigma_{1d}^3 & \kappa_{4,2}\kappa_{4,d}\sigma_{2d}^3 & \kappa_{4,3}\kappa_{4,d}\sigma_{3d}^3 & \cdots & \kappa_{4,d}^2\sigma_{dd}^3 \end{bmatrix}.$$

For simplicity, let assume that $d = 2$, then:

$$\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix} = \begin{bmatrix} \kappa_{3,1}^2\sigma_{11}^2 & \kappa_{3,1}\kappa_{3,2}\sigma_{12}^2 \\ \kappa_{3,1}\kappa_{3,2}\sigma_{12}^2 & \kappa_{3,1}^2\sigma_{22}^2 \\ \kappa_{4,1}^2\sigma_{11}^3 & \kappa_{4,1}\kappa_{4,2}\sigma_{12}^3 \\ \kappa_{4,1}\kappa_{4,2}\sigma_{12}^3 & \kappa_{4,1}^2\sigma_{22}^3 \end{bmatrix},$$

we can distinguish the following cases:

a. $\kappa_{3,j} \neq 0$ and $\kappa_{4,j} = 0$ for $j = 1,2$. In this case $\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{0}_{2\times 2} \end{bmatrix}$, therefore

$$\operatorname{rank}\left(\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}\right) = \operatorname{rank}(\mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon}) = d_3.$$

b. $\kappa_{4,j} \neq 0$ and $\kappa_{3,j} = 0$ for $j = 1,2$. Similarly to above $\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{2\times 2} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}$, therefore

$$\operatorname{rank}\left(\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}\right) = \operatorname{rank}(\mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon}) = d_4.$$

c. $\kappa_{3,j}\kappa_{4,j} \neq 0$ for only one $j \in \{1,2\}$. Assuming $j = 1$, thus $\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix} = \begin{bmatrix} \kappa_{3,1}^2\sigma_{11}^2 & 0 \\ 0 & 0 \\ \kappa_{4,1}^2\sigma_{11}^3 & 0 \\ 0 & 0 \end{bmatrix}$.

Therefore

$$\operatorname{rank}\left(\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}\right) = \operatorname{rank}(\mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon}) = \operatorname{rank}(\mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon}) = 1.$$

d. $\kappa_{3,j}\kappa_{4,j'} \neq 0$ for only one $j \neq j' \in \{1,2\}$. Assuming $j = 1$, then $\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix} = \begin{bmatrix} \kappa_{3,1}^2\sigma_{11}^2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \kappa_{4,2}^2\sigma_{11}^3 \end{bmatrix}$.

Therefore

$$\operatorname{rank}\left(\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}\right) = \operatorname{rank}(\mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon}) + \operatorname{rank}(\mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon}) = 2.$$

For higher dimensions a similar analysis can be performed, the number of cases increases but the idea is similar. Consequently

$$\operatorname{rank}\left(G_{34}^2(\lambda)\right) = \operatorname{rank}\left(\begin{bmatrix} \mathbf{v}_3^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 2}\mathbf{v}_3^{\varepsilon} \\ \mathbf{v}_4^{\varepsilon\prime}\boldsymbol{\Omega}^{\otimes 3}\mathbf{v}_4^{\varepsilon} \end{bmatrix}\right) = d_{34}$$

where d_{34} is the number of asymmetric and/or non-mesokurtic structural shocks.

5.12. Proof of Proposition 3.2

5.12.1. Asymptotic Distribution under $H_{0,1}$

At the first step ($H_{0,1}$: rank($\mathbf{\Pi}$) = 0), the asymptotic distribution for $\widehat{\boldsymbol{\pi}} = \text{vec}(\widehat{\mathbf{\Pi}})$ cannot be obtained by standard Delta method procedure. When $\widehat{\mathbf{\Pi}} = \text{Re}\left(\hat{G}_k^{2,(T)}(\lambda)\right)$, then

$$\widehat{\boldsymbol{\pi}} = \left[\left(I_d \otimes \hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda})\right)^{\otimes 2} + \left(I_d \otimes \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\right)^{\otimes 2}\right] \text{vec}(\mathbf{I}(d))$$

For understanding the idea behind our result, let assume for simplicity that $d = 1$ (univariate case). Thus

$$\widehat{\boldsymbol{\pi}} = \left[\left\{\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda})\right\}^2 + \left\{\hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\right\}^2\right]$$

and its Jacobian (gradient) evaluated at population values $g_{\text{Re},k}(\boldsymbol{\lambda})$ and $g_{\text{Im},k}(\boldsymbol{\lambda})$ under the null of joint Gaussianity:

$$\begin{aligned}\frac{\partial}{\partial\left(\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}), \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\right)}(\widehat{\boldsymbol{\pi}}) = \boldsymbol{J} &= \left[2\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) \quad 2\hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\right]\Big|_{\left(\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}), \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\right) = \left(g_{\text{Re},k}(\boldsymbol{\lambda}), g_{\text{Im},k}(\boldsymbol{\lambda})\right)} \\ &= 2\begin{bmatrix}0 & 0\end{bmatrix} = \mathbf{0}\end{aligned}$$

In consequence, the standard Delta Method (based on mean value theorem) cannot be employed. We need to employ second order approximation. That is, under joint Gaussianity

$$\widehat{\boldsymbol{\pi}} = \frac{1}{2}\text{vec}(\mathcal{H})' \begin{pmatrix}\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) \\ \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\end{pmatrix}^{\otimes 2}.$$

Notice that $\begin{pmatrix}\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) \\ \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\end{pmatrix}^{\otimes 2} = \text{vec}\left(\begin{pmatrix}\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) \\ \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\end{pmatrix}\begin{pmatrix}\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) & \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\end{pmatrix}\right)$, and since

$$a_T\left\{\begin{pmatrix}\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) \\ \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\end{pmatrix} - \begin{pmatrix}0 \\ 0\end{pmatrix}\right\} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}, \mathcal{V}(\boldsymbol{\lambda}))$$

thus

$$a_T^2\left\{\begin{pmatrix}\hat{g}_{\text{Re},k}^{(T)}(\boldsymbol{\lambda}) \\ \hat{g}_{\text{Im},k}^{(T)}(\boldsymbol{\lambda})\end{pmatrix}^{\otimes 2}\right\} \xrightarrow{d} \text{vec}\left(\mathcal{N}_2(\mathbf{0}, \mathcal{V}(\boldsymbol{\lambda}))\mathcal{N}_2'(\mathbf{0}, \mathcal{V}(\boldsymbol{\lambda}))\right) = \mathbf{X}.$$

Therefore

$$a_T^2(\widehat{\boldsymbol{\pi}}-\mathbf{0}) \xrightarrow{d} \frac{1}{2}\mathbf{X}'\operatorname{vec}(\mathscr{H})$$

When $d \geq 2$ (multivariate case), we follow the same idea and we obtain equation (4.9) in the paper:

$$a_T^2\left(\operatorname{vec}(\widehat{\boldsymbol{\Pi}})-\mathbf{0}\right) = a_T^2(\widehat{\boldsymbol{\pi}}-\mathbf{0}) = a_T^2\widehat{\boldsymbol{\pi}} \xrightarrow{d} \boldsymbol{Q},$$

with $a_T = \sqrt{H_T^{k-1}T}$, $\boldsymbol{Q} = \frac{1}{2}\left(\boldsymbol{I}_d \otimes \mathbf{X}'\right)\mathscr{H}_{\widehat{\boldsymbol{\pi}}}$ and $\mathscr{H}_{\widehat{\boldsymbol{\pi}}}$ the vectorized Hessian matrix of $\widehat{\boldsymbol{\pi}}$ evaluated at zero.

