Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Specification and Causality of Distribution Models

Autor:

Victor Emilio Troster

Director:

Carlos Velasco

Tutor:

Abderrahim Taamouti

DEPARTAMENTO DE ECONOMÍA

Getafe, junio de 2015

Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

SPECIFICATION AND CAUSALITY OF DISTRIBUTION MODELS

Autor: ***Victor Emilio Troster***

Director: **Carlos Velasco**

Codirector: **Abderrahim Taamouti**

Firma del Tribunal Calificador:

Firma

Presidente: Miguel A. Delgado

Vocal: José Olmo

Secretario: Juan Carlos Escanciano

Calificación:

Getafe, de de

Universidad Carlos III de Madrid

Specification and Causality of Distribution Models

Victor Emilio Troster

Thesis submitted to the Department of Economics of the Universidad Carlos III de Madrid for the degree of Doctor of Philosophy, Getafe, June 2015.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the Department of Economics at the Universidad Carlos III de Madrid is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Abstract

Many important economic and finance hypotheses are investigated through testing the specification of restrictions on the conditional distribution of a time series, such as conditional goodness-of-fit (Box and Pierce (1970)), conditional quantiles (Koenker and Machado (1999)), and distributional Granger non-causality (Taamouti, Bouezmarni, and El Ghouch, 2014). This PhD Thesis contributes to the study of specification and causality tests that provide a more flexible and detailed approach to evaluate economic relationships, which are useful in many relevant empirical applications.

In the first chapter, we propose a practical and consistent specification test of conditional distribution models for dependent data in a very general setting. Our approach covers conditional distribution models possibly indexed by function-valued parameters, which allows for a wide range of important empirical applications, such as the linear quantile auto-regressive, the CAViaR, and the distributional regression models. Our test statistic is based on a comparison between the estimated parametric and the empirical distribution functions. The new specification test (i) is valid for general linear and nonlinear dynamic models under parameter estimation error, (ii) allows for dynamic misspecification, (iii) is consistent against fixed alternatives, and (iv) has nontrivial power against $\sqrt{T}$-local alternatives, with T the sample size. As the test statistic is non-pivotal, we propose and theoretically justify a block bootstrap approach to obtain valid inference. Monte Carlo simulations illustrate that the proposed test has good finite sample properties for different data generating processes and sample sizes. Finally, an empirical application to models of Value-at-Risk (VaR) highlights the benefits of our approach.

The second chapter proposes a consistent parametric test of Granger-causality in quantiles. Although the concept of Granger-causality is defined in terms of the conditional distribution, the majority of papers have tested Granger-causality using conditional mean regression models in which the causal relations are linear. Rather than focusing on a single part of the conditional distribution, we develop a test that evaluates nonlinear causalities and possible causal relations in all conditional quantiles. The proposed test statistic has correct asymptotic size, is consistent against fixed alternatives and has power against Pitman deviations from the null hypothesis. The proposed approach allows us to evaluate nonlinear causalities, causal relations in conditional quantiles, and provides

a sufficient condition for Granger-causality when all quantiles are considered. As the proposed test statistic is asymptotically non-pivotal, we tabulate critical values via a subsampling approach. We present Monte Carlo evidence and an application considering the causal relation between the gold price, the USD/GBP exchange rate, and the oil price.

The last chapter of the thesis studies the co-integration relationship between industry stock returns and excess stock market returns, and it is co-authored with Prof José Penalva and Prof Abderrahim Taamouti. We find that the equilibrium error term from this co-integrating relationship has strong predictive power for excess stock returns, which is increased if combined with the previous month's excess stock returns. Our results suggest that short-term return reversals and liquidity measures are primary reasons for the negative relation between the equilibrium error and expected excess stock returns. We provide new evidence on the out-of-sample stock return predictability, in contrast to Welch and Goyal (2008), among others, who found negligible out-of-sample predictive power using standard variables. We also show that the out-of-sample explanatory power is economically meaningful for investors. Simple trading strategies implied by the proposed predictability provide portfolios with higher mean returns and Sharpe ratios than a buy-and-hold or a benchmark strategy does.

Para a minha esposa Elaine,

minha filha Helena

e minha tia Ana María.

Acknowledgements

I thank my dear wife Elaine who supported and motivated me through the whole PhD. I would not have been able to finish my thesis without her. I'd like to thank my aunt Ana María who lives in Madrid for her hospitality, kindness and support. I dedicate this thesis to them and to my lovely daughter Helena. The research presented in this thesis was supported by a scholarship from the Ministerio de Economía y Competitivad (ECO2013-46395). I thank the Department of Economics of Carlos III for providing me the Master's scholarship, which allowed me to study in Spain.

I am very grateful to my advisors Prof Carlos Velasco and Prof Abderrahim Taamouti. They were very helpful and taught me a lot. I was glad to have the opportunity to work with them.

Many other colleagues have actively helped me or commented on earlier versions of the papers. A non-exhaustive list includes: Prof Miguel Delgado, Prof Jesús Gonzalo, Prof Juan Carlos Escanciano, Prof José Penalva, Prof Pedro Gomes. My deep thanks to all of them.

I am also grateful to many of my friends that accompanied me throughout these years. I want to thank my friends at the Carlos III for making my daily life at University much more enjoyable: Cleiton Taufemback, Thomas Kansy, Miang Huang, Paula Apascaritei, Nora Wegner, Peter Eccles, Alejandro Barrachina, Lovleen Kushwah, Iana Zborshchyk, Anibal Emiliano da Silva, Emiliano Tealde, João Henrique, Carlos Gallardo, Sandra Esteve, Marco Serena, Iacopo Morchio, Alessandro Peri, Nazli Aktakke, Federico Curci, Emre Ergemen, Anett Erdmann, Sebastian Panthöfer, Robert Kirby, Xiaojun Song. I also want to thank my friends in Madrid who made my life even nicer here: Luis de la Flor, Anikó Domány, Juan Ferrari, José Luis Hellin, Dani Gordillo, Ignacion Antón, Fr. Iván Calvo, Enrique Pino, Vicente Jara. I thank my German family who received me so well during the Summers: Sarah, Christian, Willhelm, Monica, Sigfried, Angelika, Janina. I would also like to acknowledge some of my friends scattered around the world: Michel da Rocha, Israel Pacheco, Danyel Reimão, Renan Grassi, Marcelo Torres, Rafael Terra.

I want to give special thanks to Antonio Gallardo del Río for his running training

"Carlos Corre" every week. I also thank my friend Prof Francisco Marhuenda and the other participants of Carlos Corre, we really had a good time. Since I am a religious person, I also thank God for having the opportunity to study here. It probably goes without saying that any enduring project such as this, requires the support and understanding of one's family and loved ones. I thank my mother, my father, my family and my second father and mother, Gerson and Isabel. I also thank dear tio Eduardo and tia Macky, who have been really helpful, and vovó Leonor.

Contents

List of Figures

List of Tables

Chapter 1

A Specification Test of Dynamic Conditional Distributions

1.1 Introduction

Many important economic and finance hypotheses are investigated through testing the specification of restrictions on the conditional distribution of a time series, such as conditional goodness-of-fit (Box and Pierce (1970)), conditional quantiles (Koenker and Machado (1999)), and distributional Granger non-causality (Taamouti et al., 2014). After the landmark work of Hausman (1978), numerous authors have developed specification tests under i.i.d. observations. White (1982) proposed a comparison of different variance matrix estimators to detect misspecification of econometric models. Newey (1985) constructed tests of conditional moment restrictions that generalized the approach of Hausman (1978) and White (1982). Although these tests can also be applied in a time series context, none of them is consistent against all possible sources of misspecification. Despite Andrews (1997) developed a consistent test statistic for testing conditional distribution specifications, his approach can be applied only for i.i.d. data.

This paper proposes a practical and consistent specification test of conditional distribution models for dependent data in a very general setting. Our approach covers dynamic conditional distribution models possibly indexed by function-valued parameters. The difference between our approach and that taken elsewhere is motivated within the framework used by Corradi and Swanson (2006) and Rothe and Wied (2013). First, we generalize the approach of Rothe and Wied (2013) to testing the specification of dynamic conditional distribution models indexed by function-valued parameters in contexts with dependent data. This allows for a wide range of models that have been shown to be very useful for risk management and macroeconomic forecasting within a time series framework, such as the linear quantile auto-regressive, the CAViaR, and the distributional regression models. Second, we extend the validity of the block bootstrap for Kolmogorov-type conditional distribution tests proposed by Corradi and Swanson (2006) to the context of dynamic conditional distribution models indexed by function-valued parameters. Rather than analysing models indexed by finite-dimensional parameters as in Corradi and Swanson (2006), we derive a test statistic for conditional distribution models indexed by function-valued parameters that is valid under dynamic misspecification and parameter estimation error. To the best of our knowledge, it has not been developed yet a consistent specification test of conditional distribution models indexed by function-valued parameters under dependent data.

Dynamic misspecification is relevant when a dynamic specification test is developed, as one generally has the problem of defining the relevant past information $\mathcal{F}_{t-1}$ (e.g. how many lags to include), which may involve pre-testing and imply a sequential test bias. There exists dynamic misspecification when the conditional distribution of the variable of interest Y_t given a past information set X_t is not equivalent to the conditional distribution of Y_t given all the "relevant" past information set $\mathcal{F}_{t-1}$ of the conditioning variable, with $X_t \subset \mathcal{F}_{t-1}$, i.e. $Y_t|X_t$ is not equal in distribution as $Y_t|\mathcal{F}_{t-1}$. Bai (2003) developed a Kolmogorov-Smirnov type test of conditional distribution specifications for time series based on the comparison of an estimated conditional distribution function with the distribution function of a uniform on $[0, 1]$. However, Bai (2003)'s test is inconsistent as it cannot detect lag order misspecification of a linear autoregressive model with elliptically distributed innovations (see e.g., Corradi and Swanson, 2006, Delgado and Stute, 2008). Corradi and Swanson (2006) modified the approach of Bai (2003) allowing for dynamic misspecification of the past information set under the null hypothesis. They proposed a consistent test of correct specification for a given information set. In this paper, we extend the approach of Corradi and Swanson (2006) to construct a specification test for time series models that takes into account dynamic misspecification and parameter error estimation effect, in a context of conditional distribution models indexed by function-valued parameters.

Allowing the parameters to be function-valued is important for many empirical applications. For example, our approach covers the linear quantile autoregressive (QAR) of Koenker and Xiao (2006), which implies a linear structure for the inverse of the dynamic conditional distribution $F^{-1}(\tau|\theta_0, \boldsymbol{Y}_{t-p}) = \boldsymbol{Y}_{t-p}'\theta_0(\tau)$, for the quantile $\tau \in (0, 1)$, with $\boldsymbol{Y}_{t-p} = \{Y_{t-1}, \ldots, Y_{t-p}\} \in \mathcal{F}_{t-1}$, and a functional parameter $\theta_0(\tau)$ strictly monotone in τ. Our procedure also considers testing the specification of nonlinear quantile autoregressive models, such as the CAViaR model of Engle and Manganelli (2004), that directly measures the market risk of financial institutions by estimating a particular quantile of future portfolio values - the Value-at-Risk (VaR).

Our proposed test statistic checks the validity of the distributional regression model introduced by Foresi and Peracchi (1995), where the conditional distribution is modeled through a family of binary response models for the event that the variable of interest Y_t

exceeds some threshold $y \in \mathbb{R}$. The distributional regression approach uncovers higher-order multidimensional structure that cannot be found by modeling only the first two moments of the conditional distribution. This has important implications to forecasting excess stock market returns and finding an optimal portfolio (Foresi and Peracchi, 1995). Mean-variance analysis of excess stock market returns works only under special assumptions, like multivariate normality of asset returns or quadratic utility function of investors. In general, a precise definition of risk and an unambiguous ranking of portfolio strategies requires the entire distribution of future returns (Rothschild and Stiglitz, 1970). Besides, focusing on location - for example, on the conditional mean regression - may lead to overlook the impact of certain predictors of excess stock market returns, whose effect is mostly on high-order aspects of the conditional distribution. To the best of our knowledge, we are not aware of a framework to testing for the correct specification of distributional regression models under dependent data.

An additional benefit of our approach is that it permits us to test conditional quantile models over a continuum of quantiles under time series. Koenker and Machado (1999) considered tests for the specification of regression quantile location-scale models for independent observations. Koenker and Xiao (2002) applied the "Khmaladze" transformation to test the specification of linear quantile models under i.i.d data. However, none of these tests are justified for dependent data, and they do not check for the validity of the quantile regression model itself. Whang (2006) proposed a specification test of conditional quantile models for a given quantile τ for time series data, while Escanciano and Velasco (2010) generalized this approach by providing consistent tests of dynamic quantile regression models over a continuum of quantiles under dependent data. Our new test provides a further advantage: it also checks the validity of models for the whole conditional distribution and distributional regression specifications, while the framework Escanciano and Velasco (2010) considers only conditional quantile regression models. Koul and Stute (1999), Neumann and Paparoditis (2008), Andrews (2012), Bierens and Wang (2014), and Kheifets (2015), among others, have also developed consistent specification tests for conditional distribution models for dependent data, but these methods cannot be applied to evaluate models indexed by function-valued parameters. In sum, we believe that our approach is a useful alternative to existing specification methods for dynamic conditional models under dependent data because it allows for models indexed by possibly function-valued parameters, covering the setups of Corradi and Swanson (2006), Escanciano and Velasco (2010), and Rothe and Wied (2013) in a unified way.

Our test statistic is a Cramér-von-Mises (CVM) functional of the discrepancy between the empirical distribution function and a restricted estimate imposing the structure implied by the dynamic conditional distribution model, and we reject the null hypothesis of correct specification if this discrepancy is "large". Since its asymptotic distribution under general time series assumptions is non-pivotal, we propose and justify a block bootstrap resampling scheme to estimate the critical values. This is likely to be computationally intensive, but it delivers a test statistic that (i) allows for robust to dynamic misspecification, (ii) does not require the estimation of smoothing parameters or nuisance functions used in a Khmaladze transformation as in Bai (2003) or in Koenker and Xiao (2002), and (iii) is consistent against all fixed alternatives. Besides, our test statistic has nontrivial power against $\sqrt{T}$-local alternatives, with T the sample size.

As further contributions, we investigate the finite sample performance of our method on simulated data and we illustrate the empirical applicability of our setting by verifying the specification of conditional distribution models for Value-at-Risk (VaR), which is the most used measure of market risk in the financial industry. Using data on two major stock return indexes, we show that our test statistic rejects some widely used specifications of VaR models.

The plan of the paper is as follows. In Section 2, we propose a test statistic for the null hypothesis of correct specification of dynamic conditional distribution models indexed by function-valued parameters under time series and dynamic misspecification. In Section 3, we derive the asymptotic limit distribution of our test statistic under the null and the alternative hypotheses. We also prove that our test statistic has nontrivial power against $\sqrt{T}$-local alternatives, with T the sample size. In Section 4, we theoretically justify the validity of the block bootstrap in our framework. Section 5 provides some examples of conditional distribution and quantile models that are covered by our setting. Section 6 presents Monte Carlo simulation results. In Section 7, we present an empirical application of our proposed test. Finally, Section 8 concludes the paper.

1.2 A General Approach to Testing Dynamic Conditional Distributions

Suppose we observe a sample $\{(Y_t, X_t) \in \mathbb{R} \times \mathbb{R}^d, t = 1, \ldots, T\}$ from a stationary process $\{Y_t, X_t\}_{t=-\infty}^{\infty}$, with joint distribution F_{YX}, where X_t may contain lags of Y_t and/or of other variables. Let $\mathcal{F}_{t-1} := \{X_s\}_{s=-\infty}^{t}$ be the information set including all relevant past information. Let $\mathcal{G}$ be a parametric family of conditional distribution models on the support of Y given X satisfying

$$\mathcal{G} = \left\{F\left(\cdot|\theta, \cdot\right) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta)\right\}, \tag{1.1}$$

where $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ is a function-valued parameter, a class of mappings $\tau \mapsto \theta(\tau)$ such that $\theta(\tau) \in \Theta \subset \mathbb{R}^K$, for each $\tau \in \mathcal{T} \subset \mathbb{R}$. Focusing on the whole information set $\mathcal{F}_{t-1}$, the null hypothesis of correct specification could be written as $F(y|\mathcal{F}_{t-1}) = F(y|\theta_0, \mathcal{F}_{t-1})$, a.s. for all $y \in \mathbb{R}$ and for some $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$, against $\Pr[F(y|\mathcal{F}_{t-1}) \neq F(y|\theta, \mathcal{F}_{t-1})] > 0$

for some $y \in \mathbb{R}$ and for all $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$. Instead, in this paper we are interested in the distribution of Y_t given a finite dimensional vector of conditioning variables $X_t \in \mathbb{R}^d$, for $X_t \subset \mathcal{F}_{t-1}$. If $Y_t|\mathcal{F}_{t-1}$ is not equal in distribution to $Y_t|X_t$, then X_t is dynamically misspecified. However, in empirical applications we do not know a priori what is the "relevant" past information set $\mathcal{F}_{t-1}$, and finding out how much information to include may involve pre-testing (Corradi and Swanson, 2006). Moreover, the critical values for specification tests obtained under the under correct specification given $\mathcal{F}_{t-1}$ are not in general valid in the case of correct specification given X_t, for $X_t \subset \mathcal{F}_{t-1}$. Thus, we allow for dynamic misspecification of X_t and even in the presence of it, we obtain an asymptotically consistent test statistic for the correct specification of Y_t given X_t. Therefore, we want to test null hypotheses of correct specification of conditional distribution models of the form

$$\mathcal{H}_0: \; F(y|x) = F\left(y|\theta_0, x\right), \text{for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \text{ and for all } (y,x) \in \mathcal{W}, \tag{1.2}$$

versus

$$\mathcal{H}_A: \; F(y|x) \neq F\left(y|\theta, x\right), \text{for some } (y,x) \in \mathcal{W} \text{ and for all } \theta \in \mathcal{B}(\mathcal{T}, \Theta), \tag{1.3}$$

where $\mathcal{W}$ is the support of $W_t := (Y_t, X_t')'$. Under the null hypothesis of (1.2), the functional parameter $\theta_0(\cdot)$ is identified through a sequence of moment equalities. Let $\psi : \mathcal{W} \times \Theta \times \mathcal{T} \mapsto \mathbb{R}^K$ be a uniformly integrable function. For every $\tau \in \mathcal{T}$, we assume that the function-valued parameter $\theta_0(\tau)$ solves

$$\Psi(\theta_0, \tau) := \mathrm{E}\left[\psi(W_t, \theta_0, \tau)\right] = 0, \tag{1.4}$$

where $\Psi(\theta, \tau)$ is a function $\Psi : \Theta \times \mathcal{T} \mapsto \mathbb{R}^K$ that fulfills some regularity conditions described in Section 3. As in Rothe and Wied (2013), we assume that under the null hypothesis, any $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ satisfying $F(y|x) = F\left(y|\theta, x\right)$ for all $(y,x) \in \mathcal{W}$ also satisfies $\theta(\tau) = \theta_0(\tau)$, for all $\tau \in \mathcal{T}$. Thus, $\theta_0(\tau)$ is uniquely identified through the moment conditions (1.4). In this paper, we assume that under $\mathcal{H}_A$ in equation (1.3), there exists a "pseudo"-true functional parameter $\theta_1(\tau)$ solving the moment conditions (1.4), for each $\tau \in \mathcal{T}$. Chernozhukov, Fernández-Val, and Melly (2013) developed theoretical results for Z-estimators of the moment conditions of (1.4) for i.i.d. data. Rothe and Wied (2013) show that a large class of empirically relevant specifications fits into this framework in a context with i.i.d. data. We provide conditions for the estimation of function-valued parameters in a context of dependent observations in Section 3.

To test $\mathcal{H}_0$ defined in equation (1.2), we first restate our null hypothesis into an equality of unconditional distributions by integrating-up both sides of $\mathcal{H}_0$ with respect to the marginal distribution of the conditioning variable F_X; see Theorem 16.10 (iii) in Billingsley (1995). We emphasize that the idea of comparing the unrestricted and restricted joint distribution functions, under the null and the alternative, is more than twenty years old in the specification testing literature. Stute (1997) and Andrews (1997) apply this idea in the context of testing specifications of parametric conditional expectation functions and conditional distribution functions, respectively, under i.i.d. data. In a time series context, Corradi and Swanson (2006) and Neumann and Paparoditis (2008) also apply this method to consistently check for the correct specification of dynamic conditional distributions indexed by finite-dimensional parameters. However, our null hypothesis tests the validity of a conditional distributional model indexed by function-valued parameters. As $F(y|x) = E(\mathbb{1}\{Y_t \leq y\}|X_t = x)$, where $\mathbb{1}\{A\}$ is the indicator function of the event A, the null hypothesis $\mathcal{H}_0$ of (1.2) can be equivalently restated as

$$\int F(y|\bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x}) = \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x}),$$
$$\text{for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \text{ and for all } (y, x) \in \mathcal{W},$$

where $F_{YX}(y, x) := \int F(y|\bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ is the unconditional joint distribution function, and $F(y, x, \theta_0) := \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ is the unconditional distribution function implied by the parametric conditional distribution model. Let $\hat{Z}_T(y, x)$ and $\hat{F}_T(y, x, \hat{\theta}_T)$ be the joint empirical distribution function and the semi-parametric estimated distribution function of $\{Y_t, X_t\}_{t=1}^T$ respectively,

$$\hat{Z}_T(y, x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{Y_t \leq y\}\mathbb{1}\{X_t \leq x\}, \text{ for } (y, x) \in \mathbb{R}^{1+d}, \tag{1.5}$$

and

$$\hat{F}_T\big(y, x, \hat{\theta}_T\big) = \int F\big(y|\hat{\theta}_T, \bar{x}\big)\mathbb{1}\{\bar{x} \leq x\}d\hat{F}_X(\bar{x}), \text{ for } (y, x) \in \mathbb{R}^{1+d}, \tag{1.6}$$

where $\hat{F}_X(x)$ is the empirical distribution function of $\{X_t\}_{t=1}^T$,

$$\hat{F}_X(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{X_t \leq x\}, \text{ for } x \in \mathbb{R}^d. \tag{1.7}$$

Under $\mathcal{H}_0$ of (1.2), we assume there is a $\sqrt{T}$-consistent estimator $\hat{\theta}_T(\tau)$ of $\theta_0(\tau)$, for each $\tau \in \mathcal{T}$, that minimizes the empirical analog $\hat{\Psi}_T(\hat{\theta}_T, \tau)$ of the moment conditions in (1.4):

$$\left\|\hat{\Psi}_T(\hat{\theta}_T,\tau)\right\|^2 \leq \inf_{\theta\in\Theta}\left\|\hat{\Psi}_T(\theta,\tau)\right\|^2 + \hat{u}(\tau)^2, \tag{1.8}$$

where $\|\hat{u}\|_{\mathcal{T}} = o_P(T^{-1/2})$, and $\|\bullet\|$ denotes the supremum norm. Our proposed test statistic of $\mathcal{H}_0$ is the functional norm of the distance between $\hat{Z}_T(y,x)$ and $\hat{F}_T\big(y,x,\hat{\theta}_T\big)$, similar to the approach of Andrews (1997) and Rothe and Wied (2013). To this purpose we consider

$$D_T(y,x) = \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbb{1}\{Y_t \leq y\} - F\big(y|\hat{\theta}_T, X_t\big)\Big)\,\mathbb{1}\{X_t \leq x\}, \tag{1.9}$$

and to test the null hypothesis $\mathcal{H}_0$ we propose a T-scaled Cramér-von Mises functional norm of $D_T(y,x)$:

$$S_T = T\int_{\mathcal{W}}\big(D_T(y,x)\big)^2 d\hat{Z}_T(y,x). \tag{1.10}$$

The test statistic S_T should be small if the null hypothesis is correct, while "large" values of S_T imply the rejection of $\mathcal{H}_0$ in (1.2). In Section 3, we develop an asymptotic theory that covers the case of serial dependence, extending the analysis of Rothe and Wied (2013) for S_T to the specification of time series models and the approach of Corradi and Swanson (2006) to specification testing under dynamic misspecification for function-valued parameter models. It is possible to apply other functional norms to $D_T(y,x)$, such as the Kolmogorov-Smirnov functional norm: $\sqrt{T}\sup_{(y,x)\in\mathcal{W}}|D_T(y,x)|$. However, unreported simulations suggested that the S_T test statistic outperforms in terms of size and power other alternative functionals such as the Kolmogorov-Smirnov. Therefore, we focus in this paper on S_T.

1.3 Asymptotic Theory

In this section, we derive the asymptotic distributions of our test statistic S_T under the null and alternative hypothesis. Let $\{Y_{Tt} : t \leq T, T = 1,2,\ldots\}$ be a triangular array with stationary rows of random variables defined on a complete probability space $(\Omega,\mathcal{A},P)$, where T is the sample size. Let $\mathcal{A}_T(m)$ be the σ-field generated by Y_{Tt} for $t \leq m$, and $\mathcal{B}_T(m+d)$ be the σ-field generated by the variables Y_{Tt} for $t \geq m+d$. The sequence $\{Y_{Tt}\}$ is α-mixing if there is a sequence of numbers $\{\alpha(d)\}$ converging to zero for which

$$|\Pr(AB) - \Pr(A)\Pr(B)| \leq \alpha(d), \text{ for all } A \in \mathcal{A}_T(m), \text{ all } B \in \mathcal{B}_T(m+d), \text{ all } m,d,T.$$

Let $\mathcal{W}$ be the support of $W_t := (Y_t, X_t')'$ and $\mathcal{T} \subset \mathbb{R}$. Our test statistic S_T in (1.10) is based on an empirical process indexed by a class of functions $\ell^\infty(\mathcal{H})$, which is the class of real-valued functions that are uniformly bounded on $\mathcal{H}$, with $\mathcal{H} := \mathcal{W} \times \mathcal{T}$, equipped with the supremum norm $||\bullet||_{\ell^\infty(\mathcal{H})}$. To simplify notation, we use $||\bullet||$ to denote the supremum norm. The class $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a permissible class of functions that has a finite and integrable envelope function $\mathbb{F}(\theta, \tau) := \sup_{\Psi \in \mathcal{M}} |\Psi(\theta, \tau)|$ and can be covered by a finite number of elements, not necessarily in $\mathcal{M}$ (see the Appendix for more details). Let $Pf = \int f(\theta, \tau) dP(\theta, \tau)$, for $f \in \mathcal{M}$. Finally, the $\mathcal{M}$ class of functions is assumed in this paper to form a so-called Vapnik-Chervonenkis (VC) class of functions (see Dudley, 1978, Pollard, 1984).

Throughout the paper we use "$\xrightarrow{d}$" and "$\Longrightarrow$" to denote convergence in distribution of random variables and weak convergence of stochastic processes, respectively. We write $Z_T \Longrightarrow Z$ in $\ell^\infty(\mathcal{H})$ to denote weak convergence of a stochastic process Z_T to a random element Z in the function space $\ell^\infty(\mathcal{H})$ (in the Hoffmann-Jørgensen sense, according to Alexander, 1987) for the metric induced by $||\bullet||$. Let $B_\varepsilon(\theta)$ be a closed ball of radius ε centered at θ. All limits are taken as $T \to \infty$, where T is the sample size. We maintain the following main assumptions to analyse the asymptotic behavior of our test statistic:

Assumption 1. *$\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \ldots\}$ is an α-mixing triangular array with stationary rows, satisfying $E(|Y_{1,1}|^{2+\gamma}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\gamma/(4+\gamma)} < \infty$ for some $\gamma \in (0, 2)$.*

Assumption 2. *The parametric space Θ is compact in $\mathbb{R}^K$ and $\mathcal{T}$ is a compact set of some metric space.*

Assumption 3. *For each $\tau \in \mathcal{T}$, $\Psi(\theta, \tau) : \Theta \mapsto \mathbb{R}^K$ possess a unique zero at $\theta_0(\tau)$, and for some $\varepsilon > 0$, $\bigcup_{\tau \in \mathcal{T}} B_\varepsilon(\theta_0(\tau))$ is a compact subset of $\mathbb{R}^K$ contained in Θ. Moreover, the class of functions $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a permissible and VC class of measurable functions with a square integrable envelope function $\mathbb{F}$ satisfying $P(\mathbb{F})^p < \infty$, for $2 < p < \infty$.*

Assumption 4. *Let $\mathcal{I}$ be an open set containing $\mathcal{T}$. The mapping $\Psi(\theta, \tau) : \Theta \times \mathcal{I} \mapsto \mathbb{R}^K$ is continuous and $\theta \mapsto \Psi(\theta, \tau)$ is the gradient of a convex function in θ for each $\tau \in \mathcal{T}$. Besides, $\frac{\partial}{\partial \theta} \Psi(\theta, \tau) := \dot{\Psi}_{\theta,\tau}$ exists at $(\theta_0(\tau), \tau)$ and is continuous at $(\theta_0(\tau), \tau)$, for each $\tau \in \mathcal{T}$, with $\inf_{\tau \in \mathcal{T}} \inf_{\|h\|=1} ||\dot{\Psi}_{\theta_0,\tau} h|| > c_0 > 0$.*

Assumption 5. *For each $\tau \in \mathcal{T}$, the map $\theta \mapsto F(\bullet|\theta, \bullet)$ is Hadamard differentiable at all $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ with derivative $h \mapsto \dot{F}(\bullet|\theta, \bullet)[h]$.*

Assumption 1 is needed to restrict the dependence of $\{Y_{Tt}, X_{Tt}\}$ and holds for many relevant econometric models in practice, including ARMA and GARCH processes under mild additional assumptions; see e.g. Carrasco and Chen (2002). It enables us to establish weak convergence of the empirical process $Z_T(y, x)$ under a variety of situations, see Theorem 7.2 in Rio (2000). Assumptions 2-4 provide conditions to guarantee that a functional central limit theorem holds to the Z-estimator process $\tau \mapsto \sqrt{T}(\hat{\theta}_T(\tau) - \theta_0(\tau))$ for strong mixing processes. Assumption 5 is a smoothness condition required to establish a functional delta-method for the bootstrap of our test statistic (see Theorem 3.9.11 in Van der Vaart and Wellner, 2000).

In comparison with the framework of Rothe and Wied (2013), we need to impose Assumption 1 to establish the asymptotic theory of our test statistic under dependence, while this assumption is not needed in contexts with independent data. In addition, Assumption 4 requires that the class of functions $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a permissible and VC class of measurable functions, while Rothe and Wied (2013) work with Donsker class of functions in an i.i.d. setting. Assumptions 1-5 imply the following theorem, which describes the limit distribution of the proposed test statistic S_T under the null and the alternative.

Theorem 1. *Under Assumptions 1-5, the following hold:*

(i) Under the null hypothesis $\mathcal{H}_0$ in (1.2),

$$S_T \xrightarrow{d} \int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x))^2 \, dF_{YX}(y, x),$$

where $(\mathbb{H}_1, \mathbb{H}_2)$ follow a tight mean zero Gaussian process.

(ii) Under the alternative hypothesis $\mathcal{H}_A$ in (1.3), there exists an $\varepsilon > 0$ such that

$$\lim_{T \to \infty} \Pr(S_T > \varepsilon) = 1.$$

Theorem 1 shows that the asymptotic null distribution of S_T is a functional of the zero-mean Gaussian processes $(\mathbb{H}_1, \mathbb{H}_2)$. By Theorem 1, we expect that S_T is significantly positive whenever the null hypothesis $\mathcal{H}_0$ is violated. However, the asymptotic distribution of S_T varies with the conditional distribution model, the parameter $\theta_0(\cdot)$, and with the serial dependence in the data. As a result, S_T is not asymptotically pivotal and we cannot tabulate critical values. Since $\hat{Z}_T(y, x)$ is an integrating measure on $\mathcal{W}$ depending on T and on data, $\hat{Z}_T(y, x) \implies F_{YX}(y, x)$ in $\ell^\infty(\mathcal{W})$, as T goes to infinity (see Lemma A.1 in the Appendix). In Section 4, we justify a block bootstrap approach that provides critical values for S_T and does not require the estimation of nuisance functions.

