Working paper

2020-05

Statistics and Econometrics
ISSN 2387-0303

Factor extraction using Kalman filter and smoothing: This is not just another survey

Pilar Poncela, Esther Ruiz, Karen Miranda

uc3m | Universidad Carlos III de Madrid

Factor extraction using Kalman filter and smoothing: This is not just another survey

Pilar Poncela[1], Esther Ruiz*[2], and Karen Miranda†[2]

[1]Department of Economic Analysis, Quantitative Economics, Universidad Autónoma de Madrid, Spain

[2]Department of Statistics, Universidad Carlos III de Madrid, Spain.

June 24, 2020

Abstract

Dynamic Factor Models, which assume the existence of a small number of unobserved latent factors that capture the comovements in a system of variables, are the main "big data" tool used by empirical macroeconomists during the last 30 years. One important tool to extract the factors is based on Kalman filter and smoothing procedures that can cope with missing data, mixed frequency data, time-varying parameters, non-linearities, non-stationarity and many other characteristics often observed in real systems of economic variables. This paper surveys the literature on latent common factors extracted using Kalman filter and smoothing procedures in the context of Dynamic Factor Models. Signal extraction and parameter estimation issues are separately analyzed. Identification issues are also tackled in both stationary and non-stationary models. Finally, empirical applications are surveyed in both cases.

Keywords: Dynamic Factor Model, EM algorithm, Identification, State-space model.

*Corresponding author: ortega@est-econ.uc3m.es, Tel.: +34 916249851

†Financial support from the Spanish Government Projects ECO2015-70331-C2-1-R and PID2019-108079GB-C22 is gratefully acknowledged by Pilar Poncela. Esther Ruiz acknowledges financial support from projects ECO2015-70331-C2-2-R and PID2019-108079GB-C21 (MINECO/FEDER). This last project is also acknowledged by Karen Miranda.

1 Introduction

In recent decades, dynamic factor models (DFMs) have been widely used to represent comovements within large systems of macroeconomic and financial variables where the cross-sectional dimension is often relatively large compared with the time dimension; see Stock and Watson (2017) for the importance of DFM in time series econometrics research. DFMs generally assume the existence of a small number of unobserved factors capturing the comovements in the system, being the main "big data" tool used by empirical macroeconomists during the last 20 years. Diebold (2003) points out that, although DFMs do not "really" analyse big data, they represent a movement of macroeconomics in this direction. Stock and Watson (2016) also describe the DFM as a "big data" tool.

Two main types of procedures for factor extraction are popular in the related literature. First, many applications consider factors extracted using non-parametric procedures based on Principal Components (PC) which are computational simple and have well-known asymptotic properties. For example, PC is consistent under mild conditions and it is robust to the underlying dependence of common factors and idiosyncratic components as far as the factors are pervasive and the idiosyncratic dependence is weak. As a consequence, PC procedures are very popular for factor estimation and several excellent surveys are available in the literature; see, for example, Bai and Ng (2008) for a technical survey on the econometric theory for PC. However, PC procedures do not use all the information available in the data when the common factors are serially dependent and, consequently, they are not efficient. Furthermore, as there is not a particular specification of the dynamic dependence of the factors, one cannot obtain their out-of-sample forecasts.

Second, factors can be extracted using Kalman Filter and Smoothing (KFS) procedures that cast the DFM as a state-space model (SSM). One important feature of KFS procedures is that they open the door to Maximum Likelihood (ML) estimation of the model parameters and, consequently, if the assumed model specification is correct, is efficient for factor extraction. Furthermore, factor extraction based on KFS procedures allows to handle missing data and mixed frequencies in a simple way; see Camacho, Perez-Quiros and Poncela (2013) and Luciani (2017) who survey the literature on missing observations and mixing frequencies. Moreover, KFS procedures are also of interest in empirical applications because they allow incorporating restrictions on the factor loadings as, for example, in Reis and Watson (2010) and Coroneo, Giannone and Modugno (2016) who impose a block structure, or on the idiosyncratic components as in Luciani (2015). They are also attractive because it is possible to perform counterfactual exercises as in, for example, Luciani (2015) or to incorporate seasonal dependencies as in Alonso et al. (2011), Camacho, Lovcha and Perez-Quiros (2015) and Nieto, Peña and Saboyá (2016). In addition, KFS procedures have been extended to account for regime-switching nonlinearities; see, for example, Camacho, Perez-Quiros and Poncela (2015, 2018) and Camacho, Leiva-Leon and Perez-Quiros

(2016). However, KFS procedures also have drawbacks. Historically, their implementation has been numerically challenging when the cross-sectional dimension of the system under analysis was large. Furthermore, KFS procedures require full specification of the dependence of the common and idiosyncratic components. These specifications introduce potential misspecification that is not reflected in the model-based inference.

The literature about factor extraction using KFS has only been partially reviewed; see Stock and Watson (2011, 2016b), Barhoumi, Darné and Ferrara (2013), and, more recently, Doz and Fuleky (2020). In this paper, we update and complement previous surveys with a focus on open issues that still require further research. This is, by necessity, a selective review of the literature.

The rest of the paper is organized as follows. Section 2 introduces notation by presenting the SSM and the KFS algorithms for factor extraction. Section 3 deals with the representation of DFMs as SSMs and how factor extraction can be performed in this context. We consider the effect on the extracted factors and their MSEs of assuming that the DFM is exact when it is not. We also consider the effect of the dependence and number of underlying factors and of the cross-sectional dimension on the properties of the extracted factors. In section 4, we describe estimation procedures of the parameters of the DFM proposed in the literature and illustrate their performance when extracting the common factors assuming that the model specification is known. Section 5 deals with the model specification. In particular, we describe procedures for the determination of the number of factors and of the lags of factors. In Section 6, we survey empirical applications implementing the KFS in the context of stationary DFMs to describe and forecast the future evolution of variables of interest. Section 7 considers DFMs for non-stationary systems. Section 8 concludes the paper with some final remarks.

2 State-space models: Kalman filter and smoothing algorithms and estimation

In this section, we briefly describe SSMs, the algorithms to extract the unobserved states and the parameter estimators.

2.1 State-space models

SSMs were originally developed by control engineers with the attention focused on a set of m unobserved *state* variables, α_t, that evolve over time and are related with an N dimensional vector of variables, Y_t, observed at time t, for $t = 1, ..., T$. A SSM specifies a full parametric model for both Y_t and α_t and can be formulated in a variety of ways. In this paper, we follow Harvey (1989) and consider the following linear Gaussian SSM

$$Y_t = d_t + Z_t\alpha_t + \varepsilon_t \tag{1}$$

$$\alpha_t = c_t + W_t\alpha_{t-1} + R_t\eta_t \tag{2}$$

where ε_t is an $N \times 1$ white noise vector with covariance matrix H_t and η_t is a $g \times 1$ white noise vector with covariance matrix Q_t. The disturbances ε_t and η_s are uncorrelated with each other for all t and s and are uncorrelated with the initial state, α_0. The system matrices, d_t, Z_t, H_t, c_t, W_t, R_t and Q_t are $N \times 1$, $N \times m$, $N \times N$, $m \times 1$, $m \times m$, $m \times g$ and $g \times g$ are either non-stochastic or depend on past observations, so that they are known at time $t-1$. Equations (1) and (2) are known as the measurement equation and transition equation, respectively. The SSM in equations (1) and (2) is fully specified when additional assumptions about the distribution of the initial state and the distribution of the disturbances are made. Usually, both disturbances and the initial state are assumed to be Gaussian vectors, the latter with mean a_0 and covariance matrix P_0.

In the context of the DFM considered in this paper, the SSM in equations (1) and (2) can be particularized to a time-homogenous model, with the system matrices being time-invariant, and without deterministic terms, i.e. with $d_t = c_t = 0$.[1] Therefore, the SSM that we will consider from now on is given by

$$Y_t = Z\alpha_t + \varepsilon_t \tag{3}$$

$$\alpha_t = W\alpha_{t-1} + R\eta_t \tag{4}$$

where the covariance matrices of ε_t and η_t are H and Q, respectively.

2.2 Kalman filter and smoothing algorithms

The SSM opens the way for the application of the Kalman filter, a recursive procedure for computing the optimal estimator, in the sense that it minimizes the mean square error (MSE), of the state vector at time t based on observations up to and including Y_t. The Kalman filter enables the estimate of the state vector to be continually updated as new observations become available. Furthermore, the Kalman filter also provides the basis for one-step-ahead prediction and smoothing as well as out-of-sample forecast of the factors. Define $a_t = E(\alpha_t|Y_1, ..., Y_t)$ and $P_t = E\left((\alpha_t - a_t)(\alpha_t - a_t)'|Y_1, ..., Y_t\right)$. The Kalman filter updating equations are given by[2]

$$a_t = a_{t|t-1} + P_{t|t-1}Z'\Sigma_t^{-1}\left(Y_t - Za_{t|t-1}\right), \tag{5}$$

[1] There are interesting DFMs in which the system matrices are not time-invariant as, for example, the DFMs with time-varying factor loadings of Stock and Watson (2002) and Eickmeier, Lemke and Marcellino (2015) or the time-varying parameters (volatilities and constants) model proposed by Delle Monache, Petrella and Venditti (2016). Koopman, Mallee and Van der Wel (2010) allow the volatilities to be time-varying. Koop and Korobilis (2014) also consider an SSM with time-varying parameters. Furthermore, d_t and c_t could be different from zero when there are deterministic components; see, for example, Grassi, Proietti, Frale, Marcellino and Mazzi (2015) and Jungbacker and Koopman (2015), among many others. Most of the procedures described in this paper can be extended to these cases. However, we focus on the homogeneous time-invariant SSM to simplify the exposition.

[2] If the disturbances were not normaly distributed, then $a_{t|t-1}$ and a_t are the minimun mean square linear estimators (projections) of α_t given $Y_1, ..., Y_{t-1}$ and $Y_1, ..., Y_t$, respectively.

$$P_t = P_{t|t-1} - P_{t|t-1}Z^{'}\Sigma_t^{-1}ZP_{t|t-1}, \tag{6}$$

where $a_{t|t-1} = E(\alpha_t|Y_1, ..., Y_{t-1})$ and $P_{t|t-1} = E\left((\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})^{'}|Y_1, ..., Y_{t-1}\right)$ are obtained from the following prediction equations

$$a_{t|t-1} = Wa_{t-1}, \tag{7}$$

$$P_{t|t-1} = WP_{t-1}W^{'} + RQR^{'}, \tag{8}$$

and

$$\Sigma_t = ZP_{t|t-1}Z^{'} + H. \tag{9}$$

Note that Σ_t is the covariance matrix of the innovations, $\nu_t = Y_t - E(Y_t|Y_1, ..., Y_{t-1})$. Inverting Σ_t can be a difficult task when the cross-sectional dimension of Y_t, N, is large.[3] Harvey (1989) suggests two alternative possible solutions. First, using the Woodbury identity, it is possible to see that

$$\Sigma_t^{-1} = H^{-1} - H^{-1}Z\left(P_{t|t-1}^{-1} + Z^{'}H^{-1}Z\right)^{-1}Z^{'}H^{-1}, \tag{10}$$

and

$$|\Sigma_t| = |H| \times |P_{t|t-1}| \times |P_{t|t-1}^{-1} + Z^{'}H^{-1}Z|. \tag{11}$$

Expression (10) is easy to evaluate if H is diagonal. Furthermore, given that the covariance matrix of ε_t, H, is time-invariant, it only needs to be inverted once. The second solution proposed by Harvey (1989) to avoid inverting Σ_t is to use the information filter that gives a set of recursions for the information matrix, P_t^{-1}.

Note that using (10) requires the existence of $P_{t|t-1}^{-1}$. However, $P_{t|t-1}$ could be non-invertible if there are MA components in the state vector; see, for example, Ansley and Kohn (1985). Furthermore, Jungbacker and Koopman (2015) argue that (10) does not necessarilly lead to computational gains, because, when N is very large, Σ_t^{-1} and the Kalman filter recursions still remain high-dimensional. Alternatively, they propose a computationally efficient procedure for the Kalman filter recursions. The key insight is that the observed time series, Y_t can be split into a low-dimensional vector series and a high-dimensional vector series as follows

$$Y_t^* = AY_t, \tag{12}$$

where A is an $N \times N$ nonsingular matrix such that $A = [A^L A^H]'$ with $A^L = Z^{\dagger\prime}H^{-1}$ being an $n \times N$ where $n << N$ is the rank of Z and $Z^{\dagger}$ is a basis for the column space of Z. If Z is of full column rank and $n = m$, then $Z^{\dagger\prime} = Z$ and, consequently, $A^L = ZH^{-1}$.[4] Define $Y_t^L = A^L Y_t$.

[3] We assume that the inverse of Σ_t exists.

[4] The matrix A^H is not needed for filtering or estimation.

The measurement equation for Y_t^L is given by

$$Y_t^L = A^L Z\alpha_t + A^L\varepsilon_t. \tag{13}$$

For factor extraction, we need to apply the Kalman filter to the low-dimensional series, Y_t^L; see Grassi et al. (2015) for an empirical application.

The Kalman filter is in steady-state if

$$P_{t+1|t} = \bar{P}. \tag{14}$$

The important fact about a SSM being in steady-state is that the recursion for the MSEs of the state is redundant and, consequently, the covariance matrix of the innovations, Σ_t, also converges to a steady-state $\bar{\Sigma}$, as follows

$$\lim_{t\to\infty} \Sigma_t = \bar{\Sigma} = Z\bar{P}Z^{'} + H, \tag{15}$$

and

$$\bar{\Sigma}^{-1} = H^{-1} - H^{-1}Z\left(\bar{P}^{-1} + Z^{'}H^{-1}Z\right)^{-1}Z^{'}H^{-1}. \tag{16}$$

Harvey (1989) shows that, if the system is detectable[5] and stabilisable[6], and if $P_{1|0}$ is positive semi-definite, then

$$\lim_{t\to\infty} P_{t|t-1} = \bar{P}, \tag{18}$$

with $\bar{P}$ being independent of $P_{1|0}$. If the system is observable and if $P_{1|0} - \bar{P}$ is positive definite or $P_{1|0} = \bar{P}$, then the result in (18) holds although this is not sufficient for the steady state to be reached exponentially fast.

As mentioned above, on top of one-step-ahead predictions, $a_{t|t-1}$, and update filtered estimates of the state, a_t, one can also obtain smoothed estimates, given by $a_{t|T} = E[\alpha_t|Y_1, ..., Y_T]$ together with their corresponding MSEs, $P_{t|T} = E[(\alpha_t - a_{t|T})(\alpha_t - a_{t|T})^{'}|Y_1, ..., Y_T]$. The smoothed estimates of the state can be computed backwards, for $t = T-1, ..., 1$, by the following recursive formulae

$$a_{t|T} = a_t + C_t\left(a_{t+1|T} - Wa_t\right) \tag{19}$$

[5]The state vector is observable if it can be determined exactly given $Y_t, ..., Y_{t+m-1}$. The system is observable if $Rank\left[Z^{'}, W^{'}Z^{'}, ..., (W^{'})^{m-1}Z^{'}\right] = m$. Observability implies detectability.