Now, under $H_{0,1}$, it holds that $\boldsymbol{\Pi} = \mathbf{0}$ and $\mathscr{L}_0 = \mathbf{0}$, then

$$\hat{\boldsymbol{\Pi}} = \hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi} = \hat{\boldsymbol{C}}_{0,\perp}\hat{\mathscr{L}}_0\hat{\boldsymbol{D}}_{0,\perp}$$

thus

$$\begin{aligned}
\operatorname{vec}\left(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}\right) &= \left(\widehat{\boldsymbol{D}}'_{0,\perp} \otimes \widehat{\boldsymbol{C}}_{0,\perp}\right)\operatorname{vec}\left[\widehat{\mathscr{L}}_0\right] \\
\|(\widehat{\boldsymbol{\pi}}-\boldsymbol{\pi})\| &= \left\|\left(\widehat{\boldsymbol{D}}'_{0,\perp} \otimes \widehat{\boldsymbol{C}}_{0,\perp}\right)\widehat{\boldsymbol{\ell}}_0\right\| \\
&= \sqrt{\widehat{\boldsymbol{\ell}}'_0\left(\widehat{\boldsymbol{D}}_{0,\perp} \otimes \widehat{\boldsymbol{C}}'_{0,\perp}\right)\left(\widehat{\boldsymbol{D}}'_{0,\perp} \otimes \widehat{\boldsymbol{C}}_{0,\perp}\right)\widehat{\boldsymbol{\ell}}_0} \\
&= \sqrt{\widehat{\boldsymbol{\ell}}'_0\left(\widehat{\boldsymbol{D}}_{0,\perp}\widehat{\boldsymbol{D}}'_{0,\perp} \otimes \boldsymbol{I}_d\right)\widehat{\boldsymbol{\ell}}_0} = \sqrt{\widehat{\boldsymbol{\ell}}'_0(\boldsymbol{I}_d \otimes \boldsymbol{I}_d)\widehat{\boldsymbol{\ell}}_0} = \left\|\widehat{\boldsymbol{\ell}}_0\right\|
\end{aligned}$$

since $\mathbb{E}\|\widehat{\boldsymbol{\pi}}-\boldsymbol{\pi}\| = O(a_T^{-2})$, then $\mathbb{E}\left\|\widehat{\boldsymbol{\ell}}_0\right\| = O(a_T^{-2})$. Thus, $\left\|a_T^2\widehat{\boldsymbol{\ell}}_0\right\| = O_p(1)$.
Assuming that $\left\|\widehat{\boldsymbol{D}}'_{0,\perp} \otimes \widehat{\boldsymbol{C}}_{0,\perp} - \overline{\boldsymbol{D}}'_{0,\perp} \otimes \overline{\boldsymbol{C}}_{0,\perp}\right\|$ converges to 0 in probability, therefore

$$a_T^2\operatorname{vec}\left[\widehat{\mathscr{L}}_0 - \mathscr{L}_0\right] = a_T^2\widehat{\boldsymbol{\ell}}_0 \xrightarrow{d} \left(\overline{\boldsymbol{D}}_{0,\perp} \otimes \overline{\boldsymbol{C}}'_{0,\perp}\right)\boldsymbol{Q} \tag{5.58}$$

and the rank statistic is

$$KP_0^{(T)} = a_T^4\hat{\boldsymbol{\ell}}'_0\widehat{\text{aVar}}\left(\left(\overline{\boldsymbol{D}}_{0,\perp} \otimes \overline{\boldsymbol{C}}'_{0,\perp}\right)\boldsymbol{Q}\right)^{-1}\hat{\boldsymbol{\ell}}_0 \xrightarrow{d} \boldsymbol{Q}'\left(\overline{\boldsymbol{D}}'_{0,\perp} \otimes \overline{\boldsymbol{C}}_{0,\perp}\right)\operatorname{Var}\left[\left(\overline{\boldsymbol{D}}_{0,\perp} \otimes \overline{\boldsymbol{C}}'_{0,\perp}\right)\boldsymbol{Q}\right]^{-1}\left(\overline{\boldsymbol{D}}_{0,\perp} \otimes \overline{\boldsymbol{C}}'_{0,\perp}\right)\boldsymbol{Q} \tag{5.59}$$

5.13. Proof of Proposition 3.3

Given the definition of $\hat{\boldsymbol{u}}_t = \bar{\boldsymbol{\Psi}}^{(-1)}(L, \hat{\boldsymbol{\vartheta}}_f)\boldsymbol{y}_t = \sum_{j=0}^{t-1} \boldsymbol{\Psi}_j^{(-1)}(\hat{\boldsymbol{\vartheta}}_f)\boldsymbol{y}_{t-j}$ and consider the approximation of the truncated filter $\bar{\boldsymbol{\Psi}}^{(-1)}(L, \hat{\boldsymbol{\vartheta}}_f)$:

$$
\begin{aligned}
\bar{\boldsymbol{\Psi}}^{(-1)}(L, \hat{\boldsymbol{\vartheta}}_f) = \bar{\boldsymbol{\Psi}}^{(-1)}(L, \boldsymbol{\vartheta}_f) &+ \sum_{n=1}^{K} \dot{\bar{\boldsymbol{\Psi}}}_n^{(-1)}(L, \boldsymbol{\vartheta}_f)(\hat{\vartheta}_{f,n} - \vartheta_{f,n}) \\
&+ \frac{1}{2} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \ddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2}^{(-1)}(L, \boldsymbol{\vartheta}_f)(\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2}) \\
&+ \frac{1}{6} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \sum_{n_3=1}^{K} \dddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2,n_3}^{(-1)}(L, \tilde{\boldsymbol{\vartheta}}_f)(\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2})(\hat{\vartheta}_{f,n_3} - \vartheta_{f,n_3})
\end{aligned}
$$

where $\tilde{\boldsymbol{\vartheta}}_f$ is random and $\|\tilde{\boldsymbol{\vartheta}}_f - \boldsymbol{\vartheta}_f\| \leq \|\hat{\boldsymbol{\vartheta}}_f - \boldsymbol{\vartheta}_f\|$.

Besides $\dot{\bar{\boldsymbol{\Psi}}}_n^{(-1)}(L, \boldsymbol{\vartheta}_f) = \frac{\partial}{\partial \hat{\vartheta}_n} \bar{\boldsymbol{\Psi}}^{(-1)}(L, \hat{\boldsymbol{\vartheta}}_f)\Big|_{\hat{\boldsymbol{\vartheta}}_f = \boldsymbol{\vartheta}_f}$,
$\ddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2}^{(-1)}(L, \boldsymbol{\vartheta}_f) = \frac{\partial^2}{\partial \hat{\vartheta}_{n_1} \partial \hat{\vartheta}_{n_2}} \bar{\boldsymbol{\Psi}}^{(-1)}(L, \hat{\boldsymbol{\vartheta}}_f)\Big|_{\hat{\boldsymbol{\vartheta}}_f = \boldsymbol{\vartheta}_f}$ and $\dddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2,n_3}^{(-1)}(L, \tilde{\boldsymbol{\vartheta}}_f) = \frac{\partial^3}{\partial \hat{\vartheta}_{n_1} \partial \hat{\vartheta}_{n_2} \partial \hat{\vartheta}_{n_3}} \bar{\boldsymbol{\Psi}}^{(-1)}(L, \hat{\boldsymbol{\vartheta}}_f)\Big|_{\hat{\boldsymbol{\vartheta}}_f = \tilde{\boldsymbol{\vartheta}}_f}$.