1.3.1 Local Power of the Test Statistic

Now we analyze the asymptotic power of S_T against a sequence of Pitman's local alternatives converging to the null hypothesis at rate $\sqrt{T}$, where T denotes the sample size. Let $J(\cdot|\cdot)$ be an alternative conditional distribution function such that $J(\cdot|\cdot) \notin \mathcal{G}$ of (1.1). For any $0 < \delta \leq \sqrt{T}$, we consider that under a sequence of local alternatives $\mathcal{H}_{A,T}$ the data are distributed accordingly to the following conditional distribution

$$\mathcal{H}_{A,T} : F_T\left(y|x\right) = \left(1 - \frac{\delta}{\sqrt{T}}\right) F\left(y|\theta_0, x\right) + \left(\frac{\delta}{\sqrt{T}}\right) J(y|x), \tag{1.11}$$

for all $(y, x) \in \mathcal{W}$ and for some $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$. To ensure nontrivial local power of our test statistic, we make the following assumption:

Assumption 6. *Under the local alternative in (1.11), the conditional distribution under the local alternative in (1.11) implies a sequence of distribution functions* $F_T^A\big(y, x\big) = \int F_T\big(y|\bar{x}\big)\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ *that is contiguous to the distribution function* $F\big(y, x, \theta_0\big) = \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ *based on* $F(y|\theta_0, X_t)$.

Assumption 6 is standard in the study of the asymptotic power under a sequence of Pitman's local alternatives. Andrews (1997) shows that when $F\left(\cdot|\theta_0, \cdot\right)$ and $J(\cdot|\cdot)$ have density functions $f\left(\cdot|\theta_0, \cdot\right)$ and $j(\cdot|\cdot)$ with respect to the same σ-finite measure, then a sufficient condition for Assumption 6 is

$$\sup_{(y,x):f(y|\theta_0,x)>0} \frac{j(y|x)}{f(y|\theta_0, x)} < \infty.$$

Let $\Psi_J(\theta, \tau) := \mathrm{E}_J[\psi(W_t, \theta, \tau)]$ and $\Psi_F(\theta, \tau) := \mathrm{E}_F[\psi(W_t, \theta, \tau)]$, where $\mathrm{E}_J[\cdot]$ and $\mathrm{E}_F[\cdot]$ denote expectation w.r.t. $J = J(y|X_t)$ and $F = F(y|\theta_0, X_t)$, respectively in (1.11). We consider $\theta_0(\cdot)$ and $\theta_1(\cdot)$ as solutions to

$$\Psi_F(\theta_0, \tau) = 0, \tag{1.12}$$

and

$$\Psi_J(\theta_1, \tau) = 0, \tag{1.13}$$

for all $\tau \in \mathcal{T}$ respectively. Let $\frac{\partial}{\partial\theta}\Psi_F(\theta_0, \tau)$ satisfy Assumption 4 for the functional parameter θ_0 solving the moment conditions in (1.12). The following theorem sheds light on the asymptotic power of the test statistic S_T under a sequence of local alternatives satisfying (1.11).

Theorem 2. *Under the local alternative $\mathcal{H}_{A,T}$ in (1.11) and Assumptions 1-6*

$$S_T \xrightarrow{d} \int \left(\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x) + \Delta(y,x)\right)^2 dF_{YX}(y,x),$$

with $\Delta(y,x) = \delta \int (J(y|\bar{x}) - F(y|\theta_0,\bar{x}) + \dot{F}(y|\theta_0,\bar{x})[h])\mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x})$*, and h is the function* $h(\tau) = [\frac{\partial}{\partial\theta}\Psi_F(\theta_0,\tau)]^{-1}\Psi_J(\theta_0,\tau)$.

Theorem 2 implies that the test statistic S_T has non-trivial local power when $\Delta(y,x) \neq 0$. Note that the choice of θ_0 affects the asymptotic power, since $\Delta(y,x)$ is a function of θ_0. This follows because we cannot choose θ_0 under the local alternatives, and θ_1 corresponds to the value that makes $J(\cdot|\cdot)$ as "close" as possible to $F(\cdot|\theta_0,\cdot)$ in the sense of the Kullback-Leibler information distance (Andrews, 1997). For a functional parameter θ_1 solving (1.13), we may choose $F(\cdot|\theta_1,\cdot)$ as the probability limit under J to which the sequence of local alternatives $F_T(\cdot|\cdot)$ shrinks as the sample size grows. Then $[\frac{\partial}{\partial\theta}\Psi_F(\theta_0,\tau)]^{-1}\Psi_J(\theta_0,\tau) = 0$, and we have a simpler drift term

$$\Delta(y,x) = \delta \int (J(y|\bar{x}) - F(y|\theta_0,\bar{x}))\mathbb{1}(\bar{x} \leq x) dF_X(\bar{x}).$$

1.4 Bootstrap Tests

As the test statistic S_T has an asymptotic distribution under $\mathcal{H}_0$ that depends on the data-generating process, we propose a block bootstrap approach to obtain critical values. We also derive its asymptotic properties under the null and alternative hypothesis. If there were no dynamic misspecification under $\mathcal{H}_0$ of (1.2), we could apply a parametric bootstrap resampling method on $\hat{F}_T\big(y,x,\theta_0(\cdot)\big)$ to get asymptotic critical values under the null. However, in the presence of dynamic misspecification, $\hat{F}_T\big(y,x,\theta_0(\cdot)\big)$ is not independent and the covariance structure of the bootstrap statistic is not asymptotically valid. Thus, to solve this problem, we extend the block bootstrap approach proposed by Corradi and Swanson (2006) to test the specification of conditional distribution models indexed by function-valued parameters. We compare the empirical distribution of the resampled series, evaluated at the bootstrap estimator, with the empirical distribution of the actual series, evaluated at the estimator based on the actual data. This resampling method that takes into account the parameter estimation error effect and allows for dynamic misspecification.

We could consider a subsampling approach, for which similar asymptotic results can be shown to hold as well, see e.g. Chernozhukov and Fernández-Val (2005). However, we choose a block bootstrap because we expect it to have more power asymptotically and in finite samples. The block bootstrap is a resampling method with replacement extended to time series observations. It consists of splitting the data into consecutive blocks of observations with length ℓ - $(X_t, X_{t+1}, \ldots, X_{t+l-\ell})$ - and resampling the blocks with replacement from all blocks and joining them to create a bootstrap sample; for a review of block bootstrap and other resampling methods for dependent data, see Kreiss and Paparoditis (2011). Although the block bootstrap is computationally demanding, the estimated asymptotic critical values are consistent against fixed alternatives and allow for dynamic misspecification.

Block bootstrap approaches differ on whether the blocks are overlapping or non-overlapping and on whether the length of the blocks is deterministic or random. We apply a block bootstrap with an overlapping block length - since it is more efficient than the non-overlapping one - and with non-random block length, which has a smaller first order variance (Lahiri, 1999). In what follows, $P^*, E^*, F^*, \ldots$ denote probability laws, expectations, distribution functions, etc. in the block bootstrap, i.e., conditionally on the observed data. The algorithm for computing a fixed block bootstrap realization of our test statistic S_T has the following steps.

1. Let ℓ be the length of the block, $\ell \in \mathbb{N}$, $\ell << T$, where T is the sample size. At each replication, we draw $\mathbf{b}$ blocks of length ℓ from the sample $W_t = (Y_t, X_t)$, with $\mathbf{b} = [T/\ell]$. For example, for some i with probability $1/(T - \ell - 1)$, the i-th block is $W_{i+1}, W_{i+2}, \ldots, W_{i+\ell}$. Thus, the set of starting indexes of the selected blocks is described by $I_1, I_2, \ldots, I_{\mathbf{b}}$ discrete i.i.d. uniform random variables taking values in the set $\{1, 2, \ldots T - \ell\}$.

2. Conditional on the sample, we join together the uniform i.i.d. random $\mathbf{b}$ blocks to form a resampled series $W_1^*, W_2^*, \ldots, W_\ell^*, W_{\ell+1}^*, \ldots, W_T^*$, that can also be written as

$$\underbrace{W_{I_1}, W_{I_1+1}, \ldots, W_{I_1+\ell-1}}_{1^{st}\text{ block}}, \underbrace{W_{I_2}, W_{I_2+1}, \ldots, W_{I_2+\ell-1}}_{2^{nd}\text{ block}}, \ldots, \underbrace{W_{I_{\mathbf{b}}}, W_{I_{\mathbf{b}}+1}, \ldots, W_{I_{\mathbf{b}}+\ell-1}}_{\mathbf{b}^{th}\text{ block}}.$$

3. We denote $\hat{\theta}_T^*$ as the estimator obtained using the block bootstrap resampled series $\{W_t^* = (Y_t^*, X_t^*)\}$. Let $\hat{Z}_T^*(y, x)$ and $\hat{F}_T^*(y, x, \hat{\theta}_T^*)$ be the bootstrap equivalents of $\hat{Z}_T(y, x)$ and $\hat{F}_T(y, x, \hat{\theta}_T)$, respectively. Then we obtain the following re-centered bootstrap statistic S_T^*:

$$S_T^* = \sum_{t=1}^{T} \left[\left(\hat{Z}_T^*(Y_t, X_t) - \hat{F}_T^*(Y_t, X_t, \hat{\theta}_T^*) \right) - \left(\hat{Z}_T(Y_t, X_t) - \hat{F}_T(Y_t, X_t, \hat{\theta}_T) \right) \right]^2.$$

Given a significance level $\alpha \in (0, 1)$, our test rejects $\mathcal{H}_0$ if $S_T > c_T^*(\alpha)$, where the bootstrap critical value $c_T^*(\alpha)$ is the lowest value that satisfies $\Pr^* [S_T^* \leq c_T^*(\alpha)] \geq 1 - \alpha$, and this is estimated through Monte Carlo simulations. In contrast to the block bootstrap statistic of Corradi and Swanson (2006), we deal with the convergence of empirical process indexed by function-valued parameters. Thus, to justify theoretically the block bootstrap resampling in our setting, we need an additional assumption on the serial dependence on the data. We define the k-th beta mixing coefficient $\beta(k)$ as

$$\beta(k) = \frac{1}{2} \sup \sum_{(i,j) \in I \times J} |\Pr(A_i \cap B_j) - \Pr(A_i)\Pr(B_j)|,$$

where the supremum is taken over all finite measurable partitions $\{A_i\}_{i \in I}$ and $\{B_j\}_{j \in J}$ with $A_i \in \sigma(Y_m : m \leq 1)$ and $B_j \in \sigma(Y_m : m \geq 1 + k)$. We say that a sequence $\{Y_t\}$ is beta mixing if $\lim_{k \to \infty} \beta_k \to 0$. Then we impose the following assumption.

Assumption 7. *$\{Y_{Tt}, X_{Tt}, t \leq T, T \geq 1\}$ is a β-mixing triangular array with stationary rows and β-mixing coefficients satisfying*

$$\Gamma(\{\beta_k\}_{k \geq T}) \to 0, \ \text{as } T \to \infty,$$

where $\Gamma : \mathbb{R}^\infty \mapsto \mathbb{R}$ is a monotone mapping such that $a_i \leq b_i$ for $i \geq 0$ implies $\Gamma(\{a_i\}_{i \geq 0}) \leq \Gamma(\{b_i\}_{i \geq 0})$.

Assumption 7 generalizes most of the commonly used mixing conditions in time series processes. Let $P^*(\cdot)$ be the probability law in the block bootstrap, i.e., conditionally on the observed data. We follow the approach of Radulović (1996), which delivers a Block Bootstrap Central Limit Theorem for the class of M-estimators (see Theorem 2 in Radulović (1996)), and justify the block bootstrap approach for our proposed test statistic in the following theorem.

Theorem 3. *Under Assumptions 2-7, let $W_1^*, \ldots, W_T^*$ be generated according to the block bootstrap with block size $\ell := \ell(T)$, with $\ell(T) \to \infty$ as $T \to \infty$, conditional on the data $W_1, \ldots, W_T$. Let $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ be a permissible VC class of measurable functions with a square integrable envelope function $\mathbb{F}$. If we also assume:*

(i) $\limsup_{k \to \infty} k^q \beta(k) < \infty$, *for some $q > p/(p-2)$, for $2 < p < \infty$ such that $P^*(\mathbb{F})^p < \infty$, and*

(ii) $\ell(T) = O(T^\rho)$ *for some $0 < \rho < (p-2)/[2(p-1)]$,*

then:

(i) Under the null hypothesis $\mathcal{H}_0$ of (1.2),

$$\Pr\left(S_T > c_T^*(\alpha)\right) \to \alpha.$$

(ii) Under the fixed alternative hypothesis $\mathcal{H}_A$ of (1.3),

$$\Pr\left(S_T > c_T^*(\alpha)\right) \to 1.$$

(iii) Under the local alternative $\mathcal{H}_{A,T}$ of (1.11),

$$\lim_{T\to\infty} \Pr\left(S_T > c_T^*(\alpha)\right) \geq \alpha,$$

where equality holds when $\Delta(y, x) \equiv 0$ a.e., with $\Delta(y, x)$ the non-trivial shift function defined in Theorem 2.

Theorem 3 is an application of the functional delta method for bootstrap. It shows that our test based on the block bootstrap critical value has asymptotically correct size, is consistent, and is able to detect alternatives tending to the null at the parametric rate $\sqrt{T}$. Bradley (1985) showed that $P^*(\mathbb{F})^p < \infty$ and $\sum_{k=1}^{\infty} \beta(k)^{p/(2-p)}$ for some $p > 2$ is close to the weakest sufficient conditions for an original (non-bootstrap) central limit theorem for empirical processes for VC-subgraph classes of functions. As the optimal length, in terms of bias squared and variance of the block bootstrap approximation, is $\ell = CT^{1/3}$, for a constant $C > 0$ (see Künsch, 1989, Remark 3.3), the condition on the block length is not too restrictive.

1.5 Examples

In this section, we consider certain dynamic conditional distribution models whose specification can be analysed using our approach. We choose those models since they can be used in many relevant empirical applications.

1.5.1 Linear Quantile Autoregressive Processes

Under our approach, it is possible to test conditional quantile models over a continuum of quantiles under time series. Koenker and Machado (1999) and Koenker and Xiao (2002) considered tests for the specification of regression quantile location-scale models and linear quantile models under i.i.d data. However, none of these tests are justified for dependent data, and they do not check for the validity of the quantile regression model itself.

Whang (2006) proposed a specification test of conditional quantile models for a given quantile τ for time series data, while Escanciano and Velasco (2010) generalized this approach by providing consistent tests of dynamic quantile regression models over a continuum of quantiles under dependent data. Our method is complementary to Escanciano and Velasco (2010), since we provide a consistent test statistic for dynamic conditional quantile models, over a continuum of quantiles. Rather than assuming a martingale difference hypothesis and applying a subsampling resampling scheme, our method allows for testing non-markovian quantile models and uses the information of the whole sample under the block bootstrap method. Moreover, our test is consistent under dynamic misspecification. We present some comparisons with the approach of Escanciano and Velasco (2010) on Section 6.

Many papers in the literature deal with the linear quantile autoregression model, see for example Weiss (1991), Koul and Mukherjee (1994), and Hallin and Jurečková (1999). In the linear quantile autoregression model, the τ-quantile of $Y_t|X_t$ is a linear function of X_t, where X_t can take the lagged values of Y_t as arguments. Koenker and Xiao (2006) investigated quantile autoregressive models in which all of the autoregressive coefficients are τ-dependent and able to change the location, scale, and shape of the conditional densities, provided that the τ-conditional quantile of Y_t is monotone in τ. For example, the quantile autoregression (QAR) of order p of Koenker and Xiao (2006) can be written as

$$\begin{aligned} Q_\tau(Y_t|Y_{t-1},\ldots,Y_{t-p}) &= \theta_0(\tau)+\theta_1(\tau)Y_{t-1}+\ldots+\theta_p(\tau)Y_{t-p} \\ &= X_t'\theta(\tau), \text{ for some } \theta\in\mathcal{B}(\mathcal{T},\Theta), \end{aligned} \tag{1.14}$$

where $F^{-1}(\tau|Y_{t-1},\ldots,Y_{t-p},\theta(\tau)) = Q_\tau(Y_t|Y_{t-1},\ldots,Y_{t-p})$, and $X_t = (1, Y_{t-1},\ldots,Y_{t-p})'$. If the τ-conditional quantile of Y_t is correctly specified by a QAR model, then there exists a $F(y|\theta,x)\subset\mathcal{G}$ such that the null hypothesis of (1.2) is not rejected, with $\mathcal{G}$ satisfying

$$\mathcal{G} = \left\{F(\cdot|\theta,\cdot)|\ F^{-1}\left(\cdot|\theta,X_t\right) = X_t'\theta \text{ for some } \theta\in\mathcal{B}(\mathcal{T},\Theta)\right\}.$$

We consider estimators of the QAR model in (1.14) as any solution $\hat{\theta}_T(\tau)$ of the problem

$$\arg\min_{\theta\in\Theta}\sum_{t=1}^{T}\psi(W_t,\theta,\tau),$$

where $\psi(W_t,\theta,\tau) := \left(\tau-\mathbb{1}\{Y_t-X_t'\theta(\tau)\le 0\}\right)$ is the check function. Given the solutions $\hat{\theta}_T(\tau)$, the τ-quantile of $Y_t|X_t$ can be estimated by $\hat{Q}_\tau(Y_t|X_t) = X_t'\hat{\theta}_T(\tau)$. In our setup, $\hat{\theta}_T$ belong to the class of Z-estimators with $\psi(W_t,\hat{\theta}_T,\tau) = \left(\tau-\mathbb{1}\{Y_t-X_t'\hat{\theta}_T(\tau)\le 0\}\right)$. If the conditional distribution of Y_t is monotone in τ, the QAR model in (1.14) implies a conditional distribution function that can be estimated by $F(y|\hat{\theta}_T(\cdot),x) = \int_{\mathcal{T}}\mathbb{1}\{x'\hat{\theta}_T(\tau)\le y\}d\tau$. Now we establish the conditions that allows us to apply our test statistic S_T to check the specification of a QAR model.

Proposition 1. *Let:*

(i) For every $\tau \in \mathcal{T}$, $\mathrm{E}(\tau - \mathbb{1}\{Y_t - X_t'\theta(\tau) \leq 0\})$ possesses a unique zero at $\theta(\tau) = \theta_0(\tau)$, for some $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$;

(ii) $\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \ldots\}$ is an α-mixing triangular array with stationary rows satisfying $E(|Y_{1,1}|^{2+\gamma}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\gamma/(4+\gamma)} < \infty$ for some $\gamma \in (0, 2)$;

(iii) The conditional distribution function of Y_t given X_t, $F(\boldsymbol{\cdot}|\boldsymbol{\cdot})$, and its density function $f(\boldsymbol{\cdot}|\boldsymbol{\cdot})$ have continuous derivatives up to the 2^{nd}-order denoted respectively by $F^{(s)}(\boldsymbol{\cdot}|\boldsymbol{\cdot})$ and $f^{(s)}(\boldsymbol{\cdot}|\boldsymbol{\cdot})$, $s = 1, 2$;

(iv) $f(\boldsymbol{\cdot}|\boldsymbol{\cdot})$ is Lipschitz continuous and bounded away from zero on $X_t'\theta_0(\tau)$ a.s., uniformly over $\tau \in \mathcal{T}$, and $F^{(2)}(\boldsymbol{\cdot}|\boldsymbol{\cdot})$ and $f^{(2)}(\boldsymbol{\cdot}|\boldsymbol{\cdot})$ are bounded and uniformly continuous on $\mathbb{R}$ a.s.;

(v) The matrix $\mathrm{E}(X_t X_t')$ is finite and has full rank.

Then Assumptions 1-5 hold for the linear quantile autoregression model.

Proposition 1 provides conditions for identifiability of the moment conditions in (1.4) and the validity of a functional central limit for a dependent stochastic process $\sqrt{T}(\hat{\theta}_T(\tau) - \theta(\tau))$ (Andrews and Pollard, 1994, Chernozhukov et al., 2013). The Lipschitz condition in (iv) gives a sufficient condition for the class of functions $\{\psi(W_t, \theta, \tau) = \big(\tau - \mathbb{1}\{Y_t - X_t'\theta(\tau) \leq 0\}\big) : \theta \in \Theta, \tau \in \mathcal{T}\}$ to be a VC class.

1.5.2 Nonlinear Quantile Autoregressive Models

We can apply our test to check the correct specification of a nonlinear quantile regression model such as the Conditional Autoregressive Value at Risk (CAViaR) model proposed by Engle and Manganelli (2004). Value at Risk (VaR) is the standard measure of market risk used by financial institutions and market regulators. Let Y_t be a return on a portfolio series. Given a significance level τ, the VaR of a portfolio is the level of return Y_t^T over the period $[t, T)$ that is exceeded with probability τ: $VaR_t^T(\tau|x) := \inf_L\{L : \Pr(Y_t^T \leq L|x) \geq 1 - \tau\}$. Analogously, we can also write the VaR as $VaR_t^T(\tau|x) = Q_\tau(Y_t|x)$. Since the VaR is a quantile of the conditional distribution of returns, the quantile regression model is a powerful tool to model VaR, using only information pertaining to the quantiles of the distribution.

Rather than imposing a linear quantile regression model, we may assume a nonlinear functional dependence on the quantiles of $Y_t|X_t$:

$$Q_\tau(Y_t|X_t = x) = m(x, \theta(\tau)), \tag{1.15}$$

where $m : \mathbb{R}^d \times \Theta \times \mathcal{T} \mapsto \mathbb{R}$ is a known function. Under our setup, we have

$$\mathcal{G} = \left\{ F(\cdot|\theta, \cdot) |\ F^{-1}\left(\cdot|\theta(\cdot), x\right) = m(x, \theta(\cdot)) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta) \right\}.$$

Similarly to the linear QAR process, we can estimate the parameters $\hat{\theta}_T(\cdot)$ of a nonlinear quantile regression model in (1.15) by solving

$$\arg\min_{\theta \in \Theta} \sum_{t=1}^{T} \rho_\tau \left(Y_t - m(X_t, \theta(\tau)) \right), \tag{1.16}$$

with $\rho_\tau(u) = u(\tau - \mathbb{1}\{u \leq 0\})$. For sufficient conditions on $m(\cdot, \cdot)$ for the existence of a solution of (1.16), see Koenker and Park (1996). Given the solutions $\hat{\theta}_T(\cdot)$, the conditional distribution function can be obtained as $F(y|\hat{\theta}_T(\cdot), x) = \int_{\mathcal{T}} \mathbb{1}\{m(x, \hat{\theta}_T(\tau)) \leq y\} d\tau$, assuming that $F(y|\hat{\theta}_T(\cdot), x)$ is monotone in y. Nonlinear dynamic models allow the inclusion of past values of the quantiles of $Y_t|X_t$. A general CAViaR specification for the quantile regression can be the following

$$Q_\tau(Y_t|\theta(\tau), \Omega_{t-1}^p) = \theta_0(\tau) + \sum_{i=1}^{p} \theta_i(\tau) Q_\tau(Y_{t-i}|\Omega_{t-i-1}^p) + \sum_{j=1}^{q} \theta_j(\tau) \ell_{t-j}\big(\mathbf{x}_{t-j}\big), \tag{1.17}$$

where $\Omega_{t-1}^p := \{Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\}$ is the lagged-value vector of Y_t from $t-p$ up to time $t-1$, the parameter vector θ has a dimension of $r = p + q + 1$, and $\ell(\cdot)$ is a function of a vector of lagged values of observables $\mathbf{x}_{t-j} \in X_{t-j}$, which could be the lagged returns Y_{t-1} for instance. Let $X_t = (1, Y_{t-1}, \ldots, Y_{t-p})'$, then the associated estimator θ_T is in the class of Z-estimators with $\psi(W_t, \theta, \tau) = \varepsilon_t(\tau)\big(\tau - \mathbb{1}\left\{Y_t - Q_\tau(Y_t|\theta(\tau), X_t) \leq 0\right\}\big)$, where $\varepsilon_t(\tau) = Y_t - Q_\tau(Y_t|\theta(\tau), X_t)$. The following proposition provides conditions for applying our proposed test to the CAViaR model described in (1.17).

Proposition 2. *Let Assumptions C0-C7 and AN1-AN3 of Engle and Manganelli (2004) hold. Then Assumptions 2-5 hold for the CAViaR model.*

1.5.3 Distributional Regression Models

Our proposed test statistic checks the validity of the distributional regression model introduced by Foresi and Peracchi (1995), where the conditional distribution is modeled through a family of binary response models for the event that the variable of interest Y_t exceeds some threshold $y \in \mathbb{R}$. In contrast to the quantile regression model, the distributional regression model does not require the dependent variable to be continuously distributed. This can be useful in many empirical applications. Besides, the distributional regression approach uncovers higher-order multidimensional structure that cannot be found by modeling only the first two moments of the conditional distribution. This has important implications to forecasting excess stock market returns and finding an optimal portfolio (Foresi and Peracchi, 1995).

Chernozhukov et al. (2013) derive the limit theory for the continuum of binary regressions and Rothe and Wied (2013) provide specification tests for distributional regressions under i.i.d data. Our setting allows us to evaluate the specification of distributional regressions under time series data. To the best of our knowledge, we are not aware of a framework to testing for the correct specification of distributional regression models under dependent data.

In distributional regression models (DR models), the conditional distribution function of Y_t is model through a family of binary response models for the event that Y_t exceeds some threshold $y \in \mathbb{R}$, as follows:

$$F\left(y|x\right) = \Lambda\left(x'\theta(y)\right), \text{for some } \theta(y) \in \mathcal{B}(\mathbb{R}, \Theta) \subset \mathbb{R}^K \text{ and all } y \in \mathbb{R}, \tag{1.18}$$

where $\Lambda(\bullet)$ is a known strictly increasing link function (e.g., the logistic or standard normal distribution), and $\theta(\bullet)$ is a functional parameter taking values in $\mathcal{B}(\mathbb{R}, \Theta)$. The DR approach was introduced by Foresi and Peracchi (1995), and it has been analysed by Fortin, Lemieux, and Firpo (2011), Rothe (2012), Rothe and Wied (2013), and Chernozhukov et al. (2013). One can also run a distributional regression model of Y_t conditional on its lagged values:

$$F\left(y|Y_{t-1}\right) = \Lambda\left(Y'_{t-1}\theta(y)\right), \text{for some } \theta(y) \in \mathcal{B}(\mathbb{R}, \Theta) \subset \mathbb{R}^K \text{ and all } y \in \mathbb{R}, \tag{1.19}$$

For a given cut-off $y \in \mathbb{R}$, the estimator $\hat{\theta}_T(y)$ is given by

$$\begin{aligned}\hat{\theta}_T(y) := \underset{\theta \in \mathcal{B}(\mathbb{R}, \Theta)}{\arg\max} \frac{1}{T} \sum_{t=1}^{T} \Big[& \mathbb{1}\{Y_t \le y\} \ln\left[\Lambda\left(Y'_{t-1}\theta(y)\right)\right] \\ & + \left(1 - \mathbb{1}\{Y_t \le y\}\right) \ln\left[1 - \Lambda\left(Y'_{t-1}\theta(y)\right)\right] \Big]. \end{aligned} \tag{1.20}$$

Then, the conditional distribution of Y_t given Y_{t-1} is estimated as follows:

$$F_T(y|\hat{\theta}_T(y), Y_{t-1}) = \Lambda(Y'_{t-1}\hat{\theta}_T(y)), \text{ for all } y \in \mathbb{R}. \tag{1.21}$$

The following proposition provides the conditions for the distributional autoregressive model in (1.19) to satisfy the Assumptions 1-5, and hence the application of our test statistic S_T.

Proposition 3. *Let:*

(i) $\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \ldots\}$ *is an* α*-mixing triangular array with stationary rows satisfying* $E(|Y_{1,1}|^{2+\gamma}) < \infty$ *and* $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\gamma/(4+\gamma)} < \infty$ *for some* $\gamma \in (0, 2)$*. The support of Y, Supp*(Y)*, is a finite set or a bounded open subset of* $\mathbb{R}$*;*

(ii) For every $y \in Supp(Y)$*, the parameter* $\theta_0(\bullet)$ *solves*

$$\mathrm{E}\left[\mathbb{1}\{Y_t \leq y\} \ln\left(\Lambda\left(Y_{t-1}'\theta_0(y)\right)\right) + (1 - \mathbb{1}\{Y_t \leq y\}) \ln\left(1 - \Lambda\left(Y_{t-1}'\theta_0(y)\right)\right)\right] = 0,$$

such that $\theta_0(y) \in \Theta$*;*

(iii) The conditional distribution function of Y_t *given* X_t*,* $F(\bullet|\bullet)$*, has a density function* $f(\bullet|\bullet)$ *that is continuous, bounded, and bounded away from zero at all* $y \in Supp(Y)$ *a.s.;*

(iv) $\Lambda\left(Y_{t-1}'\theta(\bullet)\right)$ *is bounded away from zero and one uniformly over* $\theta \in \Theta$ *a.s.;*

(v) The matrix $\mathrm{E}(X_t X_t')$ *is finite and has full rank.*

Then Assumptions 1-5 hold for the distributional autoregressive model in (1.19).

Under Assumptions 1-5, we can apply our test statistic S_T to distributional regression models in dependent data settings, such as in (1.19).

1.6 Finite-Sample Performance

To examine the finite-sample performance of our proposed test statistic and its bootstrap procedure, we perform simulation experiments with data generating processes (DGPs) under the null and the alternative hypothesis. The data are generated from the processes below.

Size DGPs :

DGP.1 (AR(1)) : $Y_t = 0.3Y_{t-1} + u_t,$

DGP.2 (AR(2)) : $Y_t = 0.3Y_{t-1} - 0.3Y_{t-2} + u_t,$

Power DGPs :

$$\text{DGP.3 (TAR)}: \begin{cases} Y_t = 1 + 0.6Y_{t-1} + u_t, \text{ if } Y_{t-1} \leq 1, \\ Y_t = 1 - 0.5Y_{t-1} + u_t, \text{ if } Y_{t-1} \geq 1, \end{cases}$$

DGP.4 (Bilinear) : $Y_t = 0.8Y_{t-1}u_{t-1} + u_t,$

DGP.5 (Nonlinear MA) : $Y_t = 0.8u_{t-1}^2 + u_t,$

DGP.6 (Logistic Map) : $Y_t = 4Y_{t-1}(1 - Y_{t-1}),$

DGP.7 (GARCH(1,1)) : $Y_t = h_t u_t,\ h_t^2 = 0.02 + 0.06Y_{t-1}^2 + 0.93h_{t-1}^2,$

where u_t follows an i.i.d process with distribution $\mathcal{N}(0, 1)$. We want to test the null hypothesis that the quantiles of Y_t follow a AR(1) process:

$$\mathcal{H}_0 : F_{Y_t}^{-1}(\tau|\theta_0(\tau), Y_{t-1}) = \alpha + \beta Y_{t-1} + \Phi_u^{-1}(\tau), \text{ a.s.,}$$

where $\Phi_u^{-1}(\tau)$ is the τ-quantile of the standard Normal error distribution. We use DGP.1 and DGP.2, described in Corradi and Swanson (2006), to check the size performance of our test statistic. While a $QAR(1)$ model correctly specifies the conditional distribution in DGP.1, we allow for dynamic misspecification in DGP.2, as $F(y|\theta_0, Y_{t-1}) \neq F(y|\theta^0, Y_{t-1}, Y_{t-2})$ with $\theta_0 \neq \theta^0$. The DGPs 3-7 allows us to see the empirical power performance and have been considered by Hong and Lee (2003) and Escanciano and Velasco (2010). In these experiments, rejection arises because of misspecification of the conditional distribution model. DGP.4 and DGP.5 are second-order stationary, though they are not invertible (Granger and Andersen, 1978). DGP.6 follows a process similar to a white noise, but it has autocorrelations in squares similar to ARCH(1) (Granger and Teräsvirta, 2010). DGP.7 examine the power of our test against misspecifications in the conditional variance.