[6]Defining $Q = R^*R^{*'}$, the transition equation can be written as

$$\alpha_t = W\alpha_{t-1} + B\eta_t^*, \tag{17}$$

where $B = RR^*$ and $\eta_t^* = R^{*-1}\eta_t$ is such that $E(\eta_t^*) = 0$ and $E(\eta_t^*\eta_t^{*'}) = I_g$. The model is controlable if $Rank\left[B, WB, ..., W^{m-1}B\right] = m$. If B is of rank m, the controlability condition is satisfied. This condition means that from any particular value of α_t, the noises η_t^* can be chosen in such a way that any desired value for α_{t+m} can be attained. When the model is not controlable, certain elements in the state vector can only be manipulated indirectly via other elements. Conditions that are sufficient for controlability are also sufficient for stabilisability.

$$P_{t|T} = P_t + C_t \left(P_{t+1|T} - P_{t+1|t} \right) C_t' \tag{20}$$

where $C_t = P_t W' P_{t+1|t}^{-1}$.

The prediction equations of the Kalman filter in (21) and (22) can also be used recursively to obtain h-step-ahead out-of-sample forecasts of the state by using them from $h = 2, ..., H$ without the update step, as follows

$$a_{T+h|T+h-1} = W a_{T+h-1|T+h-2}, \tag{21}$$

$$P_{T+h|T+h-1} = W P_{T+h-1|T+h-2} W^{'} + RQR^{'}, \tag{22}$$

with $a_{T+1|T}$ and $P_{T+1|T}$ given by (21) and (22).

2.3 Parameter estimation

The system matrices may depend on a set of unknown parameters. One of the main statistical tasks is often the estimation of these parameters. The Kalman filter is also important because it enables the likelihood function to be calculated through the prediction error decomposition, opening the way for ML estimation of any unknown parameters of the model, Ψ. For a Gaussian model, the prediction error decomposition of the log-likelihood function is given by

$$logL(Y; \Psi) = -\frac{NT}{2} log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} log|\Sigma_t| - \frac{1}{2} \sum_{t=1}^{T} \nu_t' \Sigma_t^{-1} \nu_t, \tag{23}$$

where $Y = (Y_1, ..., Y_T)$ and both ν_t and Σ_t can be obtained from the Kalman filter. In multivariate models, it is not always clear how to compute appropriate starting values for the initial state; see de Jong (1991) for an easy to implement algorithm.

Very recently, Delle Monache and Petrella (2019) take advantage of the matrix representation of the SSM and derive closed form expressions of the log-likelihood and the smoothed estimator of the state that are computationally feasible even for very large N as far as the covariance matrices of the measurement and transition noises are invertible.

Finally, the log-likelihood can also be obtained using the transformation proposed by Jungbacker and Koopman (2015). In particular

$$logL(Y; \Psi) = logL(Y^L; \Psi) - \frac{T}{2} log \frac{|\Sigma_\varepsilon|}{|A^L \Sigma_\varepsilon A^{L\prime}|} - \frac{1}{2} \sum_{t=1}^{T} e_t' \Sigma_\varepsilon^{-1} e_t, \tag{24}$$

where $logL(Y^L; \Psi)$ can be obtained as in (23) based on the low dimensional vector $Y^L = \left(Y_1^L, ..., Y_T^L\right)$ and e_t is the Generalized Least Squares (GLS) residual of the regression of Y_t on Z with covariance matrix Σ_ε.

The numerical maximization of the likelihood can be a difficult task when N is large and the number of parameters in the model is also large. In this case, the Gaussian log-likelihood

can be maximized using the expectation maximization (EM) algorithm of Dempster, Laird and Rubin (1977) proposed by Shumway and Stoffer (1982) and Watson and Engle (1983) for ML estimation in SSMs; see Wu (1983) for the convergence properties of the EM algorithm.

A very important issue related with the estimation of the parameters of SSMs is the identification of the state vector. Given the model in (28) and (29), there is not a unique representation of the state vector. Define an arbitrary non-singular $m \times m$ matrix, B, and consider a new state vector $\alpha_t^* = B\alpha_t$. The following SSM is observationally equivalent to the SSM in equations (3) and (4)

$$Y_t = Z^* \alpha_t^* + \varepsilon_t \tag{25}$$

$$\alpha_t^* = W^* \alpha_{t-1}^* + R^* \eta_t \tag{26}$$

where $Z^* = ZB^{-1}$, $W^* = BWB^{-1}$ and $R^* = BR$. Therefore, there are m^2 unknowns in matrix B. In order to identify the model, Harvey (1989) proposes to restrict $RQR^{'} = I_m$, which means $\frac{m(m+1)}{2}$ restrictions, and the elements in Z to be such that $z_{ij} = 0$ for $j > i, i = 1, ..., m-1$, which implies the $\frac{m(m-1)}{2}$ additional restrictions needed to identify the state vector; see Trenkler and Weber (2016) for a discussion on identification issues in state space models.

Subject to certain regularity conditions, the ML estimator, $\hat{\Psi}$, has a limiting multivariate normal distribution with mean Ψ and covariance matrix $T^{-1}A^{-1}$ where $A = \lim_{T\to\infty} \frac{IM}{T}$ with IM being the information matrix; see Harvey (1989) for a discussion on the regularity conditions.

When the Kalman filter converges to a steady state exponentially fast, the properties of the ML estimator do not depend on the way in which the filter is started off.

Note that if the unknown parameters in $\bar{\Sigma}$, the steady state of Σ_t, are independent of the unknown parameters determining $E(Y_t|Y_{t-1}, ..., Y_1)$, then maximizing the likelihood is equivalent to minimizing

$$S(\Psi) = |\sum_{t=1}^{T} \nu_t \nu_t^{'}|. \tag{27}$$

3 Dynamic Factor Models and KFS factor extraction

DFMs are examples of the much larger class of SSMs, in which observable variables are expressed in terms of unobserved or latent variables, which in turn evolve according to some lagged dynamics. In this section, we describe how DFMs can be expressed as SSMs and how the KFS algorithms can be used for factor extraction.

Consider that $Y_t = (Y_{1t}, ..., Y_{Nt})^{'}$, $t = 1, ..., T$, is a stationary zero mean $N \times 1$ vector time

series generated by the following DFM[7]

$$Y_t = \chi_t + \varepsilon_t, \tag{28}$$

where $\chi_t = (\chi_{1t}, ..., \chi_{Nt})'$ and $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ are $N \times 1$ vectors representing the common and idiosyncratic components, respectively. Depending on the definition of χ_t, there are two main versions of the DFM usually considered in the literature: the "*static*" and the "*dynamic*" versions.

3.1 Static Dynamic Factor Models

The "*static*" version of the DFM (S-DFM) establishes a contemporaneous relationship between each variable in the system and the unobserved underlying factors at time t as follows

$$\chi_{it} = \lambda_i' F_t \tag{29}$$

where $\lambda_i = (\lambda_{i1}, ..., \lambda_{ir})$ is the $r \times 1$ vector of factor loadings of the Y_i variable. It is popular to assume that F_t, the $r \times 1$ vector of common factors, evolves over time following a stationary VAR(p) model given by

$$F_t = \Phi_1 F_{t-1} + \Phi_2 F_{t-2} + ... + \Phi_p F_{t-p} + u_t, \tag{30}$$

where u_t is an $r \times 1$ white noise vector with covariance matrix Σ_u. This specification of the factors have been considered in many empirical studies with the values of r and p depending on the particular application; see, for example, Proietti (2011), who specified $r = 6$ and $p = 1$, Camacho and Perez-Quiros (2010) and Scotti (2016), both with $r = 1$ and $p = 11$.[8] From now on, we will consider $p = 1$ to simplify notation. However, all results can be easily extended to models with $p > 1$.

Finally, the idiosyncratic components, ε_{it}, are often specified as AR(p_i^*) processes as follows

$$\varepsilon_{it} = \theta_{1i}\varepsilon_{it-1} + \theta_{2i}\varepsilon_{it-2} + ... + \theta_{p_i^* i}\varepsilon_{it-p_i^*} + e_{it}, \tag{31}$$

where $e_t = (e_{1t}, ..., e_{Nt})$ is the vector of idiosyncratic noises, assumed to be white noise with covariance matrix Σ_e. The autoregressive order in (31) depends on the particular application. In many studies, it is considered the same for all $i = 1, ..., N$. In this case, denote $p^* = p_i^*$. For example, Scotti (2016) assumes that $p^* = 1$ while Camacho and Perez-Quiros (2010) assume that

[7] We assume that all deterministic components have been removed from the series in Y_t previous to their analysis.

[8] Some authors also consider factors generated by Moving Average (MA) processes; see, for example, Otrok and Whiteman (1998) and Bai and Ng (2007). However, these are the exception.

$p_i^* = 1$ for all monthly variables and $p_i^* = 5$ for the quarterly variables (GDP and unemployment) of their system. Also, García-Ferrer and Poncela (2002) allow p_i^* to vary between 1 and 6. Note that these two models are small-scale with $N = 13$ and $N = 5$, respectively. As for the factors, from now ownwards, we will assume that $p^* = p_i^* = 1$ to simplify the analysis. Finally, if the idiosyncratic components, ε_t, are assumed to be cross-sectionally uncorrelated, i.e. Σ_e is diagonal, the DFM is known as "*exact*" while if the idiosyncratic noises are weakly cross-correlated, the DFM is called "*approximate*". In an exact DFM, for the purposes of explaining contemporaneous movements and making forecasts, once you know the factors, the other series provide no additional useful information.

Therefore, the S-DFM is given by

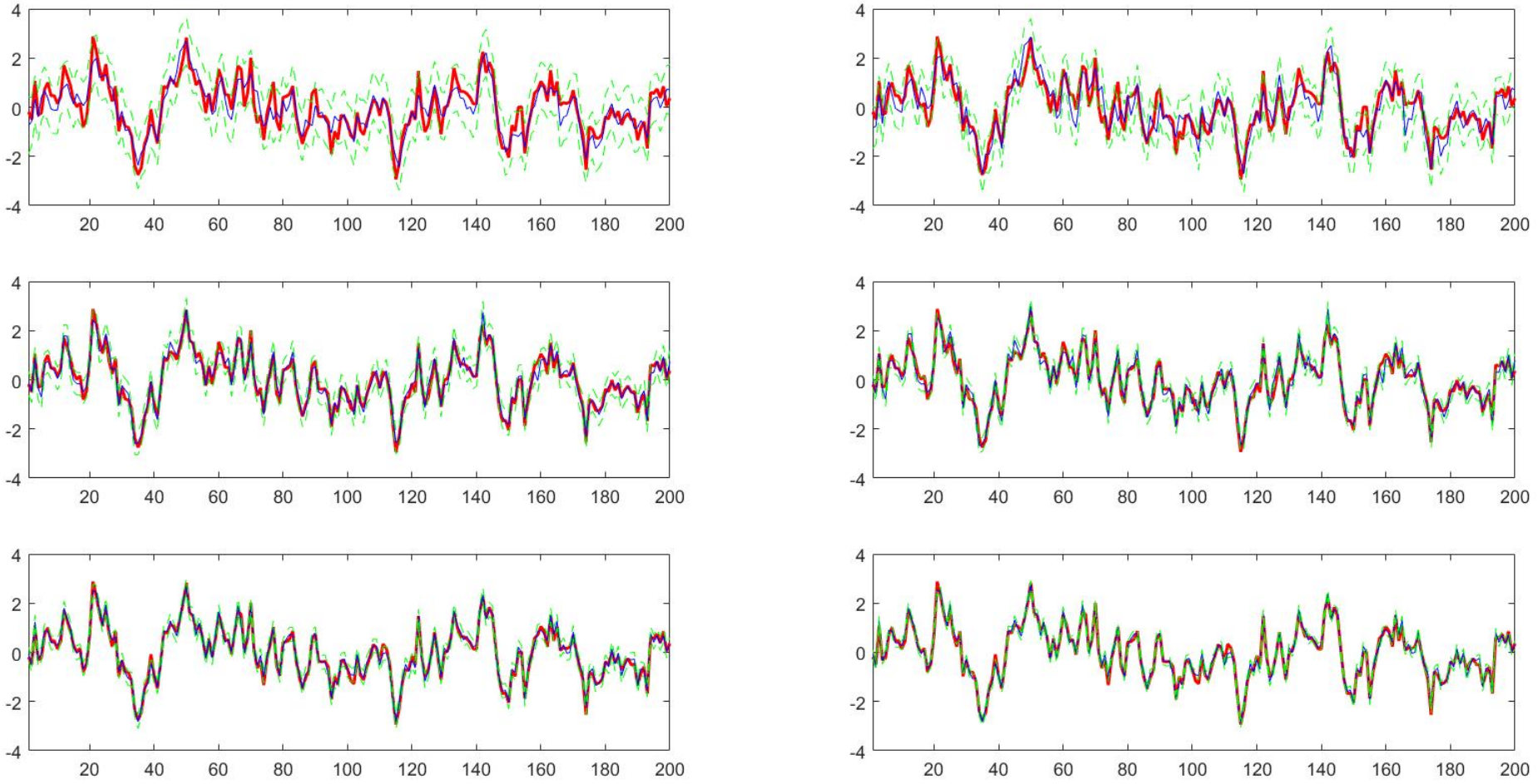
$$Y_t = \Lambda F_t + \varepsilon_t \tag{32}$$

where $\Lambda = (\lambda_1, ..., \lambda_N)'$ is the $N \times r$ matrix of factor loadings and F_t and the elements of ε_t are defined in (30) and (31), respectively.