Then

$$
\begin{aligned}
\hat{\boldsymbol{u}}_t &= \bar{\boldsymbol{\Psi}}^{(-1)}(L, \boldsymbol{\vartheta}_f)\boldsymbol{y}_t + \sum_{n=1}^{K} \dot{\bar{\boldsymbol{\Psi}}}_n^{(-1)}(L, \boldsymbol{\vartheta}_f)\boldsymbol{y}_t(\hat{\vartheta}_{f,n} - \vartheta_{f,n}) + \frac{1}{2} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \ddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2}^{(-1)}(L, \boldsymbol{\vartheta}_f)\boldsymbol{y}_t(\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2}) \\
&\quad + \frac{1}{6} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \sum_{n_3=1}^{K} \dddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2,n_3}^{(-1)}(L, \tilde{\boldsymbol{\vartheta}}_f)\boldsymbol{y}_t(\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2})(\hat{\vartheta}_{f,n_3} - \vartheta_{f,n_3}) \\
&= \tilde{\boldsymbol{u}}_t + \sum_{n=1}^{K} \xi_t^{(n)}(\hat{\vartheta}_{f,n} - \vartheta_{f,n}) + \frac{1}{2} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \boldsymbol{\chi_t^{(n)}}(\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2}) \\
&\quad + \frac{1}{6} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \sum_{n_3=1}^{K} \dddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2,n_3}^{(-1)}(L, \tilde{\boldsymbol{\vartheta}}_f)\boldsymbol{y}_t(\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2})(\hat{\vartheta}_{f,n_3} - \vartheta_{f,n_3})
\end{aligned}
$$

where $\xi_t^{(n)} \equiv \dot{\bar{\boldsymbol{\Psi}}}_n^{(-1)}(L, \boldsymbol{\vartheta}_f)\boldsymbol{y}_t$ and $\chi_t^{(n_1,n_2)} \equiv \ddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2}^{(-1)}(L, \boldsymbol{\vartheta}_f)\boldsymbol{y}_t$.

Hence

$$
\begin{aligned}
\sum_{t=0}^{T-1} \hat{\boldsymbol{u}}_t \odot e^{-i\boldsymbol{\lambda}t} = \sum_{t=0}^{T-1} \tilde{\boldsymbol{u}}_t \odot e^{-i\boldsymbol{\lambda}t} &+ \sum_{n=1}^{K} \left\{ \sum_{t=0}^{T-1} \xi_t^{(n)} \odot e^{-i\boldsymbol{\lambda}t} \right\} (\hat{\vartheta}_{f,n} - \vartheta_{f,n}) \\
&+ \frac{1}{2} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \left\{ \sum_{t=0}^{T-1} \boldsymbol{\chi_t^{(n)}} \odot e^{-i\boldsymbol{\lambda}t} \right\} (\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2}) \\
&+ \frac{1}{6} \sum_{n_1=1}^{K} \sum_{n_2=1}^{K} \sum_{n_3=1}^{K} \left\{ \sum_{t=0}^{T-1} \dddot{\bar{\boldsymbol{\Psi}}}_{n_1,n_2,n_3}^{(-1)}(L, \tilde{\boldsymbol{\vartheta}}_f)\boldsymbol{y}_t \odot e^{-i\boldsymbol{\lambda}t} \right\} (\hat{\vartheta}_{f,n_1} - \vartheta_{f,n_1})(\hat{\vartheta}_{f,n_2} - \vartheta_{f,n_2})(\hat{\vartheta}_{f,n_3} - \vartheta_{f,n_3})
\end{aligned}
$$

where $\sum_{t=0}^{T-1} \hat{\boldsymbol{u}}_t \odot e^{-i\boldsymbol{\lambda}t}$ is a vector containing finite, discrete Fourier transform of vector-valued random element $\hat{\boldsymbol{u}}_t$. The higher-order periodogram $I^{\hat{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)$ is the multiplication of any k-tuple of the elements of $\sum_{t=0}^{T-1}\hat{\boldsymbol{u}}_t \odot e^{-i\boldsymbol{\lambda}t}$. Thus:

$$I^{\hat{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)=\frac{1}{(2\pi)^{k-1}T}\prod_{j=1}^{k}\left(\sum_{t=0}^{T-1}\tilde{u}_{t,c_j}e^{-i\frac{2\pi t_j}{T}t}\right)+\boldsymbol{R}=I^{\tilde{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)+\boldsymbol{R}$$

where $\boldsymbol{R}$ contains several terms of higher order.

Therefore, the difference $\hat{g}^{\boldsymbol{u},(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda})-\hat{g}^{\hat{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda})$ is equivalent to

$$\left(\frac{2\pi}{T}\right)^{k-1}\sum_{t_1,\dots,t_k=0}^{T-1}W_T\left(\lambda_1-\frac{2\pi t_1}{T},\dots,\lambda_k-\frac{2\pi t_k}{T}\right)\left\{I^{\boldsymbol{u},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)-I^{\tilde{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)\right\}$$
$$+\left(\frac{2\pi}{T}\right)^{k-1}\sum_{t_1,\dots,t_k=0}^{T-1}W_T\left(\lambda_1-\frac{2\pi t_1}{T},\dots,\lambda_k-\frac{2\pi t_k}{T}\right)\boldsymbol{R}$$

Now, let focus on the first term and denoting $\mathbb{E}_T\equiv\left(\frac{2\pi}{T}\right)^{k-1}\sum_{t_1,\dots,t_k=0}^{T-1}W_T\left(\lambda_1-\frac{2\pi t_1}{T},\dots,\lambda_k-\frac{2\pi t_k}{T}\right)$

$$\begin{aligned}\mathbb{E}_T&\left\{I^{\boldsymbol{u},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)-I^{\tilde{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)\right\}\\&=\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\sum_{j=1}^{k}\prod_{m\in\mathscr{I}_j}z^{\boldsymbol{u},(T)}_{c_m}(\frac{2\pi t_m}{T})\prod_{m'\in\tilde{\mathscr{I}}_j}z^{\tilde{\boldsymbol{u}},(T)}_{c_{m'}}(\frac{2\pi t_{m'}}{T})\left[z^{\boldsymbol{u},(T)}_{c_j}(\frac{2\pi t_j}{T})-z^{\tilde{\boldsymbol{u}},(T)}_{c_j}(\frac{2\pi t_j}{T})\right]\right\}\right)\\&=\sum_{j=1}^{k}\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m\in\mathscr{I}_j}z^{\boldsymbol{u},(T)}_{c_m}(\frac{2\pi t_m}{T})\prod_{m'\in\tilde{\mathscr{I}}_j}z^{\tilde{\boldsymbol{u}},(T)}_{c_{m'}}(\frac{2\pi t_{m'}}{T})\left[z^{\boldsymbol{u},(T)}_{c_j}(\frac{2\pi t_j}{T})-z^{\tilde{\boldsymbol{u}},(T)}_{c_j}(\frac{2\pi t_j}{T})\right]\right\}\right)\end{aligned}$$

where $\mathscr{I}_j:=\{j'\mid j<j'\le k\}$, $\tilde{\mathscr{I}}_j:=\{j'\mid 1\le j'<j\}$ and

$$z^{\boldsymbol{u},(T)}_{c_j}(\frac{2\pi t_j}{T})-z^{\hat{\boldsymbol{u}},(T)}_{c_j}(\frac{2\pi t_j}{T})=\sum_{t=0}^{T-1}\left\{\boldsymbol{u}_{c_j,t}-\tilde{\boldsymbol{u}}_{c_j,t}\right\}e^{-i\frac{2\pi t_j}{T}t}=\sum_{t=0}^{T-1}\left[\sum_{s=t}^{\infty}\boldsymbol{\Psi}^{(-1)}_s(\boldsymbol{\vartheta})\boldsymbol{y}_{t-s}\right]_{c_j}e^{-i\frac{2\pi t_j}{T}t}$$