We also design a DGP for testing the specification of a Distributional Regression model in the form of (1.19). The data are generated as in DGP.5, a Nonlinear MA(1) model, and we are interested in testing the null hypothesis that the Distributional Regression model is correctly specified conditioning Y_t only on Y_{t-1}:

$$\mathcal{H}_0^{DR} : F\left(y|Y_{t-1}\right) = \Lambda\left(Y_{t-1}'\theta(y)\right), \text{ a.s.,} \tag{1.22}$$

where $\Lambda(\bullet)$ is specified as a logistic distribution function. For all the experiments, we consider the empirical rejection frequencies for 5% and 10% nominal level tests with different sample sizes ($T = 100$ and 300), and choose a grid $\mathcal{T} = [0.01, 0.99]$. In calculating the test statistics, we use an equally spaced grid of 100 quantiles $\mathcal{T}_n \subset \mathcal{T}$. We perform $1,000$ Monte Carlo repetitions in each of the simulations, and apply $B = 399$ block bootstrap replications in each of the simulations to calculate the critical values. Then the maximal simulation standard error for the tests empirical sizes and powers is $\max_{0\leq p\leq 1} \sqrt{p(1-p)/1000} \approx 0.016$. For each bootstrap replicate, we use three different block lengths $\ell = \{2, 4, 6\}$, which are close to the block length of $CT^{1/3}$, for a constant $C > 0$, suggested by Künsch (1989). In all the replications, we generated and discarded 200 pre-sample data values. Except for the Distributional Regression specification test, we compare our results with the test proposed by Escanciano and Velasco (2010) (EV henceforth), based on

$$EV := \int\int \left|\left(\mathbb{1}\left(Y_t - m(X_t, \hat{\theta}_T(\tau)) \leq 0\right) - \tau_j\right) \exp(\boldsymbol{i}\boldsymbol{x}' X_t)\right|^2 dW(\boldsymbol{x}) d\Phi(\alpha), \tag{1.23}$$

where W and Φ are some integrating measures on $\mathbb{R}$ and $\mathcal{T}$, and $m(X_t, \hat{\theta}_T(\tau))$ is the estimated parametric QAR(1) model for each τ-quantile, for $\tau \in \mathcal{T}$. The critical values of the test (1.23) are obtained by subsampling. In each Monte Carlo replication, $T-b-1$ subsamples of size b were generated. We apply the EV test for two different subsample sizes $b = [kT^{(2/5)}]$, for $k = 3$ and 4, following the suggestion of Sakov and Bickel (2000).

Tables 1.1 and 1.2 report the rejection frequencies of the S_T test associated with the DGPs 1-7, for sample sizes $T = 100$ and $T = 300$ respectively. The empirical level of the S_T test is generally close to the nominal level under the null hypothesis, disregarding whether there is dynamic misspecification (DGP.2) or not (DGP.1). On the other hand, the EV test of Escanciano and Velasco (2010) presents size distortions for both sample sizes, increasing in the presence of dynamic misspecification (DGP.2). Those results are robust for different subsample sizes b. Thus, our test has the correct asymptotic size even in the presence of dynamic misspecification.

In terms of power, the S_T test exhibits good power and reliable inference even when using a small sample size $T = 100$. Comparing with the EV test, the S_T test performs well: it is the most powerful test for DPG.3, DGP.4, DGP.6, and DGP.7; it has less power than the EV test only against DGP.5, when the subsample size is $b = 18$, but it still has more power than the EV test for a subsample size of $b = 25$. In addition, the power of both tests converge to 1 for $T = 300$. Our test statistic is also powerful against misspecifications in the distributional regression, as the power for testing $\mathcal{H}_0^{DR}$ in (1.22) is 1 for a small sample size of $T = 100$ (Table 1.1). To the best of our knowledge, no specification test for Distributional Regression models has been developed for a time series setting. In sum, our proposed test seems to perform quite well in finite samples.

Table 1.1. Monte Carlo empirical rejection frequencies of specification tests: $T = 100$

	$S_T(\ell = 2)$		$S_T(\ell = 4)$		$S_T(\ell = 6)$		$EV(b = 18)$		$EV(b = 25)$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.1	0.030	0.110	0.060	0.110	0.036	0.066	0.075	0.129	0.077	0.120
DGP.2	0.030	0.080	0.040	0.100	0.052	0.102	0.091	0.160	0.084	0.143
DGP.3	0.960	0.990	0.990	0.990	0.920	0.960	0.888	0.931	0.847	0.913
DGP.4	0.962	0.990	1.000	1.000	1.000	1.000	0.997	0.999	0.984	0.993
DGP.5	0.912	0.952	0.864	0.916	0.900	0.954	0.944	0.969	0.913	0.944
DGP.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.7	0.216	0.304	0.260	0.360	0.200	0.380	0.095	0.149	0.101	0.150
$\mathcal{H}_0^{DR}$	1.000	1.000	1.000	1.000	1.000	1.000	-	-	-	-

Note: S_T denotes our proposed test statistic with $B = 399$ bootstrap replications with block lengths $\ell = \{2, 4, 6\}$. EV denotes the subsampling specification test of Escanciano and Velasco (2010). The null hypothesis $\mathcal{H}_0^{DR}$ test the specification of a Distributional Regression model specified in (1.22), under DGP.5. We use 1,000 Monte Carlo repetitions based on the DGPs 1-7 described above.

Table 1.2. Monte Carlo empirical rejection frequencies of specification tests: $T = 300$

	$S_T(\ell = 2)$		$S_T(\ell = 4)$		$S_T(\ell = 6)$		$EV(b = 29)$		$EV(b = 39)$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.1	0.043	0.087	0.020	0.080	0.031	0.070	0.061	0.107	0.057	0.108
DGP.2	0.053	0.107	0.067	0.107	0.049	0.122	0.092	0.147	0.074	0.134
DGP.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.7	0.970	0.980	0.960	0.980	0.980	1.000	0.186	0.272	0.189	0.260
$\mathcal{H}_0^{DR}$	1.000	1.000	1.000	1.000	1.000	1.000	-	-	-	-

Note: S_T denotes our proposed test statistic with $B = 399$ bootstrap replications with block lengths $\ell = \{2, 4, 6\}$. EV denotes the subsampling specification test of Escanciano and Velasco (2010). The null hypothesis $\mathcal{H}_0^{DR}$ test the specification of a Distributional Regression model specified in (1.22), under DGP.5. We use 1,000 Monte Carlo repetitions based on the DGPs 1-7 described above.

1.7 An Empirical Application

Many empirical papers have proposed methods to precisely check the specification of models for Value-at-Risk (VaR). Since VaR determines the regulatory risk capital of all regulated financial institutions (see Basel Committee on Banking Supervision 1996), the outcome of a VaR model determines the multiplication factors for market risk capital requirements of financial institutions. Thus, an inaccurate VaR model leads to an underestimated multiplicative factor, that delivers an insufficient reserve of capital risk for financial institutions. Therefore, the specification of VaR models is crucial for risk managers, regulators, and financial institutions.

Since the VaR is a quantile of the portfolio returns, conditional on past information, and as the distribution of portfolio returns evolves over time, it is challenging to model time-varying conditional quantiles. An accurate VaR model satisfies $\Pr(Y_t \leq -VaR_t|\mathcal{F}_{t-1}) = \tau$, for a portfolio return series Y_t, a past information set $\mathcal{F}_{t-1}$, and a quantile $\tau \in (0,1)$. The dynamic conditional quantile regression approach specifies a conditional VaR model using only the relevant past information that influence the quantiles of interest, and many applications support this methodology (Chernozhukov and Umantsev, 2001, Engle and Manganelli, 2004, Escanciano and Olmo, 2010).

To illustrate the performance of our proposed test statistic, we test different specifications of conditional quantile regression models for estimating the VaR of stock returns. We estimate the VaR of the returns of two major stock indexes, the Frankfurt Dax Index (DAX) and the London FTSE-100 Index (FTSE-100). The DAX and the FTSE-100 daily stock indexes are two representatives of the data for which linear and non-linear quantile regression models have been widely used, see e.g. Escanciano and Velasco (2010), Iqbal and Mukherjee (2012), and Jeon and Taylor (2013). The dataset consists of 2,981 daily observations - from January 2003 to June 2014 - on Y_t, the one-day returns, and X_t, the lagged returns $(Y_{t-1}, \ldots, Y_{t-p})$.

Figure 1.1 displays the daily log-return series of the two series. It shows that both log-return series display calm as well as volatile periods and also single outlying log-return observations. Table 1.3 presents the summary statistics of the series. Both log-returns series are highly leptokurtic and present autocorrelation.

Table 1.3. Summary statistics: DAX and FTSE-100 daily log-returns

	DAX	FTSE-100
Mean	0.02	0.01
Std. Dev.	0.61	0.51
Median	0.03	0.01
Skewness	0.01	-0.12
Kurtosis	9.14	11.71
Minimum	-3.23	-4.02
Maximum	4.69	4.08
Autocorrelation	-0.01	-0.06
LB(10)	21.34	62.35

Note: The Autocorrelation is the first-order autocorrelation coefficient, and $LB(10)$ denotes the Ljung-Box Q-statistic of order 10.

For each series, we estimate a Gaussian AR(1)-GARCH(1,1) of the VaR, $VaR_t(\tau)$, as follows:

$$\begin{aligned} \text{AR(1)-GARCH(1,1): } F_{Y_t}^{-1}(\tau|\theta_0(\tau), Y_{t-1}, \sigma_t) &= \beta_0 + \beta_1 Y_{t-1} + F_{\varepsilon}^{-1}(\tau)\sigma_t, \\ \sigma_t^2 &= \gamma_0 + \gamma_1 Y_{t-1}^2 + \gamma_2 \sigma_{t-1}^2, \end{aligned}$$

where $F_{\varepsilon}^{-1}(\tau)$ is the τ-quantile of the standard Gaussian error distribution. Thus, we test the hypothesis $\mathcal{H}_0$: the VaR of the log-return Y_t follows an AR(1)-GARCH(1,1) Gaussian process. We choose this specification as GARCH models have provided appropriate specifications of the VaR of stock returns in the literature (Escanciano and Olmo, 2010). We also entertain other models: GARCH(1,1), AR(2)-GARCH(2,2), E-GARCH(1,1), AR(1)-GARCH(1,1) with Student-t5 distribution, and GARCH(1,1) with Student-t5 distribution. We apply GARCH(1,1) and AR(1)-GARCH(1,1) with a Student-t5 distribution because they are valid models for the distribution of monthly stock returns in Bai (2003) and Kheifets (2015). To present results with a different GARCH specification, we estimate an E-GARCH(1,1) model for the VaR as:

$$\begin{aligned} \text{E-GARCH(1,1): } F_{Y_t}^{-1}(\tau|\theta_0(\tau), Y_{t-1}, h_t) &= F_{\varepsilon}^{-1}(\tau) h_t, \\ \ln h_t^2 &= \alpha_0 + \alpha_1 \ln h_{t-1}^2 + \alpha_2 \left(|Y_{t-1}^2| - (2/\pi)^{\frac{1}{2}}\right) - \alpha_3 Y_{t-1}^2. \end{aligned} \tag{1.24}$$

As we want to compare our methodology with standard specification tests for conditional quantile regression models in the literature, we perform the EV test described in (1.23), with two different subsample sizes $b = [kT^{2/5}]$ for $k = 3$ and $k = 4$.

Table 1.4 shows the p-values of the specification tests for all the VaR models for the

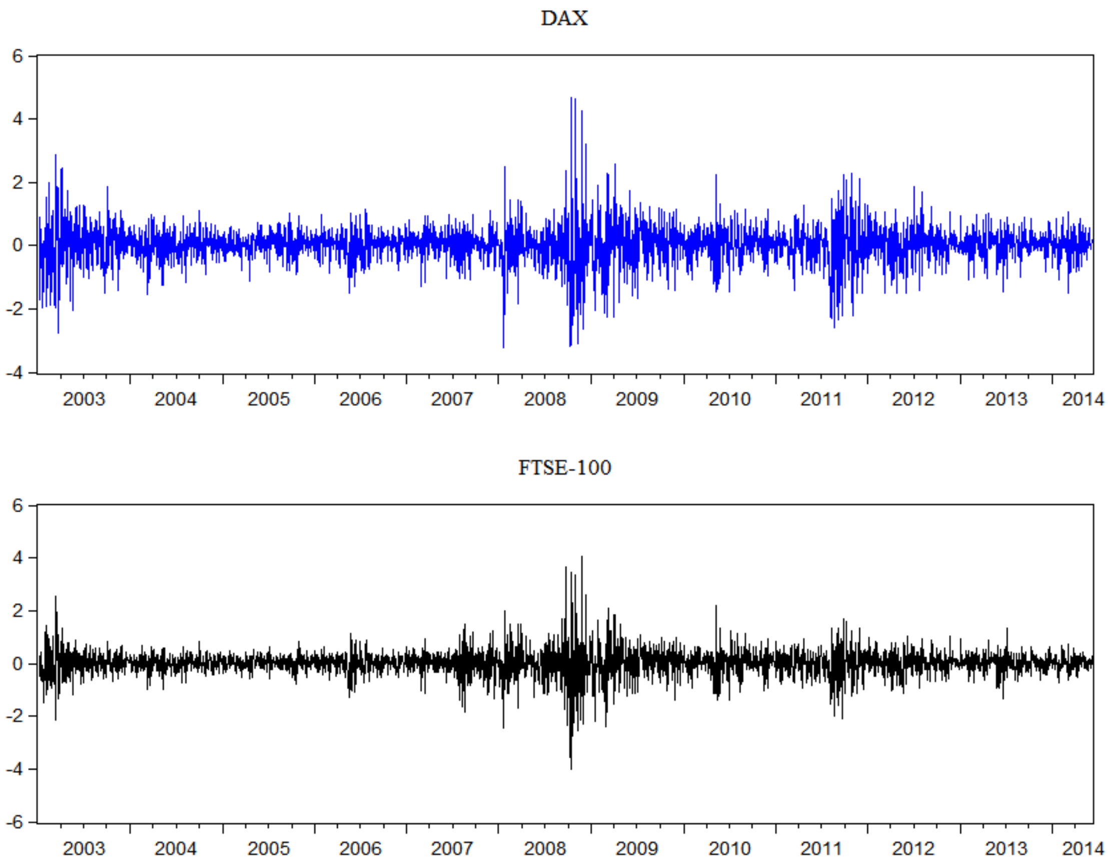

Figure 1.1. Daily log-returns of DAX and FTSE-100 indexes in the period January 6^{th}, 2003 to June 9^{th}, 2014

full sample from January 2003 to June 2014. For the DAX index series, our test S_T rejects the specifications of all proposed models to fitting a VaR for the log-returns at 1% significance level. These results are robust to three different block lengths. On the other hand, the EV test of Escanciano and Velasco (2010) do not reject an AR(1)-GARCH(1,1) specification with Student-t5 distribution at 1% significance level. Regarding the FTSE-100 series, the S_T test does not reject a AR(1)-GARCH(1,1) model at 1% significance level, while the EV test does not reject a AR(1)-GARCH(1,1) model with Student-t5 distribution at the 1% significance level. We note that the AR(1)-GARCH(1, 1) family of models is the only class of models that is not rejected for these returns series, but this result is not robust to different block lengths.

For robustness, we perform the same tests to these models using only one year of data, from June 26^{th}, 2013 to June 9^{th}, 2014. Table 1.5 displays the results for this period. While the EV test of Escanciano and Velasco (2010) rejects all models, our test S_T does not reject most of the models at the 1% significance level for the DAX daily returns series. Moreover, the AR(1)-GARCH(1,1) model with Student-t5 distribution has obtained the highest p-value and is the only model that is not rejected at the 10%

significance level. For the FTSE-100 returns, our test does not reject the GARCH(1,1), GARCH(1,1) with Student-t5 distribution and the AR(1)-GARCH(1,1) model at the 1% significance level, while the the EV test of Escanciano and Velasco (2010) does not reject only the ARCH(1,1) with Student-t5 distribution at the 1% significance level.

Thus, the empirical application shows the ability of our test to detect possibly misspecified conditional distribution models when we have a small sample size. This is useful for risk managers and financial institutions to apply a valid VaR model and obtain the correct multiplicative factors for their market risk capital requirements.

Table 1.4. Specification tests p-values of VaR models of DAX and FTSE-100 returns: January 6^{th}, 2003-June 9^{th}, 2014

DAX

	$S_{T,6}$	$S_{T,8}$	$S_{T,16}$	EV(b=98)	EV(b=122)
GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.000	0.000
GARCH(1,1)-t5 - CAViaR	0.001	0.001	0.001	0.001	0.000
AR(1)-GARCH(1,1) - CAViaR	0.002	0.001	0.001	0.000	0.000
AR(1)-GARCH(1,1)-t5 - CAViaR	0.001	0.001	0.002	0.010	0.007
AR(2)-GARCH(2,2) - CAViaR	0.001	0.001	0.001	0.000	0.000
E-GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.001	0.001

FTSE-100

	$S_{T,6}$	$S_{T,8}$	$S_{T,16}$	EV(b=98)	EV(b=122)
GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.000	0.000
GARCH(1,1)-t5 - CAViaR	0.001	0.001	0.001	0.010	0.002
AR(1)-GARCH(1,1) - CAViaR	0.002	0.011	0.003	0.009	0.004
AR(1)-GARCH(1,1)-t5 - CAViaR	0.004	0.005	0.004	0.010	0.007
AR(2)-GARCH(2,2) - CAViaR	0.009	0.004	0.003	0.000	0.000
E-GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.001	0.001

Note: $S_{T,\ell}$ is the S_T test with block length $\ell = \{6, 8, 16\}$. We denote EV as the specification test of Escanciano and Velasco (2010), with sub-sample size b. The E-GARCH(1,1) is estimated as in (1.24).

Table 1.5. Specification tests p-values of VaR models of DAX and FTSE-100 returns: June 26^{th}, 2013-June 9^{th}, 2014

DAX

	$S_{T,3}$	$S_{T,4}$	$S_{T,6}$	EV(b=27)	EV(b=36)
GARCH(1,1) - CAViaR	0.028	0.031	0.035	0.000	0.000
GARCH(1,1) - t5 - CAViaR	0.040	0.030	0.033	0.000	0.000
AR(1)-GARCH(1,1) - CAViaR	0.018	0.029	0.001	0.000	0.000
AR(1)-GARCH(1,1) - t5 - CAViaR	0.175	0.167	0.159	0.000	0.000
AR(2)-GARCH(2,2) - CAViaR	0.059	0.050	0.044	0.000	0.000
E-GARCH(1,1) - CAViaR	0.034	0.033	0.044	0.001	0.001

FTSE-100

	$S_{T,3}$	$S_{T,4}$	$S_{T,6}$	EV(b=27)	EV(b=36)
GARCH(1,1) - CAViaR	0.634	0.608	0.614	0.000	0.000
GARCH(1,1) - t5 - CAViaR	0.622	0.582	0.602	0.010	0.002
AR(1)-GARCH(1,1) - CAViaR	0.451	0.443	0.465	0.009	0.004
AR(1)-GARCH(1,1) - t5 - CAViaR	0.288	0.001	0.001	0.010	0.007
AR(2)-GARCH(2,2) - CAViaR	0.001	0.001	0.001	0.000	0.000
E-GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.001	0.001

Note: $S_{T,\ell}$ is the S_T test with block length $\ell = \{3, 4, 6\}$. We denote EV as the specification test of Escanciano and Velasco (2010), with sub-sample size b. The E-GARCH(1,1) is estimated as in (1.24).

1.8 Conclusion

In this paper, we present a practical and consistent specification test of conditional distribution and quantile models in a very general setting for dependent observations. Our setting covers conditional distribution models possibly indexed by function-valued parameters, which allows for a wide range of important empirical applications in economics and finance, such as the linear quantile auto-regressive, the CAViaR, and the distributional regression models. Based on a comparison between an estimated parametric distribution and the empirical distribution function, our proposed bootstrap test has the correct asymptotic size and is consistent against fixed alternatives. In addition, our test has non-trivial power against $\sqrt{T}$-local alternatives, with T the sample size.

Finite sample experiments suggest that our proposed test has good size and power properties, and is more powerful than other comparable specification tests in the literature against almost all alternatives. In addition, our approach has the correct asymptotic size under dynamic misspecification. An empirical application illustrates the practical importance of our setting in risk management. The use of misspecified VaR models may lead to the acceptance of a sub-optimal model for VaR, underestimating the multiplicative factors of the reserve of capital risk of financial institutions. Therefore, checking the validity of a VaR model is of crucial importance for monitoring risk of financial institutions. We observe that the AR(1)-GARCH(1, 1) family of models provided valid specifications for the VaR of two major stock returns indexes.

A possible direction for future work is to extend this study to test Granger-causality in distribution. Although the concept of Granger-causality is defined in terms of the conditional distribution, the majority of papers have tested Granger-causality using conditional mean regression models in which the causal relations are linear. As a result, a conditional mean regression model cannot assess a tail causal relation or nonlinear causalities. Our proposed approach allows us to evaluate nonlinear causalities, causal relations in conditional quantiles, and Granger-causality in distribution through an application of distributional regression in a time series context. One could also extend our approach to the class of multivariate models, providing specification tests for vector autoregressions and multivariate linear and non-linear models, see e.g. Francq and Raïssi (2007) and Escanciano, Lobato, and Zhu (2013).

1.9 Appendix

1.9.1 Tools

In this section, we introduce some auxiliary results. Let $\mathcal{M}$ be a permissible class of functions such that it can be indexed by some set $\mathcal{T}$, i.e., $\mathcal{M} = \{\Psi(\cdot, \tau) : \tau \in \mathcal{T}\}$, in such a way that the following holds: (i) $\mathcal{T}$ is a Suslin metric space (a Hausdorff topological space that is the continuous image of a Polish space) with Borel σ-field $\mathcal{B}(\mathcal{T})$, and (ii) $\Psi(\boldsymbol{\cdot},\boldsymbol{\cdot})$ is $\times\mathcal{B}(\mathcal{T})$-measurable function from $\mathbb{R}^K\times\mathcal{T}$ to $\mathbb{R}$ (see Kosorok, 2007, Section 11.6). Let $Pf = \int f(\theta,\tau)dP(\theta,\tau)$, for $f\in\mathcal{M}$. Given $\varepsilon>0$, we define the covering number $N(\varepsilon,\mathcal{M},||\boldsymbol{\cdot}||)$ as the minimal number of $L_2(P)$-balls of radius ε needed to cover $\mathcal{M}$, where a $L_2(P)$-ball of radius ε around a function $g\in L_2(P)$ is the set $\{h\in L_2(P): ||h-g||<\varepsilon\}$. We define the uniform covering numbers as $\sup_P N(\varepsilon||\mathbb{F}||,\mathcal{M},L_2(P))$, with $\mathbb{F}$ the square-integrable envelope of $\mathcal{M}$. We assume that the $\mathcal{M}$ class of functions forms a so-called Vapnik-Chervonenkis (VC) class of functions (see Dudley, 1978, Pollard, 1984). The VC class is an extension of the class of indicator functions and has the interesting property that for $1\le p<\infty$, there are constants C_1 and C_2 satisfying

$$N(\varepsilon,\mathcal{M},||\boldsymbol{\cdot}||)\le C_1\left(\frac{(P(\mathbb{F})^p)^{1/p}}{\varepsilon}\right)^{C_2},$$

for all $\varepsilon>0$ and all probability measures P (see Lemmas II.25 and II.32 in Pollard, 1984). In the following Lemma, we derive a Central Limit Theorem for strong mixing processes for the empirical distribution, $\hat{Z}_T(y,x)$, under the null and the alternative hypothesis.

Lemma 1.9.1. *Given Assumption 1, under $\mathcal{H}_0$ of (1.2) or $\mathcal{H}_A$ of (1.3),*

$$v_T(y,x) := \sqrt{T}(\hat{Z}_T(y,x)-F_{YX}(y,x)) \implies \mathbb{H}_1(y,x),\ \textit{in } \ell^\infty(\mathcal{W}),$$

where $\mathbb{H}_1$ is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$ with covariance function

$$\mathrm{Cov}(\mathbb{H}_1(y,x),\mathbb{H}_1(y',x')) = \sum_{k=-\infty}^{\infty}\mathrm{Cov}\big(\mathbb{1}\{Y_0\le y\}\mathbb{1}\{X_0\le x\},\mathbb{1}\{Y_k\le y'\}\mathbb{1}\{X_k\le x'\}\big).$$

Proof. Assumption 1 implies strong mixing coefficients $\alpha(j)=O(j^{-k})$, for some $k>1$. Then the result follows from a direct application of Theorem 7.2 in Rio (2000). $\square$

In the paper, we have a functional parameter $\tau\mapsto\theta(\tau)$, where $\tau\in\mathcal{T}$ and $\theta(\tau)\in\mathcal{B}(\mathcal{T},\Theta)$, and the true value $\theta_0(\tau)$ solves the moment equations $\Psi(\theta,\tau)=0$. The following lemma establishes a functional delta method for the empirical analog $\hat{\Psi}_T(\theta,\tau)$ of the previous moment equations and for the estimator of the functional parameter, $\hat{\theta}_T(\boldsymbol{\cdot})$.

Lemma 1.9.2. *Given Assumptions 1-5, under $\mathcal{H}_0$ of (1.2) or $\mathcal{H}_A$ of (1.3), we have*

$$r_T(\theta,\tau) := \sqrt{T}(\hat{\Psi}_T(\theta,\tau)-\Psi(\theta,\tau)) \implies \tilde{\mathbb{H}}_2(\theta,\tau),\ \textit{in } \ell^\infty(\mathcal{T}\times\Theta),$$

$$\sqrt{T}(\hat{\theta}_T(\boldsymbol{\cdot})-\theta_0(\boldsymbol{\cdot})) \implies -\dot{\Psi}^{-1}_{\theta_0,\boldsymbol{\cdot}}[\tilde{\mathbb{H}}_2(\theta_0(\boldsymbol{\cdot}),\boldsymbol{\cdot})]\ \textit{in } \ell^\infty(\mathcal{T}),$$

where $\tilde{\mathbb{H}}_2$ *is a tight mean zero Gaussian process in* $\ell^\infty(\mathcal{T} \times \Theta)$ *with covariance function*

$$\mathrm{Cov}(\tilde{\mathbb{H}}_2(\theta,\tau), \tilde{\mathbb{H}}_2(\theta',\tau')) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(\psi(W_0,\theta,\tau), \psi(W_k,\theta',\tau')).$$

Proof. First, by Lemma E.1 in Chernozhukov et al. (2013), Assumptions 2-5 imply that (i) the inverse of $\Psi(\bullet,\tau)$ defined as $\Psi^{-1}(x,\tau) := \{\theta \in \Theta : \Psi(\theta,\tau) = x\}$ is continuous at $x = 0$ uniformly in $\tau \in \mathcal{T}$ with respect to the Hausdorff distance, (ii) there exists $\dot{\Psi}_{\theta_0,\tau}$ such that $\lim_{t\to 0}\sup_{\tau\in\mathcal{T},||h||=1} |t^{-1}[\Psi(\theta_0(\tau)+th,\tau) - \Psi(\theta_0(\tau),\tau)] - \dot{\Psi}_{\theta_0,\tau}h| = 0$, where $\inf_{\tau\in\mathcal{T}}\inf_{||h||=1} ||\dot{\Psi}_{\theta_0,\tau}h|| > 0$, (iii) the maps $\tau \mapsto \theta_0(\tau)$ and $\tau \mapsto \dot{\Psi}_{\theta_0,\tau}$ are continuous, and (iv) the mapping $\tau \mapsto \theta_0(\tau)$ is continuously differentiable. Under the previous conditions, Lemma E.2 in Chernozhukov et al. (2013) holds, and the process $r_T(\theta,\tau)$ weakly converges to $\tilde{\mathbb{H}}_2(\theta,\tau)$ in $\ell^\infty(\mathcal{T}\times\Theta)$ and the map $\theta \mapsto \Psi(\theta,\bullet)$ is Hadamard differentiable at θ_0 with continuously invertible derivative $\dot{\Psi}_{\theta_0,\bullet}$. By Hadamard differentiability of the map $\theta \mapsto \Psi(\theta,\bullet)$, it follows the weak convergence of the process $\sqrt{T}(\hat{\theta}_T(\bullet) - \theta_0(\bullet))$ in $\ell^\infty(\mathcal{T})$. □

Lemma 1.9.3. *Given Assumptions 1-5, under* $\mathcal{H}_0$ *of (1.2) or* $\mathcal{H}_A$ *of (1.3), we have*

$$v_T^{\theta_0}(y,x) := \sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_0)) \implies \mathbb{H}_2(y,x) \text{ in } \ell^\infty(\mathcal{W}),$$

where $\mathbb{H}_2$ *is a tight mean zero Gaussian process in* $\ell^\infty(\mathcal{W})$.

Proof. From Lemma A.2, $\sqrt{T}(\hat{\theta}_T(\bullet) - \theta_0(\bullet)) \implies -\dot{\Psi}^{-1}_{\theta_0,\bullet}[\tilde{\mathbb{H}}_2(\theta_0(\bullet),\bullet)]$ in $\ell^\infty(\mathcal{T})$, where $\tilde{\mathbb{H}}_2$ is a Gaussian process in $\ell^\infty(\mathcal{T}\times\Theta)$. By the functional delta method, we can rewrite $v_T^{\theta_0}(y,x)$ as

$$\sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_0)) = \int (F(y|\hat{\theta}_T,\bar{x}) - F(y|\bar{x}))\mathbb{1}\{\bar{x}\le x\}\sqrt{T}dF_X(\bar{x}) + \int F(y|\bar{x})\mathbb{1}\{\bar{x}\le x\}\sqrt{T}d[\hat{F}_X(\bar{x}) - F_X(\bar{x})] + o_p(1).$$

By the Hadamard differentiability of the map $\theta \mapsto F(\bullet|\theta(\bullet),\bullet)$ in Assumption 5, we can apply the functional delta method, for fixed y and x, as follows:

$$\sqrt{T}(F(y|\hat{\theta}_T,x) - F(y|x)) \implies -\dot{F}^{-1}(y|\theta_0,x)\left[-\dot{\Psi}^{-1}_{\theta_0,\bullet}[\tilde{\mathbb{H}}_2(\theta_0(\bullet),\bullet)]\right] := \mathbb{H}_2^*(y,x) \text{ in } \ell^\infty(\mathcal{W}).$$

Similarly to Lemma A.1, under $\mathcal{H}_0$ of (1.2) or $\mathcal{H}_A$ of (1.3), given the strong mixing condition of Assumption 1, $\sqrt{T}(\hat{F}_X(\bar{x}) - F_X(\bar{x}))$ weakly converges to a tight mean zero Gaussian process. Now, let the measurable functions $\Gamma : \mathcal{W} \mapsto [0,1]$ be defined by $(y,x) \mapsto \Gamma(y,x)$ and the bounded maps $\Pi : \mathcal{H} \mapsto \mathbb{R}$ be defined by $f \mapsto \int f d\Pi$. Then it follows from Lemma D.1 in Chernozhukov et al. (2013) that the mapping $(\Gamma, \Pi) \mapsto \int \Gamma(\bullet, x) d\Pi(x)$ - with $\Gamma(\bullet, x) = \mathbb{1}\{\bullet \leq x\} F(\bullet|x)$ and $\Pi = F_X(\bullet)$ - is well defined and Hadamard differentiable at (Γ, Π). Given the Hadamard differentiability of the mapping $(\Gamma, \Pi) \mapsto \int \Gamma(\bullet, x) d\Pi(x)$, the result follows from an application of the functional delta method, where the Gaussian process $\mathbb{H}_2$ is given by

$$\mathbb{H}_2(y, x) := \int \mathbb{H}_2^*(y, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}) + \int F(y|\bar{x}) \mathbb{1}\{\bar{x} \leq x\} d\mathbb{H}_1(\infty, \bar{x}),$$

where $\mathbb{H}_1$ is the same tight mean zero Gaussian process described in Lemma A.1. □

Lemma 1.9.4. *Under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (1.11) and Assumptions 1-6,*

$$\sqrt{T}(\hat{Z}_T(y, x) - F_T^A(y, x)) \implies \mathbb{H}_1(y, x), \text{ in } \ell^\infty(\mathcal{W}),$$

$$\sqrt{T}(\hat{\Psi}_T(\theta, \tau) - \Psi_{F_T}(\theta, \tau)) \implies \tilde{\mathbb{H}}_2(\theta, \tau), \text{ in } \ell^\infty(\mathcal{T} \times \Theta),$$

where $F_T^A(y, x) = \int F_T\big(y|\bar{x}\big) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x})$, $\Psi_{F_T}(\theta, \tau) = \mathrm{E}_{F_T}\big[\psi(W_t, \theta, \tau)\big]$, and $(\mathbb{H}_1, \tilde{\mathbb{H}}_2)$ are the tight mean zero Gaussian processes derived in Lemmas A.1-A.2.