Consider first the S-DFM with serially uncorrelated idiosyncratic noises, i.e. $\theta_i = 0$ for $i = 1, ..., N$. In this case, it is straighforward to write the S-DFM as a SSM in equations (3) and (4) by considering $m = g = r$, $Z = \Lambda$, $\alpha_t = F_t$, $H = \Sigma_e$, $W = \Phi_1$, $R = I$ and $Q = \Sigma_u$. If the S-DFM is further exact, assuming that r and p as well as all parameters in the model are known, the KFS algorithms described in Section 2 can be implemented to extract the factors, regardless of the cross-sectional dimension, N, by applying the expression of the inverse of the innovation covariance matrix in (10). In the more realistic case in which Σ_e is not diagonal, one can still invert Σ_t using expression (10) by inverting Σ_e using the Cholesky decomposition $\Sigma_e = DD'$ where D is a lower triangular matrix. Inverting D is feasible even if N is very large. Alternatively, the factors can be extracted using the KFS algorithms as if Σ_e were diagonal although it is not. Obvioulsy, in this latter case, the factors extracted using KFS and the corresponding MSEs obtained are not the true conditional means (projections) and MSEs of the factors. As an illustration, we simulate a system with $r = 1$ factor characterized by an AR(1) model with autoregressive parameter $\phi = 0.7$ and variance of the noise $1 - \phi^2$. The idiosyncratic noises are serially uncorrelated, homoscedastic, with their variances being 0.5, and cross-sectionally correlated with the correlation matrix given by a Toepliz matrix whose (i, j) element is given by $\tau^{|i-j|}$ with $\tau = 0.5$. We simulate systems with $N = 5$, 50 and 150 variables, which represent small, medium and large systems, respectively, and $T = 200$ observations. The accuracy of the point factor estimates is measured by computing the sample MSEs, given by $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t - F_t\right)^2$, while the accuracy of the KFS MSEs is measured by the sample coverages computed as the percentage of times the true factor is included in the 95% confidence interval constructed using the KFS's MSEs. Table 1, which reports the sample MSEs and coverages for

the particular simulations described above, shows that, for these particular simulated systems, the increases in the MSEs of the misspecified models are 11.87%, 13.33% and 5% when compared with the corresponding MSEs of the true specification, for $N = 5, 50$ and 150, respectively. As expected, the loss of accuracy of the point estimated factors when assuming falsely that the idiosyncratic components are uncorrelated, is negligible when the cross-sectional dimension, N, is large. However, when looking at the coverages reported in Table 1, we can observe that the coverages of the intervals constructed with the miss-specified smoother are below nominal. When $N = 5$, the coverage is 90% while the coverage is 75.5% when $N = 150$. Therefore, the undercoverage seems to be larger as N increases. The smooth MSEs are clearly affected by the misspecification. In this particular example, they are smaller than they should be. To have a visual plot of this simulation, Figure 1 plots the true simulated factor together with the factor extracted using the smoothing algorithm assuming all parameters are known (left column) and the smoothed factor extracted when all the parameters are known but Σ_e is assumed to be diagonal (right column) together with their corresponding 95% confidence bounds. Figure 1 shows that the differences between both point smooth factors are only visually appreciable when $N = 5$. However, the intervals of the misspecified smoother are wider than those obtained with the correct model. When $N = 50$, the point estimates of the factors are nearly the same with the main differences appearing in the MSEs which are smaller than they should be. When N is large, the MSEs are already so close to zero that it is indifferent whether the filter is run with the true covariance matrix of the idiosyncratic components or with a diagonal matrix; see Poncela and Ruiz (2020) for further illustrations with other cross-sectional dimensions when the idiosyncratic components are treated as if they were cross-sectionally uncorrelated and Luciani (2014) who shows that accounting for cross-correlation rarely boosts the forecasting accuracy.

Figure 1: Simulated factor (red line) together with factor extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification (left column) and the true model assuming that the idiosyncratic component is serial and cross-sectional uncorrelated (right column). The systems are simuated by a DFM with $N = 5$ (first row), $N = 50$ (second row) and $N = 150$ (third row) and weakly cross-correlated idiosyncratic components.



	Known parameters				Estimated parameters			
	True model		Misspecified model		True model		Misspecified model	
	MSE	Coverage	MSE	Coverage	MSE	Coverage	MSE	Coverage
Cross-sectionally correlated idiosyncratic noises								
$N = 5$	0.219	0.965	0.245	0.900	0.218	0.945	0.266	0.670
$N = 50$	0.045	0.960	0.051	0.805	0.204	0.525	0.073	0.730
$N = 150$	0.020	0.965	0.021	0.755	4.879	0.045	0.226	0.375
Autocorrelated idiosyncratic noises								
$N = 5$	0.173	0.970	0.174	0.965	0.197	0.899	0.226	0.870
$N = 50$	0.019	0.985	0.019	0.970	0.071	0.894	0.052	0.865
$N = 150$	0.008	0.965	0.008	0.955	0.230	0.397	0.234	0.330
Cross-sectionally correlated and autocorrelation idiosyncratic noises								
$N = 5$	0.302	0.985	0.348	0.865	-	-	-	-
$N = 50$	0.051	0.990	0.055	0.755	-	-	-	-
$N = 150$	0.022	0.970	0.023	0.715	-	-	-	-

Table 1: Mean Square Errors (MSEs) of factors extracted using the true and misspecified DFM with known and estimated parameters. The coverages correspond to the intervals constructed using the KFS's MSEs with a nominal coverage of 95%.

Consider now the S-DFM in which the idiosyncratic noises are serially correlated according to (31) and denote by Θ the matrix with $\theta_i, i = 1, ..., N$ in its main diagonal. In this case, the DFM can be reformulated by augmenting the state vector with lags of the factors as follows:

$$Y_t = \Theta Y_{t-1} + \begin{bmatrix} \Lambda & -\Theta\Lambda \end{bmatrix} \begin{bmatrix} F_t \\ F_{t-1} \end{bmatrix} + e_t \tag{33}$$

$$\begin{bmatrix} F_t \\ F_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 \\ I_r & 0 \end{bmatrix} \begin{bmatrix} F_{t-1} \\ F_{t-2} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \end{bmatrix}, \tag{34}$$

where I_r is the $r \times r$ identity matrix; see Watson and Engle (1983), Stock and Watson (2005), Reis and Watson (2010), Jungbacker *et al.* (2011), Pinheiro, Rua and Dias (2013), Jungbacker and Koopman (2015) and Bai and Li (2016) for implementations of the model in (33) and (34). Defining the observations as $Y_t - \Theta Y_{t-1}$, the model in (33) and (34) can be directly cast in state space form by setting $m = 2r$, $g = m$, $Z = \begin{bmatrix} \Lambda & -\Theta\Lambda \end{bmatrix}$, $\alpha_t = \begin{bmatrix} F_t & F_{t-1} \end{bmatrix}'$, $W = \begin{bmatrix} \Phi_1 & 0 \\ I_r & 0 \end{bmatrix}$, $R = \begin{bmatrix} I_r & 0 \end{bmatrix}$ and $\eta_t = u_t$, in the SSM in (3) and (4).[9]

Even if the idiosyncratic components, ε_t, are serially correlated, the KFS can be run as if they were uncorrelated. Figure 2 illustrates the results by plotting the true factor simulated by the same model described above but with the cross-sectionally uncorrelated idiosyncratic

[9] Alternatively, one can deal with the autocorrelation of the idiosyncratic noises by augmenting the state vector by ε_t; see, for example, Bańbura and Modugno (2014), Jungbacker *et al.* (2011) and Coroneo, Giannone and Modugno (2016). The main problem associated with this alternative is that the state vector dimension, $m = g = r + N$, increases with N and can be unfeasible from a computational point of view for large cross-sectional dimensions. Both formulations lead to the same results when the initialization issues are properly accounted for.

components generated by independent AR(1) models all of them with autoregressive parameter 0.5 as in Doz, Giannone and Reichlin (2012). The factors are extracted by the KFS based on the true model and assuming that the idiosyncratic components are serially uncorrelated. Note that, in this case, the difference between the point estimated factors is even smaller than in Figure 1. Table 1, which reports the sample MSEs and coverages for these particular simulations, shows that both are nearly the same regardless of whether the autocorrelation of the idiosyncratic component is taken into account. It seems that the effects of the misspecification of the serial correlation in the idiosyncratic components are milder than the effects of misspecification of cross-sectional correlations.[10]

Finally, the same conclusions can be obtained from Figure 3 that plots the same quantities described above when the idiosyncratic noises are both serial and cross-sectionally correlated; see also the quantities reported in Table 1 that shows that the undercoverage can be mainly attributed to the lack of consideration of the cross-correlation of the idiosyncratic components.

Note that these conclusions are obtained in a very simple stationary model with a single factor that is moderately serially dependent. It is possible that if the factor and idiosyncratic noises have other alternative serial dependences, the conclusions could be different.

[10]This result could be expected when extracting the factors. However, misspecification of the serial correlation of the idiosyncratic components could have implications for forecasting; see D'Agostino and Giannone (2012) and Poncela, Senra and Sierra (2020) who conclude that the effect of the idiosyncratic dynamics on forecasting macroeconomic inflation and commodity inflation, respectively, are negligible.

Figure 2: Simulated factor (red line) together with factor extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification (left column) and the true model assuming that the idiosyncratic component is serial and cross-sectional uncorrelated (right column). The systems are simuated by a DFM with $N = 5$ (first row), $N = 50$ (second row) and $N = 150$ (third row) and serially correlated idiosyncratic components.

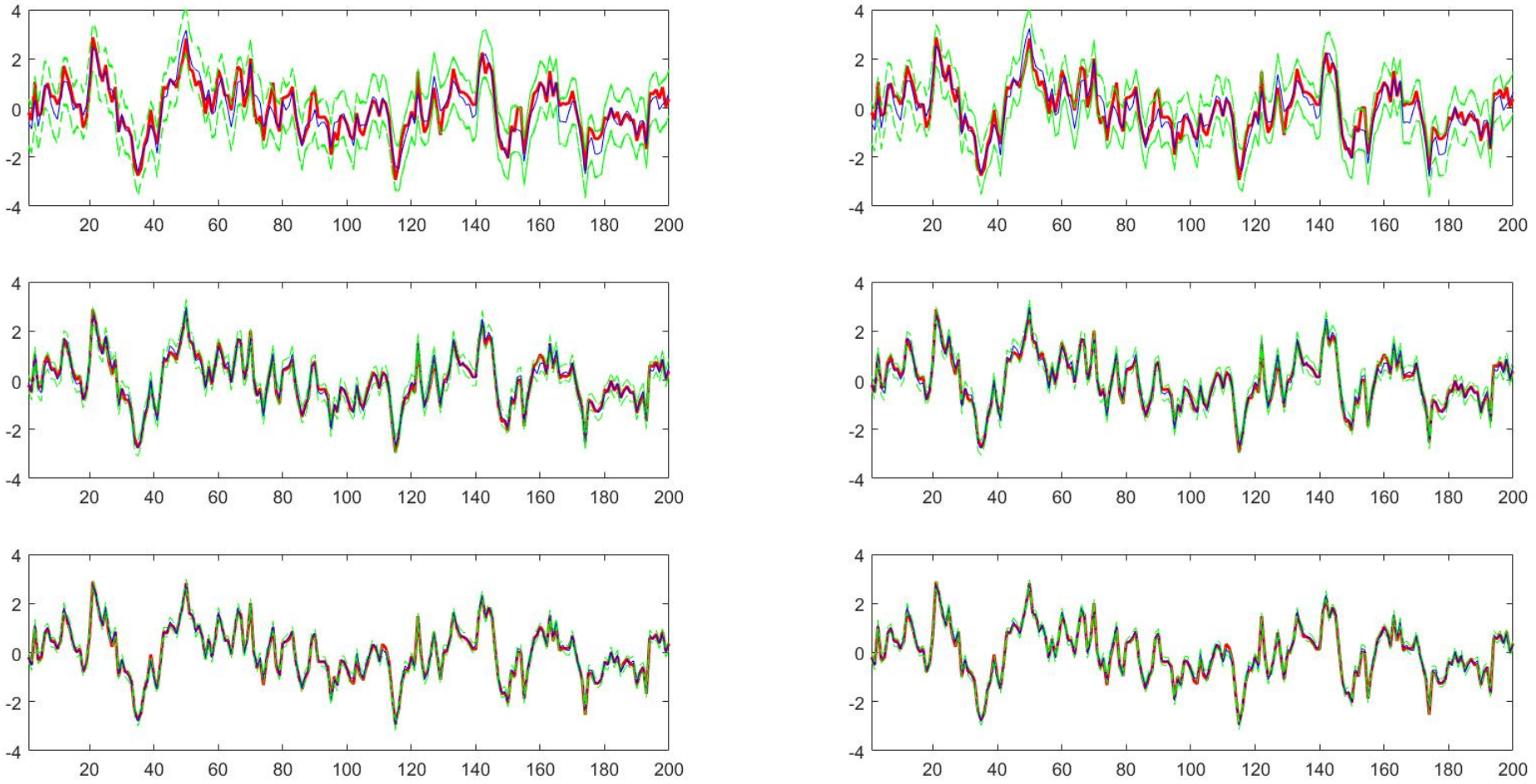
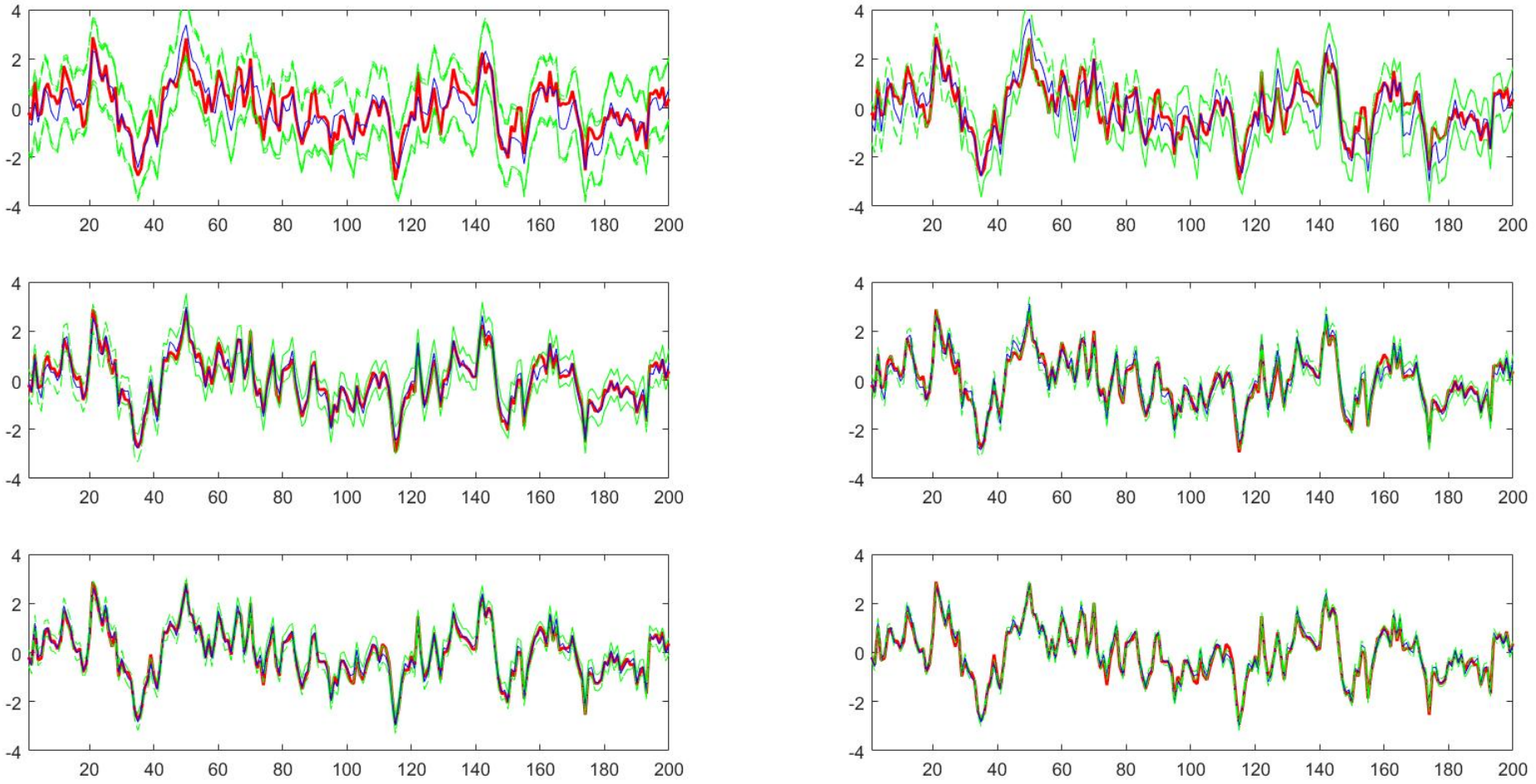


Figure 3: Simulated factor (red line) together with factor extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification (left column) and the true model assuming that the idiosyncratic component is serial and cross-sectional uncorrelated (right column). The systems are simuated by a DFM with $N = 5$ (first row), $N = 50$ (second row) and $N = 150$ (third row) and serial and cross-sectionally correlated idiosyncratic components.