calling $\boldsymbol{\nu}_t=\sum_{s=t}^{\infty}\boldsymbol{\Psi}^{(-1)}_s(\boldsymbol{\vartheta})\boldsymbol{y}_{t-s}$ with

$$\begin{aligned}\mathbb{E}\boldsymbol{\nu}_t&=\boldsymbol{0}\\\mathbb{E}\boldsymbol{\nu}_t\boldsymbol{\nu}'_t&=\sum_{s_1=t}^{\infty}\sum_{s_2=t}^{\infty}\boldsymbol{\Psi}^{(-1)}_{s_1}(\boldsymbol{\vartheta}_f)\boldsymbol{\Gamma}^{\boldsymbol{y}}_{s_1-s_2}\boldsymbol{\Psi}^{(-1)\prime}_{s_2}(\boldsymbol{\vartheta}_f)\\\mathbb{E}\boldsymbol{\nu}_t\boldsymbol{\nu}'_{t+h}&=\sum_{s_1=t}^{\infty}\sum_{s_2=t+h}^{\infty}\boldsymbol{\Psi}^{(-1)}_{s_1}(\boldsymbol{\vartheta}_f)\boldsymbol{\Gamma}^{\boldsymbol{y}}_{h+s_1-s_2}\boldsymbol{\Psi}^{(-1)\prime}_{s_2}(\boldsymbol{\vartheta}_f)\end{aligned}$$

given that $\boldsymbol{\Gamma}^{\boldsymbol{y}}_h=\sum_{s=-\infty}^{\infty}\boldsymbol{\Psi}_s(\boldsymbol{\vartheta}_0)\boldsymbol{\Psi}'_{s+h}(\boldsymbol{\vartheta}_0)$ and our data is generated by an SVARMA mode,

$\left\|\mathbf{\Gamma}_h^{\boldsymbol{y}}\right\| \le C\sum_{s=-\infty}^{\infty}|s+h|^{-\eta_0}|s|^{-\eta_0} = C\sum_{s=1}^{\infty}|s+h|^{-\eta_0}|s|^{-\eta_0} \le C|h|^{-\eta_0}$ for $\eta_0>1$. Then, it is straightforward to show

$$\left\|\mathbb{E}\boldsymbol{\nu}_t\boldsymbol{\nu}_t'\right\| \le C\left\|\sum_{\Delta=0}^{\infty}\sum_{s_1=t}^{\infty}\mathbf{\Psi}_{s_1}^{(-1)}(\boldsymbol{\vartheta}_f)\mathbf{\Gamma}_\Delta^{\boldsymbol{y}}\mathbf{\Psi}_{s_1-\Delta}^{(-1)\prime}(\boldsymbol{\vartheta}_f)\right\| \le C\sum_{\Delta=0}^{\infty}\left\|\mathbf{\Gamma}_\Delta^{\boldsymbol{y}}\right\|\sum_{s_1=t}^{\infty}\left\|\mathbf{\Psi}_{s_1}^{(-1)}(\boldsymbol{\vartheta}_f)\mathbf{\Psi}_{s_1-\Delta}^{(-1)\prime}(\boldsymbol{\vartheta}_f)\right\|$$
$$\le C\sum_{\Delta=0}^{\infty}\left\|\mathbf{\Gamma}_\Delta^{\boldsymbol{y}}\right\|\sum_{s_1=t}^{\infty}|s_1|^{-\eta_0}|s_1-\Delta|^{-\eta_0} \le C\sum_{\Delta=0}^{\infty}\left\|\mathbf{\Gamma}_\Delta^{\boldsymbol{y}}\right\|\sum_{s_1=t}^{\infty}|s_1|^{-\eta_0} \le C\sum_{\Delta=0}^{\infty}\left\|\mathbf{\Gamma}_\Delta^{\boldsymbol{y}}\right\||t|^{1-\eta_0} \le Ct^{1-\eta_0}$$
$$\left\|\mathbb{E}\boldsymbol{\nu}_t\boldsymbol{\nu}_{t+h}'\right\| \le Ct^{1-\eta_0}|h|^{1-\eta_0}$$

then $\left\|\mathbb{E}\boldsymbol{\nu}_t\boldsymbol{\nu}_t'\right\| \to 0$ as $t\to\infty$ and $\left\|\mathbb{E}\boldsymbol{\nu}_t\boldsymbol{\nu}_{t+h}'\right\| \to 0$ as $h\to\infty$ for $\eta_0>1$. This implies that $\boldsymbol{\nu}_t$ is stationary and converges in probability to zero.

Hence

$$\mathbb{E}_T\left\{I_{\boldsymbol{c},k}^{\boldsymbol{u},(T)}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right) - I_{\boldsymbol{c},k}^{\tilde{\boldsymbol{u}},(T)}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)\right\}$$
$$= \sum_{j=1}^{k}\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m\in\mathscr{I}_j}z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\prod_{m'\in\tilde{\mathscr{I}}_j}z_{c_{m'}}^{\tilde{\boldsymbol{u}},(T)}(\frac{2\pi t_{m'}}{T})\left[\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_j}e^{-i\frac{2\pi t_j}{T}t}\right]\right\}\right)$$

then

$$\mathbb{E}\left|\sqrt{B_T^{k-1}T}\,\mathbb{E}_T\left\{I_{\boldsymbol{c},k}^{\boldsymbol{u},(T)}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right) - I_{\boldsymbol{c},k}^{\tilde{\boldsymbol{u}},(T)}\left(\frac{2\pi t_1}{T},\dots,\frac{2\pi t_k}{T}\right)\right\}\right|$$
$$\le CB_T^{\frac{k-1}{2}}T^{\frac{1}{2}}\,\mathbb{E}\left|\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m=2}^{k}z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\left[\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_1}e^{-i\frac{2\pi t_1}{T}t}\right]\right\}\right)\right|$$

in order to bound $\mathbb{E}\left|\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m=2}^{k}z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\left[\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_1}e^{-i\frac{2\pi t_1}{T}t}\right]\right\}\right)\right|$, it is necessary to bound its variance.

$$\mathbb{V}\mathrm{ar}\left\{\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m=2}^{k}z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\left[\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_1}e^{-i\frac{2\pi t_1}{T}t}\right]\right\}\right)\right\} =$$
$$\mathbb{E}_T\,\mathbb{E}_T\,\frac{1}{(2\pi)^{2k-2}T^2}\,\mathbb{C}\mathrm{ov}\left\{\prod_{m=2}^{k}z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}),\ \prod_{m=2}^{k}z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m'}{T})z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1'}{T})\right\}$$

with

$$\mathbb{C}\mathrm{ov}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}), \prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\} =$$
$$\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}) \cdot \prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\}$$
$$-\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T})\right\}\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\}$$

and

$$\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}) \cdot \prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\} =$$
$$\sum_{\tau_1,\ldots,\tau_{2k}=0}^{T-1} e^{-i\left(\sum_{m=1}^{k}\frac{2\pi t_m}{T}+\sum_{m=1}^{k}\frac{2\pi t'_m}{T}\right)}\mathbb{E}\left([\boldsymbol{\nu}_{\tau_1}]_{c_1}[\boldsymbol{\nu}_{\tau_2}]_{c_1}\prod_{m=2}^{k}[\boldsymbol{u}_{\tau_{m+1}}]_{c_m}[\boldsymbol{u}_{\tau_{m+k}}]_{c_m}\right)$$