Proof. First, under Assumption 6, $F_T^A(y, x)$ is contiguous to $F\big(y, x, \theta_0\big)$, then the convergence of the process $v_T^{\theta_0}(y, x) := \sqrt{T}(\hat{F}_T(y, x, \hat{\theta}_T) - F_T(y, x, \theta_0))$ on Lemma A.3 imply that $\sqrt{T}(\hat{Z}_T(y, x) - F_T^A(y, x)) \implies \mathbb{H}_1(y, x)$ in $\ell^\infty(\mathcal{W})$. Under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (1.11) and Assumptions 1-6, $F_T(y|X_t)$ of (1.11) is a linear combination of two measures that are VC class with a square integrable envelope. From the convergence of the process $\sqrt{T}(\hat{\Psi}_T(\theta, \tau) - \Psi(\theta, \tau))$ in Lemma A.2 and an application of Lemma 2.8.7 in Van der Vaart and Wellner (2000), we have that $\sqrt{T}(\hat{\Psi}_T(\theta, \tau) - \Psi_{F_T}(\theta, \tau))$ weakly converges to $\tilde{\mathbb{H}}_2(\theta, \tau)$ in $\ell^\infty(\mathcal{T} \times \Theta)$. □

We define weak convergence conditional on the data in probability ($\underset{M}{\overset{\mathbb{P}}{\Longrightarrow}}$ -convergence) in the Hoffmann-Jørgensen sense, i.e., $\hat{X}_n \underset{M}{\overset{\mathbb{P}}{\Longrightarrow}} X$ in a metric space $\mathbb{D}$ denotes conditional bootstrap convergence in probability under $\mathbb{P}$, that is, $\sup_{f \in \ell^\infty(\mathcal{H})} |E_M f(\hat{X}_n) - E f(X_n)| \overset{\mathbb{P}}{\to} 0$. The subscript M denotes taking the expectation conditional on the data. The following lemma derives the convergence of the block bootstrap of empirical process for dependent observations.

Lemma 1.9.5. *Let $W_t = \{Y_{Tt}, X_{Tt}\}$ be a $(1 + d)$-dimensional triangular array with stationary rows satisfying Assumption 7 with marginal distribution P, and let $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ be a permissible VC class of measurable functions with a square integrable envelope function $\mathbb{F}$ satisfying $P(\mathbb{F})^p < \infty$, for $2 < p < \infty$. Conditional on the* *data* $W_1, \ldots, W_T$, *let* $W_1^*, \ldots, W_T^*$ *be generated according to the block bootstrap with block length* $\ell := \ell(T)$, *with* $\ell(T) \to \infty$ *as* $T \to \infty$. *Let* $v_T^*(y,x) := \sqrt{T}(\hat{Z}_T^*(y,x) - \hat{Z}_T(y,x))$ *be the block bootstrap version of the empirical process* $v_T(y,x) = \sqrt{T}(\hat{Z}_T(y,x) - F_{YX}(y,x))$. *Suppose that*

$$\limsup_{k\to\infty} k^q \beta(k) < \infty \text{ for some } q > p/(p-2) \text{ and that } P^*(\mathbb{F})^p < \infty \text{ for some } p > 2.$$

Assume that the block length $\ell(T)$ *also satisfies*

$$\ell(T) = O(T^\rho) \text{ for some } 0 < \rho < (p-2)/[2(p-1)].$$

Then

$$v_T^*(y,x) \underset{M}{\overset{\mathbb{P}}{\Longrightarrow}} \mathbb{H}_1(y,x), \text{ in } \ell^\infty(\mathcal{W}),$$

where $\mathbb{H}_1$ *is a tight mean zero Gaussian process as defined in Lemma A.1.*

Proof. The result follows directly from an application of Theorem 1 in Radulović (1996) or Theorem 11.26 in Kosorok (2007), slightly modified to address measurability. □

Lemma 1.9.6. *Under Assumptions 2-7, under* $\mathcal{H}_0$ *of (1.2), or* $\mathcal{H}_A$ *of (1.3), or under the local alternative* $\mathcal{H}_{A,T}$ *of (1.11),*

$$\sqrt{T}(\hat{F}_T^*(y,x,\hat{\theta}_T^*) - \hat{F}_T(y,x,\hat{\theta}_T)) \underset{M}{\overset{\mathbb{P}}{\Longrightarrow}} \mathbb{H}_2(y,x) \text{ in } \ell^\infty(\mathcal{W}),$$

where $\mathbb{H}_2$ *is the tight mean zero Gaussian process defined in Lemma A.3.*

Proof. Since $F(\cdot|\theta,\cdot)$ is Hadamard differentiable, by the chain rule for the Hadamard derivative and bootstrap convergence result of Lemma A.5 we can apply a functional delta-method for bootstrap in probability defined in Theorem 3.9.11 of Van der Vaart and Wellner (2000) that yields the result. □

1.9.2 Proofs

Proof of Theorem 1. To prove part (i), we consider the empirical processes $v_T(y,x) = \sqrt{T}(\hat{Z}_T(y,x) - F_{YX}(y,x))$ and $v_T^{\theta_0}(y,x) = \sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_0))$ defined in Lemma A.1 and Lemma A.3, respectively. Under $\mathcal{H}_0$ of (1.2), $F_{YX}(y,x) \equiv F(y,x,\theta_0)$, and we have

$$
\begin{aligned}
S_T =& T \int (\hat{Z}_T(y,x) - \hat{F}_T(y,x,\hat{\theta}_T))^2 d\hat{Z}_T(y,x) \\
=& T \int (\hat{Z}_T(y,x) - \hat{F}_T(y,x,\hat{\theta}_T) \pm F_{YX}(y,x))^2 d\hat{Z}_T(y,x) \\
=& \int (v_T(y,x) - v_T^{\theta_0}(y,x))^2 d\hat{Z}_T(y,x) \\
=& \int (v_T(y,x) - v_T^{\theta_0}(y,x))^2 dF_{YX}(y,x) \\
+& \int (v_T(y,x) - v_T^{\theta_0}(y,x))^2 d(\hat{Z}_T(y,x) - F_{YX}(y,x)).
\end{aligned}
$$

By Lemma A.1, we have $\sqrt{T}(\hat{Z}_T(y,x) - F_{YX}(y,x)) \implies \mathbb{H}_1(y,x)$ that is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$. Then

$$
S_T = \int (v_T(y,x) - v_T^{\theta_0}(y,x))^2 dF_{YX}(y,x) + o_P(1).
$$

By Lemmas A.1 and A.3, $(v_T(y,x), v_T^{\theta_0}(y,x)) \implies (\mathbb{H}_1(y,x), \mathbb{H}_2(y,x))$ in $\ell^\infty(\mathcal{W} \times \mathcal{W})$. Then the result follows by an application of the continuous mapping theorem.

In part (ii), under the alternative hypothesis $\mathcal{H}_A$ of (1.3), $F_{YX}(y,x) \neq F(y,x,\theta_1)$ for some $(y,x) \in \mathcal{W}$ and for all $\theta_1 \in \mathcal{B}(\mathcal{T},\Theta)$. Now the process $v_T^{\theta_0}(y,x)$ becomes $v_T^{\theta_0}(y,x) = \sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_1))$. Then

$$
\begin{aligned}
S_T = & \; T \int \left(\hat{Z}_T(y,x) - \hat{F}_T(y,x,\hat{\theta}_T) \pm F_{YX}(y,x) \pm F(y,x,\theta_1) \right)^2 dF_{YX}(y,x) \\
= & \int \left(v_T(y,x) - v_T^{\theta_0}(y,x) + \sqrt{T}\left(F_{YX}(y,x) - F(y,x,\theta_1)\right) \right)^2 dF_{YX}(y,x) + o_P(1).
\end{aligned}
$$

By Lemmas A.1 and A.3, $(v_T(y,x), v_T^{\theta_0}(y,x)) \implies (\mathbb{H}_1(y,x), \mathbb{H}_2(y,x))$ in $\ell^\infty(\mathcal{W} \times \mathcal{W})$. Therefore, for any fixed constant $\varepsilon > 0$, $\lim_{T\to\infty} \Pr(S_T > \varepsilon) = 1$ and the result follows. $\square$

Proof of Theorem 2. Under the local alternative $\mathcal{H}_{A,T}$ in (1.11), consider the empirical processes

$$
\begin{aligned}
v_T^1(y,x) &= \sqrt{T}\left(\hat{Z}_T(y,x) - \int F(y|\theta_0,\bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}) \right), \text{ and} \\
r_T^1(\theta,\tau) &= \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \mathrm{E}_F[\psi\left(W_t,\theta,\tau\right)]\big),
\end{aligned}
$$

where $\Psi_F(\theta,\tau) := \mathrm{E}_F[\psi(W_t,\theta,\tau)]$ as defined in (1.12). Then

$$\begin{aligned}
v_T^1(y,x) =& \sqrt{T}\left(\hat{Z}_T(y,x) - \int F(y|\theta_0,\bar{x})\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right)\\
=& \sqrt{T}\left(\hat{Z}_T(y,x) - \int \left[F_T(y|\bar{x}) + \frac{\delta}{\sqrt{T}}\left(F(y|\theta_0,\bar{x}) - J(y|\bar{x})\right)\right]\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right)\\
=& \sqrt{T}\left(\hat{Z}_T(y,x) - F_T^A(y,x) + \frac{\delta}{\sqrt{T}}\int \left(J(y|\bar{x}) - F(y|\theta_0,\bar{x})\right))\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right).
\end{aligned}$$

Thus, it follows from Lemma A.4 that

$$v_T^1(y,x) \implies \mathbb{H}_1(y,x) + \delta\int (J(y|\bar{x}) - F(y|\theta_0,\bar{x}))\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x}),$$

where $\mathbb{H}_1$ is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$ defined in Lemma A.1. Now we have that

$$\begin{aligned}
r_T^1(\theta,\tau) =& \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \mathrm{E}_F[\psi\left(W_t,\theta,\tau\right)]\big)\\
=& \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \{\mathrm{E}_{F_T}[\psi\left(W_t,\theta,\tau\right)] + \delta\,\mathrm{E}_F[\psi(W_t,\theta,\tau)] - \delta\,\mathrm{E}_J[\psi(W_t,\theta,\tau)]\}\,\big)\\
=& \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \Psi_{F_T}(\theta,\tau) + \delta\left[\mathrm{E}_J[\psi(W_t,\theta,\tau)] - \mathrm{E}_F[\psi(W_t,\theta,\tau)]\right]\big),
\end{aligned}$$

where $\Psi_J(\theta,\tau) := \mathrm{E}_J[\psi(W_t,\theta,\tau)]$ as defined in (1.13). Thus, by Lemma A.4, we have

$$r_T^1(\theta,\tau) \implies \tilde{\mathbb{H}}_2(\theta,\tau) + \delta\big[\,\mathrm{E}_J[\psi(W_t,\theta,\tau)] - \mathrm{E}_F[\psi(W_t,\theta,\tau)]\big],$$

where $\tilde{\mathbb{H}}_2$ is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{T}\times\Theta)$ defined in Lemma A.2. Now, we consider the empirical process $v_T^{1\theta_0}(y,x)$

$$v_T^{1\theta_0}(y,x) = \sqrt{T}\left(\int F(y|\hat{\theta}_T,\bar{x})\mathbb{1}\{\bar{x}\le x\}d\hat{F}_X(\bar{x}) - \int F(y|\theta_0,\bar{x})\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right).$$

Thus, by Lemma A.3, we have that

$$v_T^{1\theta_0}(y,x) \implies \mathbb{H}_2(y,x) + \delta\int \dot{F}(y|\bar{x})[h]\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x}),$$

with $h(\tau) = [\frac{\partial}{\partial\theta}\Psi_F(\theta_0,\tau)]^{-1}\Psi_J(\theta_0,\tau)$. Therefore, under $\mathcal{H}_{A,T}$ of (1.11), we have

$$
\begin{aligned}
S_T =& T\int\left(\hat{Z}_T(y,x) - \hat{F}_T(y,x,\hat{\theta}_T) \pm \int F(y|\theta_0,\bar{x})\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right)^2 d\hat{Z}_T(y,x)\\
=& \int (v_T^1(y,x) - v_T^{1\theta_0}(y,x))^2 d\hat{Z}_T(y,x)\\
=& \int (v_T^1(y,x) - v_T^{1\theta_0}(y,x))^2 dF_{YX}(y,x)\\
+& \int (v_T^1(y,x) - v_T^{1\theta_0}(y,x))^2 d(\hat{Z}_T(y,x) - F_{YX}(y,x))\\
=& \int (v_T^1(y,x) - v_T^{1\theta_0}(y,x))^2 dF_{YX}(y,x) + o_P(1),
\end{aligned}
$$

then the result follows from the continuous mapping theorem. □

Proof of Theorem 3. For part (i), by Lemma A.6, $\hat{c}_T^*(\alpha) = c(\alpha) + o_P(1)$, where $c(\alpha)$ satisfies $\Pr(S_T > c(\alpha)) = \alpha + o(1)$. Then as $T\to\infty$, $\Pr(S_T > \hat{c}_T^*(\alpha)) = \alpha + o(1)$. For part (ii), there exists a fixed constant $C > 0$ such that

$$
\begin{aligned}
\Pr(S_T \le \hat{c}_T^*(\alpha)) &= \Pr(S_T \le \hat{c}_T^*(\alpha), S_T \le C) + \Pr(S_T \le \hat{c}_T^*(\alpha), S_T > C)\\
&\le \Pr(S_T \le C) + \Pr(\hat{c}_T^*(\alpha) > C)\\
&\le o(1) + \varepsilon + o(1),
\end{aligned}
$$

where the first element of the third line follows from Theorem 1 - $\Pr(S_T \le C) = o(1)$ - and the rest of the third line is due to Lemmas A.5-A.6, that imply the block bootstrap critical value $\hat{c}_T^*(\alpha)$ is bounded in probability under fixed alternatives, i.e., for any $\varepsilon > 0$, there exists a fixed constant C such that $\Pr(\hat{c}_T^*(\alpha) > C) < \varepsilon + o(1)$. The result follows from an arbitrary choice of $\varepsilon > 0$. Part (iii) follows from an application of Theorem 4 of Andrews (1997) and Anderson's Lemma in Ibragimov and Has'minskii (1981). By Anderson's Lemma, since $\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x)$ has mean zero $\forall(y,x)\in\mathcal{W}$, under $\mathcal{H}_0$ we have

$$
\begin{aligned}
&\Pr\left(\int (\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x) + \Delta(y,x))^2 dF_{YX}(y,x) \ge c(\alpha)\right)\\
&\ge \Pr\left(\int (\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x))^2 dF_{YX}(y,x) \ge c(\alpha)\right)\\
&= \Pr\left(S_T \ge c(\alpha)\right) = \alpha.
\end{aligned}
$$

Thus, under a sequence of local alternatives, we have $\Pr\left(S_T > c(\alpha)\right) \geq \alpha + o(1)$. Under Assumption 6, the conditional distribution under a local alternative $F_T(\boldsymbol{\cdot}|\boldsymbol{\cdot})$ implies a sequence of distribution functions $Z_T(y,x)$ that is contiguous to the distribution function $F\big(y,x,\theta_0\big)$ given by $\int F(y|\theta_0,\bar{x}))\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$, under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (1.11). Since contiguity preserves convergence in probability to constants, under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (1.11) we have

$$\Pr\left(\int \left(\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x) + \Delta(y,x)\right)^2 dF_{YX}(y,x) \geq \hat{c}^*_T(\alpha)\right)$$

$$= \Pr\left(\int \left(\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x) + \Delta(y,x)\right)^2 dF_{YX}(y,x) \geq c(\alpha)\right) + o(1)$$

$$\geq \Pr\left(\int \left(\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x)\right)^2 dF_{YX}(y,x) \geq c(\alpha)\right)$$

$$= \Pr\left(S_T \geq c(\alpha)\right) \geq \alpha,$$

where equality holds when $\Delta(y,x) \equiv 0$ a.e., with $\Delta(y,x)$ the non-trivial shift function defined in Theorem 2. □

Proof of Proposition 1. Condition (ii) is equivalent to Assumption 1. Condition (i) ensures that, for each $\tau \in \mathcal{T}$, $\Psi(\theta,\tau) : \Theta \mapsto \mathbb{R}^K$ possess a unique zero at $\theta_0(\tau)$. By Lemma D.1 of Chernozhukov et al. (2013), Conditions (iii) and (iv) imply Hadamard differentiability of the map $\theta \mapsto F(\boldsymbol{\cdot}|\theta,\boldsymbol{\cdot})$, for each $\tau \in \mathcal{T}$ and for all $\theta \in \mathcal{B}(\mathcal{T},\Theta)$. Condition (iii) provides conditions for the check function $\psi(W_t,\hat{\theta}_T,\tau) = \big(\tau - \mathbb{1}\{Y_t - X_t'\hat{\theta}_T(\tau) \leq 0\}\big)$ to be differentiable, and thus for Assumption 4. Finally, The Lipschitz Condition (iv) ensures that the class of functions $\{\psi(W_t,\theta,\tau) = \big(\tau - \mathbb{1}\{Y_t - X_t'\theta(\tau) \leq 0\}\big) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a VC class. □

Proof of Proposition 2. Conditions C0-C7 and AN1-AN3 of Engle and Manganelli (2004) assure that conditions (i),(iii)-(v) of Proposition 1 hold for the distribution of Y_t given X_t implied by the CAViaR model. Thus, it follows from the proof of Proposition 1. □

Proof of Proposition 3. Condition (i) provides conditions for Assumption 1. Condition (ii) ensures that, for each $\tau \in \mathcal{T}$, $\Psi(\theta,\tau) : \Theta \mapsto \mathbb{R}^K$ possess a unique zero at $\theta_0(\tau)$. Let the check function be $\psi(W_t,\theta(y),\tau) = \mathbb{1}\{Y_t \leq y\}\ln(\Lambda(Y_{t-1}'\theta(y))) + (1 - \mathbb{1}\{Y_t \leq y\})\ln(1 - \Lambda(Y_{t-1}'\theta(y)))$. Then conditions (ii)-(v) imply that the mapping $\Psi(\theta,\tau) : \Theta \times \mathcal{I} \mapsto \mathbb{R}^K$ is continuous, where $\mathcal{I}$ is an open set containing $\mathcal{T}$. Besides, $\frac{\partial}{\partial\theta}\Psi(\theta,\tau) := \dot{\Psi}_{\theta,\tau}$ exists at $(\theta_0(\tau),\tau)$ and is continuous at $(\theta_0(\tau),\tau)$, for each $\tau \in \mathcal{T}$, with $\inf_{\tau\in\mathcal{T}} \inf_{\|h\|=1} ||\dot{\Psi}_{\theta_0,\tau}h|| > 0$. By Lemma E.1 of Chernozhukov et al. (2013), the mapping $\tau \mapsto \theta(\tau)$ is continuously differentiable. By Lemma E.2 of Chernozhukov et al. (2013), we have Hadamard differentiability of the map $\theta \mapsto F(\boldsymbol{\cdot}|\theta,\boldsymbol{\cdot})$, for each $\tau \in \mathcal{T}$ and for all $\theta \in \mathcal{B}(\mathcal{T},\Theta)$. □

Chapter 2

Testing for Granger-Causality in Quantiles

2.1 Introduction

The Granger-causality definition proposed by Granger (1969) is the fundamental concept for studying dynamic relationships between time series. According to this definition, a series Y_t is said to Granger-causes X_{t+l} if it incorporates information about the predictability for X_{t+l} encompassed nowhere else in some large information set, which includes X_{t-j} , $j \geq 0$. Although the concept of Granger-causality is defined in terms of the conditional distribution, the majority of papers have tested Granger-causality using conditional mean regression models in which the causal relations are linear. As a result, a conditional mean regression model cannot assess a tail causal relation or nonlinear causalities.

This paper proposes a parametric omnibus test of Granger-causality in quantiles. The proposed approach allows us to evaluate nonlinear causalities, causal relations in conditional quantiles, and provides a sufficient condition for Granger-causality when all quantiles are considered. The quantile regression approach provides a more detailed and flexible analysis of the entire conditional distribution than the conditional mean-regression analysis, that focus only on a single part of the conditional distribution. In addition, a quantile causal relation may contrast with a causality in the mean of the conditional distribution. While a relationship with mean-causality shifts at least a non-negligible number of quantiles, a tail causal relation does not necessarily imply a causality in the mean. For example, Lee and Yang (2012) show that money-income Granger-causality in the conditional mean is quite weak and unstable, while it is significant in tail quantiles in most data sets. Finally, the proposed test is equivalent to testing Granger-causality in distribution when all quantiles are considered. Rather than checking a necessary condition for Granger-causality, our approach analyses a continuous space of conditional quantile functions that fully characterizes the concept of Granger-causality in distribution.

Our test is an extension of the method proposed by Escanciano and Velasco (2010) in the context of testing parametric conditional quantile restrictions over a range of quantiles. The intuition is to specify, for each quantile of the conditional distribution, a parametric conditional quantile model for a series X_t contained in an information set without Y_{t-1}, and check if the innovations of this model are correlated with the series Y_{t-1}, included in a larger information set that contains X_{t-j} for $j \geq 1$. To our knowledge, testing for Granger-causality in quantiles by parametric methods in a flexible specification setting has not been analysed in the literature before. Our test statistic is a Cramér-von Mises (CvM) functional norm of quantile-marked empirical processes that characterizes the null hypothesis of Granger non-causality. We reject the null hypothesis that Y_{t-1} does not Granger-causes X_t whenever our test statistic is significantly different from zero, for any quantile over a continuum of quantile levels.

As the proposed test statistic is asymptotically non-pivotal and depends on the data generating process, we tabulate critical values via a subsampling method. The subsampling approach allows us to apply non-linear conditional quantile regression models. Although our proposed test is computationally demanding, it has many interesting theoretical features: it does not require the choice of smoothing parameters, is consistent against all fixed alternatives, and is asymptotically strictly unbiased against a sequence of Pitman's local alternatives.

Chuang, Kuan, and Lin (2009) and Yang, Tu, and Zeng (2014) estimated the quantile causal effects by quantile regressions and tested the hypothesis of Granger non-causality by performing the Sup-Wald test of Koenker and Machado (1999) in all quantiles. We extend their method in two ways. First, our test provides an omnibus type of property: it requires only a model for the marginal quantile regression (under the null of no causality), and then searches for rejections of the null hypothesis in every direction, while the Sup-Wald test requires a particular model specification for the quantile regression under the alternative hypothesis of causality. In addition, we allow for non-linear specifications of the quantile regression model under the null. Many causal relations are non-linear, see for instance Bouezmarni, Rombouts, and Taamouti (2012). Therefore, a test based on a linear quantile regression model cannot be applied to testing nonlinear causality.

Regarding nonparametric approaches, Hong, Liu, and Wang (2009) proposed a nonparametric test of causality in Value-at-Risk (VaR), but their method provides only a necessary condition for Granger-causality. Jeong, Härdle, and Song (2012) extended the idea of Zheng (1998) to transform conditional quantile restrictions into conditional mean restrictions to testing causality in quantiles; more recently, Taamouti et al. (2014) proposed a nonparametric test for conditional density based Granger-causality. However, both testing procedures of the Granger-causality null hypothesis require beta-mixing conditions in the data generating process and are based on kernel methods. We provide two advantages. First, except for the application of the subsampling, our proposed causality test does not need mixing assumptions, and it requires only α-mixing assumptions for the validity of the subsampling, which are less restrictive than beta-mixing conditions. Moreover, our approach does not require the choice of smoothing parameters. Eventually, our parametric test is able to identify all the patterns of causality in the conditional distribution for flexible linear and nonlinear models, while nonparametric methods hardly provide a clear interpretation of the causal relations.

As further contributions, we investigate the finite sample performance of our method on simulated data and we illustrate the empirical applicability of our setting by verifying the causal relation between the gold price, the USD/GBP exchange rate, and the oil price.

The rest of the paper is organized as follows. In Section 2, we propose a test statistic for the null hypothesis of non Granger-causality in quantiles. In Section 3, we derive the asymptotic limit distribution of our test statistic under the null and the alternative hypotheses. We also prove that our test statistic has nontrivial power against $\sqrt{T}$-local alternatives, with T the sample size, and we theoretically justify the validity of the subsampling approach in our framework. Section 4 presents Monte Carlo simulation results. In Section 5, we show an empirical application of our proposed test, and Section 6 concludes the paper.

2.2 An Omnibus Test for Granger-Causality in Quantiles

2.2.1 Testing Problem

Let $\{(Y_t, Z_t) : \Omega \times \Omega \mapsto \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t \in \mathbb{Z}\}$ be a strictly stationary and ergodic stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}_t$ is the $\sigma-$field $\mathcal{F}_t = \{(Y_s, Z_s), s \leq t\}$, with joint distribution function $F_{Y,Z}(y, z)$. Let $F_Y(y|Z)$ be the conditional distribution function of Y given Z, and we assume it is continuous for all $y \in \mathbb{R}$. For simplicity, we examine only univariate Markov processes of order one and Granger-causality in lags, but we can extend our results to multivariate Markov processes of order $d > 1$ and/or to instantaneous Granger-causality. We define the information set available at time t as $\boldsymbol{I}_t \equiv \left(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right)$, where $\boldsymbol{I}_t^Y := (Y_{t-1}, \dots Y_{t-T+1})' \in \mathbb{R}^{T-1}$ and $\boldsymbol{I}_t^Z := (Z_{t-1}, \dots Z_{t-T+1})' \in \mathbb{R}^{T-1}$, and A' denotes the transpose matrix of A.

According to Granger (1969), a random variable Z does not Granger cause another random variable Y when we are not better able to predict Y using all available information than if the information apart from the past of Z until $t-1$ had been used. We characterize the null hypothesis of Granger non-causality from Z to Y as follows:

$$\mathrm{H}_0^{Z \nrightarrow Y} : F_Y\left(y \middle| \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right) = F_Y\left(y \middle| \boldsymbol{I}_t^Y\right), \text{ for all } y \in \mathbb{R}. \tag{2.2.1}$$

We denote the null hypothesis of (2.2.1) as Granger non-causality in distribution. Since the estimation of the conditional distribution may be complicated in practice, many papers have tested Granger non-causality in mean, that is only a necessary condition (2.2.1). In this case, Z does not Granger cause Y in mean if

$$E\left(Y_t \middle| \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right) = E\left(Y_t \middle| \boldsymbol{I}_t^Y\right) \text{ a.s.}, \tag{2.2.2}$$

where $E\left(Y_t \middle| \mathcal{F}\right)$ denotes the mean of $F_Y\left(\boldsymbol{.} \middle| \mathcal{F}\right)$. Granger non-causality in mean of (2.2.2) can be easily extended to higher order moments, see for example Cheung and Ng (1996). However, causality in mean (or in higher moments) overlooks the dependence that may appear in conditional tails of the distribution. On the other hand, the Granger non-causality distribution of (2.2.1) does not inform us about the level where the causality exists, if (2.2.1) is rejected. Thus, we propose to test Granger non-causality in conditional quantiles, since it allows us to determine the pattern of causality and it provides a sufficient condition for testing Granger non-causality in distribution of (2.2.1), as the quantiles completely characterize a distribution. Let $Q_\alpha^{\mathcal{F}}\left(\boldsymbol{.}|\mathcal{F}\right)$ be the α-quantile of $F_Y\left(\boldsymbol{.} \middle| \mathcal{F}\right)$, we can equally test (2.2.1) as:

$$\mathrm{H}_0^{QC:Z \nrightarrow Y} : Q_\alpha^{Z,Y}\left(Y_t \middle| \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right) = Q_\alpha^{Y}\left(Y_t \middle| \boldsymbol{I}_t^Y\right), \text{ a.s. for all } \alpha \in \mathcal{T}, \tag{2.2.3}$$

where $\mathcal{T}$ is a compact set such that $\mathcal{T} \subset [0,1]$ and the conditional α-quantiles of Y satisfy the restrictions below

$$\begin{aligned}
&\Pr\left\{Y_t \leq Q_\alpha^{Y}\left(Y_t \middle| \boldsymbol{I}_t^Y\right) \middle| \boldsymbol{I}_t^Y\right\} := \alpha, \text{for all } \alpha \in \mathcal{T}, \\
&\Pr\left\{Y_t \leq Q_\alpha^{Z,Y}\left(Y_t \middle| \boldsymbol{I}_t^Z, \boldsymbol{I}_t^Y\right) \middle| \boldsymbol{I}_t^Z, \boldsymbol{I}_t^Y\right\} := \alpha, \text{for all } \alpha \in \mathcal{T}.
\end{aligned} \tag{2.2.4}$$

Since $\Pr\left\{Y_t \leq Q_\alpha^{\mathcal{F}}\left(Y_t \middle| \mathcal{F}\right) \middle| \mathcal{F}\right\} = E\left\{\mathbb{1}\left[Y_t \leq Q_\alpha\left(Y_t \middle| \mathcal{F}\right)\right] \middle| \mathcal{F}\right\}$, where $\mathbb{1}(a \leq b)$ is an indicator function of the event that a is less or equal than b, (2.2.3) is equivalent to

$$E\left\{\mathbb{1}\left[Y_t \leq Q_\alpha^{Z,Y}\left(Y_t \middle| \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right)\right] \middle| \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right\} = E\left\{\mathbb{1}\left[Y_t \leq Q_\alpha^{Y}\left(Y_t \middle| \boldsymbol{I}_t^Y\right)\right] \middle| \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right\}, \text{ a.s. for all } \alpha \in \mathcal{T}, \tag{2.2.5}$$

where the left-hand side of (2.2.5) is equal to the α-quantile of $F_Y\left(\boldsymbol{.}|\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right)$ by definition. We postulate a parametric model to estimate the α-th quantile of $F_Y\left(\boldsymbol{.}|\mathcal{F}\right)$, where we assume that $Q_\alpha^{\mathcal{F}}\left(\boldsymbol{.}|\mathcal{F}\right)$ is correctly specified by a parametric model $m(\boldsymbol{\cdot}, \theta(\alpha))$ belonging to a family of functions $\mathcal{M} = \left\{m(\boldsymbol{\cdot}, \theta(\alpha)) \middle| \theta(\boldsymbol{\cdot}) : \tau \mapsto \theta(\tau) \in \Theta \subset \mathbb{R}^p, \text{ for } \tau \in \mathcal{T} \subset [0,1]\right\}$.

Let $\mathcal{B} \subset \mathcal{M}$ be a family of uniformly bounded functions $\tau \mapsto \theta(\tau)$ such that $\theta(\tau) \in \Theta \subset \mathbb{R}^p$. Then, under the null hypothesis in (2.2.3), the α-conditional quantile $Q_\alpha^Y\left(\boldsymbol{.}|\boldsymbol{I}_t^Y\right)$ is correctly specified by a parametric model $m(\boldsymbol{I}_t^Y, \theta_0(\alpha))$, for some $\theta_0 \in \mathcal{B}$, using only the restricted information set $\boldsymbol{I}_t^Y$, and we redefine our testing problem in (2.2.3) as:

$$\mathcal{H}_0^{Z \nrightarrow Y} : E\left[\mathbb{1}\left(Y_t \leq m\left(\boldsymbol{I}_t^Y, \theta_0(\alpha)\right)\right)|\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right] = \alpha, \text{ a.s. for all } \alpha \in \mathcal{T}, \tag{2.2.6}$$

versus

$$\mathcal{H}_A^{Z \nrightarrow Y} : E\left[\mathbb{1}\left(Y_t \leq m\left(\boldsymbol{I}_t^Y, \theta_0(\alpha)\right)\right)|\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right] \neq \alpha, \text{ for some } \alpha \in \mathcal{T}, \tag{2.2.7}$$

with $m(\boldsymbol{I}_t^Y, \theta_0(\alpha))$ as the only element of $\mathcal{M}$ that is a possible candidate equivalent for the true conditional quantile $Q_\alpha\left(Y_t|\boldsymbol{I}_t^Y\right)$, for all $\alpha \in \mathcal{T}$. To simplify notation, we rewrite (2.2.6) as $\mathcal{H}_0^{Z \nrightarrow Y} : E\left[\Psi_{\alpha,t}(\theta_0)|\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right] = 0$ almost surely, for all $\alpha \in \mathcal{T}$, where

$$\Psi_{\alpha,t}(\theta_0) := \mathbb{1}\left(Y_t - m\left(\boldsymbol{I}_t^Y, \theta_0(\alpha)\right) \leq 0\right) - \alpha. \tag{2.2.8}$$

The null hypothesis implies the moment condition $E\left[\Psi_{\alpha,t}(\theta_0)\, w\left(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right)\right] = 0$ for all measurable functions $w\left(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right)$ such that $E\left[|w\left(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right)|\right] < \infty$ and all $\alpha \in \mathcal{T}$. Following Escanciano and Velasco (2010), we characterize, under a proper measure-theoretic argument, the null hypothesis (2.2.6) by the infinite set of unconditional moment restrictions as follows:

$$E\left\{\Psi_{\alpha,t}(\theta_0)\exp\left(i\mathbf{x}'\boldsymbol{I}_t^Y\boldsymbol{I}_t^Z\right)\right\} = 0, \text{ for all } \mathbf{x} \in \mathbb{R}^{T-1} \text{ and for all } \alpha \in \mathcal{T}, \tag{2.2.9}$$

where $w\left(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right) = \exp\left(i\mathbf{x}'\boldsymbol{I}_t^Y\boldsymbol{I}_t^Z\right)$ was chosen because it has obtained better power properties than other weighting functions, and $i = \sqrt{-1}$ is the imaginary root. We base our test on the sample analog of the moment restriction of (2.2.9)

$$v_T(\mathbf{x}, \alpha) := \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Psi_{\alpha,t}(\theta_n)\exp\left(i\mathbf{x}'\boldsymbol{I}_t^Y\boldsymbol{I}_t^Z\right), \tag{2.2.10}$$

where $\theta_n(\alpha)$ is a $\sqrt{T}$-consistent estimator of $\theta_0(\alpha)$, for all $\alpha \in \mathcal{T}$. Our framework applies for any $\sqrt{T}$-consistent estimator of $\theta_n(\alpha)$ satisfying some mild conditions (described in the next section) such as the quantile regression estimator by Koenker and Bassett (1978), the quantile autoregressive estimator by Koenker and Xiao (2006), and the CAViaR estimator by Engle and Manganelli (2004). Given our sample $\{(Y_t, Z_t) : 1 \leq t \leq T\}$, we define $v_T(\mathbf{x}, \alpha)$ as the quantile marked-residual process, indexed by $\mathbf{x} \in \mathbb{R}^{T-1}$ and $\alpha \in \mathcal{T}$. Our proposed test statistic GCQ_T is a Cramér-von Mises (CvM) functional norm of $v_T(\mathbf{x}, \alpha)$ defined as

$$GCQ_T := \int_{\mathcal{T}} \int_{\mathcal{X}} \left| v_T(\mathbf{x}, \alpha) \right|^2 dF_{\mathbf{x}}(\mathbf{x}) dF_{\alpha}(\alpha) \tag{2.2.11}$$

$$= \frac{1}{m(T-1)} \sum_{j=1}^{m} \sum_{i=1}^{T-1} \left| v_T(\mathbf{x}_i, \alpha_j) \right|^2,$$

where $F_{\mathbf{x}}(\centerdot)$ and $F_{\alpha}(\centerdot)$ are some integrating measures on $\mathcal{X}$ and $\mathcal{T}$ respectively, $\mathcal{X}$ is a generic compact subset of $\mathbb{R}^{T-1}$ containing the origin, and m is the size of a deterministic grid of equidistributed quantiles, $\{\alpha_j\}_{j=1}^m = \mathcal{T}_m$, used in the estimation of the parametric model $m\left(\centerdot, \theta_n(\alpha)\right)$. We may also estimate the test statistic of (2.2.11) when $m \to \infty$ and the grid $\{\alpha_j\}_{j=1}^m$ is obtained independently from a distribution on $\mathcal{T}$, see Escanciano and Velasco (2010) for more details. We chose the CvM functional norm because unreported simulations suggested that the Cramér-von Mises type statistics provide better size and power results than the ones implied by other continuous functional norms such as the Kolmogorov norm.