3.2 Dynamic version of the dynamic factor model

In the "*dynamic*" version of the DFM (D-DFM), each variable in the system at time t is related with the unobserved factors at time t, $t-1$,...,$t-s$.[11] Following Bai and Ng (2007) and Stock and Watson (2016), among others, we express the common component of the D-DFM as follows

$$\chi_t = \Lambda(L)G_t, \tag{35}$$

where $\Lambda(L) = \Lambda_0 + \Lambda_1 L + ... + \Lambda_s L^s$ and G_t is the $q \times 1$ vector of unobserved factors. Assuming for simplicity that the factors, G_t, follow a VAR(1) and $s = 1$, the D-DFM with the common component defined as in (35) can be written as a S-DFM with restrictions as follows[12]

$$Y_t = \begin{bmatrix} \Lambda_0 & \Lambda_1 \end{bmatrix} \begin{bmatrix} G_t \\ G_{t-1} \end{bmatrix} + \varepsilon_t \tag{36}$$

$$\begin{bmatrix} G_t \\ G_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 \\ I_q & 0 \end{bmatrix} \begin{bmatrix} G_{t-1} \\ G_{t-2} \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \end{bmatrix}, \tag{37}$$

where $F_t = \begin{bmatrix} G_t \\ G_{t-1} \end{bmatrix}$ can be treated as an $r = q(s+1)$ vector of static factors; see Bai and Ng (2007) for the expression of the D-DFM as an S-DFM in a more general context. The covariance matrix of the disturbances of the static factors, Σ_u has rank q.

Alternatively, the D-DFM can be written as a S-DFM as follows

$$Y_t = (\Lambda_0\Phi + \Lambda_1)G_{t-1} + \Lambda_0 v_t + \varepsilon_t \tag{38}$$

$$\begin{bmatrix} G_{t-1} \\ v_t \end{bmatrix} = \begin{bmatrix} \Phi & I_q \\ 0 & 0 \end{bmatrix} \begin{bmatrix} G_{t-2} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ I_q \end{bmatrix} v_t \tag{39}$$

Once more, the D-DFM is writen as a S-DFM with two common factors. It is important to point out that, when the D-DFM is written as a S-DFM as in equations (36) and (37), it is not possible to assume that the static factors are orthogonal. Alternatively, if the D-DFM is written as a S-DFM as in equations (38) and (39), the factors can be assumed to be orthogonal but the matrix of loading is of rank q and, consequently, it is not possible to find a rotation such that $\Lambda'\Lambda$ is diagonal.[13]

If the idiosyncratic noises are serially uncorrelated, then the model in (36) and (37) can be cast in SSM and the KFS can be implemented to extract the factors; see Pinheiro, Rua and Dias

[11] Bai and Ng (2007) claim that, for forecasting purposes, little is to be gained from a distinction between static and dynamic factors.

[12] The cases in which $p > 1$ and/or $s > 1$ follow straightforwardly but they are notationally more cumbersome.

[13] Note that many procedures based on PC require that simultaneously the factors and loadings are orthogonal.

(2013) for an empirical implementation. If the idiosyncratic noises are serially correlated as in (31), then the D-DFM with $s = p = 1$ can be expressed as follows

$$Y_t = \Theta Y_{t-1} + \Lambda_0 G_t + (\Lambda_1 - \Theta\Lambda_0) G_{t-1} - \Theta\Lambda_1 G_{t-2} + e_t \tag{40}$$

$$\begin{bmatrix} G_t \\ G_{t-1} \\ G_{t-2} \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 & 0 \\ I_r & 0 & 0 \\ 0 & I_r & 0 \end{bmatrix} \begin{bmatrix} G_{t-1} \\ G_{t-2} \\ G_{t-3} \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \\ 0 \end{bmatrix}. \tag{41}$$

The D-DFM in (40) and (41) can be cast as a SSM and the factors extracted using KFS without further issues.

As explained above, when describing the KFS factor extraction in S-DFMs, the factors of a D-DFM can also be extracted using KFS as if the idiosyncratic noises were serial and cross-sectionally uncorrelated even if they are not.

4 Estimation of parameters of DFMs

The KFS factor extraction described in the previous section assumes known parameters. However, in practice, the parameters are unknown and need to be estimated before running the KFS algorithms. In this section, we survey the estimators of the parameters of the DFM based on ML when the model specification, i.e. s, r and p, is known.[14] The first issue faced when estimating the parameters of a DFM is related with parameter identification. In this section, we first describe the parameter identification and then their ML-based and Least Squares (LS)-based estimation.

4.1 Identification

As explained in Section 2, in a SSM as that in equations (3) and (4), one needs to impose m^2 restrictions to identify the m unobserved states; see Anderson and Rubin (1956) for an excellent discussion on identification issues in the context of static factor models. Consequently, in the S-DFM with serially uncorrelated idiosyncratic noises, one needs to impose r^2 restrictions to identify the factors. For many applications, including macro-monitoring and forecasting, it is necessary only to identify the space spanned by the factors. Consequently, it is popular to solve the lack of identification by imposing mathematically convenient normalizations. It is common to

[14] In this paper, we focus on ML-based estimators. Other estimators have been proposed in the literature as Bayesian estimators (Otrok and Whiteman (1998), Lopes and West (2004), Kose, Otrok and Whiteman (2003), Jackson, Kose, Otrok and Owyang (2016) and Kaufmann and Schumaker (2019)), MCMC procedures as in Moench, Ng and Potter (2013) and the frequency-domain version of the EM algorithm (Fiorentini, Galesi and Sentana, 2018). Finally, Kapetanios and Marcellino (2009) propose estimating the parameters using subspace algorithms; see Eickmeier and Ziegler (2008) for a comparison of the predictive performance of factors estimated using the subspace estimator with alternative estimators of the factors.

assume that the factor noise covariance matrix, $\Sigma_u = I_r$ or the covariance of the factors, $\Sigma_F = I_r$ that amounts to $\frac{r(r+1)}{2}$ restrictions and that $\lambda_{ij} = 0$, $j > i$ that imposes the $\frac{r(r-1)}{2}$ additional restrictions needed. For example, Jungbacker and Koopman (2015) assume instead that Σ_u is diagonal and the diagonal elements of the r top rows of the loading matrix Λ are restricted to be one and Solberger and Spanberg (2020) assume that $\Sigma_F = I_r$ with additional restrictions on the matrix of loadings. Alternatively, many authors assume that $\Lambda = [I_r \Lambda_{(N-r)}]'$ (r^2 restrictions); see, for example, Stock and Watson (2011), Proietti (2011), Bai and Ng (2013) and Coulombe et al. (this issue). This latter restriction is denoted as the "*named factor*" restriction by Stock and Watson (2011).[15] Alternatively, Reis and Watson (2010) assume that $\Lambda'\Lambda$ is diagonal and that the columns of Λ sum up to zero. Note that these are $\frac{r(r+1)}{2}$ restrictions. Althuogh, they did not say it explicitly, additional $\frac{r(r-1)}{2}$ restrictions are needed for identification; see Bai and Li (2012, 2016) who consider five sets of restrictions and discuss how the distribution of a ML estimator depends on the identification restrictions.

Note that, when the idiosyncratic errors are correlated and/or the DFM has dynamic factors, even if the state vector is extended, the number of restrictions is still r^2 given that all new elements in the transition matrix, W, and in the covariance matrix Q are known. However, it is important to note that when dealing with the D-DFM, there is a further identification issue that affects the specification of the model and that we will consider latter in this paper.

4.2 Estimation

Estimation of the parameters of the DFM depends crucially on whether the model is static or dynamic and on the specification of the idiosyncratic components. We describe first the ML estimator and then the two-step estimators based on LS.

4.2.1 Maximum Likelihood

Consider first, the S-DFM in (32) with serial and contemporaneously uncorrelated idiosyncratic errors. In this case, estimation of the parameters can be carried out, after assuming normality, by ML. Given that, in this case, Σ_ε is diagonal, the innovation covariance matrix, Σ_t can be easily inverted using (10) and the Kalman filter (KF) can be used to compute the innovation decomposition form of the Gaussian likelihood in (23), which can be maximized using numerical optimization algorithms; see, for example, Engle and Watson (1981) and Aruoba, Diebold and Scotti (2009). As explained above, instead of numerical maximization of the likelihood obtained using the recursions of the Kalman filter, one can use the matrix form proposed by Delle Monache and Petrella (2019) who give details about how to implement it to the S-DFM with serially

[15] Note that, in this type of identification restrictions, the variable ordering could matters for parameter estimation given that the leading series determine the factors; see Lopes and West (2004) and Chan, Leon-Gonzalez and Strachan (2018) in the context of a factor model in which the factors have not dynamic dependence.

uncorreted errors and show results for N up to 200. Alternatively, given that finding numerically the maximum of the log-likelihood can be unfeasible when N is large due to the very large number of parameters to be estimated, the Gaussian log-likelihood can be maximized using the expectation maximization (EM) algorithm proposed by Shumway and Stoffer (1982) and Watson and Engle (1983) for ML estimation in SSMs.[16] The EM algorithm is based on the following decomposition of the log-likelihood conditional on the initial conditions for the factors

$$\begin{aligned} logL = -\frac{T}{2}log|\Sigma_u| - \frac{1}{2}\sum_{t=1}^{T}\left(F_t - \Phi F_{t-1}\right)'\Sigma_u^{-1}\left(F_t - \Phi F_{t-1}\right) \\ -\frac{T}{2}log|R| - \frac{1}{2}\sum_{t=1}^{T}\left(Y_t - \Lambda F_t\right)'\Sigma_\varepsilon^{-1}\left(Y_t - \Lambda F_t\right). \end{aligned}$$

The expected value of the log-likelihood conditional on the observations $Y_1, ..., Y_T$ is given by

$$\begin{aligned} E\left(logL|Y_1, ..., Y_T\right) = -\frac{T}{2}log|\Sigma_u| - \frac{1}{2}\sum_{t=1}^{T}tr\left\{E\left[\left(F_t - \Phi F_{t-1}\right)\left(F_t - \Phi F_{t-1}\right)'|Y_1, ..., Y_T\right]\Sigma_u^{-1}\right\} \\ -\frac{T}{2}log|R| - \frac{1}{2}\sum_{t=1}^{T}tr\{E\left[\left(Y_t - \Lambda F_t\right)\left(Y_t - \Lambda F_t\right)'|Y_1, ..., Y_T\right]\Sigma_\varepsilon^{-1}\}, \end{aligned}$$

where

$$E\left[\left(Y_t - \Lambda F_t\right)\left(Y_t - \Lambda F_t\right)'|Y_1, ..., Y_T\right] = Y_tY_t' + \Lambda P_{t|T}\Lambda' + \Lambda f_{t|T}f_{t|T}'\Lambda' - 2Y_tf_{t|T}'\Lambda', \tag{42}$$

and

$$\begin{aligned} E\left[\left(F_t - \Phi F_{t-1}\right)\left(F_t - \Phi F_{t-1}\right)'|Y_1, ..., Y_T\right] = P_{t|T} + f_{t|T}f_{t|T}' + \Phi P_{t-1|T}\Phi' + \\ \Phi f_{t-1|T}f_{t-1|T}'\Phi' - 2\Phi\left[f_{t-1|T}f_{t|T}' + C_t\right], \end{aligned}$$

with $f_{t|T}$ being the smoothed estimate of F_t obtained using (19), $P_{t|T}$ being its MSE given in (20) and $C_t = E\left[\left(F_t - f_{t|T}\right)\left(F_{t-1} - f_{t-1|T}\right)'|Y_1, ..., Y_T\right]$ that can be obtained by the Kalman smoother if the state vector is augmented to include F_{t-1}.

The conditional expectation of the log-likelihood is maximized by

$$\hat{\Lambda} = \sum_{t=1}^{T}Y_tf_{t|T}'\left(\sum_{t=1}^{T}\left(f_{t|T}f_{t|T}' + P_{t|T}\right)\right)^{-1} \tag{43}$$

$$\hat{\Phi} = \sum_{t=1}^{T}\left(f_{t|T}f_{t-1|T}' + C_t\right)\left(\sum_{t=1}^{T}\left(f_{t-1|T}f_{t-1|T}' + P_{t-1|T}\right)\right)^{-1}. \tag{44}$$

[16] Comparing the properties of parameter estimates obtained based on maximizing the likelihood based on the EM algorithm and on the procedure proposed by Delle Monache and Petrella (2019) could be of interest for further research.

Note that the estimators in (43) and (44) can be substituted when necessary by restricted versions after imposing the adequate restrictions. Furthermore, if the VAR order $p > 1$, the estimator can be modified accordingly. Finally, the corresponding estimators of the covariance matrices are given by

$$\hat{\Sigma}_e = diag\left\{\frac{1}{T}\sum_{t=1}^{T}\hat{e}_t\hat{e}_t'\right\} \tag{45}$$

$$\hat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t' \tag{46}$$

where $\hat{e}_t = Y_t - \hat{\Lambda}F_t$ and $\hat{u}_t = F_t - \hat{\Phi}F_{t-1}$.