$$\mathbb{E}\left([\boldsymbol{\nu}_{\tau_1}]_{c_1}[\boldsymbol{\nu}_{\tau_2}]_{c_1}\prod_{m=2}^{k}[\boldsymbol{u}_{\tau_{m+1}}]_{c_m}[\boldsymbol{u}_{\tau_{m+k}}]_{c_m}\right) =$$
$$\mathbb{E}\left(\left[\sum_{s=\tau_1}^{\infty}\boldsymbol{\Psi}_s^{(-1)}\boldsymbol{y}_{\tau_1-s}\right]_{c_1}\left[\sum_{s=\tau_2}^{\infty}\boldsymbol{\Psi}_s^{(-1)}\boldsymbol{y}_{\tau_2-s}\right]_{c_1}\prod_{m=2}^{k}\left[\sum_{s=0}^{\infty}\boldsymbol{\Psi}_s^{(-1)}\boldsymbol{y}_{\tau_{m+1}-s}\right]_{c_m}\left[\sum_{s=0}^{\infty}\boldsymbol{\Psi}_s^{(-1)}\boldsymbol{y}_{\tau_{m+k}-s}\right]_{c_m}\right)$$
$$\left|\mathbb{E}\left([\boldsymbol{\nu}_{\tau_1}]_{c_1}[\boldsymbol{\nu}_{\tau_2}]_{c_1}\prod_{m=2}^{k}[\boldsymbol{u}_{\tau_{m+1}}]_{c_m}[\boldsymbol{u}_{\tau_{m+k}}]_{c_m}\right)\right| \leq C|\tau_1|^{1-\eta_0}|\tau_2|^{1-\eta_0}$$

given that the filter $\boldsymbol{\Psi}^{-1}(L,\boldsymbol{\vartheta}_f)$ is absolute summable and that $\mathbb{E}(\boldsymbol{y}_t^{\otimes 2k})<\infty$. Besides, it is easily to show that

$$\left|\mathbb{E}\left([\boldsymbol{\nu}_{\tau_1}]_{c_1}[\boldsymbol{\nu}_{\tau_2}]_{c_1}\prod_{m=2}^{k}[\boldsymbol{u}_{\tau_{m+1}}]_{c_m}[\boldsymbol{u}_{\tau_{m+k}}]_{c_m}\right)\right| \leq \left|\mathbb{E}\left(\prod_{m=1}^{k}[\boldsymbol{u}_{\tau_m}]_{c_m}[\boldsymbol{u}_{\tau_{m+k}}]_{c_m}\right)\right|$$

with strict inequality if $\max\{\tau_1,\tau_2\}\neq 0$.

Then

$$\left|\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}) \cdot \prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\}\right| \leq C\left(T^{2k-2}+T^{2k+2-2\eta_0}\right)$$
$$\left|\mathbb{E}\left\{\prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) \cdot \prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T})\right\}\right| \leq CT^{2k}$$
$$\left|\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}) \cdot \prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\}\right| = o\left(\left|\mathbb{E}\left\{\prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) \cdot \prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T})\right\}\right|\right)$$

where the last is for all $\eta_0 > 1$.

Similarly

$$\left|\mathbb{E}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T})\right\}\right| = o\left(\left|\mathbb{E}\left\{\prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\right\}\right|\right)$$

and

$$\mathbb{C}\mathrm{ov}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t_1}{T}), \prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T}) z_{c_1}^{\boldsymbol{\nu},(T)}(\frac{2\pi t'_1}{T})\right\} = o\left(\mathbb{C}\mathrm{ov}\left\{\prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T}), \prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t'_m}{T})\right\}\right)$$

$$\mathbb{V}\mathrm{ar}\left\{\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m=2}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\left[\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_1} e^{-i\frac{2\pi t_1}{T}t}\right]\right\}\right)\right\} = o\left(\mathbb{V}\mathrm{ar}\left\{\mathbb{E}_T\left(\frac{1}{(2\pi)^{k-1}T}\left\{\prod_{m=1}^{k} z_{c_m}^{\boldsymbol{u},(T)}(\frac{2\pi t_m}{T})\right\}\right)\right\}\right)$$

Therefore

$$\mathbb{E}\left|\sqrt{B_T^{k-1}T}\,\mathbb{E}_T\left\{I_{\boldsymbol{c},k}^{\boldsymbol{u},(T)}\left(\frac{2\pi t_1}{T},\ldots,\frac{2\pi t_k}{T}\right) - I_{\boldsymbol{c},k}^{\tilde{\boldsymbol{u}},(T)}\left(\frac{2\pi t_1}{T},\ldots,\frac{2\pi t_k}{T}\right)\right\}\right| = o(1)$$

Now, one of the elements in $\mathbb{E}_T\boldsymbol{R}$ is

$$\begin{aligned}
\boldsymbol{R}_1 &= \mathbb{E}_T\frac{1}{(2\pi)^{k-1}T}\left\{\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_1} e^{-i\frac{2\pi t_1}{T}t}\right\}\left\{\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_2} e^{-i\frac{2\pi t_2}{T}t}\right\}\left\{\sum_{n=1}^{K}\left\{\sum_{t=0}^{T-1}[\boldsymbol{\xi}_t^{(n)}]_{c_3} e^{-i\frac{2\pi t_3}{T}t}\right\}(\hat{\vartheta}_{f,n}-\vartheta_{f,n})\right\} \\
&= \sum_{n=1}^{K}(\hat{\vartheta}_{f,n}-\vartheta_{f,n})\mathbb{E}_T\frac{1}{(2\pi)^{k-1}T}\left\{\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_1} e^{-i\frac{2\pi t_1}{T}t}\right\}\left\{\sum_{t=0}^{T-1}[\boldsymbol{\nu}_t]_{c_2} e^{-i\frac{2\pi t_2}{T}t}\right\}\left\{\sum_{t=0}^{T-1}[\boldsymbol{\xi}_t^{(n)}]_{c_3} e^{-i\frac{2\pi t_3}{T}t}\right\} \\
&\leq CO_p(T^{-1/2})O(1)
\end{aligned}$$

hence

$$\sqrt{H_T^{k-1}T}\,\mathbb{E}|\boldsymbol{R}_1| \leq CH_T^{\frac{k-1}{2}}$$

The most problematic term in $\boldsymbol{R}$ is

$$\begin{aligned}
\boldsymbol{R}_2 = \sum_{n_1,n_2,n_3}^{K}\prod_{h=1}^{3}(\hat{\vartheta}_{f,n_h}-\vartheta_{f,n_h})\mathbb{E}_T\frac{1}{(2\pi)^{k-1}T}&\left\{\sum_{t=0}^{T-1}\left[\sum_{s=t}^{\infty}\boldsymbol{\Psi}_s^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{y}_{t-s}\right]_{c_1} e^{-i\frac{2\pi t_1}{T}t}\right\}\left\{\sum_{t=0}^{T-1}\left[\sum_{s=t}^{\infty}\boldsymbol{\Psi}_s^{(-1)}(\boldsymbol{\vartheta})\boldsymbol{y}_{t-s}\right]_{c_2} e^{-i\frac{2\pi t_2}{T}t}\right\} \\
&\times\left\{\sum_{t=0}^{T-1}\left[\dddot{\boldsymbol{\Psi}}_{n_1,n_2,n_3}^{(-1)}(L,\tilde{\boldsymbol{\vartheta}}_f)\boldsymbol{y}_t\right]_{c_3} e^{-i\frac{2\pi t_3}{T}t}\right\}
\end{aligned}$$

thus

$$\sup_{\tilde{\boldsymbol{\vartheta}}_f \in \mathcal{V}_f} |\boldsymbol{R}_2| \leq C \left| \prod_{h=1}^{3} (\hat{\vartheta}_{f,n_h} - \vartheta_{f,n_h}) \right|$$

$$\times \sup_{\tilde{\boldsymbol{\vartheta}}_f \in \mathcal{V}_f} \left| \mathbb{E}_T \frac{1}{(2\pi)^{k-1} T} \left\{ \sum_{t=0}^{T-1} [\boldsymbol{\zeta}_t]_{c_1} e^{-i \frac{2\pi t_1}{T} t} \right\} \left\{ \sum_{t=0}^{T-1} [\boldsymbol{\zeta}_t]_{c_2} e^{-i \frac{2\pi t_2}{T} t} \right\} \left\{ \sum_{t=0}^{T-1} \left[\dddot{\boldsymbol{\Psi}}^{(-1)}_{n_1,n_2,n_3}(L, \tilde{\boldsymbol{\vartheta}}_f) \boldsymbol{y}_t \right]_{c_3} e^{-i \frac{2\pi t_3}{T} t}\right.$$

$$\mathbb{E} \sup_{\tilde{\boldsymbol{\vartheta}}_f \in \mathcal{V}_f} |\boldsymbol{R}_2| \leq CO(T^{-3/2}) \times O(T^1) = CO(T^{-1/2})$$

where $[\zeta_t]_{c_1} = \left[\sum_{s=t}^{\infty} \boldsymbol{\Psi}_s^{(-1)}(\boldsymbol{\vartheta}) \boldsymbol{y}_{t-s} \right]_{c_1}$.