Under the assumptions described in the next section, the test statistic GCQ_T weakly converges to zero under the null hypothesis (2.2.6), and to a probability limit different than zero under the alternative (2.2.7). We reject the null hypothesis whenever we observe "large" values of GCQ_T.

2.2.2 Subsampling Critical Values

The null distribution of test statistic GCQ_T is asymptotically non-pivotal and depends on the data generating process (DGP), then we implement a subsampling procedure to calculate critical values for GCQ_T. Subsampling is a resampling method that provides an asymptotic inference under general conditions on the DGP, including the time series case. We can compute a subsampling realization of our test statistic GCQ_T as follows:

1. Draw a subsample of the variables $\{(Y_{b,t}, Z_{b,t}), 1 \leq t \leq T\}$ without replacement from the realized sample $\{(Y_t, Z_t), 1 \leq t \leq T\}$;

2. Using the subsampling data $\{(Y_{b,t}, Z_{b,t}), 1 \leq t \leq T\}$, compute estimates $v_{b,T}(\mathbf{x}, \alpha)$ of $v_T(\mathbf{x}, \alpha)$ and calculate the correspondent subsampling realization of the test statistic:

$$GCQ_{b,T} = \frac{1}{mb} \sum_{j=1}^{m} \sum_{i=1}^{b} \left| v_{b,T}(\mathbf{x}_i, \alpha_j) \right|^2.$$

We approximate the CDF of GCQ_T, $F_{GCQ_T}(w) = \Pr(GCQ_T \leq w)$, from the distribution of the realizations of $v_{b,T}(\mathbf{x}, \alpha)$ over the different subsamples of size $T - b$:

$$F^b_{GCQ_T}(w) = \frac{1}{T-b} \sum_{i=1}^{T-b} \mathbb{1}\left(GCQ^i_{b,T} \leq w\right), \ w \geq 0. \tag{2.2.12}$$

Our proposed test statistic GCQ_T rejects the null hypothesis (2.2.6) if $GCQ_T > c_{b,T}(1-\alpha)$ for some significance level $\alpha \in (0,1)$, where the critical value $c_{b,T}(1-\alpha)$ is the $(1-\alpha)$-th sample quantile of (2.2.12).

2.3 Asymptotic Theory

In this section, we derive the asymptotic distributions of our test statistic GCQ_T under the null and alternative hypothesis. We consider the process $v_T(\mathbf{x}, \alpha)$ of (2.2.10) as a mapping from $(\Omega, \mathcal{F}, P)$ taking values in $\ell^\infty(\mathcal{X} \times \mathcal{T})$, that is the set of all complex-valued uniformly bounded functions defined with the supremum metric, d_∞, and $\mathcal{B}_{d_\infty}$ is its Borel σ-algebra. Hereafter "$\Longrightarrow$" denotes the weak convergence on $(\mathcal{B}_{d_\infty}, d_\infty)$, and C is a fixed constant. Let $\mathcal{F}_t = \sigma\left(\boldsymbol{I}^Y_t, \boldsymbol{I}^Y_{t-1}, \ldots\right)$ be the σ-field generated up to time t, we define the α-quantile innovation, for each $t \in \mathbb{Z}$, as $\varepsilon_t(\alpha) := Y_t - Q_\alpha\left(\boldsymbol{I}^Y_t\right)$ and the parametric quantile error as $e_t(\theta(\alpha)) := Y_t - m\left(\boldsymbol{I}^Y_t, \theta(\alpha)\right)$. In addition, f_x denotes the density function of a conditional distribution function F_x. All limits are taken as $T \to \infty$, where T is the sample size. We maintain the following main assumptions to analyse the asymptotic behavior of our test statistic:

Assumption 8. *$\{(Y_t, Z_t) : t \in \mathbb{Z}\}$ is a strictly stationary and ergodic process, with $E\left[|\boldsymbol{I}^Y_0|^2\right] < C$. Under $\mathcal{H}^{Z \nrightarrow Y}_0$ of (2.2.6), $\{\Psi_{\alpha,t}(\theta_0(\alpha)), \mathcal{F}_t\}$ is a martingale difference sequence for all $\alpha \in \mathcal{T}$. The parametric family $m(\bullet, \theta_0(\alpha))$ is nondecreasing in α a.s. The family of distributions functions $\{F_x, x \in \mathbb{R}^{T-1}\}$ has Lebesgue densities $\{f_x, x \in \mathbb{R}^{T-1}\}$ that are equicontinuous and uniformly bounded away from zero for the quantiles of interest.*

Assumption 9. *For each $\theta_1 \in \mathcal{B}$,*

(a) There exists a vector of functions $g_{t-1} : \Theta \mapsto \mathbb{R}^{T-1}$, for $g_{t-1}(\theta_1(\alpha))$ $\mathcal{F}_{t-1}-$measurable for each $t \in \mathbb{Z}$ satisfying, for all $k < \infty$,

$$\sup_{1 \leq t \leq T, \ \|\theta_1 - \theta_2\|_{\mathcal{B}} \leq kT^{-1/2}} T^{1/2} \|m(\boldsymbol{I}^Y_t, \theta_2) - m(\boldsymbol{I}^Y_t, \theta_1) - (\theta_2 - \theta_1)' g_{t-1}(\theta_1)\|_{\mathcal{B}} = o_p(1).$$

(b) For all sufficiently small $\delta > 0$,

$$E\left[\sup_{\|\theta_1-\theta_2\|_{\mathcal{B}}\leq\delta}\left|\mathbb{1}(Y_t \leq m(\boldsymbol{I}_t^Y, \theta_1(\alpha))) - \mathbb{1}(Y_t \leq m(\boldsymbol{I}_t^Y, \theta_2(\alpha)))\right|\right] \leq C\delta, \text{ for all } \alpha \in \mathcal{T},$$

and

$$E\left[\sup_{|\alpha_1-\alpha_2|\leq\delta}\left|m(\boldsymbol{I}_t^Y, \theta_1(\alpha_1)) - m(\boldsymbol{I}_t^Y, \theta_1(\alpha_2))\right|\right] \leq C\delta.$$

(c) Uniformly in $\alpha \in \mathcal{T}$, $E\left|g_{t-1}(\theta_1(\alpha))\right|^2 < \infty$, *and uniformly in* $(x', \alpha)' \in \mathcal{X} \times \mathcal{T}$,

$$\left|\frac{1}{T}\sum_{t=1}^{T} g_{t-1}(\theta_0(\alpha))\exp(i\boldsymbol{x}'\boldsymbol{I}_t^Y\boldsymbol{I}_t^Z)f_{\boldsymbol{I}_t^Y}(m(\boldsymbol{I}_t^Y, \theta_0)) - E[g_{t-1}(\theta_0(\alpha))\exp(i\boldsymbol{x}'\boldsymbol{I}_t^Y\boldsymbol{I}_t^Z)f_{\boldsymbol{I}_t^Y}(m(\boldsymbol{I}_t^Y, \theta_0))]\right| = o_p(1).$$

Assumption 10. *Let* $N_{[\bullet]}(\delta, \mathcal{G}, \|.\|_{\mathcal{B}})$ *be the* δ*-bracketing number of a class of functions* $\mathcal{G}$ *with respect to a norm*$\|\bullet\|$*. The parametric space* Θ *is compact in* $\mathbb{R}^{T-1}$*. The true parameter* $\theta_0(\alpha)$ *belongs to the interior of* Θ *for each* $\alpha \in \mathcal{T}$*, and* $\theta_0 \in \mathcal{B}$*. The class* $\mathcal{B}$ *satisfies*

$$\int_0^\infty \left(\log\left(N_{[\bullet]}\left(\delta^2, \mathcal{B}, \|.\|_{\mathcal{B}}\right)\right)\right)^{1/2} d\delta < \infty.$$

Assumption 11. *The estimator* θ_n *satisfies that* $\Pr(\theta_n \in \mathcal{B}) \to 1$ *as* $T \to \infty$*, and the following asymptotic expansion under* $\mathcal{H}_0^{Z\nrightarrow Y}$ *of (2.2.6), uniformly in* $\alpha \in \mathcal{T}$,

$$\begin{aligned} Q_n(\alpha) =& \sqrt{T}(\theta_n(\alpha) - \theta_0(\alpha)) \\ =& \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\ell_\alpha(Y_t, \boldsymbol{I}_t^Y, \theta_0(\alpha)) + o_p(1), \end{aligned}$$

where $E\left[\ell_\alpha(Y_1, \boldsymbol{I}_0^Y, \theta_0(\alpha))\right] = 0$, $E\left[\ell_\alpha(Y_1, \boldsymbol{I}_0^Y, \theta_0(\alpha))\ell'_\alpha(Y_1, \boldsymbol{I}_0^Y, \theta_0(\alpha))\right]$ *exists and it is positive definite, and* $E\left[\ell_\alpha(Y_t, \boldsymbol{I}_t^Y, \theta_0(\alpha))\Psi_{\alpha,s}(\theta_0)\right] = 0$ *if* $t \neq s$*. As a process in* $\ell^\infty(\mathcal{T})$, $Q_n(\alpha)$ *converges weakly to a Gaussian process* $Q(\bullet)$ *with zero mean and covariance function*

$$K_Q(\alpha_1, \alpha_2) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[\ell_{\alpha_1}(Y_t, \boldsymbol{I}_t^Y, \theta_0(\alpha_1)) \times \ell_{\alpha_2}(Y_s, \boldsymbol{I}_s^Y, \theta_0(\alpha_2))\right].$$

Assumption 12. *Under the alternative hypothesis* $\mathcal{H}_A^{Z\nrightarrow Y}$ *of (2.2.7):*

(a) There exists a $\theta_1 \in \mathcal{B}$ *such that* $\|\theta_n - \theta_1\|_{\mathcal{B}} = o_p(1)$*;*

(b) $E\left\{\Psi_{\alpha,t}(\theta_0)\exp\left(i\boldsymbol{x}'\boldsymbol{I}_t^Y\boldsymbol{I}_t^Z\right)\right\} \neq 0$ *in a subset with positive Lebesgue measure on* $\mathcal{X}\times\mathcal{T}$.

Assumptions 1-5 are very similar to Assumptions A1-A5 of Escanciano and Velasco (2010), but we consider most of them in a context of a restricted information set $\boldsymbol{I}_t^Y$ rather than in the full one $(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z)$. Assumptions 1-3 provide conditions for the validity of a functional central limit theorem for empirical processes in the dependent data case. Assumption 4 is required to guarantee the asymptotic distribution of our test when nonlinear quantile regression models are applied, see for example Mukherjee (1999). Assumption 5 provides a sufficient condition for the estimator θ_n to be consistent under a fixed alternative hypothesis, see Angrist, Chernozhukov, and Fernández-Val (2006) for conditions to satisfy Assumption 5(a); Assumption 5(b) holds if $(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z)$ is bounded. Theorem 1 below is an application of Theorems 1-3 of Escanciano and Velasco (2010) and the continuous mapping theorem.

Theorem 1. *Under Assumptions 1-5, we have*

(i) Under the null hypothesis $\mathcal{H}_0^{Z\nrightarrow Y}$ *of (2.2.6),*

$$GCQ_T \xrightarrow{d} \int_{\mathcal{T}} \int_{\mathcal{X}} \left|\mathbb{G}_1(\boldsymbol{x}, \alpha)\right|^2 dF_{\boldsymbol{x}}(\boldsymbol{x}) dF_\alpha(\alpha),$$

where $\mathbb{G}_1$ *is a tight mean zero Gaussian process.*

(ii) Under the alternative hypothesis $\mathcal{H}_A^{Z\nrightarrow Y}$ *of (2.2.7), there exists an* $\varepsilon > 0$ *such that*

$$\lim_{T\to\infty} \Pr\left(GCQ_T > \varepsilon\right) = 1.$$

Theorem 1 shows that the asymptotic null distribution of GCQ_T is a functional of a zero-mean Gaussian process $\mathbb{G}_1$. By Theorem 1, we expect that GCQ_T is significantly positive whenever the null hypothesis $\mathcal{H}_0^{Z\nrightarrow Y}$ is violated.

2.3.1 Local Alternatives and Subsampling Validity

Now we analyze the asymptotic distribution of GCQ_T against a sequence of Pitman's local alternatives converging to the null hypothesis at rate $\sqrt{T}$, where T denotes the sample size. Under a sequence of local alternatives $\mathcal{H}_{A,T}^{Z\nrightarrow Y}$, we have

$$\mathcal{H}_{A,T}^{Z\nrightarrow Y} : E\left[\Psi_{\alpha,t}\left(\theta_0\right) | \boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z\right] = \delta_\alpha/\sqrt{T}, \text{ a.s. for all } \alpha \in \mathcal{T}, \tag{2.3.13}$$

where δ_α is a function satisfying the following assumption.

Assumption 13. *The function* $\delta_\alpha : \mathbb{R}^{T-1} \mapsto \mathbb{R}$ *has the following properties:*

(i) $E\{\sup_{\alpha\in\mathcal{T}} |\delta_\alpha(\boldsymbol{I}_t^Y)|\} < \infty$*;*

(ii) There exists a $\boldsymbol{I}_t^Y$-measurable random variable h_{t-1} with $E\left[h_{t-1}^2\right] < \infty$ such that, for all $t \in \mathbb{Z}$ and for all α_1, $\alpha_2 \in \mathcal{T}$, $|\delta_{\alpha_1}(\boldsymbol{I}_t^Y) - \delta_{\alpha_2}(\boldsymbol{I}_t^Y)| \leq h_{t-1}|\alpha_1 - \alpha_2|$, a.s..

Assumption 6 is analogous to Assumption A6 of Escanciano and Velasco (2010), but we consider just the restricted information set $\boldsymbol{I}_t^Y$. To ensure nontrivial local power of our proposed test, we need to impose an assumption on the estimator θ_n under a local alternative as given in (2.3.13). Then we modify Assumption 4 as follows:

Assumption 4'. *Under a local alternative $\mathcal{H}_{A,T}^{Z\nrightarrow Y}$ in (2.3.13),*

$$\sqrt{T}\left(\theta_n(\alpha) - \theta_0(\alpha)\right)) = \eta_a(\alpha) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \ell_\alpha\left(Y_t, \boldsymbol{I}_t^Y, \theta_0(\alpha)\right) + o_p(1),$$

uniformly in α, where ℓ_α satisfies the same conditions as in Assumption 4 and $\eta_a(\alpha) \in \mathbb{R}^{T-1}$ for each $\alpha \in \mathcal{T}$.

Assumption 4' can be applied to most quantile regression estimators in the literature, see for example Mukherjee (1999). Theorem 2 demonstrates that under a local alternative $\mathcal{H}_{A,T}^{Z\nrightarrow Y}$ of (2.3.13) the asymptotic distribution of GCQ_T has an extra shift function implying consistency against $\sqrt{T}$-alternatives. Theorem 2 follows from Theorem 4 of Escanciano and Velasco (2010) and the continuous mapping theorem.

Theorem 2. *Under the local alternatives $\mathrm{H}_{A,n}^{QC}$ in (2.3.13), Assumptions 1-3, 4' and 5, we have*

$$GCQ_T \xrightarrow{d} \int_{\mathcal{T}}\int_{\mathcal{X}} \left|\mathbb{G}_1(\boldsymbol{x}, \alpha) + \Delta(\boldsymbol{x}, \alpha)\right|^2 dF_{\boldsymbol{x}}(\boldsymbol{x})dF_\alpha(\alpha),$$

where $\Delta(\boldsymbol{x}, \alpha)$ is a non-trivial shift function.

Now we derive the asymptotic validness of the subsampling critical values described in Section 2.2. Although no mixing conditions are required for the convergence of GCQ_T, we need another assumption on the serial dependence of the data generating process to validate the subsampling theoretically. According to Politis, Romano, and Wolf (1999), Assumption 7 below is sufficient for the asymptotic validity of the critical values generated by the subsampling approach.

Assumption 14. *$\{(Y_t, Z_{t+1})' : t \in \mathbb{Z}\}$ is a strictly stationary strong mixing process with $\alpha-$mixing coefficients satisfying $\sum_{m=1}^{T} \alpha(m) = o(T)$, with*

$$\alpha(m) = \sup_{T\in\mathbb{Z}} \sup_{B\in\mathcal{F}_T, A\in\mathcal{P}_{T+m}} |\Pr(A \cap B) - \Pr(A)\Pr(B)|,$$

for $m \geq 1$, where $\mathcal{F}_T := \sigma\left(\boldsymbol{I}_t^Y, t \leq T\right)$ and $\mathcal{P}_T := \sigma(\boldsymbol{I}_t^Y, t \geq T)$.

The next result allows us to establish the asymptotic distribution of the subsampling test statistics. Since it is an application of Theorem 2 of Whang (2006) and Theorem 5 of Escanciano and Velasco (2010), we omit the proof.

Theorem 3. *Under Assumptions 1-7, $b/T \to 0$ and $b \to \infty$ as $T \to \infty$, we have:*

(i) Under the null hypothesis $\mathcal{H}_0^{Z \nrightarrow Y}$ in (2.2.6),

$$\lim_{T\to\infty} \Pr\left(GCQ_T > c_{b,T}(\alpha)\right) = \alpha.$$

(ii) Under the fixed alternative hypothesis $\mathcal{H}_A^{Z \nrightarrow Y}$ in (2.2.7),

$$\lim_{T\to\infty} \Pr\left(GCQ_T > c_{b,T}(\alpha)\right) = 1.$$

(iii) Under the local alternative $\mathcal{H}_{A,T}^{Z \nrightarrow Y}$ in (2.3.13),

$$\lim_{T\to\infty} \Pr(GCQ_T > c_{b,T}(\alpha)) \geq \alpha.$$

Theorem 3 shows that our test based on the subsampling critical value has asymptotically correct size, is consistent, and is able to detect alternatives tending to the null at the parametric rate $\sqrt{T}$. Since the asymptotic properties of the subsampling tests depend on the choice of the subsample b, we follow the approach of Sakov and Bickel (2000) and we choose a subsample of size $b = \left[kT^{2/5}\right]$, for different values of k, where $[\cdot]$ is the integer part of a number.

2.4 Monte Carlo Experiments

In this section, we perform Monte Carlo simulation experiments with data generating processes (DGPs) under the null and the alternative hypothesis to evaluate the finite-sample performance of our proposed test statistic. The data are generated from the following data-generating processes (DGPs):

$$\text{DGP1: } Y_t = 0.5Y_{t-1} + cZ_{t-1} + \varepsilon_{1t}, \text{ and } Z_t = \varepsilon_{2t}, \tag{2.4.14}$$

$$\text{DGP2: } Y_t = 0.5Y_{t-1} + cZ_{t-1} + \varepsilon_{1t}, \text{ and } Z_t = 1 + 0.8Z_{t-1} + \varepsilon_{2t}, \tag{2.4.15}$$

$$\text{DGP3: } Y_t = 0.5Y_{t-1} + cZ_{t-1}^2 + \varepsilon_{1t}, \text{ and } Z_t = 1 + 0.8Z_{t-1} + \varepsilon_{2t}, \tag{2.4.16}$$

where $\varepsilon_{it} \sim$ i.i.d. $N(0,1)$, for $i = 1, 2$. For all DGPs above, under the null hypothesis $c = 0.00$, where the coefficient c captures the degree of causality from past values of Z_t to Y_t, thus a higher absolute value of c implies a stronger causality. The coefficients of the DGPs above assure the generated time series are stationary. We consider the empirical rejection frequencies for 5% nominal level tests for different sample sizes T, sub-sample sizes b, conditional quantile parametric models $m(\cdot,\cdot)$, and causality parameters c. We consider the sample sizes $T = 100$, $T = 250$ and $T = 500$. To show that our test is robust to the choice of the sub-sample size, we use three different sub-sample sizes $b = \left[kT^{2/5}\right]$ for each sample T, where $[\cdot]$ is the integer part of a number, for $k = 3, 4$ and 5. Thus for $k = 3, 4$ and 5, we have $b = 18, 25$ and 31 for $T = 100$, $b = 27, 36$ and 45 for $T = 250$ and $b = 36, 48$ and 60 for $T = 500$. We propose three different parametric quantile auto-regressive specifications $m(\cdot)$ - quantile AR(1), AR(2) and AR(3) - for modeling the quantiles of Y_t, for all $\alpha \in \mathcal{T}$, as follows:

$$
\begin{aligned}
&m^1(\boldsymbol{I}_t^Y, \theta_n(\alpha)) = \mu_0(\alpha) + \mu_1(\alpha)Y_{t-1} + \sigma_t\Phi_\varepsilon^{-1}(\alpha),\\
&m^2(\boldsymbol{I}_t^Y, \theta_n(\alpha)) = \mu_0(\alpha) + \mu_1(\alpha)Y_{t-1} + \mu_2(\alpha)Y_{t-2} + \sigma_t\Phi_\varepsilon^{-1}(\alpha),\\
&m^3(\boldsymbol{I}_t^Y, \theta_n(\alpha)) = \mu_0(\alpha) + \mu_1(\alpha)Y_{t-1} + \mu_2(\alpha)Y_{t-2} + \mu_3(\alpha)Y_{t-3} + \sigma_t\Phi_\varepsilon^{-1}(\alpha),
\end{aligned}
\tag{2.4.17}
$$

where the parameters $\theta_n(\alpha) = (\mu_0(\alpha), \mu_1(\alpha), \mu_2(\alpha), \mu_3(\alpha), \sigma_t)'$ are estimated by maximum likelihood in an equally spaced grid of 20 quantiles on the interval $\mathcal{T} = [0.10, 0.90]$. For each of the models, we denote our test statistic as $GCQ_{T,d}$ for $\boldsymbol{I}_t^Y = \{Y_{t-1}, Y_{t-d}\}$, for $d = 1, 2, 3$. For $c = 0.00$, there is no causality from past values of Z_t to Y_t and the rejection rates denote the empirical sizes. For $c \neq 0.00$, there is causality from lagged values of Z_t to Y_t and the rejection rates yield the empirical power of our test statistic. We apply 1,000 Monte Carlo replications in each of the simulations, which implies a maximal simulation standard error for the empirical sizes and powers of the test of $\max_p \sqrt{p(1-p)/1000} \approx 0.016$.

Table 1 shows the rejection frequencies of the GCQ_T test[1]. The proposed test has good power even when the degree of causality c is low. Besides, the GCQ_T has small size distortions even when sample size is small. As DGP3 is presented in Jeong et al. (2012), Table 1 shows that our test not only outperforms their test for $T = 500$, but also that it obtains reliable results for a smaller sample size of $T = 100$. For three different DGPs and conditional quantile regression models, the power of the GCQ_T test increases with the sample size, and these results are also robust to different sub-sample sizes.

We also compare our results with the Sup-Wald test statistic proposed by Koenker and Machado (1999). To calculate the Sup-Wald test statistic, we include the lagged values $\{Z_{t-1}, \ldots, Z_{t-d}\}$ in the linear conditional quantile regression models in (2.4.17). Without loss of generality, we assume that the quantile regression model is correctly specified if we include Z_{t-1} in the quantile regression model. Then we consider the following

[1]We do not include the results for $T = 250$ to save space.

specifications for the quantiles of Y_t:

W1: $m_*^1(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z, \theta_n(\alpha)) = \mu_0(\alpha) + \mu_1(\alpha)Y_{t-1} + \boldsymbol{\beta}_1(\alpha)Z_{t-1} + \sigma_t\Phi_\varepsilon^{-1}(\alpha)$,

W2: $m_*^2(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z, \theta_n(\alpha)) = \mu_0(\alpha) + \mu_1(\alpha)Y_{t-1} + \boldsymbol{\beta}_1(\alpha)Z_{t-1} + \mu_2(\alpha)Y_{t-2} + \sigma_t\Phi_\varepsilon^{-1}(\alpha)$,

W3: $m_*^3(\boldsymbol{I}_t^Y, \boldsymbol{I}_t^Z, \theta_n(\alpha)) = \mu_0(\alpha) + \mu_1(\alpha)Y_{t-1} + \boldsymbol{\beta}_1(\alpha)Z_{t-1} + \mu_2(\alpha)Y_{t-2} + \mu_3(\alpha)Y_{t-3} + \sigma_t\Phi_\varepsilon^{-1}(\alpha)$,

Given a conditional linear model in W1-W3, testing $\mathcal{H}_0^{Z \nrightarrow Y}$ of (2.2.6) consists in testing H_0^{SW} : $\boldsymbol{\beta}_1(\alpha) = 0$, for all $\alpha \in \mathcal{T}$. Table 2 gives the results for the Granger-causality tests based on the Sup-Wald test. A drawback of the Sup-Wald test is that the critical values do not have the correct nominal size in small samples, as the empirical sizes are always smaller than the 5% nominal level of the test. The results also suggest that the subsampling GCQ_T test considerably outperforms the Sup-Wald procedure in terms of power. For the DGPs considered, even using a small sub-sample size, $b = [3T^{(2/5)}]$, our test presents powerful and reliable inference. In addition, the subsampling GCQ_T test is robust to changes in the sub-sample size. In sum, our proposed test seems to perform quite well in finite samples.

Table 2.1. Empirical rejection frequencies for 5% subsampling $GCQ_{T,d}$ test

			$GCQ_{T,1}$ b			$GCQ_{T,2}$ b			$GCQ_{T,3}$ b		
DGP	T	c	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$	$k=3$	$k=4$	$k=5$
1											
	100	0.00	6.8	7.1	6.1	6.4	6.5	6.9	6.1	6.4	6.6
		0.01	6.2	6.9	8.6	7.2	6.9	6.4	7.0	6.8	6.4
		0.03	6.7	6.7	8.0	7.1	6.0	6.6	7.0	7.2	7.2
		0.06	7.4	7.4	7.4	8.9	8.9	7.5	6.9	8.1	8.1
		0.12	10.1	10.7	9.9	10.2	9.8	10.4	9.9	9.8	9.5
		0.24	24.4	23.4	21.3	22.7	20.8	18.9	21.0	20.3	18.4
		0.50	72.3	68.1	65.1	72.3	67.0	61.9	71.3	65.9	62.5
	500	0.00	5.2	4.9	5.0	5.2	5.0	5.1	5.4	5.3	5.3
		0.01	5.1	5.4	5.2	4.8	4.3	4.3	5.8	6.0	5.6
		0.03	6.2	5.8	6.4	7.4	7.4	7.4	5.9	5.4	5.2
		0.06	10.6	10.6	10.6	10.9	10.8	9.5	11.0	10.7	10.6
		0.12	33.1	32.4	29.6	29.5	27.9	26.6	33.4	30.7	28.4
		0.24	92.0	89.8	87.1	89.2	88.9	87.2	90.9	89.0	86.5
		0.50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2											
	100	0.00	6.2	6.7	7.2	6.0	6.0	5.9	6.2	6.3	6.7
		0.01	7.2	7.4	7.6	7.1	7.1	8.2	6.8	6.8	6.1
		0.03	6.7	7.0	7.3	7.2	7.5	7.4	6.1	6.1	6.5
		0.06	9.1	9.2	8.8	9.4	8.8	9.4	8.4	8.9	9.3
		0.12	17.6	14.8	16.0	15.8	15.2	15.9	15.8	14.7	14.7
		0.24	42.5	36.2	35.8	42.1	36.6	35.1	42.6	39.3	36.8
		0.50	87.3	81.6	79.2	84.8	78.6	73.5	87.4	81.1	77.2
	500	0.00	5.1	5.2	5.5	5.2	5.2	5.2	4.7	4.6	5.5
		0.01	5.7	5.5	5.7	6.5	5.8	5.8	5.0	4.9	5.1
		0.03	8.6	8.4	7.9	7.6	7.3	6.7	8.4	8.1	8.4
		0.06	20.8	20.3	20.1	18.5	18.2	16.9	20.2	20.4	18.7
		0.12	67.3	65.3	62.5	67.3	64.4	62.7	64.8	62.5	60.9
		0.24	99.8	99.8	99.7	99.7	99.6	99.3	99.9	100.0	99.9
		0.50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
3											
	100	0.00	6.3	6.0	6.1	6.1	5.9	7.3	6.4	5.8	6.2
		0.01	15.1	14.1	13.8	13.5	12.0	12.0	13.2	13.2	12.4
		0.03	56.7	52.4	50.2	56.5	53.8	50.0	54.8	49.5	46.0
		0.06	91.0	87.0	82.7	90.2	86.8	83.9	90.9	87.3	84.7
		0.12	98.6	97.7	96.4	97.4	96.6	95.3	98.5	98.5	95.2
		0.24	98.6	97.8	97.4	95.1	92.1	90.5	96.2	92.3	91.1
		0.50	91.6	88.9	87.7	83.4	79.5	77.8	83.8	78.2	75.6
	500	0.00	4.9	5.0	5.1	5.3	5.4	5.1	5.3	5.5	5.5
		0.01	52.6	50.0	49.6	52.2	50.2	48.5	50.3	49.0	47.1
		0.03	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		0.06	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		0.12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		0.24	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		0.50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: We use 1,000 Monte Carlo repetitions based on the DGPs 1-3 of equations (2.4.14)-(2.4.16); $b = [kT^{2/5}]$; $\alpha \in [0.10; 0.90]$.