The EM algorithm works iteratively. Given starting values for the parameters, Ψ^0, the expectation (E) step consists in computing the smoothed estimates of the factors and their MSEs. Once these smoothed expectations are obtained, the maximization (M) step consists in estimating the parameters using (43) to (46). These steps are iterated until convergence. Note that the EM algorithm can also be adopted to obtain ML estimates based on the likelihood decomposition proposed by Jungbacker and Koopman (2015) in (24). For each EM step, the Kalman smoothing is based on the low-dimensional model; see Grassi et al. (2015) for an empirical implementation. The parameters of the exact S-DFM can be estimated by ML using the EM algorithm regardless of N; see, among many others, Stock and Watson (1989, 1991) and Coulombe et al. (this issue) with $N = 4$, Bańbura et al. (2013) with $N = 24$, Quah and Sargent (1993) with $N = 60$ and Proietti (2011) with $N = 148$. The EM algorithm is convenient because, in state space models, it is straightforward to compute the expected value of α_t conditional on Y_T and to maximize the log-likelihood using standard regression formulae. However, the EM algorithm has also some disadvantages. First, the matrix of second partial derivatives is not available and, consequently, standard errors of the parameter estimates cannot be obtained directly. However, these partial derivatives can be approximated by perturbing the likelihood function in the neighborhood of the maximum. Computation of the information matrix via recursions is also possible as in Harvey (1989) or Cavanaugh and Shumway (1996). Versions of the information matrix, obtained from outputs arising naturally in the EM algorithm, such as in Meng and Rubin (1991) or Oakes (1999), are either hard to compute, as in the former, or will involve relatively untractable derivatives as in the latter.[17]

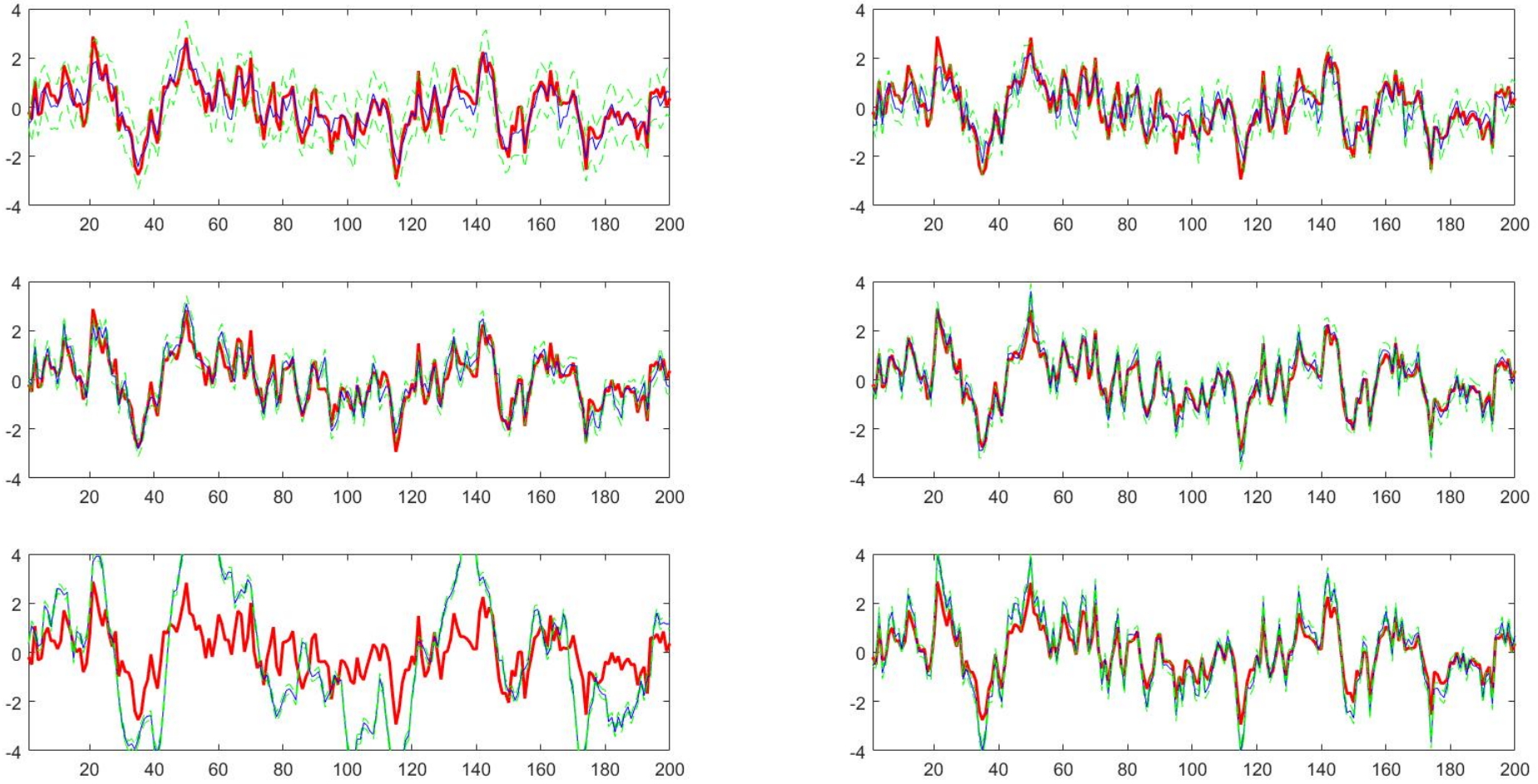
Finally, note that the EM estimators in (43) and (44) can be substituted when necessary by restricted versions after imposing the adequate restrictions. Furthermore, if the VAR order $p > 1$, the EM estimator can also be modified accordingly.

[17] A compromise that is easy to apply and will be robust toward distributional assumptions is the bootstrap, as derived in Stoffer and Wall (1991). However, it is not clear whether bootstrap can be implemented in the context of large cross-sectional dimension, N. In the context of non-stationary DFMs, Peña and Poncela (2006) estimate the model parameters using the EM procedure with a final pass using the scoring algorithm to obtain uncertainty measures of the estimated parameters.

If the idiosyncratic noises are weakly cross-correlated, ML estimation is still feasible if N is not very large; see, for example, Coulombe et al. (this issue). However, if the cross-sectional dimension is large, ML is not feasible due to the extremely large number of parameters. In any case, Doz, Giannone and Reichlin (2012) prove consistency, when both N and T diverge to infinity, of smoothed factors extracted when the DFM parameters are substituted by estimates obtained using the EM algorithm assuming wrongly that the idiosyncratic components have neither cross-sectional nor temporal correlations. This estimator is known in the related literature as Quasi-ML (QML). The $min\left(\sqrt{N},\sqrt{T}\right)$- consistency and asymptotic normality of the estimates of the loadings, factors and common components have been proved by Barigozzi and Luciani (2019a) who derive the conditions under which the asymptotic distribution can still be used for inference in case of miss-specification.[18] As an illustration of the performance of the factors extracted using KFS with estimated parameters instead of the true parameters in a DFM with cross-sectionaly correlated errors, we consider the same systems generated in the previous section with $N = 5$, 50 and 150 series and $T = 200$ temporal observations and implement the EM algorithm to estimate the parameters. Table 1, which reports the sample MSEs and coverages, shows that, when the true model is estimated with $N = 5$, the results are very similar to those obtained with known parameters. Although the point estimates of the estimated misspecified model are only slightly worse, the coverage in this case is very low (only a 67% when the nominal is 95%). As N increases, and given that T is fixed, the quality of the estimated parameters decreases and, also the quality of the point and interval estimates of the factors extracted with the estimated true model. For example, when $N = 50$, the sample MSE of the factor extracted with the estimated true model increases 353.3% with respect to that of the factor extracted using the true parameters. Furthermore, the coverage is 52.5% instead of 96%. When $N = 150$, it is not possible to estimate all the parameters in the model just using $T = 200$ observations. The MSE is huge while the coverage is 4.5%. The true factor is almost always outside the interval. Figure 4 plots the true factor together with the factor extracted when the parameters are estimated assuming that the true model specification is known (left column) together with the 95% intervals for the factor obtained when the MSE of the smoothed estimates is obtained using the estimated parameters. The first conclusion from Figure 4 is that, if one wants to estimate the full covariance matrix of the idiosyncratic errors, the sample size, T, should be large enough when compared with the cross-sectional dimension. When, $N = 150$ and $T = 200$, there is not enough information as to estimate the full covariance matrix of the idiosyncratic errors and, consequently, the extrated factor is far away from the true factor. However, when T is large enough, the factor extracted with estimated parameters, has a similar behaviour to the factor extracted with known parameters; compare with the first column in Figure 1 when $N = 5$ or 50.

[18] Barigozzi and Luciani (2019a) compare the loadings, factors and common components estimated using PC and QML estimators and conclude that, in static DFMs, both procedures are rather similar.

Figure 4: Simulated factor (red line) together with factor extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification (left column) and the true model assuming that the idiosyncratic component is serial and cross-sectional uncorrelated (right column), both with the parameters estimated by ML using the EM algorithm. The systems are simuated by a DFM with $N = 5$ (first row), $N = 50$ (second row) and $N = 150$ (third row) and serial and cross-sectionally correlated idiosyncratic components.



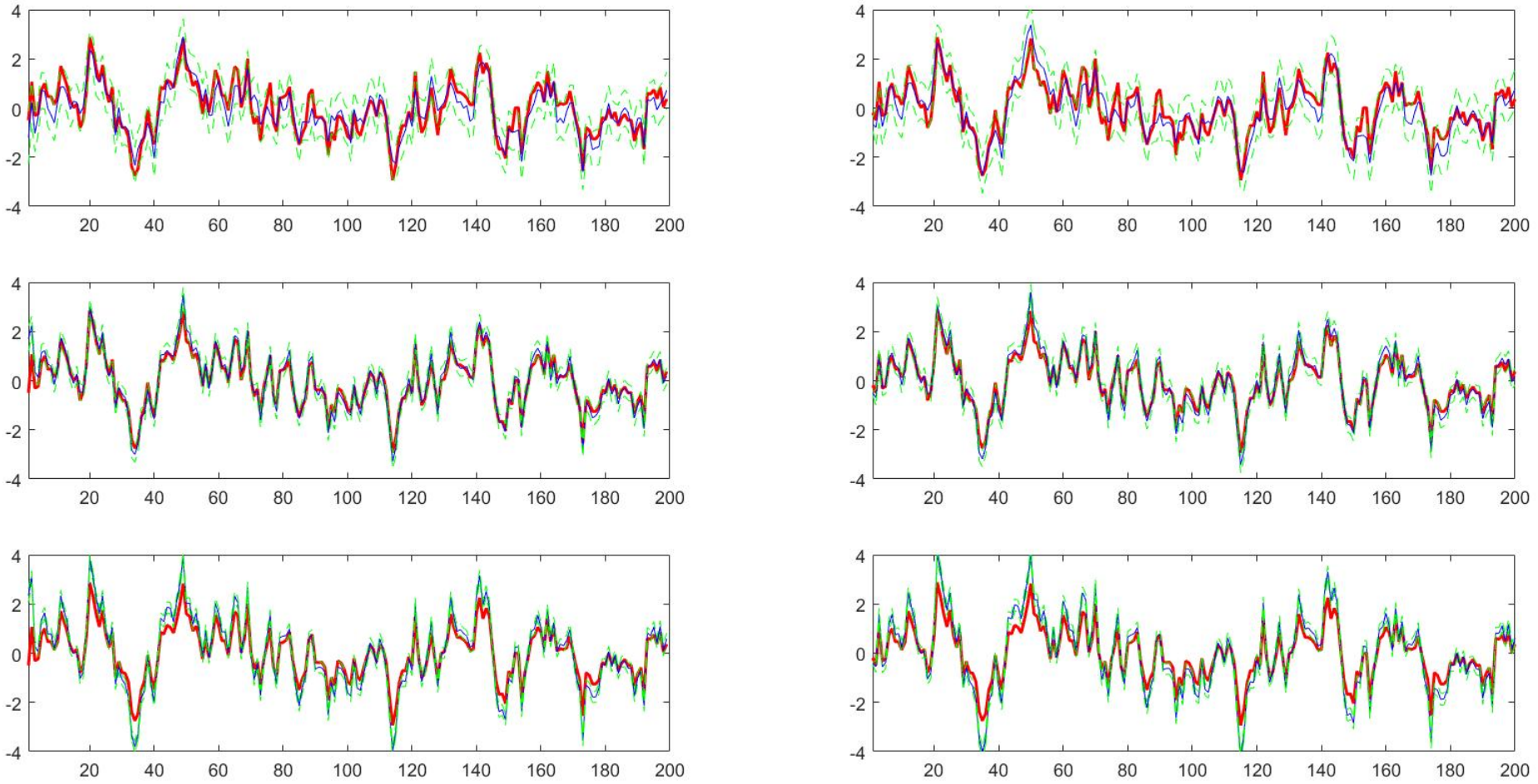
Regardless of N, the parameters of the exact S-DFM in equations (33) and (34), with serially correlated idiosyncratic errors specified as a diagonal VAR model, can still be estimated by ML using the EM algorithm modified with Cochrane-Orcutt iterations to estimate Θ conditional on Λ and Λ conditional on Θ; see Reis and Watson (2010) for an empirical implementation in a system with $N = 187$, Grassi et al. (2015) consider a system with $N = 170$ series and Bai and Li (2016) who conjecture what the limiting distribution of the parameters and factors should be.[19] Note that, in this case, the innovation covariance matrix can still be inverted easily using (10) due to the diagonality of Σ_e.[20] Table 1 reports the sample MSE of the factor extracted from the same simulated systems described above generated by the DFM with autocorrelated idiosyncratic components when the parameters are estimated. Regardless of whether the true or misspecified models are estimated, we can observe that, if $N = 5$, the sample MSEs (coverages) are only slightly larger (smaller) than those of the corresponding models with known parameters. However, the sample MSEs (coverages), which are similar regardless of whether the true or the misspecified models are estimated, increase (decrease) with respect to those of the corresponding models with known parameters. When $N = 150$, the sample MSEs are very large and the coverage of the confident intervals very low (remember that we are estimating the parameters with just $T = 200$ observations); see Figure 5 that illustrates the figures reported in Table 1.[21]

[19] The procedure implemented by Reis and Watson (2010) is only valid when no missing values occur in Y_t; see Jungbacker, Koopman and van der Wel (2011) for a detailed discussion and computationally feasible solution of the missing value problem.

[20] The parameters of the S-DFM with autocorrelated idiosyncratic noises written, as in footnote 9, by extending the state vector with the idiosyncratic noises have been estimated by Coroneo, Giannone and Modugno (2016) by EM adding a small noise to the measurement equation.

[21] We do not consider the estimation of the parameters of the DFM with cross-sectional and serially correlated idiosyncratic noises because it is infeasible due to the very large number of parameters to be estimated with just $T = 200$ observations.

Figure 5: Simulated factor (red line) together with factor extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification (left column) and the true model assuming that the idiosyncratic component is serial and cross-sectional uncorrelated (right column), both with the parameters estimated by ML using the EM algorithm, with the first based on the Cochrane-Orcutt specification. The systems are simuated by a DFM with $N = 5$ (first row), $N = 50$ (second row) and $N = 150$ (third row) and serial and cross-sectionally correlated idiosyncratic components.