Therefore

$$\mathbb{E} \sqrt{H_T^{k-1} T} \left| \hat{g}^{\boldsymbol{u},(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda}) - \hat{g}^{\hat{\boldsymbol{u}},(T)}_{\boldsymbol{c},k}(\boldsymbol{\lambda}) \right| \leq C H_T^{\frac{k-1}{2}}$$

APPENDIX - CHAPTER III

5.14. Proof of Proposition 4.1

Wlog, let assume that $\mathscr{I}_{SR} = \{1, 2, \ldots, d_{SR}\}$, i.e. the first d_{SR} structural innovations are subject to sign-restrictions. Since Assumption 4.2 holds, we know that the unrestricted identified set ($\mathscr{B}_{0,n}$) contains all the signed-permutations of one solution of cumulant conditions, $\boldsymbol{B}_1$; and the cardinality of this set is equal to $2^d \times d!$.

Now, consider the sign restrictions for each structural innovation $i \in \mathscr{I}_{SR}$:

$$\boldsymbol{S}_i \boldsymbol{\Psi}(i)' \boldsymbol{b}_i \geq 0$$

From what we have been discussed above, this problem can be translated into:

$$\boldsymbol{S}_i \tilde{\boldsymbol{\Psi}}(i)' \boldsymbol{q}_i \geq 0$$

where $\tilde{\boldsymbol{\Psi}}(i)' = \boldsymbol{\Psi}(i)' \boldsymbol{B}_1$ and $\boldsymbol{q}_i$ is the i-th column of a signed-permutation matrix $P \in \mathscr{P}(d)$.

Let consider the following elements:

1. First,

$$\tilde{\boldsymbol{\Psi}}'_i = \begin{bmatrix} \tilde{\psi}^i_{11} & \tilde{\psi}^i_{12} & \cdots & \tilde{\psi}^i_{1n} \\ \tilde{\psi}^i_{21} & \tilde{\psi}^i_{22} & \cdots & \tilde{\psi}^i_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\psi}^i_{r_i1} & \tilde{\psi}^i_{r_i2} & \cdots & \tilde{\psi}^i_{r_in} \end{bmatrix}$$

   then $\tilde{\psi}^i_{jl}$ represents the value of the j-th row and l-th column of $\tilde{\boldsymbol{\psi}}'_i$. Since $r_i = 1$, $\tilde{\boldsymbol{\Psi}}'_i$ is row vector. Thus, the sign restriction looks like:

$$s_i \begin{bmatrix} \tilde{\psi}^i_{11} & \tilde{\psi}^i_{12} & \ldots & \tilde{\psi}^i_{1n} \end{bmatrix} \begin{bmatrix} q_{1i} \\ q_{2i} \\ \vdots \\ q_{ni} \end{bmatrix} \geq 0$$

2. $\theta^R_i = \{\boldsymbol{q} \,|\, \boldsymbol{S}_i \tilde{\boldsymbol{\Psi}}'_i \boldsymbol{q} \geq 0, \boldsymbol{q} \in \pi(\boldsymbol{e}^{\pm}_i)\}$ is the set of signed permutations of the elements of the unitarian vector $\boldsymbol{e}_i$ that satisfies the sign restrictions for innovation i. Let $m_i = \#(\theta^R_i)$.

3. Let $\tilde{m}_i$ be an n-vector which j-th entry, $\tilde{m}_{i,j}$, denotes the number of elements of the set $\{\boldsymbol{e}_j, -\boldsymbol{e}_j\}$ that satisfy the sign restriction. Thus, $\tilde{m}_{i,j}$ could take only the values $\{0,1,2\}$. Since $r_i = 1$, it is straightforward to notice that $\tilde{m}_{i,j}$ is equal to 1 if $\tilde{\psi}^i_{1j} \neq 0$, or 2 if $\tilde{\psi}^i_{1j} = 0$. Therefore, $m_i = \sum_{j=1}^n \tilde{m}_{i,j}$.
   For instance, if $\boldsymbol{S}_i \tilde{\boldsymbol{\Psi}}'_i = [a, 0, \ldots, 0]$ with $a > 0$, hence $\theta^R_i = \{\boldsymbol{e}_1, \pm\boldsymbol{e}_2, \cdots \pm \boldsymbol{e}_n\}$, i.e. the sign restriction discriminates only the permutation vectors for the first dimension. In this example, $\tilde{m}_i = [1, 2, \ldots, 2]$. It is easy to check that $\sum_{j=1}^d \tilde{m}_{i,j} = 2d - 1 = \#(\theta^R_i)$. This discussion leads to the following step.

4. Let $\mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i) = [\mathbb{1}(\tilde{\psi}^i_{11} \neq 0), \ldots, \mathbb{1}(\tilde{\psi}^i_{1n} \neq 0)]$ and $\mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i)^c = [\mathbb{1}(\tilde{\psi}^i_{11} = 0), \ldots, \mathbb{1}(\tilde{\psi}^i_{1n} = 0)]$. Then

$$\tilde{m}_i = \mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i) + 2\mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i)^c$$

   since $\mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i) + \mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i)^c = \boldsymbol{\iota}$, where ι is a vector of ones. Hence:

$$\tilde{m}_i = 2!\boldsymbol{\iota} - \mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i)$$

   In our previous example, $\mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i) = [1, 0, \ldots, 0]$, then $2!\boldsymbol{\iota} - \mathbb{1}(\tilde{\boldsymbol{\Psi}}'_i) = [1, 2, \ldots, 2]$, which is $\tilde{m}_i$. Additionally, for any $j \in \complement\mathscr{I}_{SR}$, $\tilde{m}_j = 2\boldsymbol{\iota}$.