Table 2.2. Empirical rejection frequencies for 5% Sup-Wald test

DGP	T	c	$W1$	$W2$	$W3$
1					
	100	0.00	2.80	2.50	1.30
		0.01	2.30	1.90	1.60
		0.03	2.20	2.50	3.00
		0.06	5.30	3.30	3.30
		0.12	8.30	6.10	7.70
		0.24	29.40	28.40	30.40
		0.50	92.40	90.60	90.40
	500	0.00	0.90	1.20	1.00
		0.01	0.60	0.60	0.50
		0.03	1.80	2.30	1.30
		0.06	6.80	6.40	4.70
		0.12	31.10	32.70	27.80
		0.24	95.40	95.50	95.30
		0.50	100.00	100.00	100.00
2					
	100	0.00	2.30	3.40	3.00
		0.01	2.00	2.40	2.70
		0.03	2.60	3.60	2.40
		0.06	6.50	6.20	6.80
		0.12	19.10	16.70	17.60
		0.24	64.60	62.30	62.70
		0.50	99.10	99.50	99.30
	500	0.00	1.50	1.60	1.00
		0.01	1.30	1.60	1.00
		0.03	4.90	3.80	5.50
		0.06	19.00	18.60	18.20
		0.12	80.00	80.30	78.70
		0.24	100.00	100.00	100.00
		0.50	100.00	100.00	100.00
3					
	100	0.00	2.20	2.30	1.90
		0.01	12.80	13.60	12.60
		0.03	82.20	79.90	78.50
		0.06	99.70	99.90	99.90
		0.12	100.00	100.00	100.00
		0.24	100.00	100.00	100.00
		0.50	100.00	100.00	100.00
	500	0.00	1.70	1.20	1.10
		0.01	60.60	63.10	58.20
		0.03	100.00	100.00	100.00
		0.06	100.00	100.00	100.00
		0.12	100.00	100.00	100.00
		0.24	100.00	100.00	100.00
		0.50	100.00	100.00	100.00

Note: $W1 - W3$ are the parametric quantile regression specifications described above; $\alpha \in [0.1, 0.9]$.

2.5 Empirical Application

To illustrate the applicability of our approach, we analyse the causality between gold prices, oil prices, and USD/GBP exchange rate. The gold and oil market are the main representatives of the large commodity markets. Gold is a valuable asset and can maintain its value in turbulent times. There are evidence that the gold and oil markets have a close interaction. For instance, both gold and crude oil prices entered into a boom time in 2002 due to US dollar depreciation, global inflation, and oil supply manipulation by the OPEC; and both commodity prices collapsed together in the financial crisis of 2008 (Zhang and Wei, 2010). Therefore it is important to study how gold and oil prices variate, and their causal relationship. While standard tests evidence a positive mean causal relation between oil and gold prices, our main goal is to evaluate such a relation on each quantile of the distribution.

We apply our GCQ_T test to check the relationship between the S&P gold prices (per ounce) and Brent crude oil prices (per barrel). Under our approach, we can discriminate between causality affecting the median and the tails of the conditional distribution. Then the empirical analysis should provide a more complete description of the causal relation between gold and oil prices. We also evaluate the effect of the USD/GBP exchange rate on gold prices to compare the performance of our parametric test with the nonparametric approach proposed by Jeong et al. (2012). The data consist of 3,440 daily observations - from July 2000 to September 2013 and all series were obtained from Datastream.

Figure 1 displays the daily log and log-difference series. It shows that the three return series display calm as well as volatile periods and also single outlying return observations. Besides, the graphs of the log series evidence the series are non-stationary and follow a common pattern. Table 3 presents the summary statistics of the series. The gold and oil prices are very volatile, and all series are positively skewed and leptokurtic. We apply the GCQ_T test on the log-difference of the series, as Dickey-Fuller and KPSS unit root tests show that the three log series are non-stationary.

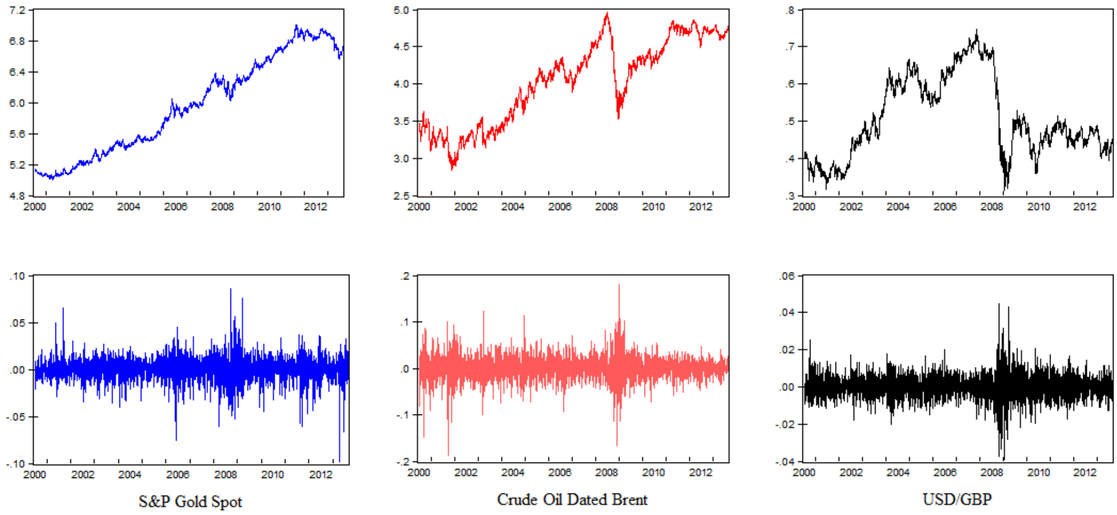

Figure 2.1. Log of the series (upper panel) and first-difference of the logs (lower panel).

Table 2.3. Summary statistics:

	Gold Prices	Oil Prices	USD/GBP
Mean	466.89	65.31	1.67
Std. Dev.	282.83	33.23	0.18
Median	382.74	62.06	1.61
Skewness	0.64	0.32	0.49
Kurtosis	2.04	1.81	2.05
Minimum	149.39	17.00	1.37
Maximum	1101.48	143.60	2.11

Note: Gold is the S&P GSCI Gold Spot price index; Oil price is the price adjusted default Crude Oil Dated Brent in US dollars per barrel; USD/GBP is the exchange rate of US dollars to UK british pounds; The data covers the period that spans 03 July 2000 to 06 September 2013.

We estimate three quantile auto-regressive QAR models as in (2.4.17) for each dependent variable on the GCQ_T test. Tables 4 and 5 report the subsampling p-values of our GCQ_T test. If we take into account all quantiles, the results suggest that variations in the oil prices Granger-cause variations in the gold prices, and vice-versa, at the 1% significance level. However, if we perform a median-regression and consider only $\alpha = 0.50$, we do not reject the null hypothesis that variations in the oil prices do not Granger-cause variations in the gold prices. Therefore, our approach is robust to detect tail causalities that could possibly be ignored by using a standard conditional regression model analysis.

In addition, changes in the USD/GBP Granger-cause changes in the gold and oil prices at the 1% significance level, for all quantiles. If we consider only the extreme tails of the conditional distribution, we cannot always reject at 1% significance level that variations of USD/GBP Granger-cause variations in gold prices and in oil prices.

However, for $\alpha = 0.10$ and $\alpha = 0.90$, we often do not reject that changes of USD/GBP do not Granger-cause variations in gold prices and in oil prices at 1% significance level. These results are consistent with the ones obtained by Jeong et al. (2012), who showed that USD/GBP exchange rate changes do not cause the gold price change if $\alpha < 0.22$ or $\alpha > 0.80$.

Table 2.4. Causality between ΔUSD/GBP and ΔOil prices to ΔGold prices - subsampling p-values

T=3,440	α	$GCQ_{T,1}$	$GCQ_{T,2}$	$GCQ_{T,3}$
ΔOil to ΔGold	$[0.10; 0.90]$	0.000	0.000	0.000
	0.10	0.000	0.000	0.000
	0.50	0.294	0.265	0.323
	0.90	0.000	0.000	0.000
ΔUSD/GBP to ΔGold	$[0.10; 0.90]$	0.000	0.000	0.000
	0.10	0.011	0.010	0.006
	0.50	0.004	0.006	0.007
	0.90	0.010	0.007	0.013

Note: For $\alpha \in [0.10, 0.90]$.

Table 2.5. Causality between ΔUSD/GBP and ΔGold prices to ΔOil prices - subsampling p-values

T=3,440	α	$GCQ_{T,1}$	$GCQ_{T,2}$	$GCQ_{T,3}$
ΔGold to ΔOil	$[0.10; 0.90]$	0.000	0.000	0.000
	0.10	0.000	0.000	0.000
	0.50	0.461	0.345	0.384
	0.90	0.000	0.000	0.000
ΔUSD/GBP to ΔOil	$[0.10; 0.90]$	0.000	0.000	0.000
	0.10	0.011	0.010	0.006
	0.50	0.005	0.006	0.006
	0.90	0.010	0.006	0.012

Note: For $\alpha \in [0.10, 0.90]$.

2.6 Conclusions

Many important policy and financial analyses are investigated through testing for Granger-causality between economic time series. However, most of the results in the literature were obtained in the context of Granger-causality in mean. In this paper, we present a consistent parametric test of Granger-causality in quantiles. Rather than focusing on a single part of the conditional distribution, we develop a test that evaluates possible causal relations in all conditional quantiles. The proposed test statistic has correct asymptotic size, is consistent against fixed alternatives and has power against Pitman deviations from the null hypothesis. In addition, the proposed approach allows us to evaluate nonlinear causalities, causal relations in conditional quantiles, and provides a sufficient condition for Granger-causality when all quantiles are considered.

Finite sample experiments suggest that our proposed test has good size and power properties, and is more powerful than other comparable test in the literature against almost all alternatives. An empirical application highlights the practical importance of our setting considering the causal relation between the gold price, the USD/GBP exchange rate, and the oil price. We illustrate that oil price, USD/GBP, and gold price changes presented a different causal relationship in the tail and in the center of the distribution.

A possible direction for future work is to extend this method to analyse the effect of misspecifications in the quantile regression model to Granger-causality. A possibly misspecified quantile regression model may lead to over-rejections of the Granger-noncausality null hypothesis.

Chapter 3

Stock Market Equilibrium Error and Expected Excess Stock Returns

3.1 Introduction

Many studies have investigated the predictability of stock returns in time series data. Campbell and Shiller (1988), and Fama and French (1988) found that valuation ratios such as the dividend-price ratio or earnings-price ratio are positively related to subsequent stock returns and that the implied predictability is large at longer horizons. Fama and Schwert (1977), Campbell (1987), and Fama and French (1989) found that variables such as the term premium, the default premium and the yield on corporate bonds forecast subsequent stock returns. Other papers suggested new predictor variables using information from interest rates (Hodrick, 1992), the consumption-wealth ratio (Lettau and Ludvigson, 2001), and the relative valuations of high- and low-beta stocks (Polk, Thompson, and Vuolteenaho, 2006).

However, many authors cast doubt on the evidence of predictability of stock returns. Nelson and Kim (1993) and Stambaugh (1999) showed that many predictor variables in the literature are persistent, which lead to biased coefficients in forecasting models if innovations in the predictor variable are correlated with stock returns. Besides, under these conditions, the t-test for predictability is biased (Cavanagh, Elliott, and Stock, 1995). Kilian (1999), Campbell and Yogo (2006) and Jansson and Moreira (2006), among others, propose alternative econometric methods for addressing the size bias and performing valid inference under persistence. Another criticism on the stock returns predictability question the poor out-of-sample performance of predictive regressions (Bossaerts and Hillion, 1999, Goyal and Welch, 2003, Welch and Goyal, 2008). Welch and Goyal (2008) compare predictive regressions with a benchmark of historical average stock returns and show that predictive regressions almost never provide superior stock return predictability.

Although Inoue and Kilian (2004) argue that in-sample tests are more powerful and not necessarily less reliable than out-of-sample tests, in this paper we provide new evidence on the out-of-sample predictability of stock returns. We take up the challenge of Welch and Goyal (2008) and Campbell and Thompson (2008), and we compare the forecasting performance of some variables with the benchmark of historical average stock returns. We show that predictive regressions that include variables like the Fama-French factors, the previous month's return and the equilibrium error term from the co-integrating relation with stock market return enhance the out-of-sample predictability of stock returns and provide profitable market-timing portfolio strategies.

We use the Fama-French 30 industry portfolio returns and obtain the equilibrium error factor, $EE_{i,t}$, as the error term from the co-integration relationship between industry stock returns and excess stock market returns. We find that the equilibrium error factor ($EE_{i,t}$) leads to remarkable out-of-sample forecasting abilities, which are increased if the previous month's excess industry stock returns is included in the predictive regression. Our results show that the omission of the previous month's excess industry stock returns may lead to a biased relation between the stock returns and the equilibrium error factor.

We evaluate the economic benefits of stock returns predictability of our forecasting models, as in Johannes, Polson, and Stroud (2002) and Guo (2006), among others. First, we take the perspective of an investor who uses predictability from a model of time-varying expected returns to sequentially build portfolios. Following Breen, Glosten, and Jagannathan (1989) and Pesaran and Timmermann (1995), we take the case of an investor who holds stocks of the i-th industry if the predicted excess industry return is positive and holds bonds if there is no positive expected excess industry return. We also consider a model of time-varying expected returns and volatility. For each period, an investor allocates his wealth between the i-th industry stock according to an optimal portfolio rule, derived from an extension of Stein's lemma (Johannes et al., 2002). We compare the generated returns to those implied by a model without predictability and to the excess stock market return.

We find that strategies based on time-varying expected returns and volatility provide higher annualized mean returns and Sharpe ratios than historical mean average returns or the market. For example, an investor with a risk aversion parameter of 5 who adopts the optimal portfolio strategy obtain an annualized Sharpe Ratio of 45.1%, compared with 37.3% for the no predictability strategy, 41.0% for the predictability strategy without the equilibrium error factor in the predictive regression. Moreover, the optimal portfolio strategy generates an annualized certainty equivalence gain of 2.7% relative to the model of historical average returns.

We provide further tests to demonstrate the economic gains of the stock return predictability. Following Cumby and Modest (1987) and Breen et al. (1989), we reject the null hypothesis that our predicted excess returns have no market timing ability. The Jensen's α test for supports that our predicted returns cannot be explained neither by the CAPM nor by the Fama and French (1993) model. Moreover, these results are are robust in the presence of transaction costs when the investor pays a fee for switching his portfolio. Therefore, our results are consistent with Pesaran and Timmermann (1995), Johannes et al. (2002), and Guo (2006), who find economic gains from time-varying trading strategies.

Our results differ with those of Bossaerts and Hillion (1999), Goyal and Welch (2003) and Welch and Goyal (2008), who found that there is no predictability of stock returns. A possible reason for these contrasts might be that our forecasting variables may discard the variables used by those authors and the equilibrium error factor is a panel variable that uses more information than time series variables. However, we include the stochastically detrended risk-free rate ($RREL$) suggested by Campbell, Lo, and MacKinlay (1997), which was used by these authors and provide substantial information about subsequent stock returns.

We choose our forecasting variables motivated by the common view that expected stock returns have a mean-reverting component (Campbell and Shiller, 2001, Merton, 1971). This mean-reverting component may be captured by the equilibrium error factor, which may reflect a short-term reversal, momentum or liquidity premium effect. The equilibrium error factor appears to be a pervasive variable that captures systematic movements of stock returns. For example, we find that the equilibrium error factor is always significantly negative in the predictive regression of subsequent stock returns.

We compare the equilibrium error term with aggregate liquidity measures and short-term reversal measures, which forecast stock returns (Amihud, 2002, Jones, 2002). We find the equilibrium error and short term reversal measures capture similar forecasting information of stock returns. Therefore, the equilibrium error is an omitted short term reversal factor that is negatively related to stock returns. We include the three Fama-French factors in our model as we need to apply a pricing model to be consistent with the methodology used to risk-adjust the returns. Over the period Jul.1968-Dec.2007, we find that although the equilibrium error factor and the previous month's excess stock return, $R_i(-1)$, have negligible forecasting power in the in-sample regression, they jointly provide a significant predictor of excess stock returns. We find very similar results using subsamples. Moreover, their predictive abilities are also statistically significant in the out-of-sample tests.

Another contribution of this paper is the control for industry effects, as we estimate the returns using fixed-effects panel data methods. The industry control reduces forecast biases that are constant across stocks within the same industry (Da, Liu, and Schaumburg, 2014). Besides, the industry control eliminates common trends between expected returns and discount rate news. Moskowitz and Grinblatt (1999) show that there is a significant momentum effect in industry components of stock returns. Thus, the industry control increases the short-term return reversal effect, by removing the industry moment effect.

The remainder of this paper is organized as follows. Section 2 discusses the data and reports in-sample and out-of-sample forecast results. Section 3 presents the analysis of trading strategies using the out-of-sample predictability of stock returns from our forecasting models. In Section 4, we investigate whether $EE_{i,t}$ is related to short-term return reversals and aggregate stock market liquidity. Section 5 concludes the paper.

3.2 Forecasting Excess Stock Returns

3.2.1 Data

Our data include monthly returns of NYSE, AMEX, and NASDAQ common stocks from July 1965 to December 2007. We use the data from the 30 industry portfolios of Kenneth French's website, where each NYSE, AMEX, and NASDAQ stock is assigned to an industry portfolio at the end of June of year t based on its four-digit SIC code at that time. To calculate the market return at time t, $R_{M,t}$, we use the value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ at the beginning of month t. We use the one-month Treasury Bill rate as the risk-free rate at time t, $R_{F,t}$, and the cumulative market returns over the six months for month $t-6$ to month $t-1$ as the momentum at time t, WML_t, and the difference between the nominal risk-free rate and its last four-quarter average as the stochastically detrended risk-free rate at time t, $RREL_t$. We define the excess return for the industry i in month t as $R_{i,t} - R_{F,t}$.

To compute the expected returns, we use a pricing model. There is a long debate about pricing models in the literature. While the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) suffers from a number of limitations to explain patterns in average stock returns, called as "asset pricing anomalies", Fama and French (1993) suggested that the CAPM should be augmented with two additional factors, HML and SMB, and showed that their three-factor ($F\&F$) model explains well stock returns. HML is the return on a portfolio that is long in stocks with high book-to-market value ratios and short in stocks with low book-to-market value ratios, and SMB is the return on a portfolio that is long in small stocks and short in big stocks. The monthly Fama and French factors, HML and SMB, were obtained from Kenneth French at Darmouth College.

We use the Fama-French 30 industry portfolio returns and the excess stock market return, $R_{M,t} - R_{F,t}$, to obtain the equilibrium error factor, $EE_{i,t}$, that is the error term from the co-integration relationship between industry stock returns and excess stock market returns. We calculate $EE_{i,t}$ in two steps. First, we perform the Augmented Dickey-Fuller test (with constant and trend) on the cumulative industry stock returns and on the cumulative excess stock market returns. If we do not reject the null hypothesis of unit root at the 1% significance level for the cumulative i-th industry stock returns and the cumulative excess stock market returns, we proceed to Johansen (1988, 1995) 's co-integration test. If we cannot reject at the 5% level that the cumulative i-th industry stock returns is co-integrated with the cumulative excess stock market returns, we compute $EE_{i,t}$ as the error term from the vector error-correction model (VECM) between those variables, with 18 lags to assure that there is no serial correlation of the residuals.

For example, given a cumulative i-th industry portfolio returns that is I(1), we define $\mathbf{y}_t = (R_{i,t}, R_{M,t} - R_{F,t})'$ to apply the co-integration test of Johansen (1988, 1995) using a finite-order vector error-correction model (VECM) as follows:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}\mathbf{y}_{t-1} + \boldsymbol{\mu}) + \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j \Delta \mathbf{y}_{t-j} + \boldsymbol{\gamma} + \varepsilon_t, \tag{3.2.1}$$

where $\Delta\mathbf{y}_t$ is the L-operator applied to the vector $\mathbf{y}_t$, and $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Gamma}_j, \boldsymbol{\gamma}$ are 2×1 vectors of parameters, and p is the lag order of the VECM. Then for each i-th industry return co-integrated with the excess stock market return, we define the the equilibrium error factor, $EE_{i,t}$, as the error term ε_t from the VECM in (3.2.1) between the cumulative industry and the cumulative excess stock market return.

Table 3.1 presents summary statistics of the excess industry stock return at time $t+1$, $R_{i,t+1} - R_{F,t+1}$, and of the main forecasting variables used in this paper, for the full sample and two subsamples. For all series of cumulative industry returns and for the cumulative excess stock market return, we do not reject the null hypothesis of unit root at the 1% significance level. We found that there are 15 out of 30 industry portfolio returns co-integrated with the excess stock market return, at the 5% significance level. We excluded the remaining 15 industry excess returns series from our analysis as we want to deal only with the industry returns that are co-integrated excess stock market return. Regarding the subsamples, we divided the full sample before and after Dec.1987 due to the stock market crash in 1987. We found 17 and 8 co-integrated industry returns with the excess stock market return for the first and second subsample, respectively.

Table 3.1 shows that the excess stock returns of the industry i, $R_{i,t+1} - R_{F,t+1}$, are positively correlated with the previous excess stock market return and with the SMB_t factor. On the other hand, there is a negative correlation between $R_{i,t+1} - R_{F,t+1}$ and the equilibrium error at time t, $EE_{i,t}$, reflecting a mean-reverting effect of $EE_{i,t}$. The correlations among the excess industry returns and the forecasting factors are always below 0.10 in the full sample. We found some different results in the two subsamples. First, the previous month's excess stock market return, $R_{M,t} - R_{F,t}$, and SMB are more positively correlated with $R_{i,t+1} - R_{F,t+1}$ in the first subsample (Panel B) than in the second subsample (Panel C). Besides, while $EE_{i,t}$ is negatively related to $R_{i,t+1} - R_{F,t+1}$

in Panel C, the two are positively related in the first subsample (Panel B). Finally, the stochastically detrended risk-free rate, $RREL_t$, is negatively related with $R_{i,t+1} - R_{F,t+1}$ in Panel B, while they are positively related in Panel C.

3.2.2 In-Sample Forecasting

We propose the following panel-data forecasting model for excess industry returns:

$$R_{i,t+1} - R_{F,t+1} = \alpha_i + \boldsymbol{\beta}'\mathbf{X}_t + \beta_i^{EE} EE_{i,t} + \beta_i^R R_i(-1) + \varepsilon_{i,t}, \qquad t = 1, \ldots, T, \tag{3.2.2}$$

where $\mathbf{X}_t = [R_{M,t} - R_{F,t}, HML_t, SMB_t, RREL_t]$, $R_i(-1)$ is the previous month's excess industry stock return, $R_i(-1) = R_{i,t} - R_{F,t}$, α_i is the industry effect, and $\varepsilon_{i,t}$ is the resulting residual. We may also include the momentum factor, WML_t, in $\mathbf{X}_t$. We apply a fixed effects model to estimate (3.2.2), since standard errors are biased in the presence of a firm effect (e.g., $\mathrm{Cov}(EE_{i,t}\varepsilon_{i,t}, EE_{i,t-k}\varepsilon_{i,t-k}) \neq 0$) when estimated by OLS, White, Newey-West (modified for panel data sets), Fama-MacBeth, or Fama-MacBeth corrected for first-order autocorrelation (Petersen, 2009).

Now we discuss the in-sample forecasting results. Although significant in-sample evidence of predictability does not imply significant out-of-sample predictability, Inoue and Kilian (2004) show that in-sample tests are more powerful than out-of-sample tests, with no presumption that in-sample tests of predictability suffer from greater size distortions than out-of-sample tests.

Table 3.2 reports the in-sample least squares regression results of the fixed effects model (3.2.2), with heteroscedastic-corrected standard errors in parentheses. Panel A is the full sample spanning from Jul.1965 to Dec.2007. Row 1 shows that $R_{M,t} - R_{F,t}$ is not significative for predicting future excess stock returns. This result might be explained by an omitted variables problem. As suggested by Huang, Liu, Rhee, and Zhang (2010), the omission of the previous month's industry excess stock return might lead to a omitted variable bias in estimating the coefficient on the excess stock market return, $R_{M,t} - R_{F,t}$. There is a negative first-order correlation in monthly stock returns and it is regarded as short-term return reversals of individual stocks, first noted by Jegadeesh (1990). Row 2 provides evidence that $EE_{i,t}$ is a significant predictor of industry excess stock returns. It is likely that $EE_{i,t}$ forecasts industry excess stock returns because there is a correlation between $EE_{i,t}$ and some widely used forecasting variables. The previous month's $R_{M,t} - R_{F,t}$ becomes significant if the previous month's industry excess stock return, $R_i(-1)$, is included in the forecasting regression with a higher adjusted R^2(row 3). The coefficient on HML_t and SMB_t also change with the inclusion of $R_i(-1)$. These results confirm that there is an omitted-variable bias in rows 1 and 2. Row 4 shows that $RREL_t$ has significative forecasting power for future industry excess stock returns, which is consistent with the results of Campbell et al. (1997). However, the inclusion of $RREL_t$ has a small effect on the forecasting power of the other variables (row 5). Finally, the momentum at $t-1$, WML_{t-1}, provides negligible information besides the other variables for predicting industry excess stock returns (rows 6-7).

We report the in-sample forecasting results using two subsamples, Jul.1965-Dec.1987 and Jan.1988-Dec.2007, in Panels B and C, respectively. There are some differences between the two subsamples. First, while HML_t is not significative in the first subsample, SMB_t and $R_{M,t} - R_{F,t}$ are not significative in the second subsample. Besides, $EE_{i,t}$ has a positive sign in the first subsample and a higher predictive power, though it becomes insignificant when $R_i(-1)$ is included (rows 10, 12 and 14). However, in the second subsample, $EE_{i,t}$ is negatively related to $R_{i,t+1} - R_{F,t+1}$, while $R_i(-1)$ is not statistically significant (rows 17, 19 and 21). Finally, $RREL_t$ is positively related to $R_{i,t+1} - R_{F,t+1}$ in Panel C, while the two are negatively related in Panel B.

Table 3.1. Summary Statistics

	$R_{i,t+1} - R_{F,t+1}$	$R_{M,t} - R_{F,t}$	HML_t	SMB_t	$EE_{i,t}$	$RREL_t$
		A. Jul.1965-Dec.2007				
Mean	.011	.004	.004	.002	.000	.000
Standard Deviation	.061	.044	.029	.033	1.308	.001
Median	.011	.008	.004	.001	-.005	.000
		Correlation Matrix				
$R_{i,t+1} - R_{F,t+1}$	1.000					
$R_{M,t} - R_{F,t}$	.035	1.000				
HML_t	-.043	-.415	1.000			
SMB_t	.058	.303	-.279	1.000		
$EE_{i,t}$	-.030	.275	-.139	.078	1.000	
$RREL_t$	-.037	-.166	.082	-.132	-.018	1.000
		B. Jul.1965-Dec.1987				
Mean	.008	.003	.005	.005	.000	.000
Standard Deviation	.059	.048	.025	.029	.106	.001
Median	.008	.004	.003	.001	-.002	.000
		Correlation Matrix				
$R_{i,t+1} - R_{F,t+1}$	1.000					
$R_{M,t} - R_{F,t}$	.079	1.000				
HML_t	-.014	-.307	1.000			
SMB_t	.087	.380	-.107	1.000		
$EE_{i,t}$	.067	.486	-.171	.096	1.000	
$RREL_t$	-.095	-.248	.119	-.143	-.109	1.000
		C. Jan.1981-Dec.2007				
Mean	.010	.007	.003	.001	.000	.000
Standard Deviation	.060	.040	.030	.033	2.146	.001
Median	.012	.012	.000	.000	-.011	.000
		Correlation Matrix				
$R_{i,t+1} - R_{F,t+1}$	1.000					
$R_{M,t} - R_{F,t}$	.030	1.000				
HML_t	-.074	-.451	1.000			
SMB_t	.005	.201	-.339	1.000		
$EE_{i,t}$	-.080	.354	-.140	.115	1.000	
$RREL_t$	.028	.017	-.031	-.119	-.001	1.000

This table presents summary statistics for the i-th excess industry return at time $t+1$, $R_{i,t+1} - R_{F,t+1}$; the excess stock market return at time t, $R_{M,t} - R_{F,t}$; the return on a portfolio that is long in small stocks and short in big stocks at time t, SMB_t; the return on a portfolio that is long in stocks with high book-to-market value ratios and short in stocks with low book-to-market value ratios at time t, HML_t; the equilibrium error of the i-th excess industry return with the excess stock market return at time t, $EE_{i,t}$; and the stochastically detrended risk-free rate at time t, $RREL_t$.

Table 3.2. In-Sample Forecasting Monthly Excess Stock Returns

Models	$R_{M,t} - R_{F,t}$	HML_t	SMB_t	$EE_{i,t}$	$R_i(-1)$	$RREL_t$	WML_t	$\bar{R}^2$
				A. Jul.1965-Dec.2007				
(1)	1.26	**-5.28**	**8.91**					.004
	(2.02)	(1.94)	(1.98)					
(2)	2.96	**-5.48**	**8.94**	**-0.20**				.005
	(2.27)	(2.28)	(2.07)	(0.07)				
(3)	**-5.45**	**-6.72**	**8.36**	**-0.32**	**9.01**			.009
	(1.90)	(1.94)	(1.88)	(0.08)	(2.14)			
(4)	2.41	**-5.48**	**8.49**	**-0.20**		**-156.96**		.006
	(2.13)	(2.27)	(1.99)	(0.07)		(54.16)		
(5)	**-5.98**	**-6.72**	**7.91**	**-0.31**	**9.00**	**-156.17**		.009
	(1.84)	(1.93)	(1.81)	(0.08)	(2.14)	(52.73)		
(6)	1.96	**-6.28**	**8.54**	**-0.20**		**-165.66**	**-3.34**	.006
	(2.18)	(2.10)	(1.99)	(0.07)		(53.11)	(1.32)	
(7)	**-6.34**	**-7.46**	**7.96**	**-0.31**	**8.93**	**-164.41**	**-3.15**	.009
	(1.89)	(1.77)	(1.82)	(0.08)	(2.09)	(51.11)	(1.40)	
				B. Jul.1965-Dec.1987				
(8)	7.09	2.50	**13.62**					.010
	(1.87)	(2.51)	(2.07)					
(9)	4.15	2.75	**14.58**	**2.58**				.011
	(2.61)	(2.51)	(2.01)	(0.87)				
(10)	**-9.41**	1.14	**10.92**	0.53	**15.83**			.017
	(2.48)	(2.69)	(2.15)	(1.15)	(3.16)			
(11)	2.05	3.62	**13.69**	**2.63**		**-370.89**		.016
	(2.44)	(2.50)	(1.92)	(0.85)		(51.63)		
(12)	**-11.70**	2.00	**9.97**	0.55	**16.02**	**-375.71**		.023
	(2.47)	(2.70)	(2.12)	(1.12)	(3.16)	(49.05)		
(13)	2.04	4.27	**14.44**	**2.58**		**-361.45**	**3.47**	.017
	(2.45)	(2.51)	(1.99)	(0.86)		(50.76)	(1.33)	
(14)	**-11.86**	2.73	**10.79**	0.47	**16.20**	**-364.92**	**3.98**	.023
	(2.51)	(2.67)	(2.11)	(1.13)	(3.17)	(48.42)	(1.34)	
				B. Jan.1988-Dec.2007				
(15)	-0.79	**-16.37**	-3.82					.004
	(3.46)	(5.04)	(3.27)					
(16)	4.93	**-15.95**	-2.99	**-0.28**				.013
	(3.17)	(5.76)	(3.58)	(0.09)				
(17)	3.18	**-16.33**	-3.09	**-0.31**	2.19			.012
	(4.97)	(5.99)	(3.48)	(0.12)	(5.37)			
(18)	4.89	**-15.59**	-2.36	**-0.28**		**190.86**		.013
	(3.17)	(5.84)	(3.55)	(0.09)		(69.83)		
(19)	3.18	**-15.96**	-2.46	**-0.30**	2.13	**188.73**		.012
	(4.93)	(6.07)	(3.44)	(0.12)	(5.32)	(66.99)		
(20)	3.73	**-16.04**	-1.57	**-0.28**		**190.89**	-4.68	.013
	(3.76)	(5.50)	(3.26)	(0.09)		(69.82)	(4.60)	
(21)	**2.49**	**-16.30**	-1.68	-0.30	1.60	**189.29**	-4.52	.013
	(4.42)	(5.78)	(3.16)	(0.13)	(5.69)	(67.25)	(5.20)	

This table reports the least squares regression results of the fixed effects model for the one-month-ahead excess stock returns, $R_{i,t+1} - R_{F,t+1}$, on some variables. The heteroscedastic-corrected standard errors are reported in parentheses, and bold denotes significance at the 5% level. $R_{M,t} - R_{F,t}$ is the excess stock market return. HML_t is the high-minus-low factor. SMB_t is the small-minus-big factor. $EE_{i,t}$ is the equilibrium error of the i-th industry excess stock return with the market excess return. $R_i(-1)$ is the i-th industry excess stock return during the previous month. $RREL_t$ is the stochastically detrended risk-free rate. WML_t is the momentum factor. $\bar{R}^2$ is the adjusted-R^2.

* Scaled by 100.