Following Doz, Giannone and Reichlin (2012), extracting the factors using the KFS based on QML parameter estimates is very popular in empirical applications; see, for example, Bańbura, Giannone and Reichlin (2011) and den Reijer and Johansson (2019). This procedure has been extended by Bańbura and Modugno (2014) to deal with missing observations and idiosyncratic dynamics; see Scotti (2016) for an empirical application. Marcellino and Sivec (2016) extend the procedure proposed by Doz, Giannone and Reichlin (2012) to a mixed-frequency factor augmented VAR (MF-FAVAR) model.

Moving now to the D-DFM in equations (36) and (37), note that QML estimation based on the EM algorithm is still possible as far as the "*regressions*" in the maximization step are conveniently restricted; see, for example, Pinheiro, Rua and Dias (2013) and Jungbacker and Koopman (2015). Barigozzi and Luciani (2019a) extend the asymptotic results to the D-DFM written as a S-DFM in which the covariance matrix of the factor noises is singular.[22] Stock and Watson (2016) also suggest to estimate the D-DFM by imposing the restriction that Σ_u has rank q.

Finally, closely related to te QML estimator of Doz, Giannone and Reichlin (2012), Bai and Li (2016) propose a three-step estimator and derive the limiting distribution of the smoothed factors and estimate the dynamics of the idiosyncratic components. In the first step, Bai and Li (2016) propose the joint estimation of the loadings and idiosyncratic variances avoiding the specification of the factor dynamics by maximizing the following log-likelihood[23]

$$logL(Y;\Psi) = \frac{1}{2N} log|\Sigma_Y| - \frac{1}{2N} tr\left(M_Y \Sigma_Y^{-1}\right) \tag{47}$$

where $\Sigma_Y = \Lambda M_F \Lambda' + \Omega$ with $M_F = \frac{1}{T}\sum_{t=1}^{T}\left(F_t - \bar{F}\right)\left(F_t - \bar{F}\right)'$ and $\Omega = diag\left(E\left(\varepsilon\varepsilon'\right)\right)$, where ε is the $N \times T$ matrix given by $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)$, and $M_Y = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \bar{Y}\right)\left(Y_t - \bar{Y}\right)'$. Note that Σ_Y is an approximation of $E\left(M_Y\right)$ because Ω is restricted to be diagonal. The parameters to be estimated are $\Psi = (\Lambda, \Omega, M_Y)$. In the second step, the factors are estimated by GLS as follows

$$\hat{F} = Y'\hat{\Omega}^{-1}\hat{\Lambda}\left(\hat{\Lambda}'\hat{\Omega}^{-1}\hat{\Lambda}\right)^{-1} \tag{48}$$

and the parameters of the VAR(p) model are estimated based on $\hat{F}$. Finally, in the last step, the Kalman smoother is evaluated using the previous parameter estimates.

[22] Barigozzi and Luciani (2019a) show that, when the number of shocks is smaller than the number of factors, the QML estimates of the loadings are worse when estimated by PC than when estimated by QML while the factors are hardly affected.

[23] Bai and Li (2016) point out that, under fixed N, if cross-sectional heteroscedasticity exists but it is not allowed in the estimation, then the estimated factor loadings are inconsistent.

4.2.2 Least Squares estimation

Doz, Giannone and Reichlin (2011) prove the consistency, when both N and T diverge to infinity, of the smoothed factors extracted when the parameters of the stationary S-DFM are substituted by two-step estimates. In the first step, the r factors are extracted using PC. Denote the PC extracted factors by $\hat{f}_t^{PC}$ and by $\hat{\Lambda}^{PC}$ the corresponding estimated loadings. In the second step, assuming that the dynamic dependence of the factors is described by a VAR(1) process, the autoregressive parameters can be estimated by OLS as follows

$$\hat{\Lambda}^{PC} = \sum_{t=1}^{T} Y_t \hat{f}_t^{PC\prime} \left(\sum_{t=1}^{T} \hat{f}_t^{PC} \hat{f}_t^{PC\prime} \right)^{-1} \tag{49}$$

$$\hat{\Phi}^{PC} = \sum_{t=1}^{T} \hat{f}_t^{PC} \hat{f}_{t-1}^{PC\prime} \left(\sum_{t=1}^{T} \hat{f}_{t-1}^{PC} \hat{f}_{t-1}^{PC\prime} \right)^{-1} \tag{50}$$

The corresponding estimators of the covariance matrices are given by equations (45) and (46), where $\hat{e}_t = Y_t - \hat{\Lambda}^{PC} \hat{f}_t^{PC}$ and $\hat{u}_t = \hat{f}_t^{PC} - \hat{\Phi}^{PC} \hat{f}_{t-1}^{PC}$. Doz, Giannone and Reichlin (2011) show that the smoothed factors extracted using the PC estimates of the parameters are consistent even if the idiosyncratic component is wrongly assumed to be temporal and cross-sectionally uncorrelated due to the misspecification error vanishing as N and T diverge to infinity; see Giannone, Reichlin and Sala (2005), Giannone, Reichlin and Small (2008), Angelini et al. (2011) and Bańbura and Rünstler (2011) for implementations of this estimator. Koop and Korobilis (2014) extend the procedure proposed by Doz, Giannone and Reichlin (2011) to deal with time-varying parameters and stochastic volatility.

It is important to point out that when estimating using this two-step procedure the restrictions imposed to identify the factors in PC and those imposed in the SSM representation of the DFM should be the same. For example, very recently, Solberger and Spanger (2020) implement in E-Views the KFS factor extraction based on the two-step estimator of the parameters of the S-DFM. Although Solberger and Spanger (2020) say that they can estimate the D-DFM, this is not the case, as they are not taking into account the specification in the static representation of the D-DFM in (36) and (37); see, for example, Bai and Ng (2007) for the relation between the static factors extracted using PC and the dynamic factors.

Finally, note that the two-step estimator of the parameters has often been used to obtain starting values of the parameters in the EM algorithm; see, for example, Proietti (2011) and Doz, Giannone and Reichlin (2012). Joseph, Kalamara, Potjagailo and Kapetanios (this issue) also use the PC factors as starting values in the context of the estimator proposed by Kapetanios and Marcellino (2009).

5 Model specificacion

For KFS factor extraction to be efficient is crucial that the SSM is correctly specified. In this section, we describe some selected procedures used in the literature to specify the model, in particular, to determine the number of static and dynamic factors and their lags. We also describe procedures to decide about the dependence structure of factors and idiosyncratic noises. In many works implmenting KFS to extract latent factors in the context of DFMs, the specification of the factors is *ad hoc*. For example, in many works the factors and idiosyncratic components are assumed to follow a VAR(p_1) and VAR(p_2) model respectively, with p_1 and p_2 being fixed *a priori*; see, for example, Aruoba, Diebold and Scotti (2009) and Proietti (2011) who consider DFMs with $p_1 = 1$ and $p_2 = 0$ or Frale et al. (2011) who are interested in just one common factor, i.e. $r = 1$. Alternatively, p_1 can be determined by using information criteria as, for example, in Joseph, Kalamara, Potjagailo and Kapetanios (this issue). In other works, the number of factors is determined *a priori* without using any specific criteria; see, for example, Aruoba, Diebold and Scotti (2009) and Camacho and Perez-Quiros (2010) who, in the context of small-scale DFMs, fix $r = 1$ or Chauvet and Senyuz (2016) who fix $r = 2$. Koopman, Malle and Van der Wel (2010) and Delle Monache, Petrella and Venditti (2016) also chose an ad hoc number of factors. Joseph, Kalamara, Potjagailo and Kapetanios (this issue) determine the number of factors as that equal to the lowest number of factors which explains 50% of the variance in the data. Also, in many works the specification of the model is choosen by minimizing a Root Mean Square Error (RMSE) criterion; see, for example, Rünstler et al. (2009) and Hindrayanto, Koopman and de Winter (2016).

Alternatively, many authors implement methods for determining the number of factors designed in the context of Principal Components (PC) methods. For example, in the context of S-DFMs, the number of static factors can be selected by using one of the several procedures proposed in the literature; see Breitung and Eickmeier (2006), Stock and Watson (2011) and Barhoumi *et al.* (2013), among others for more detailed and complete reviews. The most widely used methods to determine the number of factors in S-DFMs are due to Bai and Ng (2002); see, for example, Proietti (2011). The Bai and Ng (2002) criteria are based on modifications of the Akaike (AIC) and Bayesian (BIC) information criteria, taking into account the cross-sectional and temporal dimensions of the dataset as arguments of the function penalizing overparametrization. However, these criteria usually detect too many factors being quite sensitive to the choice of $r_{\max}$, the maximum number of factors; see, for example, the Monte Carlo results in Ahn and Horestein (2013) and the arguments by Hyndrayanto, Koopman and de Winter (2016). Alternatively, Alessi, Barigozzi and Capasso (2010) refine the AIC and BIC criteria proposed by Bai and Ng (2002) by multiplying the penalty function by a constant that tunes the penalizing power of the function itself. Furthermore, Alessi, Barigozzi and Capasso (2010) suggest estimating the

number of factors using different subsamples. These criteria are linked to the eigenvalues of the sample covariance matrix of the variables in the system. In particular, the number of factors is selected as the number of eigenvalues larger than a threshold specified by a penalty function.

Another useful procedure to select the number of static factors is due to Onatski (2010) who proposes to select r using the difference between eigenvalues of $\frac{1}{T}Y'Y$ and proves that this selection is consistent. Note that, in spite of the lack of consistency of the eigenvalues when both N and T go together to infinity, pointed out by Lam and Yao (2012) through a simulation exercise, the empirical evidence and simulations results obtained by Corona, Poncela and Ruiz (2017) indicate that the difference between adjacent eigenvalues is a good estimator of the number of common factors. The procedure works even when the proportion of the variance attributed to the factors is small relative to the variance due to the idiosyncratic noises or when these are substantially correlated. Under the assumption of Normality, both cross-sectional and temporal dependence in the idiosyncratic noise are allowed. In a way, this procedure formalizes the eyeball decisions taken when examining the scree plot of the eigenvalues as proposed by Cattell (1966). The intuition behind the method is as follows: the line connecting two zero adjacent eigenvalues has zero slope, while that connecting eigenvalues linked to common factors should have a slope different from zero. This procedure is based on determining a sharp threshold that consistently separates the bounded and diverging eigenvalues of the sample covariance matrix of the observed series.

Finally, Lam and Yao (2012) propose a procedure to determine the number of factors based on separating "strong" and "weak" factors that require the idiosyncratic components to be white noise. Following ideas in Peña and Poncela (2006), they propose to determine the number of factors by the number of nonzero eigenvalues of the following matrix

$$S_Y = \sum_{k=1}^{K} C_Y(k) C_Y'(k), \tag{51}$$

where $C_Y(k) = Cov(Y_t, Y_{t-k})$ and K should be small as the autocorrelation is often at its strongest at the small lags.

Finally, Choi and Jeong (2019) propose some further new criteria to determine r. They also carry out very detailed Monte Carlo experiments to compare the performance of the more popular criteria available in the literature. They show that it is difficult to conclude which criterion performs best and advice that, in empirical applications, one should consider several criteria at the same time. In the same line, Stock and Watson (2011) point out that, when dealing with empirical systems, different methods frequently determine a different number of static factors with limited research comparing the performance of the different methods. Consequently, they suggest to augment the statistical estimators of r with inspection of the scree plots and with judgement informed by the application at hand; see, for example, Hyndrayanto, Koopman and de Winter

(2016) and Schiavoni, Palm, Smeekes and van der Brakel (2019) for empirical implementations.

In the context of D-DFMs, the first issue faced when trying to identify the number of dynamic factors is related with the simultaneous identification of the lag order of the VAR model for the factors, p, and the number of lags, s. To illustrate this problem, consider the following D-DFM in which the common component is given in equation (30)

$$Y_t = \Lambda(L)G_t + \varepsilon_t \tag{52}$$

$$G_t = \Phi_1 G_{t-1} + ... + \Phi_p G_{t-p} + v_t. \tag{53}$$

The D-DFM can be written as follows

$$Y_t = \Pi(L)v_t + \varepsilon_t \tag{54}$$

where $\Pi(L) = \Lambda(L)\Phi(L)^{-1}$ with $\Phi(L) = I_r - \Phi_1 L - \Phi_2 L^2 - ... - \Phi_p L^p$. The model in equation (54) is known as Generalized DFM (G-DFM) and the parameters of the infinite lag polynomial $\Pi(L)$ matrix cannot be estimated by ML or QML. Several authors propose estimating them using Dynamic Principal Components based on frequency-domain procedures; see Forni, Hallin, Lippi and Reichlin (2004, 2005). Recently, Peña, Smucler and Yohai (2019) propose estimating the G-DFM by linear generalized PCs in the time domain.[24] The identification problem appears because from $\Pi(L)$ it is not possible to recover the polynomials $\Lambda(L)$ and $A(L)$ in a unique form without imposing restrictions; see, for example, the discussion in Lütkepohl (2005). In order to identify the model, Forni, Hallin, Lippi and Reichlin (2000) consider restrictions on the eigenvalues of the spectral density matrix and Hallin and Liska (2007) and Onastki (2009) propose procedures to determine q in the context of the generalized DFM.

Alternatively, Stock and Watson (2005) propose to determine the number of common dynamic factors, q, by considering the static representation. Consider, for example, the D-DFM in equations (36) and (37) that can be written as follows

$$Y_t = \Lambda F_t + \varepsilon_t \tag{55}$$

$$F_t = \Phi F_{t-1} + B v_t. \tag{56}$$

where $\Lambda = [\Lambda_0 \Lambda_1]$ and $\Phi = \begin{pmatrix} \Phi_1 & 0 \\ I_q & 0 \end{pmatrix}$ and $B = \begin{pmatrix} I_q \\ 0 \end{pmatrix}$. After substituting (56) into (55), define

[24] Stock and Watson (2005) point out that Dynamic Principal Components produces two-sided estimates of the factors and, consequently, these estimates are not suitable for forecasting. Rünstler et al. (2009) compare the forecast performance of G-DFMs with those of alternative DFMs and find evidence in favour of the G-DFM with the gains in forecast precision against the alternative factor models being rather small and statistically insignificant. D'Agostino and Giannone (2012) also compare static and dynamic PCs by forecasting inflation and conclude that both are similar. Recently, Peña, Smucler and Yohai (2019) propose one-sided dynamic PCs that is appropriate for forecasting.