5. Once $\{\theta^R_i\}_{i=1}^n$ is determined, we can find the elements of $\mathscr{B}^{SR}_{0,n}$ as the admissible combinations of the elements of $\{\theta^R_i\}_{i=1}^n$. An admissible combination is defined as the n-tuple $\{p_1, \ldots, p_n\}$, with $p_i \in \theta^R_i$, such that $p_i \perp p_j$ for $i \neq j$. Then, $\mathscr{B}^R_{0,n} \subset \times_{i=1}^n \theta^R_i$. For example, if $n = 3$ and $\theta^R_i = \{\boldsymbol{e}_i, -\boldsymbol{e}_i\}$ for all $i = 1, 2, 3$. Then, it is straightforward that the number of admissible combinations is 8. Using the previous defined elements, in this example we have $\tilde{m}_1 = [2, 0, 0]$, $\tilde{m}_2 = [0, 2, 0]$ y $\tilde{m}_3 = [0, 0, 2]$. Then, the total number of admissible combinations is equal to:

$$\tilde{m}_{1,1} \times \tilde{m}_{2,2} \times \tilde{m}_{3,3} + \tilde{m}_{1,1} \times \tilde{m}_{2,3} \times \tilde{m}_{3,2} + \tilde{m}_{1,2} \times \tilde{m}_{2,1} \times \tilde{m}_{3,3} \ldots$$
$$+ \tilde{m}_{1,2} \times \tilde{m}_{2,3} \times \tilde{m}_{3,1} + \tilde{m}_{1,3} \times \tilde{m}_{2,1} \times \tilde{m}_{3,2} + \tilde{m}_{1,3} \times \tilde{m}_{2,2} \times \tilde{m}_{3,1} = 8$$

6. The previous result can be generalized. Let $\mathscr{J}$ be the set of all n-tuples of mutually different indexes $\{1, 2, \ldots, n\}$. For instance, if $n = 3$, then

$$\mathscr{J} = \{(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1)\}$$

That is, $\mathscr{J}$ is the set of all permutations of the set $\{1,\dots,d\}$ of size d. Hence

$$\#(\mathscr{B}_{0,n}^{SR}) = \sum_{a\in\mathscr{J}}\left(\prod_{i=1}^{d}\tilde{m}_{i,a_i}\right) \tag{5.60}$$

where $a = [a_1, a_2, \dots, a_n]$ is a vector of indexes. This formula is equivalent to

$$\#(\mathscr{B}_{0,n}^{SR}) = 2^{d-d_{SR}}\sum_{a\in\mathscr{J}}\left(\prod_{i\in\mathscr{I}_{SR}}\tilde{m}_{i,a_i}\right)$$

adding the convention that if $\mathscr{I}_{SR} = \emptyset$ (unrestricted case), then $\prod_{i\in\mathscr{I}_{SR}}\tilde{m}_{i,a_i} = 1$. Thus, $\#(\mathscr{B}_{0,n}) = 2^d \times d!$.

Now, we can analyze the two extremes cases we can face in the restricted case. First, the least informative case in the unrestricted case occurs when $\tilde{m}_i = \tilde{m}$ for all $i \in \mathscr{I}_{SR}$, where

$$\tilde{m} = \begin{cases} 1, & \text{for one } j \in \{1,\dots,d\} \\ 2, & \text{for all } j' \neq j \end{cases}$$

This represent the case where $\boldsymbol{S}_i\tilde{\boldsymbol{\Psi}}'$ has a unique non-zero entry in the same position j for all $i \in \mathscr{I}_{SR}$. Then, if we apply the formula given above we have that the total number of possible permutations in this least informative case is

$$\begin{aligned}
\#(\mathscr{B}_{0,n}^{SR}) = 2^{d-d_{SR}} &= 2^{d-d_{SR}} \times \left[2^{d_{SR}-1}(d-1)!d_{SR} + 2^{d_{SR}}(d! - (d-1)!d_{SR})\right] \\
&= 2^{d-1} \times [(d-1)!\{d_{SR} + 2(d-d_{SR})\}] \\
&= 2^{d-1} \times [(d-1)!(2d-d_{SR})] = 2^d \times (d-1)!\left(d - \frac{d_{SR}}{2}\right) \\
&= 2^d \times (d-1)!n\left(1 - \frac{d_{SR}}{2d}\right) = 2^d \times d!\left[1 - \frac{d_{SR}}{2d}\right]
\end{aligned}$$

Since we restrict at least one structural innovation, we have $d_{SR} \geq 1$. Thus, $1 - \frac{d_{SR}}{2d}$ is strictly smaller than 1, therefore $\#(\mathscr{B}_{0,n}^{SR}) < \#(\mathscr{B}_{0,n})$.

On the other hand, the most informative case is $\tilde{m}_i = \boldsymbol{\iota}$ for all $i \in \mathscr{I}_{SR}$, where $\boldsymbol{\iota}$ is a vector of ones. Then, the cardinality of the restricted identified set is

$$\#(\mathscr{B}_{0,n}^{SR}) = 2^{d-d_{SR}} \times [1 \times n!] = 2^{d-d_{SR}} \times n!$$

5.15. Proof of Proposition 4.4

Wlog, let consider that $l^\star = 1$, i.e. the sign-restricted shock is the first structural error. The system of inequalities implied by the sign-restrictions is $\varsigma_1^{SR} = \{\boldsymbol{S}_1 \boldsymbol{\Psi}'(1)\boldsymbol{B}_1 \mathbf{x} \geq 0\}$. By hypothesis, $\boldsymbol{\Psi}'(1)\boldsymbol{B}_1$ is an L-matrix, therefore by Lemma 4.1

$$\mathscr{E} \bigcap \varsigma_1^{SR} \neq \emptyset.$$

Now, by condition $\operatorname{rank}\left(\boldsymbol{\Psi}'(1)\boldsymbol{B}_1\right) = r_1 = d$. Implying that $\boldsymbol{\Psi}'(1)\boldsymbol{B}_1$ is full-rank L-matrix. Besides, by hypothesis $\boldsymbol{\Psi}'(1)\boldsymbol{B}_1$ has all entries different from zero. Then, being an L-matrix and being full rank with $r_l = d$ implies that its column and row rank are equal. Therefore, if we assume that $\boldsymbol{S}_1\boldsymbol{\Psi}'(1)\boldsymbol{B}_1$ has more than one mono-signed column, i.e. they would be positive or negative element-wise. However, in such scenario $\boldsymbol{\Psi}'(1)\boldsymbol{B}_1$ would not satisfy the property of being an L-matrix, because there may exists a real-valued matrix in the class $\mathscr{Q}(\boldsymbol{S}_1\boldsymbol{\Psi}'(1)\boldsymbol{B}_1)$ that is not full rank.

Therefore, matrix $\boldsymbol{S}_1\boldsymbol{\Psi}'(1)\boldsymbol{B}_1$ has only one a mono-signed column. Then

$$\left|\mathscr{E} \bigcap \varsigma_1^{SR}\right| = 1.$$

5.16. Proof of Proposition 4.6

When the structural shocks in the system includes more than one Gaussian shocks, i.e. $\operatorname{rank}(\mathbf{v}_3^{\varepsilon}) = d_{ng} \leq d-2$ or $\operatorname{rank}(\mathbf{v}_{34}^{\varepsilon}) = d_{ng} \leq d-2$. For simplicity, we use only second and third order cumulants and wlog assume the skewed non-Gaussian components are the first d_{ng} elements of $\boldsymbol{\varepsilon}_t$, i.e. $\kappa_{3,j}^{\varepsilon} \neq 0$ with $j = 1,\ldots,d_{ng}$. Besides, $\kappa_{3,j}^{\star} \neq 0$ denotes the skewness coefficient of rotated (permuted) structural shocks.

The cumulants conditions are

$$\begin{aligned} q_{ij}\left(\kappa_{3,i}^{\star}\boldsymbol{e}_i - \kappa_{3,j}^{\varepsilon}\boldsymbol{q}_j\right) &= 0, \quad \forall\, i \in \{1,\ldots,d\}, j \in \{1,2,\ldots,d_{ng}\} && (5.61)\\ \kappa_{3,i}^{\star} q_{ij}\boldsymbol{e}_i &= 0, \quad \forall\, i \in \{1,\ldots,d\}, j \in \{d_{ng}+1,\ldots,n\} && (5.62) \end{aligned}$$

Equation 5.61 is equivalent to:

$$\begin{aligned} -q_{ij}\kappa_{3,j}^{\star} q_{sj} &= 0, \quad \forall\, s \neq i, j \in \{1,2,\ldots,d_{ng}\} \\ q_{ij}(\kappa_{3,i}^{\star} - \kappa_{3,j}^{\varepsilon} q_{ij}) &= 0, \quad \forall\, i \in \{1,\ldots,d\}, j \in \{1,2,\ldots,d_{ng}\} && (5.63) \end{aligned}$$

Hence, this discard the possibility of having $q_{ij} \neq 0$ for all i and $j \in \{1,2,\ldots,d_{ng}\}$. This means that the first d_{ng} columns of Q cannot be any column of a orthogonal matrix. Although, if $\boldsymbol{q}_j = \pm\boldsymbol{e}_i$, we have that $\kappa^{\star}_{3,i} = \pm\kappa^{\varepsilon}_{3,j}$ for $j \in \{1,2,\ldots,d_{ng}\}$.