3.2.3 Out-of-Sample Forecast Performance

This section provides the analysis of the out-of-sample forecasting performance of our proposed models. Bossaerts and Hillion (1999), Goyal and Welch (2003), and Welch and Goyal (2008) question the in-sample evidence of stock return predictability, as they showed that even the best prediction models have no out-of-sample forecasting power. On the other hand, Inoue and Kilian (2004) show that out-of-sample tests are not necessarily more reliable than in-sample tests. To analyse this point, we compare the out-of-sample performance of our proposed model with a model that does not include the $EE_{i,t}$ and with a benchmark model of historical average returns. We perform two analyses. First, we assume that investors know the co-integration parameters of $EE_{i,t}$, estimated using the full sample. In the second analysis, the co-integration parameters are estimated recursively using only information available at the time of the forecast. This analysis is more realistic and has more applicability, since investors can use only the data available at the time of the forecast to make decisions.

A. Fixed Co-integrating Factors

Table 3.3 evaluates the out-of-sample performance of three models: (i) a model including $R_{M,t} - R_{F,t}$, SMB_t, HML_t, and $RREL_t$; (ii) an augmented model including also $EE_{i,t}$ and the previous month's returns of $R_{i,t} - R_{F,t}$, $R_i(-1)$; and (iii) a benchmark model of the historical average excess return estimated through period t, $\bar{R}_{i,t} - \bar{R}_{F,t}$. We present five forecast performance statistics: (i) the root mean squared error (RMSE), (ii) the mean absolute percentage error (MAPE), (iii) the Theil's U inequality coefficient (U), (iv) the out-of-sample R^2 statistic (R^2_{OS}), and (v) the correlation between the actual and the predicted value of the industry excess stock return (ρ). The out-of-sample R^2 statistic (R^2_{OS}) can be compared with the in-sample R^2 statistic and is computed as

$$R^2_{OS} = 1 - \frac{\sum_{t=T}^{T+h} (r_t - \hat{r}_t)^2}{\sum_{t=T}^{T+h} (r_t - \bar{r}_t)^2},$$

where r_t is the industry excess stock return, $\hat{r}_t$ is the predicted value from a predictive regression estimated through period T, $\bar{r}_t$ is the historical average of the industry excess stock return estimated through period T, h is the number of out-of-sample periods, and T is the sample size. In the out-of-sample forecasts, we first run an in-sample regression using data from Jul.1965 until Jun.1968 and then we forecast the returns $R_{i,t} - R_{F,t}$ of Jul.1968. After computing the forecast, we update the sample from Jul.1965 until Jul.1968 and we perform a forecast for Aug.1968 and so forth. We estimate the historical average return in the benchmark model recursively.

Panel A of Table 3.3 shows that the augmented-model including the $EE_{i,t}$ presents better forecasting performance than the other two models, except for the MAPE criteria, for the sample from Jul.1965 to Dec.2007. For example, the augmented-model has the smallest RMSE and the highest R^2_{OS} between the three models. These results are consistent with the in-sample analysis in Table 3.2, where the inclusion of $EE_{i,t}$ and $R_i(-1)$ provides additional forecasting power. Panel B of Table 3.3 displays the out-of-sample results for Jan.1988-Dec.2007. Consistent with the full sample results, the augmented-model including the $EE_{i,t}$ has better forecasting abilities for subsequent excess industry stock returns than the other two models in almost all criteria.

Table 3.3. Out-of-Sample Forecasting - Fixed Co-integrating Factors

	$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + R_i(-1) + EE_{i,t}$ (2)	Historical Average (3)
		A. Jul.1965-Dec.2007	
RMSE	.0423	.0390	.0619
MAPE	2.7491	2.5403	1.6959
U	.6751	.6236	.9892
R^2_{OS}	.5342	.6026	-
ρ	.7304	.7828	-.0341
		B. Jan.1988-Dec.2007	
RMSE	.0476	.0415	.0610
MAPE	3.0832	2.7882	1.5407
U	.7707	.6721	.9878
R^2_{OS}	.3913	.5370	-
ρ	.6242	.7357	-.0255

$F\&F_t$ denotes the three Fama-French factors: the excess stock market return, $R_{M,t} - R_{F,t}$, the high-minus-low factor, HML_t, and the small-minus-big factor, SMB_t. $EE_{i,t}$ is the equilibrium error of the i-th industry excess stock return with the excess stock market return. $R_i(-1)$ is the i-th previous month's industry excess stock return. $RREL_t$ is the stochastically detrended risk-free rate. Historical Average denotes a benchmark model of the historical average excess return estimated through period t, $\bar{R}_{i,t} - \bar{R}_{F,t}$.

Figure 3.1 plots the recursive MSE ratio of the augmented model including $EE_{i,t}$ (column 2 of Table 3.3) to the benchmark model of historical average returns (column 3 of Table 3.3) and to a model including $R_{M,t} - R_{F,t}$, SMB_t, HML_t, and $RREL_t$, but excluding the equilibrium error $EE_{i,t}$ (column 1 of Table 3.3). The horizontal line is the initial forecasting date; for instance, the MSE ratio of Jul.1972 corresponds to the forecast period Jul.1972-Dec.2007. We use at least 36 observations for the in-sample estimation; thus, we use the range Jul.1968-Jul.2004 for the starting forecast date. Figure 3.1 shows that the augmented model including $EE_{i,t}$ has a better out-of-sample forecasting power than the benchmark model of historical average returns, as the dashed line is always smaller than one. In comparison with a model including $R_{M,t} - R_{F,t}$, SMB_t, HML_t, and $RREL_t$, the equilibrium error adds substantial forecasting power, with a MSE ratio always smaller than 1. These results are consistent with the MSE-F test in Table 3.5. In sum, we find evidence that the augmented model with the equilibrium error beats two competing models for predicting subsequent excess industry stock returns.

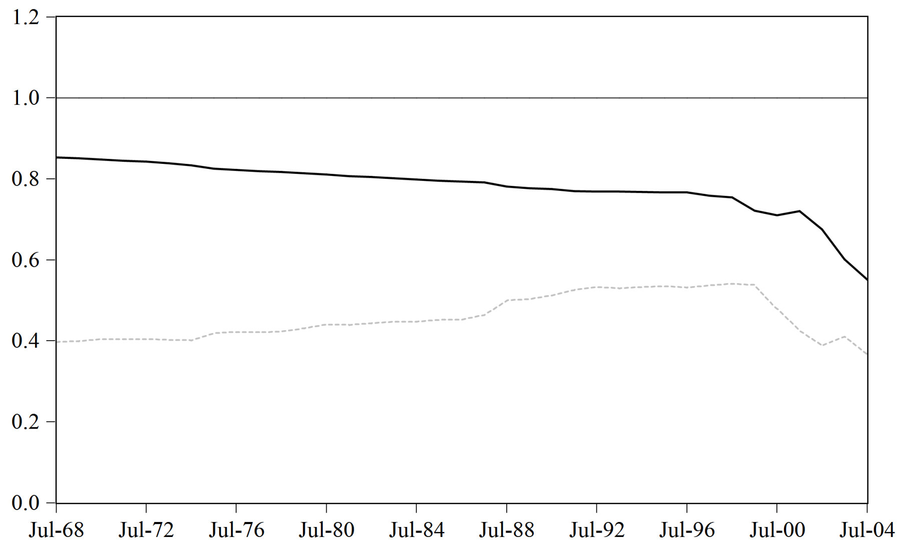

Figure 3.1. MSE ratio of augmented $EE_{i,t}$ to $F\&F_t$ model (solid line) and to historical average returns (dashed line).

B. Recursive Co-integrating Factors

Table 3.4 evaluates the out-of-sample performance of our proposed models using recursively estimated $EE_{i,t}$. The analysis is the same in the case of fixed co-integrating factors, except that the equilibrium error factor, $EE_{i,t}$, is estimated recursively using only information available at the time of the forecast. For instance, we first calculate the equilibrium error factor from the co-integration relationship between industry stock returns and excess stock market returns using data from Jul.1965 until Jun.1968. Then we run an in-sample regression using data from Jul.1965 until Jun.1968 and we make a forecast of the excess returns $R_{i,t} - R_{F,t}$ for Jul.1968. After computing the forecast, we update the sample from Jul.1965 until Jul.1968, recalculate $EE_{i,t}$ and make a forecast for Aug.1968 and so forth. The results are similar to those in Table 3.3. The predictability of the augmented model with $EE_{i,t}$ is slightly weaker in Table 3.4 than in Table 3.3, though the augmented model with $EE_{i,t}$ still remains with the best overall forecasting performance among the three models.

For the period from Jul.1965 to Dec.2007, the augmented model with $EE_{i,t}$ has the smallest RMSE and the highest R^2_{OS} among the three models. Those results are robust in the subsample from Jan.1988 to Dec.2007, where the augmented model with $EE_{i,t}$ still provides better forecasting abilities than the other two models.

Figure 3.2 plots the recursive MSE ratio of the augmented model including $EE_{i,t}$ (col. 2 of Table 3.4) to the benchmark model of historical average returns (col. 3 of Table 3.4) and to a model including $R_{M,t} - R_{F,t}$, SMB_t, HML_t, and $RREL_t$, but excluding the equilibrium error $EE_{i,t}$ (col. 1 of Table 3.4). The horizontal line is the initial forecasting date.

The results are similar to those in Figure 3.1. The augmented model including $EE_{i,t}$ has better out-of-sample forecasting abilities than the benchmark model of historical average returns, as the dashed line is always smaller than one. Besides, the augmented model including $EE_{i,t}$ adds substantial information to the model excluding $EE_{i,t}$, since the MSE ratio between them is always smaller than 1 (solid line of Figure 3.2). Overall, these results indicate the augmented model with recursively estimated $EE_{i,t}$ has substantial forecasting abilities for subsequent excess industry returns.

Table 3.4. Out-of-Sample Forecasting - Recursive Co-integrating Factors

	$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + R_i(-1) + EE_{i,t}$ (2)	Historical Average (3)
		A. Jul.1965-Dec.2007	
RMSE	.0425	.0398	.0621
MAPE	2.7525	2.6271	1.7428
U	.6822	.6387	.9954
R^2_{OS}	.5304	.5883	-
ρ	.7287	.7789	-.0595
		B. Jan.1988-Dec.2007	
RMSE	.0473	.0423	.0590
MAPE	2.9127	2.4763	1.8912
U	.7963	.7127	.9925
R^2_{OS}	.3563	.4844	-
ρ	.6010	.7077	-.0368

$F\&F_t$ denotes the three Fama-French factors: the excess stock market return, $R_{M,t} - R_{F,t}$, the high-minus-low factor, HML_t, and the small-minus-big factor, SMB_t. $EE_{i,t}$ is the equilibrium error of the i-th industry excess stock return with the excess stock market return that is recursively estimated using only data available at the time of forecast. $RREL_t$ is the stochastically detrended risk-free rate. Historical Average denotes a benchmark model of the historical average excess return estimated through period t, $\bar{R}_{i,t} - \bar{R}_{F,t}$.

C. Testing Out-of-Sample Forecasting Performance

We present three test-statistics to evaluate the out-of-sample forecasting power of our proposed models. Following Guo and Savickas (2006), we use the mean squared forecasting error (MSE) ratio, the encompassing test (ENC-NEW) of Clark and McCracken (2001), and the equal forecast accuracy test (MSE-F) proposed by McCracken (1999). The encompassing test (ENC-NEW) tests the null hypothesis that the benchmark model all of the information about the next month's industry excess stock return against the alternative that the augmented model adds information. The equal forecast accuracy test (MSE-F) tests the null hypothesis that the benchmark model has a MSE less than or equal to that of the augmented model against the alternative hypothesis that the augmented model has smaller MSE. The MSE-F and ENC-NEW have the best power and size properties among the many possible tests of out-of-sample forecasting performance in the literature (Clark and McCracken, 2001).

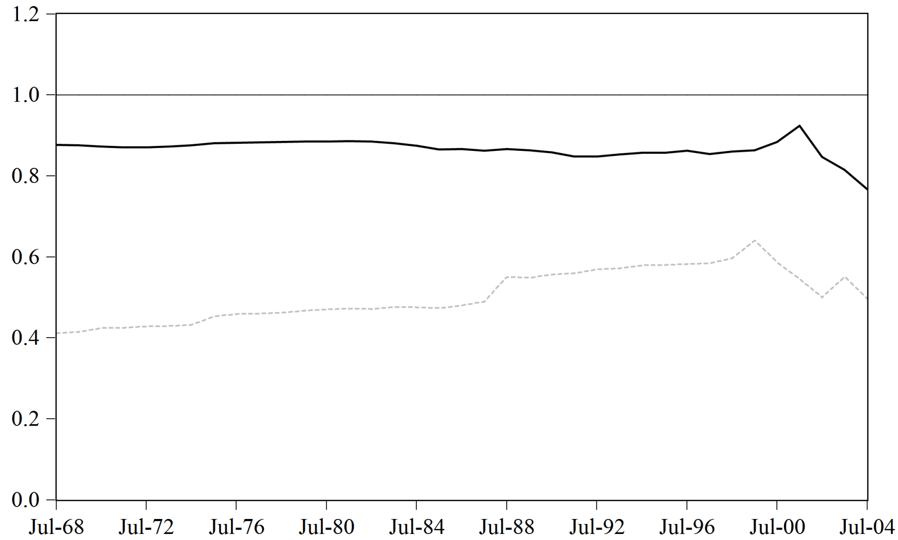

Figure 3.2. MSE ratio of augmented $EE_{i,t}$ to $F\&F_t$ model (solid line) and to historical average returns (dashed line) - recursive co-integrating $EE_{i,t}$.

Table 3.5 presents the out-of-sample forecast test statistics. To compare with the asymptotic critical values of the test statistic, we need to use a large in-sample period of estimation. Thus, we estimate the in-sample regression using one-third of the observations and make the out-of-sample forecasts recursively for the rest of the sample. We use the observations from Jul.1965 to Aug.1979 to forecast the out-of-sample industry excess returns of Sep.1979 and update the sample recursively to make the forecast for the next month. The column MSE_A/MSE_B reports the MSE ratio of the proposed model to that of the benchmark model. For the ENC-NEW and MSE-F tests, Asy. CV denote the 95% critical values derived by Clark and McCracken (2001) and McCracken (1999), for a ratio of out-of-sample to in-sample periods of 2.

In Panel A, we estimate $EE_{i,t}$ using the full sample. The augmented model has a smaller MSE than the benchmark model, as the MSE ratio is smaller than one (rows 1 and 2, Table 3.5). Consistent with the MSE ratio, the MSE-F test rejects the null hypothesis that the benchmark model has a MSE smaller than the augmented model at the 5% of significance. Besides, the ENC-NEW rejects the null hypothesis that EE_i, t contains no additional information about the predictability of future industry excess stock returns at the 5% of significance. We also add the previous month's industry excess stock returns, $R_i(-1)$, to check if the equilibrium error has significative forecasting power. Row 2 of table 3.5 shows that the augmented model with $EE_{i,t}$ has smaller MSE than the benchmark model and significative additional information for forecasting future industry excess stock returns at the 5% significance level, as indicated by the MSE-F and ENC-NEW tests.

In Panel B, $EE_{i,t}$ is estimated recursively using only the information available at the time of the forecast. The augmented model still provides a smaller MSE ratio than the benchmark model. We find evidence that the augmented model with $EE_{i,t}$ beats the other two models at the 5% significance level.

Table 3.5. Out-of-Sample Monthly Forecasts of Excess Stock Market Returns: Performance Tests

Models	MSE_A/MSE_B	ENC-NEW Statistic	ENC-NEW Asy. CV	MSE-F Statistic	MSE-F Asy. CV
	A.Fixed Co-Integrating Factors				
(1) Bench. + $R_i(-1)$ + $EE_{i,t}$ vs. Bench.	0.81	89.86	3.56	77.33	1.61
(2) Bench. + $R_i(-1)$ + $EE_{i,t}$ vs. Bench. + $R_i(-1)$	0.81	90.34	2.09	77.94	1.52
	B.Recursive Co-Integrating Factors				
(1) Bench. + $R_i(-1)$ + $EE_{i,t}$ vs. Bench.	0.88	120.20	2.71	38.64	1.91
(2) Bench. + $R_i(-1)$ + $EE_{i,t}$ vs. Bench. + $R_i(-1)$	0.92	123.72	1.58	67.09	1.55

This table presents the mean-squared forecasting error ratio of the augmented model to the benchmark model (MSE_A/MSE_B), the encompassing test ENC-NEW proposed by Clark and McCracken (2001), and the MSE-F test derived by McCracken (1999). We assume that the benchmark model includes the three Fama-French factors - $R_{M,t} - R_{F,t}$, HML_t, and SMB_t - and the stochastically detrended risk-free rate, $RREL_t$, in rows 1 and 3, and also the i-th industry excess stock return during the previous month, $R_i(-1)$, in rows 2 and 4. We augment the benchmark model with $R_i(-1)$ + $EE_{i,t}$ in rows 1 and 3, and with $EE_{i,t}$ in rows 2 and 4. The ENC-NEW tests if the benchmark model encompasses all the relevant information about the next month's excess stock market return, against the alternative hypothesis that the augmented model includes additional relevant information. MSE-F tests if the benchmark model has a smaller mean-squared forecasting error than the augmented model. The in-sample period estimation spans from Jul.1965 through Aug.1979 and then the forecasting errors are generated forecasts recursively for excess stock returns over the period Sep.1979-Dec.2007. The variable $EE_{i,t}$ is estimated using the full sample in panel A and recursively estimated using only data available at the time of forecast in panel B. Columns 4 and 6 display the asymptotic 95% critical values provided by McCracken (1999) and Clark and McCracken (2001).

3.3 Economic Value of Forecasting

According to Leitch and Tanner (1991), traditional measures of forecasting performance, such as the RMSE, may not be closely related to a forecast's profit. Using profit measures, they find only very weak relationships between such summary error statistics and forecast value. If these results are robust, then least-squares regression analysis may not be appropriate for many studies of economic behavior. We analyse in this section if the observed forecasting power can be applied to generate higher returns with lower volatility than the returns implied by a buy-and-hold strategy.

3.3.1 Switching Portfolio

To check if our recursive out-of-sample forecasts could have been used to generate a higher mean return than that earned from adopting a buy-and-hold strategy, we follow the approach of Pesaran and Timmermann (1995) and use our forecasts in a simple switching strategy, which has been widely used in the literature. According to this strategy, an investor holds stocks in periods where the business cycle suggest that stock returns are going to outperform bond returns (i.e., the predicted excess industry stock return is positive), and otherwise holds bonds. We do not include the short-selling of assets and we do not assume that an investor can use leverage when selecting his portfolio.

Table 3.6 reports the mean, the standard deviation (S.D.), the Sharpe ratio, and the Adjusted Sharpe ratio for the annualized returns on portfolios based on three forecast models analyzed in the previous sections. As in Graham and Harvey (1997), Johannes et al. (2002), and Guo (2006), we adjust the return on the managed portfolio to have the same standard deviation as the stock market return. The realized adjusted return is used to calculate the Adjusted Sharpe ratio in a regular way. The Adjusted Sharpe ratio helps to weaken the effect of leverage on the portfolio selection without affecting the Sharpe ratio calculation. For example, if the portfolio had a 10% mean return and 11% volatility, and the market volatility is 15%, we multiply the mean return by the ratio of the market volatility to portfolio volatility, (0.15/0.11), which will give a risk-adjusted return of 13.63%. Then we calculate the Adjusted Sharpe ratio based on this risk-adjusted return.

Table 3.6 shows that the managed portfolio based on an augmented model including $EE_{i,t}$ has annualized return of higher mean and Sharpe ratio than those implied by the two competing models, over the period Jul.1965-Dec.2007. For instance, the switching portfolio based on augmented forecast model of column 2 provides an annual mean return of 31.5% with a volatility of 70.0% compared with 25.6% and 68.6% respectively, for a switching strategy based on a benchmark of historical average returns. Besides, the Sharpe ratio of the augmented model is 120% higher than the benchmark portfolio. Thus, the equilibrium error $EE_{i,t}$ is not only statistically significant in terms of out-of-sample forecasting ability, but also economically important. The additional information provided by the equilibrium error is used effectively in the switching portfolio.

Our results are robust in the three subsample periods presented in Panels B-D of Table 3.6. For all subsample periods, the managed portfolio based on an augmented model including $EE_{i,t}$ has the highest annualized mean return and Sharpe ratio among the three models. Consistent with Pesaran and Timmermann (1995) and Guo (2006), the performance of the managed portfolio relative to the benchmark varies over time. For example, the managed portfolio has an Adjusted Sharpe ratio of 38.1% for the period Jul.1965-Dec.1979, compared with 31.1% for the benchmark portfolio. However, the managed portfolio generates an Adjusted Sharpe ratio of 39.2% (48.2%) for the period Jan.1980-Dec.1994 (Jan.1995-Dec.2007), compared with 35.5% (30.2%) for the benchmark portfolio

Table 3.6. Performance Measures for the Switching Portfolio - No Transaction Costs

	$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + EE_{i,t} + R_i(-1)$ (2)	Historical Average (3)
	A. Jul.1965-Dec.2007		
Mean Return	.2835	.3155	.2562
S.D.	.6917	.6999	.6863
Sharpe Ratio	.4099	.4508	.3733
Adj. Sharpe Ratio	.3498	.3809	.3109
	B. Jul.1965-Dec.1979		
Mean Return	.2486	.2631	.2498
S.D.	.8600	.8545	.8582
Sharpe Ratio	.2891	.3078	.2910
Adj. Sharpe Ratio	.2404	.2533	.2422
	C. Jan.1980-Dec.1994		
Mean Return	.2819	.2957	.2695
S.D.	.6162	.6189	.6242
Sharpe Ratio	.4574	.4778	.4318
Adj. Sharpe Ratio	.3817	.3922	.3552
	D. Jan.1995-Dec.2007		
Mean Return	.2492	.3478	.1975
S.D.	.5300	.6139	.5265
Sharpe Ratio	.4702	.5666	.3752
Adj. Sharpe Ratio	.4003	.4820	.3018

This table display returns on switching portfolios, where an investor holds stocks if the predicted industry excess stock return is positive and holds bonds otherwise. All the statistics are for the annualized returns. As in Graham and Harvey (1997) and Guo (2006), we adjust the return on the managed portfolio to have the same standard deviation as the stock market return to calculate the Adjusted Sharpe Ratio. The variable $EE_{i,t}$ is recursively estimated using only data available at the time of forecast.

Allowing for "high" transaction costs of 1.0 of a percent on switching from bonds to stocks and 0.1 of a percent on switching from stocks to bonds, Table 3.7 presents the effect of transaction costs on the switching portfolio. Investors have to pay 1% of the return on stocks if they switch from bonds to stocks and 0.1% of the return on bonds if they switch from stocks to bonds. As a 25-basis-point fee is in the upper range of transaction costs for the market index (Balduzzi and Lynch, 1999), we assure that we are imposing a high fee of 100-basis-point fee. The effects of imposing transaction costs on the switching portfolios is negligible on the performance of the trading strategies (Table 3.7). The strategy based on the augmented model with $EE_{i,t}$ still has higher mean and lower volatility than other strategies.

Table 3.7. Performance Measures for the Switching Portfolio - High Transaction Costs

	$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + EE_{i,t} + R_i(-1)$ (2)	Historical Average (3)
		A. Jul.1965-Dec.2007	
Mean Return	.2819	.3139	.2562
S.D.	.6879	.6962	.6863
Sharpe Ratio	.4098	.4509	.3733
Adj. Sharpe Ratio	.3497	.3811	.3109
		B. Jul.1965-Dec.1979	
Mean Return	.2465	.2612	.2498
S.D.	.8535	.8487	.8582
Sharpe Ratio	.2888	.3077	.2910
Adj. Sharpe Ratio	.2401	.2533	.2422
		C. Jan.1980-Dec.1994	
Mean Return	.2802	.2942	.2695
S.D.	.6126	.6157	.6242
Sharpe Ratio	.4573	.4778	.4318
Adj. Sharpe Ratio	.3813	.3920	.3552
		D. Jan.1995-Dec.2007	
Mean Return	.2478	.3461	.1975
S.D.	.5280	.6111	.5265
Sharpe Ratio	.4693	.5663	.3752
Adj. Sharpe Ratio	.3995	.4819	.3018

This table display returns on switching portfolios, where an investor holds stocks if the predicted industry excess stock return is positive and holds bonds otherwise. All the statistics are for the annualized returns. As in Graham and Harvey (1997) and Guo (2006), we adjust the return on the managed portfolio to have the same standard deviation as the stock market return to calculate the Adjusted Sharpe Ratio. The variable $EE_{i,t}$ is recursively estimated using only data available at the time of forecast. We assume that investors pay 1.0% on switching from bonds to stocks and 0.1% on switching from stocks to bonds.

3.3.2 Optimal Portfolio Weights

Now we allocate wealth between stocks and bonds using the optimal portfolio weight approach, taken in Kandel and Stambaugh (1996), Stambaugh (1999), Pástor and Stambaugh (2000), Pástor (2000), and Johannes et al. (2002). The investor solves a single-period optimal portfolio problem:

$$\max_{\omega_t} E\left[U(W_{t+1})|R^t\right] := \max_{\omega_t} \int U(W_{t+1}) \Pr(R_{t+1}|R^t) dR_{t+1},$$

where R^t is a vector of observed compounded returns up to time t, $W_{t+1} = W_t[\omega_t(R_{i,t} + (1-\omega_t)R_{F,t}]$ is the next period's wealth, $\Pr(R_{t+1}|R^t)$ is the predictive distribution of future returns, and the maximization is subject to the usual budget constraint. We assume that the utility function, $U(W_{t+1})$, is strictly increasing, twice differentiable and concave in the portfolio weight. Solving for the single-period optimal portfolio gives, we have

$$\omega_t = \frac{1}{\gamma} \frac{E\left[R_{i,t+1} - R_{F,t+1} | R^t\right]}{E\left[\sigma^2_{i,t+1} | R^t\right]},$$

where γ denotes the investor's relative risk aversion, $E\left[R_{i,t+1} - R_{F,t+1} | R^t\right]$ is the forecast industry excess stock return, and $E\left[\sigma^2_{i,t+1} | R^t\right]$ is the forecast conditional variance of $R_{i,t+1} - R_{F,t+1}$. We focus on the single-period portfolio problem. The difference between single period and multi-period problems is hedging demands. Ang and Bekaert (2002), Chacko and Viceira (2005), and Pástor and Stambaugh (2000), among others, found that hedging demands are typically extremely small components of the optimal allocation and are important only for long-horizon investors such as the infinitely lived investors in Campbell, Chan, and Viceira (2003).

We forecast the conditional variance at $t+1$, $E\left[\sigma^2_{i,t+1} | R^t\right]$, from an AR(2) model of $\sigma^2_{i,t}$, for each i-th excess industry stock return. For simplicity, we do not allow for the short-selling of assets or borrowing from bond markets, i.e. $\omega_t \in [0, 1]$, and we do not take into account the estimation uncertainty. The optimal portfolio weight justifies a mean-variance rule for investing in stocks, where the risk-aversion parameter γ takes into account returns that are generated by a fat-tailed stochastic volatility distribution. While the switching strategy just gives information on the signs of predicted excess industry stock returns, this investment strategy also includes information on the magnitude of the forecast excess returns normalized by its forecast conditional variance.

Table 3.8 provides the summary statistics for the annualized returns from an optimal portfolio strategy weight based on three different forecasting models. We assume that $\gamma = 5$ in the calculation of the optimal weights, but the results are robust to different choices of γ. The portfolio based on the augmented model with $EE_{i,t}$ has higher annualized mean return and Sharpe ratios than those reported in Table 3.6 for a switching strategy. For example, over the period Jul.1965-Dec.2007, the Adjusted Sharpe ratio is 69.6% if an investor allocates portfolio weight optimally, compared with 38.9% for the switching strategy. However, the results are similar to those presented in Table 3.6. Optimal portfolio weighting based on augmented models using $EE_{i,t}$ provides return of higher annualized mean and Sharpe ratios than portfolio based on the other two models, over the full sample and the three subsample periods. Besides, the relative performance of optimal portfolio weighting strategies varies over time.

Table 3.8. Choosing Optimal Portfolio Weights with No Transaction Costs

	$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + EE_{i,t} + R_i(-1)$ (2)	Historical Average (3)
		A. Jul.1965-Dec.2007	
Mean Return	.3960	.5290	.1982
S.D.	.5667	.6434	.5137
Sharpe Ratio	.6987	.8221	.3858
Adj. Sharpe Ratio	.6028	.6960	.3203
		B. Jul.1965-Dec.1979	
Mean Return	.4482	.5116	.1085
S.D.	.7174	.7871	.4533
Sharpe Ratio	.6247	.6500	.2395
Adj. Sharpe Ratio	.5060	.5171	.1955
		C. Jan.1980-Dec.1994	
Mean Return	.3902	.5001	.2585
S.D.	.5060	.5571	.5879
Sharpe Ratio	.7711	.8977	.4397
Adj. Sharpe Ratio	.7335	.8440	.3585
		D. Jan.1995-Dec.2007	
Mean Return	.2999	.5268	.1856
S.D.	.3881	.5630	.5062
Sharpe Ratio	.7727	.9357	.3666
Adj. Sharpe Ratio	.7027	.8170	.2930

This table presents the returns for an optimal weighting strategy, where an investor allocates an optimal weight of the total wealth in stocks:

$$\omega_t = \frac{1}{\gamma}\frac{E\left[R_{i,t+1} - R_{F,t+1}|R^t\right]}{E\left[\sigma^2_{i,t+1}|R^t\right]},$$

where R^t is a vector of observed compounded returns up to time t, γ denotes the investor's relative risk aversion, $E\left[R_{i,t+1} - R_{F,t+1}|R^t\right]$ is the forecast industry excess stock return, and $E\left[\sigma^2_{i,t+1}|R^t\right]$ is the forecast conditional variance of $R_{i,t+1} - R_{F,t+1}$ based on a AR(2) model of $\sigma^2_{i,t}$, for each i-th excess industry stock return. The variable $EE_{i,t}$ is recursively estimated using only data available at the time of forecast. We assume that $\gamma = 5$, $\omega_t \in [0, 1]$, and we ignore the estimation uncertainty.

3.3.3 Market Timing Ability Test

In this section, we check the forecasting power of our model by testing whether the expected excess industry stock returns during forecast up markets is different from that during forecast down markets. This was first proposed by Cumby and Modest (1987) and it is called the market timing ability test. It consists on testing the null hypothesis $a_1 = 0$ in the regression

$$R_{i,t+1} - R_{F,t+1} = a_0 + a_1 I_t + v_{t+1},$$

where $R_{i,t+1} - R_{F,t+1}$ are the observed excess industry stock returns, and I_t is one if the forecasting model predicts $R_{i,t+1} - R_{F,t+1}$ to be positive and is equal zero otherwise. The

Table 3.9. Choosing Optimal Portfolio Weights with High Transaction Costs

	$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + EE_{i,t} + R_i(-1)$ (2)	Historical Average (3)
		A. Jul.1965-Dec.2007	
Mean Return	.3937	.5251	.1982
S.D.	.5605	.6360	.5137
Sharpe Ratio	.7025	.8256	.3858
Adj. Sharpe Ratio	.6065	.6993	.3203
		B. Jul.1965-Dec.1979	
Mean Return	.4446	.5077	.1085
S.D.	.7064	.7755	.4533
Sharpe Ratio	.6293	.6546	.2395
Adj. Sharpe Ratio	.5103	.5215	.1955
		C. Jan.1980-Dec.1994	
Mean Return	.3878	.4971	.2585
S.D.	.5005	.5510	.5879
Sharpe Ratio	.7748	.9022	.4397
Adj. Sharpe Ratio	.7374	.8484	.3585
		D. Jan.1995-Dec.2007	
Mean Return	.2987	.5242	.1856
S.D.	.3865	.5600	.5062
Sharpe Ratio	.7729	.9361	.3666
Adj. Sharpe Ratio	.7032	.8177	.2930

This table presents the returns for an optimal weighting strategy, with the same specifications as in Table 3.8. We assume that investors pay 1.0% on switching from bonds to stocks and 0.1% on switching from stocks to bonds.

market timing ability analyses only the first moment, but investors may care about other moments of the return distribution. Following Breen et al. (1989), we also investigate the forecast ability of the variance of the excess industry stock market returns during forecast up and down markets. Thus, we test the null $b_1 = 0$ in the regression

$$v_{t+1}^2 = b_0 + b_1 I_t + \eta_{t+1},$$

where v_{t+1} are the squared residuals of the first regression. Table 3.10 reports the results of the market timing ability test for the first and second moment based on three different forecasting models. We reject the null hypothesis of no market timing ability for all the three models (Panel A) at the 5% significance level. Besides, we find evidence that the variance of the excess industry stock returns are slightly positive related to the market index based on the augmented model including $EE_{i,t}$ (col.2, Panel B).