$X_t = Y_t - \Lambda\Phi F_{t-1}$ which can be writen as follows

$$X_t = \Pi v_t + \varepsilon_t, \tag{57}$$

where $\Pi = \Lambda B$. In equation (57), X_t is represented as a factor model with q serially uncorrelated factors that correspond to the shocks, v_t. Were X_t observed data, Stock and Watson (2005) argue that q could be consistently determined using the Bai and Ng (2002) information criteria. However, given that this is unfeasible, they propose applying Bai and Ng (2002) to $\hat{X} = Y_t - \hat{\Xi}\hat{F}_{t-1}$, where $\hat{\Xi}$ is an estimator of $\Lambda\Phi$ and $\hat{F}_{t-1}$ is an estimator of F_{t-1}. Stock and Watson (2005) propose estimating $\hat{\Xi}$ and $\hat{F}_{t-1}$ by estimating first Λ and F_t by PC and components and then Φ by the OLS estimator of the regression of $\hat{F}_t$ onto $\left(\hat{F}_{t-1}, ..., \hat{F}_{t-p}\right)$, where p can be choosen using the BIC criteria.[25] Amengual and Watson (2007) prove the consistency of this procedure. Alternatively, Bai and Ng (2007) propose determining q by computing the eigenvalues of the covariance matrix of the residuals of the regression of $\hat{F}_t$ onto $\left(\hat{F}_{t-1}, ..., \hat{F}_{t-p}\right)$ and testing whether these eigenvalues satisfy an asymptotically shrinking bound that reflects sampling error. Note that the results of the test may depend on this shrinking bound.[26] Breitung and Pigorsch (2013) propose determining the number of dynamic factors, q, using canonical correlation analysis of the current and past values of the common factors in the static representation instead of PCs. The procedure proposed by Breitung and Pigorsch (2013) also depend on tunning parameters. Zhao, Cui and Wang (2017) carry out a Monte Carlo comparison of alternative procedures to determine the number of dynamic factors and conclude that the procedures proposed by Bai and Ng (2007) and Hallin and Liska (2007) perform better.

These procedures are very popular. For example, Angelini et al. (2011) determine r using Bai and Ng (2002), p using the SIC criteria and q using Bai and Ng (2007).[27] However, it is important to recall that the procedures for determining q have been designed in the context of PCs factor extraction and that the assumptions behind PC could not be satisfied when the D-DFM is writen as a S-DFM. Furthermore, when the data follow a D-DFM and we fit a S-DFM, the number of factors increase from q to $r = q(s+1)$ but these r factors are not mutually independent at all lags. Consequently, it could be expected that the number of "static" factors determined by the procedures designed for truly S-DFMs is going to be such that $r < q(s+1)$; see the arguments in Peña and Tsay (2020).

The discussion about the determination of r and q shows that, in practice, this is a difficult problem.[28] Furthermore, in the case of the D-DFM, even if we were able to chose r and q, we

[25] Alternatively, Stock and Watson (2005) propose estimating directly $\hat{X}i$ using the OLS estimator from the regression of Y_t onto $\left(\hat{F}_{t-1}, ..., \hat{F}_{t-p}\right)$.

[26] The procedure proposed by Bai and Ng (2007) is a useful cross-check of the more informal procedure proposed by Giannone, Reichlin and Sala (2005).

[27] In an application to euro area data, they chose $r = 5$, $q = 3$ and $p = 1$.

[28] It is important to point out that, in an empirical exercice, D'Agostino and Giannone (2012) conclude that

do not have enough information as to recover the specification in equations (36) and (37). The procedures to determine the number of factors, r, when the D-DFM is written as a "static" DFM, described above, are based on a different specification, in which the autoregressive matrices of the factors are not restricted. Therefore, when estimating the parameters and running the KFS algorithms, the estimated model is not exactly the true specification.

The difficulty in choosing the correct specification of the DFM that is then used for factor extraction (after estimating the unknown parameters) calls for adequate procedures for testing for the specification of the model. Harvey and Koopman (1992) propose some specification tests based on auxiliary residuals. However, these tests were implemented to univariate models. Extensions to multivariate models as those of interest in this survey are needed in the literature. Recently, Fiorentini and Sentana (2019) derive score tests of misspecification in DFM based on frequency domain techniques.

6 Forecasting using KFS in stationary DFMs: Empirical applications

The number of empirical applications of stationary DFMs using KFS for factor extraction is very extense. In this subsection, we just survey some empirical applications as an example to show the large range of possibilities of these models and procedures. We apologize for those importnat works that have not been refered in this subsection.

One of the first implementations of KFS algorithms to extract factors in the context of DFMs was Engle and Watson (1981) who estimate the unobserved metropolitan wage rate for Los Angeles based on observations of sectoral wages within the Standard Metropolitan Statistical Area. Afterwards, in the last 40 years, the number of applications of KFS algorithms in DFMs has been increasing over time. In this section, we aim to describe just some of the main contributions.

Among the many applications of forecasting using KFS procedures within the context of factor extraction in DFMs, one of the most active areas is real-time macro-monitoring; see Stock and Watson (2016) for a description of empirical applications. The first relevant contribution in this area is due to Stock and Watson (1989, 1991) who construct the experimental coincident economic indicator (CEI), which was released monthly through the National Bureau of Economic Research from May 1989 to December 2003. The CEI was the Kalman filter estimate of the common factor among the big monthly indicators often used in dating the business cycle. The DFM was estimated by ML using a SSM; see Carriero and Marcellino (2007) who compare several parametric and non-parametric procedures to extract one-single common factor in the context of estimating composite coincident and leading indexes for the UK. Since then, several further

there is no evident improvement in the forecast accuracy when allowing for the number of dynamic factors to be smaller than the number of static factors.

applications have appeared in the literature. Aruoba, Diebold and Scotti (2009) track the high-frequency evolution of real activity by constructing an index based on indicators measured at different frequencies, even vey high (daily) data. This index, updated daily by the Federal Reseve Bank of Philadelphia, has been added to Bloomberg's real-time data that can be followed on its platform. Based on the index proposed by Aruoba, Diebold and Scotti (2009), Scotti (2016) constructs uncertainty indexes based on smoothed weights estimated as proposed by Koopman and Harvey (2003). Camacho and Perez-Quiros (2010) evaluate the short term forecasts from a factor model in a truly real-time set up for the euro area. They deal with ragged edges using the proposal by Giannone, Reichlin and Small (2008), with mixed frequencies using the filter proposed by Mariano and Murasawa (2003) and with data revisions using Evans (2005). Aruoba and Diebold (2010) is an extension of the DFM suggested by Stock and Watson (1991). Grassi et al. (2015) construct monthly indicators of economic activity for the euro area and its largest member countries.

Closely related with the construction of business cycle indexes is the problem of nowcasting which is important since forecasting improvements of GDP with respect to naive models is mainly limited to the current quarter; see Bańbura, Giannone and Reichlin (2011) for a survey on nowcasting. Factors extracted using KFS algorithms can be used to forecast in the context of the diffusion indexes of Stock and Watson (2002). Bańbura and Rünstler (2011) propose obtaining prediction weights of the individual time series in the system by an extension of the Kalman filter using the estimation procedure of Doz, Giannone and Reichlin (2011). Rünstler (2016) proposes using these prediction weights to refine the data set by eliminating uninformative series when forecasting using DFMs. KFS has been implemented to exploit large information to bridge monthly and quarterly variables with different publication lags, by combining predictors in few common factors which are then used as regressors in bridge equations via the Kalman filter. The procedure, originally proposed by Giannone et al. (2008), was first applied on US data at the Board of Governors of the Federal Reserve and is also regularly implemented at the ECB; see Rünstler et al. (2009) for a comparison of forecasts of GDP obtained in ten countries of the euro area, Matheson (2010), Angelini, Camba-Mendez, Giannone, Reichlin and Rünstler (2011) to obtain early estimates of current quarter GDP in the euro area, D'Agostino, McQuinn and O'Brien (2012) for Irish quarterly GDP and Aastveit and Trovit (2012) who investigate the properties of financial assets to forecast growth in Norway. Hindrayanto, Koopman and de Winter (2016) compare the performance of four different estimators of the factors in a pseudo real-time competition to forecasts GDP growth in the euro area and five of its largest countries. In particular, they compare forecasts when the factors are estimated using PC and the two procedures of Doz, Giannone and Reichlin (2011), Doz, Giannone and Reichlin (2012) and Bräuning and Koopman (2014). These last authors propose adopting a low-dimensional SSM which deals with the target variable and the factors jointly. More recently, Jansen and de Winter (2018) ex-

plore the effects of combining model-based nowcasting GDP forecasts and quarterly judgmental consensus forecasts for the G7 countries. It is important to note that, due to the difficulty in chosing the correct specification of the model, several authors propose to pool forecasts obtained from different models with different number of factors; see, for example, Kuzin, Marcellino and Schumacher (2013), Hindrayanto, Koopman and de Winter (2016) and Jansen, Jiu and de Winter (2016). Camacho and Martinez-Martin (2014) is an application in which a stationary small-scale DFM is implemented to forecast US GDP with a system of variables that include survey data and financial indicators.

Apart from monitoring and forecasting real economic activity, central banks, researchers and analysts have widely used DFMs to track and forecast the evolution of inflation or to detect comovements in prices, especially in goods or commodities. Just to cite a few applications, Arouba and Diebold (2010) use small scale DFMs to monitor both real economic activity and inflation taking also into account their interactions. Reis and Watson (2010) use quarterly inflation for $N = 187$ sectors in US personal consumption expenditures to extract what they call pure inflation and separate it from other components of inflation. Inflation in commodities have been analysed in Poncela, Senra and Sierra (2020) and Delle Chiaie, Ferrara and Giannone (2017).

Giannone, Reichlin and Small (2008) and Bańbura and Modugno (2014) use state space models for nowcasting, the problem of predicting the present, the very near future and the very recent past, which is related with both missing observations and mixing frequencies. Marcellino and Schumacher (2010) and Foroni and Marcellino (2014) compare different methods to deal with mixed frequencies when forecasting German GDP and nowcasting Euro area macroeconomic aggregates, respectively.

Another important application of factor extraction in forecasting is the Factor Augmented VAR (FAVAR) model. FAVAR models have enjoyed increasing popularity in forecasting macroeconomic variables. Koop and Korobilis (2014) propose a FAVAR model with time-varying parameters (TVP-FAVAR) to construct a financial conditions index to track expectations about growth in key US macroeconomic variables. Marcellino and Sivec (2016) implement the a MF-FAVAR model to evaluate the effects of monetary, oil and fiscal shocks. Very recently, Joseph, Kalamara, Potjagailo and Kapetanios (this issue) implement FAVAR models to forecast UK inflation comparing several alternative procedures.

KFS procedures have also been implemented when forecasting in the context of financial variables. One of the most popular applications is to forecast the yield curve. Diebold and Li (2006) and Diebold, Rudebusch and Aruoba (2006) describe the yield curve by framing the Nelson-Siegel model into a SSM with three unobserved factors that represent the level, slope and curvature and that are modeled as VAR processes with the loadings depending on a single parameter. Koopman, Mallee and Van der Wel (2010) forecast interest rates over different

maturities based on this model extended by allowing the loading parameter to vary over time and the variances to follow GARCH models.[29] Jungbacker, Koopman and Van der Wel (2014) impose smoothness restrictions on factor loadings by using a cubic spline function that depends on time to maturity and Koopman and Van der Wel (2013) include macroeconomic factors in the dynamic yield curve model and conclude that macroeconomic variables can lead to more accurate yield curve forecasts.[30] In a different context, Coroneo, Giannone and Modugno (2016) implement the methodology of Doz, Giannone and Reichlin (2012) to forecast monthly excess bond returns based on out-of-sample forecasts of the underlying yield and macro-economic factors. Another strand of the literature focus on extracting information from the components of the yield curve to forecast business cycle turning points; see Chauvet and Senyaz (2016) who fit a DFM with $r = 2$ with this purpose.

Mariano and Murasawa (2010), Camacho and Perez-Quiros (2010), Camacho, dal Bianco and Martinez-Martin (2015a, 2015b) and Jungbacker et al. (2011) also implement state space models to deal with missing data. Very recently, Schiavoni, Palm, Smeekes and van der Brakel (2019) incorporate Google trends information into the estimation of unobserved components to improve unemployment official statistics using KFS with mixed frequencies. Martinez-Martin and Rusticelli (in press) also implement the methodology of Mariano and Murasawa (2010).

Proietti (2011) consider a large-scale ($N = 149$, $T = 150$) static stationary exact DFM to extract factors for the euro-area in wich the series in the system are observed monthly and quarterly dealing with the aggregation problem.

Finally, there are also a large number of applications to non-economic data. Just to name one included in this special issue, Coulombe et al. (this issue) extract the common factor from a system of four sea ice measures to obtain an optimal composite measure.

7 Non-stationary DFMs

So far, DFMs have been assumed to be stationary. However, panel time series are often non-stationary due to the presence of unit roots. In this case, it is very popular to differentiate individually each series in a univariate fashion in the system until stationarity is reached and then fit the corresponding DFM to the differentiated system to extract the factors and, finally, obtain the recumulated factors if needed; see, for example, Mariano and Murosawa (2003, 2010) and Grassi et al. (2015), among many others. This solution might derive in loss of information, missclassification of factors in stationary/non-stationary and inefficient estimation of the param-

[29] Note that models with time-varying loadings could be considered as non-stationary. However, given that the factors are stationary, their treatment, from the point of view of factor extraction, is the same as in stationary models.

[30] Poncela (2013) note that this increase in accuracy may be the effect of the model specification instead of the effect of the added macroeconomic variables.

eters and factors; see, for instance, Bariggozzi and Luciani (2019b) and Corona, Poncela and Ruiz (2020). Alternatively, the DFM can be implemented directly to the original non-stationary system. This section surveys some relevant issues specific to non-stationary DFMs when the KFS algorithms are implemented to extract the factors. First, we consider the case of known model. Second, we consider that the DFM specification is known but the parameters need to be estimated. Finally, we will deal with specification issues. The last subsection reviews selected empirical applications of non-stationary DFMs in which KFS procedures are implemented to extract the factors.

7.1 Factor extraction in non-stationary DFMs with known parameters

If the model specification and the parameters are known, the KFS can be implemented to extract the factors in non-stationary DFMs without new issues on top of those already described for their stationary counterparts. There are only two considerations worth to mention. First, the role of the starting values for the state vector (factors) that should be different from those usually used in stationary models. Second, it is important to investigate the performance of the KFS extraction when the idiosyncratic noises are non-stationary but their temporal dependence is ignored when filtering. Consider first that $Y_t = (Y_{1t}, \cdots, Y_{Nt})$, $t = 1, \cdots, T$, is a non-stationary $I(1)$ vector of time series generated by the following DFM

$$Y_t = \Lambda F_t + \epsilon_t. \tag{58}$$

where F_t is an $r \times 1$ vector of $I(1)$ common factors and the idiosyncratic components are stationary.[31] Therefore, there are N non-stationary $I(1)$ time series generated by $r < N$ non-stationary $I(1)$ common trends. For simplicity, consider that the VAR model for the r common non-stationary factors is as in equation (30) with $r = r_1$, $p = 1$ and $\Phi_1 = I_r$. The error u_t is white noise with diagonal variance matrix Σ_u. We also assume that the idiosyncratic components are white noise with diagonal variance matrix R. The Kalman filter can be applied in order to extract the common factors using the initial conditions for the state vector suggested in Peña and Poncela (2006) with $f_0 = \Lambda^- Y_1$ being Λ^- a generalized inverse of Λ and with $P_0 = kI_r$ where the scalar k is such that $k \to \infty$. Additional possibilities of initializing the Kalman filter in the presence of non-stationary latent variables are suggested, for instance, in de Jong (1991) and Durbin and Koopman (2001). To illustrate the performance of the KFS in this case, we

[31] We assume that there are not deterministic components. In systems where there are, Bai (2004) suggests to eliminate all the deterministic components before applying nonstationary factor analysis to the series in levels. Furthermore, Banerjee et al. (2017) argue that the assumption of stationary idiosyncratic noises is in accordance with macroeconomic systems that exhibit a high rejection of the hypothesis of a unit root of the estimated idiosyncratic components; see, for example, Poncela and García-Ferrer (2014) and Schiavani, Palm, Smeekes and van der Brakel (2019) for empirical implementations with stationary idiosyncratic components. Finally, Bariggozzi and Luciani (2019b) also contemplate the possibility of lags of the common factors in the measurement equation. This can be dealt with as in the stationary case.