The previous discussion determines d_{ng} of the skewness coefficients of the rotated structural innovations. The question now will be if it is possible to have more than k skewness coefficients different from zero, that is whether or not $\text{rank}(\mathbf{v}_3^{0,\star}) > \text{rank}(\mathbf{v}_3^{0})$. The equation 5.62 implies:

$$\kappa^{\star}_{3,i} q_{ij} = 0, \quad \forall\, i \in \{1,\ldots,n\}, j \in \{d_{ng}+1,\ldots,n\} \tag{5.64}$$

Then, $q_{ij} = 0$ for all $i \in \{1,\ldots,d\}$, $j \in \{d_{ng}+1,\ldots,d\}$. Implying that orthogonality condition of Q is violated. Something similar can be done for discarding the case of having less skewed rotated innovation. Therefore $\text{rank}(\mathbf{v}_3^{\star}) = \text{rank}(\mathbf{v}_3^{\varepsilon})$.

Finally, for a clear exposition of the idea behind the result, let assume $\boldsymbol{q}_j = \pm\boldsymbol{e}_i$ for all $i, j \in \{1,2,\ldots,d_{ng}\}$, that is the first d_{ng} columns of Q have the form $\left[P^{\pm,\prime}_{d_{ng}}, \mathbf{0}_{d_{ng}\times(d-d_{ng})}\right]'$. Now, the orthogonality condition for Q ($Q'Q = \boldsymbol{I}_d$) implies:

$$\boldsymbol{q}_j'\boldsymbol{q}_j = 1, \quad j = 1,2,\ldots,d \tag{5.65}$$

$$\boldsymbol{q}_j'\boldsymbol{q}_s = 0, \quad s \neq j \tag{5.66}$$

This is satisfied for all $\boldsymbol{q}_j$ with $j \in \{1,2,\cdots,d_{ng}\}$. Besides, it implies that the first d_{ng} elements of $\boldsymbol{q}_s$ are zero for all $s \in \{d_{ng}+1,\ldots,d\}$, leaving the $d - d_{ng}$ components unrestricted. Therefore:

$$Q = \begin{bmatrix} P^{\pm}_{d_{ng}} & \mathbf{0}_{d_{ng}\times(d-d_{ng})} \\ \mathbf{0}_{(d-d_{ng})\times d_{ng}} & \tilde{Q}_{d-d_{ng}} \end{bmatrix}$$

with $\tilde{Q}'_{d-d_{ng}}\tilde{Q}_{d-d_{ng}} = \tilde{Q}_{d-d_{ng}}\tilde{Q}'_{d-d_{ng}} = \boldsymbol{I}_{d-d_{ng}}$.

This means that, identification of the non-Gaussian components up to permutations persists, while the Gaussian part will not be identified without further restrictions.

5.16.1. Proof of Lemma 4.1

Given the finiteness of the coefficient matrices of moving-average representation of $\boldsymbol{y}_t$, assured by assumption on the location of roots of AR polynomial.

Part (a) in 4.1 is a direct application of results in Ralph T Rockafellar (1969). Notice $F(\varsigma^{SR}_{l_i})$ is the dual cone of K_i, which is the cone generated by the columns $-\boldsymbol{B}_1'\boldsymbol{\Psi}(l_i)\boldsymbol{S}_{l_i}$, and by definition

$$F(\varsigma_{l_i}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}'\boldsymbol{y} \leq 0 \ \forall \ \boldsymbol{y} \in K_i\}.$$

Part (b) in 4.1 is straightforward. If $\mathbf{x}_1$ and $\mathbf{x}_2$ belongs to $F(\varsigma_{l_i}^{SR})$, then any $\tilde{\mathbf{x}} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, with $\lambda \in (0,1)$ also belongs to $F(\varsigma_{l_i}^{SR})$. Besides, the unbounded-ness follows from the former discussion by taking multiples of $\mathbf{x}_1$, $\mathbf{x}_2$ and $\tilde{\mathbf{x}}$.

Part (c): When the $|\mathscr{I}_{SR}| = 1$, then existence of $F(\varsigma_l^{SR})$ is sufficient for the existence of $\mathscr{B}_2^{Q,SR}$. Because for any $\mathbf{x} \in F(\varsigma_l^{SR})$ it is feasible to find a tuple of $d-1$ components such that it is mutually orthogonal.

When the $|\mathscr{I}_{SR}| = 2$, existence of $F(\varsigma_l^{SR})$ is not sufficient for the existence of the identified set, because the existence of each solution set of the inequality system ς_l^{SR} does not assure the existence of having at least one pair $(\mathbf{x}_{l_1}, \mathbf{x}_{l_2}) \in F(\varsigma_{l_1}^{SR}) \times F(\varsigma_{l_2}^{SR})$ that is mutually orthogonal.

Thus, if identified set $\mathscr{B}_2^{Q,SR}$ exists, this implies that there exists $(\mathbf{x}_{l_1}, \mathbf{x}_{l_2}) \in F(\varsigma_{l_1}^{SR}) \times F(\varsigma_{l_2}^{SR})$ such that $\mathbf{x}_{l_1}'\mathbf{x}_{l_2} = 0$. Then, the Gram-Schmidt algorithm apply to the pair $(\mathbf{x}_{l_1}, \mathbf{x}_{l_2})$ delivers the pair $(\mathbf{x}_{l_1}, \mathbf{x}_{l_2})$.

Now, consider the set $GS\left(F(\varsigma_{l_1}^{SR}) \times F(\varsigma_{l_2}^{SR})\right)$. When $GS\left(F(\varsigma_{l_1}^{SR}) \times F(\varsigma_{l_2}^{SR})\right) \bigcap F(\varsigma_{l_1}^{SR}) \times F(\varsigma_{l_2}^{SR}) \neq \emptyset$, then a pair $(\mathbf{x}_{l_1}, \tilde{\mathbf{x}}_{l_2})$ belongs to $F(\varsigma_{l_1}^{SR}) \times F(\varsigma_{l_2}^{SR})$, i.e. $\mathbf{x}_{l_1} \in F(\varsigma_{l_1}^{SR})$ and $\tilde{\mathbf{x}}_{l_2} \in F(\varsigma_{l_2}^{SR})$. The rest of components can obtained by finding orthogonal vectors to the pair $(\mathbf{x}_{l_1}, \tilde{\mathbf{x}}_{l_2})$, because they are not sign-restricted.

5.16.2. Proof of Lemma 4.2

Provided the existence of solution set $F(\varsigma)$ of the system $\varsigma = \{\boldsymbol{A}'\mathbf{x} \leq 0\}$. If $\boldsymbol{A}'$ is an L-matrix, then for any row sign-changing matrix $\boldsymbol{D}$, $\boldsymbol{D}\boldsymbol{A}'$ has mono-signed column. Besides, the set $\mathscr{E} = \mathscr{E}^+ \cup \mathscr{E}^-$ contains all the vectors that selects a column or changes its sign. Therefore:

$$F(\varsigma) \bigcap \mathscr{E} \neq \emptyset.$$