Table 3.10. Market Timing Ability Test: Jul.1968-Dec.2007

	$F\&F_t$ + $RREL_t$ + $R_i(-1)$ (1)	$F\&F_t$ + $RREL_t$ + $R_i(-1)$ + $EE_{i,t}$ (2)	Historical Average (3)
	A. $R_{i,t+1} - R_{F,t+1} = a_0 + a_1 I_t + v_{t+1}$		
a_0	**-0.034**	**-0.035**	.000
	(16.37)	(19.37)	(1.04)
a_1	**0.058**	**0.058**	**.016**
	(20.30)	(23.93)	(3.15)
	B. $v^2_{t+1} = b_0 + b_1 I_t + \eta_{t+1}$		
b_0	**.003**	**0.003**	**0.004**
	(45.41)	(19.45)	(7.17)
b_1	.000	**0.001**	-.003
	(1.5)	(4.30)	(1.31)

This table presents a panel version of the market timing ability test developed by Cumby and Modest (1987) on the excess stock returns in Panel A and on the variance of the excess stock return in Panel B as suggested by Breen et al. (1989). The regression coefficients were estimated by a fixed effects method, where dependent variable is the excess stock market return of the i-th industry at $t+1$, $R_{i,t+1} - R_{F,t+1}$, and the regressor is an indicator function, I_t, that is equal to one if $R_{i,t+1} - R_{F,t+1}$ is expected to be positive at t and zero otherwise. $F\&F_t$ denotes the three Fama-French factors: the excess stock market return, $R_{M,t} - R_{F,t}$, the high-minus-low factor, HML_t, and the small-minus-big factor, SMB_t. $EE_{i,t}$ is the equilibrium error of the i-th industry excess stock return with the market excess return. $R_i(-1)$ is the i-th industry excess stock return during the previous month. $RREL_t$ is the stochastically detrended risk-free rate. The heteroscedastic-corrected t-statistics are reported in parentheses, and bold denotes significance at the 5% level.

3.3.4 Additional Tests

In this section, we follow the approach of Fleming, Kirby, and Ostdiek (2001) and measure the volatility timing of our forecasting strategies. For each one of the forecasting-based strategy, we compare its performance with an unconditional mean-variance efficient static strategy that would have the same target expected return and volatility. If the volatility timing implied by our strategies has no value, then their performance should be no different from an unconditional mean-variance efficient static strategy. To make this comparison, we use a performance measure that evaluates the trade-off between risk and return. Assuming a fixed parameter of the investor's relative risk aversion, γ, we use an utility function

$$U(\bullet) = W_0 \left(\sum_{t=0}^{T-1} R_{i,t+1} - \frac{\gamma}{2(1+\gamma)} R^2_{i,t+1} \right), \tag{3.3.3}$$

where W_0 is the investor's initial wealth. We calculate the certainty equivalent gain Δ by equating the utility for two different portfolios. Then Δ is the maximum performance fee that an investor would be willing to pay to switch from a strategy to another. In our approach, we compare each forecast-based strategy that pays a rate of $R_{i,t+1}$ with the market portfolio that pays $R_{M,t+1}$. Thus, to estimate the certainty equivalent Δ, we find the value of Δ that satisfies

$$\sum_{t=0}^{T-1}(R_{i,t+1}-\Delta)-\frac{\gamma}{2(1+\gamma)}(R_{i,t+1}-\Delta)^2=\sum_{t=0}^{T-1}R_{M,t+1}-\frac{\gamma}{2(1+\gamma)}R^2_{M,t+1}. \tag{3.3.4}$$

Table 3.11 illustrates the certainty equivalent gain from holding a portfolio based on each of the three forecasting strategies. The strategy based on the augmented model including the $EE_{i,t}$ always provides a higher certainty equivalent than the one obtained using the other forecast models, and it gives a value of $0.7\%-2.7\%$ that is not significantly affected by transaction costs.

Table 3.11. Certainty Equivalence Gain (Δ): Jul.1968-Dec.2007

$F\&F_t + RREL_t + R_i(-1)$ (1)	$F\&F_t + RREL_t + R_i(-1) + EE_{i,t}$ (2)	Historical Average (3)
	1. Switching Strategies	
.0055	.0073	.0037
	2. Switching Strategies with Transaction Costs	
.0055	.0073	.0037
	3. Optimal Portfolio Weighting Strategy	
.0199	.0270	.0033
	4. Optimal Portfolio Weighting Strategy with Transaction Costs	
.0198	.0268	.0033

This table presents the average annualized certainty equivalent (Δ) that an investor with quadratic utility defined in equation (3.3.3) and constant relative risk aversion of $\gamma = 5$ would be willing to pay to switch from the static portfolios paying a market portfolio, $R_{M,t+1}$, to the strategies based on the three forecasting models, as in equation (3.3.4). The variable $EE_{i,t}$ is recursively estimated using only data available at the time of forecast.

We also check if the returns generated by a forecasting strategy can be explained by the CAPM and the Fama-French model. We run the following regressions:

$$R_{P,t+1}-R_{F,t+1}=\alpha^{CAPM}+\beta^{MKT}MKT_{t+1}+u_{t+1}, \tag{3.3.5}$$

$$R_{P,t+1}-R_{F,t+1}=\alpha^{FF}+\beta^{MKT}MKT_{t+1}+\beta^{SMB}SMB_{t+1}+\beta^{HML}HML_{t+1}+v_{t+1}, \tag{3.3.6}$$

where $R_{P,t+1} - R_{F,t+1}$ are the portfolio excess returns and $MKT_{t+1} = R_{M,t+1} - R_{F,t+1}$. This is called the Jensen's α test for the portfolio returns. Under the null hypothesis, the constant term of (3.3.5) or (3.3.6) is not significantly different from zero, implying that each of these models are correct for explaining the portfolio excess returns. Table 3.12 shows that both CAPM and Fama-French model cannot explained the returns based on the three forecasting strategies at the 5% significance level. Besides, the results are robust to the presence of transaction costs (Panels 3 and 4).

Table 3.12. Jensen's α Test for Portfolio Returns: Jul.1968-Dec.2007

	$F\&F_t + RREL_t + R_i(-1)$	$F\&F_t + RREL_t + R_i(-1) + EE_{i,t}$	Historical Average
	(1)	(2)	(3)
	1. Switching Strategies		
α^{CAPM}	**.022**	**.026**	**.005**
	(27.33)	(31.93)	(12.76)
α^{FF}	**.022**	**.026**	**.005**
	(25.40)	(30.42)	(11.81)
	2. Switching Strategies with Transaction Costs		
α^{CAPM}	**.023**	**.024**	**.006**
	(26.62)	(29.05)	(10.12)
α^{FF}	**.023**	**.024**	**.006**
	(24.81)	(27.00)	(9.23)
	3. Optimal Portfolio Weighting Strategy		
α^{CAPM}	**.022**	**.026**	**.005**
	(27.33)	(31.93)	(12.76)
α^{FF}	**.022**	**.026**	**.005**
	(25.40)	(30.42)	(11.81)
	4. Optimal Portfolio Weighting Strategy with Transaction Costs		
α^{CAPM}	**.022**	**.029**	**.004**
	(27.46)	(32.86)	(5.13)
α^{FF}	**.021**	**.028**	**.003**
	(25.01)	(32.34)	(3.71)

This table presents the estimated constant of an OLS regression of the excess portfolio returns, $R_{P,t+1} - R_{F,t+1}$, on solely the excess stock market returns (CAPM), $R_{M,t+1} - R_{F,t+1}$, and on the excess stock market returns plus the other two Fama-French factors (FF), SMB_{t+1} and HML_{t+1}, as in equations (3.3.5) and (3.3.6). The variable $EE_{i,t}$ is recursively estimated using only data available at the time of forecast. Bold denotes significance at the 5% level.

3.4 Equilibrium Error, Stock Market Liquidity and Return Reversals

In this section, we check whether the equilibrium error term ($EE_{i,t}$) is related to stock market liquidity and return reversals. According to Pástor and Stambaugh (2003), stock market liquidity is a broad and elusive concept that generally denotes the ability to trade large quantities at low cost, and without moving the price. Therefore, many concepts have been proposed in the literature to define stock market liquidity. We follow the approach of Pástor and Stambaugh (2003) and focus on an aspect of liquidity associated with transitory price fluctuations implied by order flow. We also investigate whether $EE_{i,t}$ is related to the funding liquidity risk (FL) measure of Fontaine and Garcia (2012), obtained from a panel of U.S. Treasury security pairs across a range of maturities. The elements of each pair have identical maturities, similar cash flows, but may have different ages. The funding factor (FL) can be interpreted as a measure of liquidity risk by relating FL to future repo spreads and by linking FL to broader measures of funding conditions.

To analyse if $EE_{i,t}$ is related to stock market return reversals, we investigate whether $EE_{i,t}$ varies with the momentum factor, WML. The momentum factor is calculated as the cumulative stock market return from month $t-6$ to $t-1$, with the previous month being t and the current month being $t+1$. For robustness, we also include other measures of return reversals, like the Short-Term (STR) and Long-Term return reversal (LTR) factors from French's data library. The Short-Term and Long-Term reversal factors are defined as the average return on the two low prior return portfolios minus the average return on the two high prior return portfolios, or $1/2(SmallLow + BigLow) - 1/2(SmallHigh + BigHigh)$. A stock must have a price for the end of month $t-2$ and a good return for $t-1$ to be included in a portfolio for month t in the calculation of STR, while LTR only includes stock that have a price for the end of month $t-61$ and a good return for $t-13$ to be included in a portfolio for month t.

Table 3.13 presents summary statistics of the liquidity and return reversals measures. There are four liquidity measures: the levels of aggregate liquidity ($ALiq$), innovations in aggregate liquidity or the non-traded liquidity factor ($InLiq$), the value-weighted return on the 10-1 portfolio from a sort on historical liquidity betas or the traded liquidity betas ($TLiq$), and the funding liquidity factor (FL). The liquidity measures $ALiq$, $InLiq$, and $TLiq$ were proposed by Pástor and Stambaugh (2003) and obtained from Lubos Pastor's website. The funding liquidity factor,FL, was available on the website of Jean-Sébastien Fontaine. Panel A of Table 3.13 shows that the $ALiq$, $InLiq$, and $TLiq$ are positively correlated with $EE_{t,i}$, while the funding liquidity factor presents no correlation with the equilibrium error term. In contrast, $EE_{i,t}$ is negatively related to the Momentum factor WML and positively correlated with STR, whereas there is no correlation with LTR.

Panel B of Table 3.13 displays the results of forecasting one-month-ahead excess industry stock returns. The equilibrium error factor, $EE_{i,t}$, is insignificant individually, but it becomes significant when the previous month's excess industry return, $R_i(-1)$, is included in the regression. Besides, $EE_{i,t}$ and $R_i(-1)$ never become insignificant when combined with other variables. The aggregate liquidity factor, $ALiq$, and the non-traded liquidity factor, $InLiq$, of Pástor and Stambaugh (2003) does not forecast excess industry stock market returns over our monthly sample (rows 3 and 6). However, $ALiq$ and $InLiq$ become significant when combined with $EE_{i,t}$ and $R_i(-1)$. These results suggest that $EE_{i,t}$ has some forecast abilities close to those of $ALiq$ and $InLiq$.

We find that the traded liquidity factor $TLiq$ of Pástor and Stambaugh (2003) and the FL of Fontaine and Garcia (2012) are positively and significantly related to future excess industry stock market returns. $TLiq$ and FL are still significant after we control for $EE_{i,t}$ and $R_i(-1)$ (rows 11 and 14 of Table 3.13). In sum, these results indicate that $EE_{i,t}$ does not share the forecasting abilities of $TLiq$ and FL.

The momentum factor, WML, is negative and significant alone in the forecasting equation (row 15). However, it becomes statistically insignificant after including $EE_{i,t}$ (row 16) and $EE_{i,t}$ with the previous month's excess industry return $R_i(-1)$ (row 17). The short-term reversal factor, STR, is positive and significant related to future excess stock industry returns (row 18). However, it is insignificant when combined with $EE_{i,t}$ and $R_i(-1)$ (row 20). In contrast, LTR is negative and statistically significant by itself (row 21) and when combined with $EE_{i,t}$ and $R_i(-1)$ (row 23). Finally, Table 3.13 provides evidence that $EE_{i,t}$ shares some similar information about one-month-ahead excess industry stock returns with the short-term return reversals measures WML and STR.

In sum, our analysis suggests that $EE_{i,t}$ is related to some liquidity and short-term reversals measures.

Table 3.13. Equilibrium Error, Stock Market Liquidity and Return Reversals Measures

			$ALiq$	$InLiq$	$TLiq$	FL	WML	STR	LTR	
A. Summary statistics										
Mean			-.032	-.001	.005	-.108	.009	.006	.003	
S.D.			.062	.056	.032	.984	.041	.031	.025	
Corr. with $EE_{i,t}$			.072	.061	.044	.009	-.042	.048	-.004	
B. Forecasting one-month-ahead excess stock returns										
	$EE_{i,t}$	$R_i(-1)$	$ALiq$	$InLiq$	$TLiq$	FL	WML	STR	LTR	$\bar{R}^2$
(1)	-.049									.000
	(-.888)									
(2)	**-.315**	**9.458**								.007
	(-5.428)	(7.345)								
(3)			.618							.000
			(.845)							
(4)	-.045		-.728							.000
	(-.783)		(-.730)							
(5)	**-.319**	**10.068**	**-2.567**							.008
	(-5.596)	(7.630)	(-2.556)							
(6)				-1.417						.000
				(-1.836)						
(7)	-.037			**-2.861**						.001
	(-.646)			(-2.334)						
(8)	**-.334**	**10.959**		**-5.564**						.009
	(-5.802)	(8.281)		(-4.369)						
(9)					**3.367**					.000
					(3.341)					
(10)	-.054				**3.663**					.000
	(-.973)				(2.429)					
(11)	**-.315**	**9.310**			**4.035**					.007
	(-5.473)	(7.444)			(2.655)					
(12)						**.373**				.003
						(7.212)				
(13)	-.065					**.324**				.002
	(-1.236)					(4.470)				
(14)	**-.225**	**5.913**				**.305**				.004
	(-4.212)	(3.847)				(4.284)				
(15)							**-3.940**			.001
							(-3.624)			
(16)	-.052						-1.563			.000
	(-.959)						(-.981)			
(17)	**-.315**	**9.436**					-.276			.007
	(-5.401)	(7.204)					(-.166)			
(18)								**6.967**		.001
								(4.877)		
(19)	-.059							**5.970**		.001
	(-1.089)							(3.189)		
(20)	**-.310**	**9.119**						3.027		.007
	(-5.380)	(7.393)						(1.767)		
(21)									**-17.402**	.005
									(-14.771)	
(22)	-.054								**-23.356**	.009
	(-1.036)								(-16.632)	
(23)	**-.291**	**8.480**							**-21.472**	.014
	(-5.078)	(6.722)							(-16.075)	

This table reports the Fixed-Effect regression results of the one-month-ahead excess stock returns on some variables. The heteroscedastic-corrected t-statistics are reported in parentheses and bold denotes significance at the 5% level. The liquidity measures $ALiq$, $InLiq$, and $TLiq$ were proposed by Pástor and Stambaugh (2003). $ALiq$ denotes the levels of aggregate liquidity. $InLiq$ denotes innovations in aggregate liquidity or the non-traded liquidity factor. $TLiq$ is the traded liquidity factor. FL is the funding liquidity factor proposed by Fontaine and Garcia (2012). $R_i(-1)$ is the i-th industry excess stock return during the previous month. WML is the momentum factor. STR and LTR are the short and long-term reversal factors from French's data library. $\bar{R}^2$ is the adjusted-R^2. $EE_{i,t}$ is the recursively estimated equilibrium error factor.
*Scaled by 100.

3.5 Conclusion

In this paper, we find that the equilibrium error, the error term from the co-integration relationship between industry stock returns and excess stock market returns, has strong forecasting abilities for excess stock returns, which are increased if combined with the previous month's excess stock returns. Besides, our results suggest that short-term return reversals and liquidity measures are primary reasons for the negative relation between $EE_{i,t}$ and excess stock returns in the subsequent month. This relation is robust after the previous month's excess stock returns is included to account for return reversals. In general, the equilibrium error factor appears to be a pervasive variable that captures systematic movements of stock returns.

We provide new evidence on the out-of-sample stock return predictability, in contrast to Bossaerts and Hillion (1999), Goyal and Welch (2003), and Welch and Goyal (2008), among others, who found negligible out-of-sample predictive power using standard variables. This difference might be due to our forecasting variables discard the variables used by those authors, and the equilibrium error factor is a panel variable that uses more information than time series variables. We also show that the out-of-sample explanatory power is economically meaningful for investors. Simple trading strategies implied by the proposed predictability provide portfolios with higher mean returns and Sharpe ratios than a buy-and-hold or a benchmark strategy does.

In future work, a number of extensions is possible. First, our results seem to be consistent with two different hypotheses, that $EE_{i,t}$ is a proxy for liquidity premium and that $EE_{i,t}$ is a proxy for short-term return reversals. However, it is not possible to discriminate between these two hypotheses, as liquidity and short-term return reversals are two related concepts (Da et al., 2014). Using standard economic theories, we may develop more powerful tests to discriminate between the two hypotheses. Second, the relationship among $EE_{i,t}$, short-term return reversals, and liquidity reveals a connection between market microstructure and general equilibrium theory, as shown in O'Hara (2003). A joint analysis of these two approaches may have important implications for asset pricing.

Bibliography

Alexander, K. S. (1987), "The Central Limit Theorem for Empirical Processes on Vapnik-Cervonenkis Classes," *The Annals of Probability*, 15, 178–203.

Amihud, Y. (2002), "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," *Journal of Financial Markets*, 5, 31–56.

Andrews, D. W. K. (1997), "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1128.

Andrews, D. W. K. and Pollard, D. (1994), "An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes," *International Statistical Review*, 62, 119–132.

Andrews, I. (2012), "A Mean Likelihood Ratio Specification Test," *Working Paper*.

Ang, A. and Bekaert, G. (2002), "International Asset Allocation with Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006), "Quantile Regression Under Misspecification, with an Application to the US Wage Structure," *Econometrica*, 74, 539–563.

Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models," *The Review of Economics and Statistics*, 85, 531–549.

Balduzzi, P. and Lynch, A. W. (1999), "Transaction Costs and Predictability: Some Utility Cost Calculations," *Journal of Financial Economics*, 52, 47–78.

Bierens, H. J. and Wang, L. (2014), “Weighted Simulated Integrated Conditional Moment Tests for Parametric Conditional Distributions of Stationary Time Series Processes,” *Forthcoming in Econometric Reviews.*

Billingsley, P. (1995), *Probability and Measure*, John Wiley & Sons.

Bossaerts, P. and Hillion, P. (1999), “Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?” *Review of Financial Studies*, 12, 405–428.

Bouezmarni, T., Rombouts, J. V., and Taamouti, A. (2012), “Nonparametric Copula-Based Test for Conditional Independence with Applications to Granger Causality,” *Journal of Business & Economic Statistics*, 30, 275–287.

Box, G. E. P. and Pierce, D. A. (1970), “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models,” *Journal of the American statistical Association*, 65, 1509–1526.

Bradley, R. C. (1985), “On the Central Limit Question under Absolute Regularity,” *The Annals of Probability*, 13, 1314–1325.

Breen, W., Glosten, L. R., and Jagannathan, R. (1989), “Economic Significance of Predictable Variations in Stock Index Returns,” *The Journal of Finance*, 44, 1177–1189.

Campbell, J. Y. (1987), “Stock Returns and the Term Structure,” *Journal of Financial Economics*, 18, 373–399.

Campbell, J. Y., Chan, Y. L., and Viceira, L. M. (2003), “A Multivariate Model of Strategic Asset Allocation,” *Journal of Financial Economics*, 67, 41–80.

Campbell, J. Y., Lo, A. W.-C., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Campbell, J. Y. and Shiller, R. J. (1988), “Stock Prices, Earnings, and Expected Dividends,” *The Journal of Finance*, 43, 661–676.

— (2001), "Valuation Ratios and the Long-Run Stock Market Outlook: an Update," Tech. rep., National Bureau of Economic Research.

Campbell, J. Y. and Thompson, S. B. (2008), "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21, 1509–1531.

Campbell, J. Y. and Yogo, M. (2006), "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60.

Carrasco, M. and Chen, X. (2002), "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17–39.

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995), "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147.

Chacko, G. and Viceira, L. M. (2005), "Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets," *Review of Financial Studies*, 18, 1369–1402.

Chernozhukov, V. and Fernández-Val, I. (2005), "Subsampling inference on quantile regression processes," *Sankhya*, 67, 253–276.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013), "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V. and Umantsev, L. (2001), "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics*, 26, 271–292.

Cheung, Y.-W. and Ng, L. K. (1996), "A Causality-in-Variance Test and its Application to Financial Market Prices," *Journal of Econometrics*, 72, 33–48.

Chuang, C. C., Kuan, C. M., and Lin, H. Y. (2009), "Causality in Quantiles and Dynamic Stock Return-Volume Relations," *Journal of Banking & Finance*, 33, 1351–1360.

Clark, T. E. and McCracken, M. W. (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110.

Corradi, V. and Swanson, N. R. (2006), "Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification," *Journal of Econometrics*, 133, 779–806.

Cumby, R. E. and Modest, D. M. (1987), "Testing for Market Timing Ability: A Framework for Forecast Evaluation," *Journal of Financial Economics*, 19, 169–189.

Da, Z., Liu, Q., and Schaumburg, E. (2014), "A Closer Look at the Short-Term Return Reversal," *Management Science*, 60, 658–674.

Delgado, M. A. and Stute, W. (2008), "Distribution-Free Specification Tests of Conditional Models," *Journal of Econometrics*, 143, 37–55.

Dudley, R. M. (1978), "Central Limit Theorems for Empirical Measures," *The Annals of Probability*, 6, 899–929.

Engle, R. and Manganelli, S. (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business & Economic Statistics*, 22, 367–381.

Escanciano, J. and Velasco, C. (2010), "Specification Tests of Parametric Dynamic Conditional Quantiles," *Journal of Econometrics*, 159, 209–221.

Escanciano, J. C., Lobato, I. N., and Zhu, L. (2013), "Automatic Specification Testing for Vector Autoregressions and Multivariate Nonlinear Time Series Models," *Journal of Business & Economic Statistics*, 31, 426–437.

Escanciano, J. C. and Olmo, J. (2010), "Backtesting Parametric Value-at-Risk with Estimation Risk," *Journal of Business & Economic Statistics*, 28, 36–51.

Fama, E. F. and French, K. R. (1988), "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

— (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of financial economics*, 25, 23–49.

— (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F. and Schwert, G. W. (1977), "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

Fleming, J., Kirby, C., and Ostdiek, B. (2001), "The Economic Value of Volatility Timing," *The Journal of Finance*, 56, 329–352.

Fontaine, J.-S. and Garcia, R. (2012), "Bond Liquidity Premia," *Review of Financial Studies*, 25, 1207–1254.

Foresi, S. and Peracchi, F. (1995), "The Conditional Distribution of Excess Returns: An Empirical Analysis," *Journal of the American Statistical Association*, 90, 451–466.

Fortin, N., Lemieux, T., and Firpo, S. (2011), "Decomposition Methods in Economics," in *Handbook of Labor Economics*, eds. Ashenfelter, O. and Card, D., North Holland, Amsterdam, vol. 4 A, 1st ed., pp. 1–102.

Francq, C. and Raïssi, H. (2007), "Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors," *Journal of Time Series Analysis*, 28, 454–470.

Goyal, A. and Welch, I. (2003), "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49, 639–654.

Graham, J. and Harvey, C. (1997), "Grading the Performance of Market-Timing Newsletters," *Financial Analysts Journal*, 53, 54–66.

Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–438.

Granger, C. W. J. and Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoeck und Ruprecht Göttingen.

Granger, C. W. J. and Teräsvirta, T. (2010), *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

Guo, H. (2006), "On the Out-of-Sample Predictability of Stock Market Returns," *Journal of Business*, 79, 645–670.

Guo, H. and Savickas, R. (2006), "Idiosyncratic Volatility, Stock Market Volatility, and Expected Stock Returns," *Journal of Business & Economic Statistics*, 24, 43–56.

Hallin, M. and Jurečková, J. (1999), "Optimal Tests for Autoregressive Models Based on Autoregression Rank Scores," *The Annals of Statistics*, 27, 1385–1414.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Hodrick, R. J. (1992), "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial studies*, 5, 357–386.

Hong, Y. and Lee, T. H. (2003), "Diagnostic Checking for Adequacy of Nonlinear Time Series Models," *Econometric Theory*, 19, 1065–1121.

Hong, Y., Liu, Y., and Wang, S. (2009), "Granger Causality in Risk and Detection of Extreme Risk Spillover Between Financial Markets," *Journal of Econometrics*, 150, 271–287.

Huang, W., Liu, Q., Rhee, S. G., and Zhang, L. (2010), "Return Reversals, Idiosyncratic Risk, and Expected Returns," *Review of Financial Studies*, 23, 147–168.

Ibragimov, I. A. and Has'minskii, R. Z. (1981), *Statistical Estimation: Asymptotic Theory*, New York: Springer.

Inoue, A. and Kilian, L. (2004), "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews*, 23, 371–402.

Iqbal, F. and Mukherjee, K. (2012), "A Study of Value-at-Risk Based on M-Estimators of the Conditional Heteroscedastic Models," *Journal of Forecasting*, 31, 377–390.

Jansson, M. and Moreira, M. J. (2006), "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74, 681–714.

Jegadeesh, N. (1990), "Evidence of Predictable Behavior of Security Returns," *The Journal of Finance*, 45, 881–898.

Jeon, J. and Taylor, J. W. (2013), "Using CAViaR Models with Implied Volatility for Value-at-Risk Estimation," *Journal of Forecasting*, 32, 62–74.

Jeong, K., Härdle, W. K., and Song, S. (2012), "A Consistent Nonparametric Test For Causality In Quantile," *Econometric Theory*, 28, 861–887.

Johannes, M. S., Polson, N., and Stroud, J. R. (2002), "Sequential Optimal Portfolio Performance: Market and Volatility Timing," Unpublished Working paper, University of Chicago, Graduate School of Business.

Johansen, S. (1988), "Statistical analysis of cointegration vectors," *Journal of economic dynamics and control*, 12, 231–254.

— (1995), *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford University Press.

Jones, C. M. (2002), "A Century of Stock Market Liquidity and Trading Costs," Unpublished Working Paper, Columbia University.

Kandel, S. and Stambaugh, R. F. (1996), "On the Predictability of Stock Returns: An Asset-Allocation Perspective," *The Journal of Finance*, 51, 385–424.

Kheifets, I. L. (2015), "Specification Tests for Nonlinear Dynamic Models," *The Econometrics Journal*, 18, 67–94.

Kilian, L. (1999), "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions," *Journal of Applied Econometrics*, 14, 491–510.

Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R. and Machado, J. A. F. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296–1310.

Koenker, R. and Park, B. J. (1996), "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265–283.

Koenker, R. and Xiao, Z. (2002), "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

— (2006), "Quantile Autoregression," *Journal of the American Statistical Association*, 101, 980–990.

Kosorok, M. R. (2007), *Introduction to Empirical Processes and Semiparametric Inference*, Springer.

Koul, H. and Mukherjee, K. (1994), "Regression Quantiles and Related Processes Under Long Range Dependent Errors," *Journal of multivariate analysis*, 51, 318–337.

Koul, H. L. and Stute, W. (1999), "Nonparametric Model Checks for Time Series," *Annals of Statistics*, 27, 204–236.

Kreiss, J.-P. and Paparoditis, E. (2011), "Bootstrap Methods for Dependent Data: A Review," *Journal of the Korean Statistical Society*, 40, 357–378.

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 3, 1217–1241.

Lahiri, S. N. (1999), "Theoretical Comparisons of Block Bootstrap Methods," *Annals of Statistics*, 27, 386–404.

Lee, T.-H. and Yang, W. (2012), "Money-income Granger-causality in quantiles," in *Advances in Econometrics*, Emerald Publishers, pp. 383–407.

Leitch, G. and Tanner, J. E. (1991), "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, 81, 580–590.

Lettau, M. and Ludvigson, S. (2001), "Consumption, Aggregate Wealth, and Expected Stock Returns," *The Journal of Finance*, 56, 815–849.

Lintner, J. (1965), "Security Prices, Risk, and Maximal Gains from Diversification," *The Journal of Finance*, 20, 587–615.

McCracken, M. W. (1999), "Asymptotics for Out-of-Sample Tests of Causality," Unpublished Working paper, Louisiana State University.

Merton, R. C. (1971), "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3, 373–413.

Moskowitz, T. J. and Grinblatt, M. (1999), "Do Industries Explain Momentum?" *The Journal of Finance*, 54, 1249–1290.

Mukherjee, K. (1999), "Asymptotics of Quantiles and Rank Scores in Nonlinear Time Series," *Journal of Time Series Analysis*, 20, 173–192.

Nelson, C. R. and Kim, M. J. (1993), "Predictable Stock Returns: The Role of Small Sample Bias," *The Journal of Finance*, 48, 641–661.

Neumann, M. H. and Paparoditis, E. (2008), "Goodness-of-Fit Tests for Markovian Time Series Models: Central Limit Theory and Bootstrap Approximations," *Bernoulli*, 14, 14–46.

Newey, W. K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.

O'Hara, M. (2003), "Presidential Address: Liquidity and Price Discovery," *The Journal of Finance*, 58, 1335–1354.

Pástor, L. (2000), "Portfolio Selection and Asset Pricing Models," *The Journal of Finance*, 55, 179–223.

Pástor, L. and Stambaugh, R. F. (2000), "Comparing Asset Pricing Models: An Investment Perspective," *Journal of Financial Economics*, 56, 335–381.

— (2003), “Liquidity Risk and Expected Stock Returns,” *Journal of Political Economy*, 111, 642–85.

Pesaran, M. H. and Timmermann, A. (1995), “Predictability of Stock Returns: Robustness and Economic Significance,” *The Journal of Finance*, 50, 1201–1228.

Petersen, M. A. (2009), “Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches,” *Review of Financial studies*, 22, 435–480.

Politis, D., Romano, J., and Wolf, M. (1999), *Subsampling*, Springer-Verlak, New York.

Polk, C., Thompson, S., and Vuolteenaho, T. (2006), “Cross-Sectional Forecasts of the Equity Premium,” *Journal of Financial Economics*, 81, 101–141.

Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer, New York.

Radulović, D. (1996), “The Bootstrap for Empirical Processes Based on Stationary Observations,” *Stochastic Processes and Their Applications*, 65, 259–279.

Rio, E. (2000), *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, vol. 31, Berlin: Springer-Verlag.

Rothe, C. (2012), “Partial Distribution Policy Effects,” *Econometrica*, 80, 2269–2301.

Rothe, C. and Wied, D. (2013), “Misspecification Testing in a Class of Conditional Distributional Models,” *Journal of the American Statistical Association*, 108, 314–324.

Rothschild, M. and Stiglitz, J. E. (1970), “Increasing Risk: I. A Definition,” *Journal of Economic Theory*, 2, 225–243.

Sakov, A. and Bickel, P. (2000), “An Edgeworth Expansion for the m Out of n Bootstrap Median,” *Statistics & Probability Letters*, 49, 217–223.

Sharpe, W. F. (1964), “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *The Journal of Finance*, 19, 425–442.

Stambaugh, R. F. (1999), “Predictive Regressions,” *Journal of Financial Economics*, 54, 375–421.

Stute, W. (1997), "Nonparametric Model Checks for Regression," *The Annals of Statistics*, 25, 613–641.

Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014), "Nonparametric Estimation and Inference for Conditional Density Based Granger Causality Measures," *Journal of Econometrics*, 180, 251–264.

Van der Vaart, A. and Wellner, J. (2000), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.

Weiss, A. (1991), "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation," *Econometric Theory*, 7, 46–68.

Welch, I. and Goyal, A. (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

Whang, Y.-J. (2006), "Smoothed Empirical Likelihood Methods for Quantile Regression Models," *Econometric Theory*, 22, 173–205.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Yang, Z., Tu, A., and Zeng, Y. (2014), "Dynamic Linkages between Asian Stock Prices and Exchange Rates: New Evidence from Causality in Quantiles," *Applied Economics*, 46, 1184–1201.

Zhang, Y.-J. and Wei, Y.-M. (2010), "The Crude Oil Market and the Gold Market: Evidence for Co-integration, Causality and Price Discovery," *Resources Policy*, 35, 168–177.

Zheng, J. X. (1998), "A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 14, 123–138.