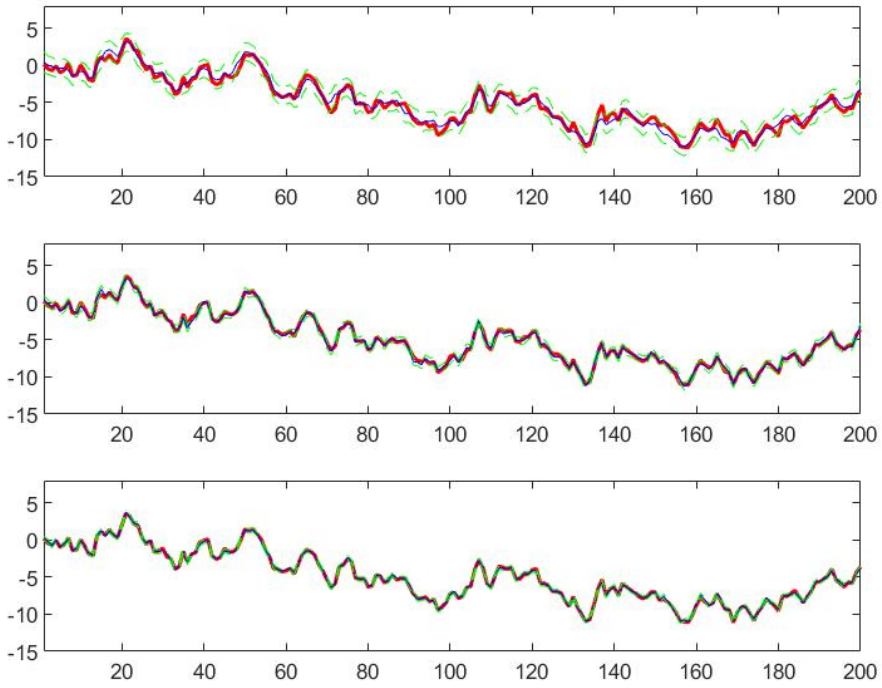


Figure 6: Simulated factor (red line) together with factor extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification and known parameters for a random walk common factor. The systems are simuated by a DFM with $N = 5$ (first row), $N = 50$ (second row) and $N = 150$ (third row).

simulate a system with $r = 1$ factor characterized by a random walk with zero initial condition, $F_0 = 0$, and $\sigma_u^2 = 1$. The idiosyncratic noises are serial and cross-sectionally uncorrelated and homoscedastic with variances equal to 1. As in the stationary case, we simulate systems with $T = 200$ observations and $N = 5$, 50 and 150 variables to replicate small, medium and large systems, respectively. The factor loadings are drawn from a uniform distribution $U(0, 1)$. Figure 6 plots the true simulated factor together with the factor extracted using the the KFS assuming all parameters known and using the initial conditions obtained as suggested by Peña and Poncela (2006). As for the stationary case, the differences between the extracted and population factors are only visually appreciable when N = 5.

Instead of considering a DFM with all factors being $I(1)$, it is possible to face situations in which some factors are non-stationary while others are stationary; see García-Ferrer and Poncela (2002), Peña and Poncela (2006) and Bariggozzi and Luciani (2017), among others. Assume that there are r_1 factors that are $I(1)$ and r_0 stationary factors with $r = r_1 + r_0$ being the total number of common factors.[32] The model is given by (28) to (30) with r_1 roots of the determinantal equation, $|\Phi(L)| = |I_r - \Phi_1 L - \cdots - \Phi_p L^p| = 0$, on the unit circle and r_0 outside it. Usually, the variance associated to the common non-stationary factors is higher than that

[32]Higher orders of integration have been considered, for instance, in Peña and Poncela (2006).

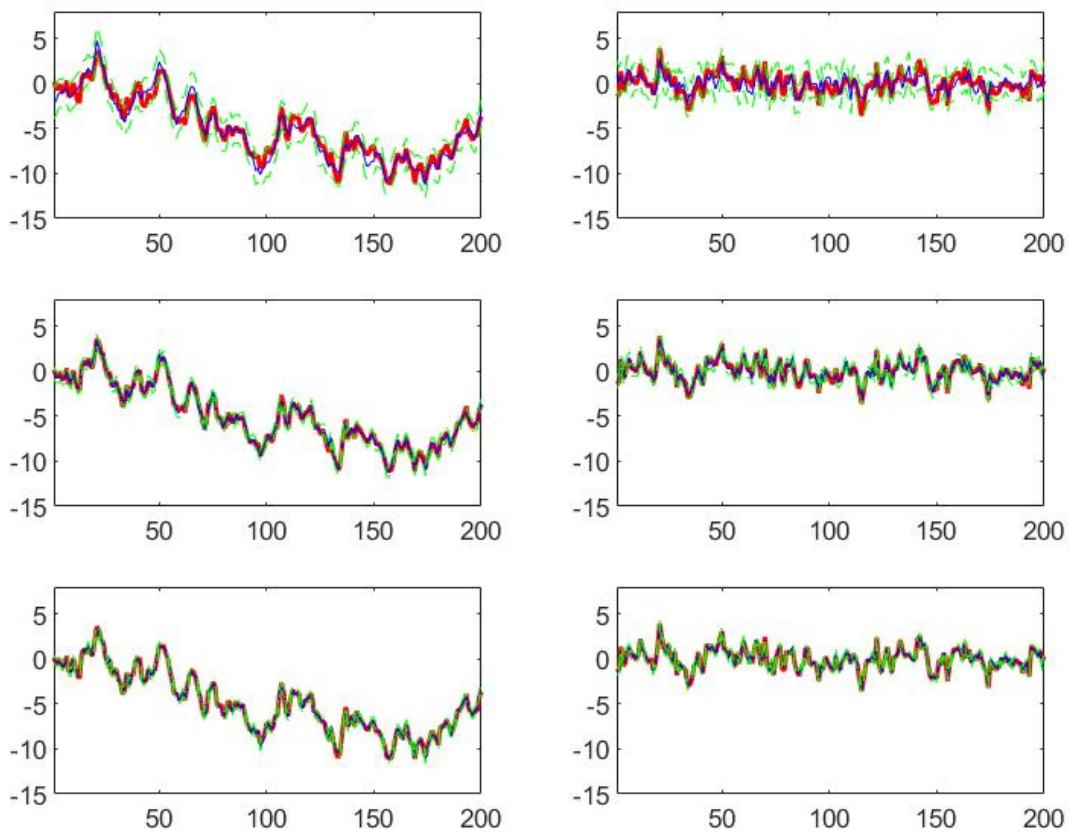


Figure 7: Simulated factors (red line) together with factors extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification and known parameters for a random walk (left column) and a stationary AR(1) (right column) common factor. The systems are simuated by a DFM with N = 5 (first row), N = 50 (second row) and N = 150 (third row).

linked to the stationary ones. In order to check how both types of common factors are extracted, we perform a simulation considering $r = 2$ with one factor generated by a random walk with $\sigma^2_{u_1} = 1$ and the other by a stationary AR(1) model with autoregressive parameter $\phi = 0.5$ and $\sigma^2_{u_2} = 1$. Therefore, in (29) $p = 1$ and $\Phi_1 = diag(1; 0.5)$. Figure 7 plots the true simulated factors together with their corresponding extracted factors and 95% confidence bounds. Notice that, as before, if the model is known, the differences between the extracted and population factors are only visually appreciable when $N = 5$ for both factors.

Finally, we consider other sources of non-stationarity in the system besides the $I(1)$ common factors, which may render spurious common factors if they are not properly taken into consideration. In particular, we consider the case of just $r = r_1$ non-stationary factors with some or all of the idiosyncratic components being also random walks. If all the idiosyncratic components were $I(1)$, the observed time series are not cointegrated, while if just $g_1 < N$ idiosyncratic components are $I(1)$, the observed time series are cointegrated. Notice that if all the idiosyncratic components are random walks, the observed time series can be univaritely differenced and the

analysis can be performed in first differences. For illustrative purposes consider now the model given in equations (28) through (31) where the VAR for the common non-stationary factors is as in (30) with $r = r_1$, $p = 1$ and $\Phi_1 = I_r$. The error u_t associated to the common factors is white noise with diagonal variance matrix Σ_u. As for the stationary case, from now on, we will assume that $p^* = p_i^* = 1$ in equation (31) in order to simplify the analysis where either $\theta_{1i} = 1$ when the idiosyncratic components are also nonstationary or $|\theta_{1i}| < 1$ if they are stationary. For simplicity, we assume that the idiosyncratic components are cross-sectionally uncorrelated. Barigozzi, Lippi and Luciani (2020) argue that when $\theta_{1i} < 1$, the dynamics of the $I(0)$ idiosyncratic components do not need to be specified in order to obtain consistent estimates of the common factors. For illustrative purposes, we simulate a system generated by only one common random walk ($r = r_1 = 1$) where the first $N/2$ idiosyncratic components are also random walks if N is even and the first $N/2 - 1$ if N is odd. The remaining idiosyncratic components are stationary with parameter $\theta_{1i} = 0.5$. The model can be written as in equations (33) and (34) with $\Phi_1 = 1$ and $\Theta = diag(\theta_{11}, \cdots, \theta_{1N})$. We extract the common factor assuming that we know the true model. The results are presented in Figure 8, left column. Only for $N = 5$, we can see by visual inspection the difference between the estimated and the true common factor.[33] This is illustrated in Figure 8, centered column, where we have replaced the true AR parameter of the idiosyncratic component whenever it was stationary (that is when $\theta_{1i} = 0.5$) by 0. As it can be seen differences between the true and estimated common factors are only visually appreciable for $N = 5$ as for the case of using the true model. Finally, we ignore all dynamics in the idiosyncratic component, stationary or not, and extract the common nonstationary factor using a further misspecified model considering that the specific components where all stationary white noise. As pointed out by Bai (2004) and Onatski and Wang (2020) for principal components and Barigozzi and Luciani (2019b) for the KFS methods, the extracted factor in this case is not consistent as it is reflected in the illustration presented in Figure 8, right comlumn, where we can see important differences between the true and estimated common factors for all cross-sectional dimensions $N = 5, 50$ and 150. Therefore, misspecification of the idiosyncratic components still renders consistent estimates of the common factor as long as the misspecification is only associated to the stationary part while misspecifications related to the nonstationary idiosyncratic components breaks the consistency of the extracted common factors. The adverse effects seem to be less important as the cross sectional dimension increases in the simulated example.

[33] Barigozzi and Luciani (2019b) argue that estimating a misspecified model ignoring the dynamics of the idiosyncratic components when they are stationary does not affect the consistency results for the estimated common components.

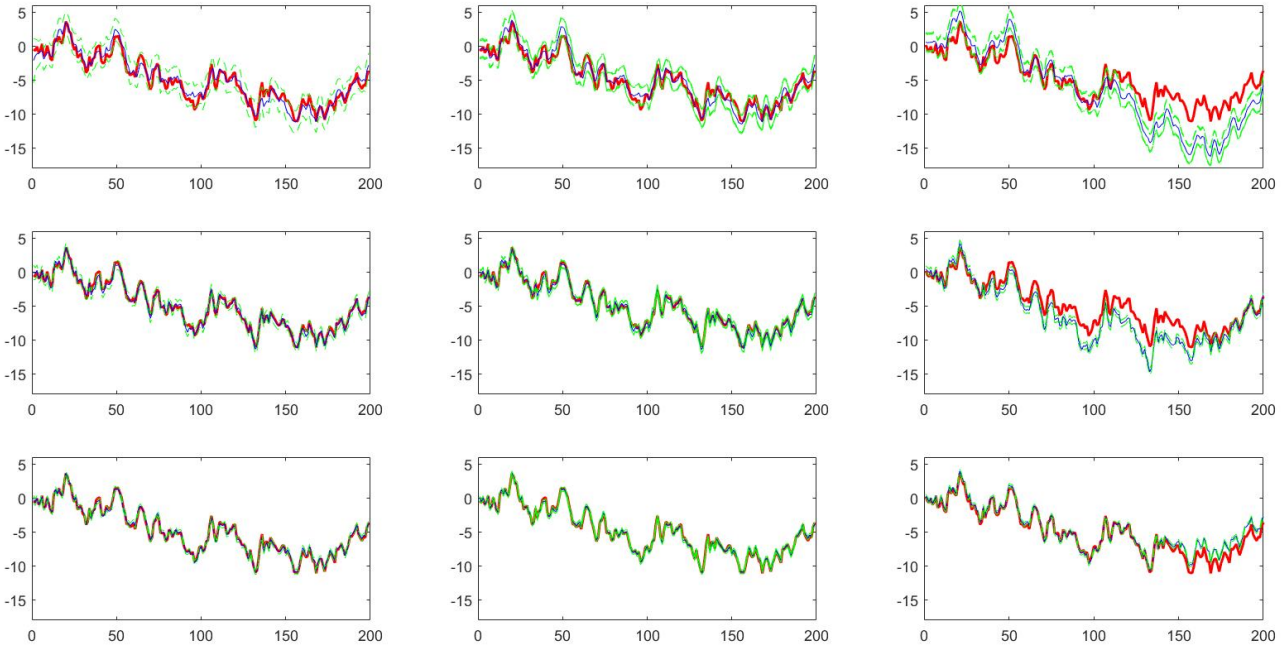


Figure 8: Simulated factors (red line) together with factors extracted by the Kalman smoother (blue lines) and 95% confidence bounds (green lines) obtained using the true specification and known parameters for a random walk and possibly nonstationary idiosyncratic errors when using the true model (left column), a misspecified model only for the stationary idiosyncratic components (center column) assuming that they are white noise and a misspecified model for the idiosyncratic components, stationary or nonstationary, assuming that they are white noise (right column). The systems are simuated by a DFM with N = 5 (first row), N = 50 (second row) and N = 150 (third row).

7.2 Factor extraction in non-stationary DFMs with unknown parameters

As for the estationary case, usually the parameters of the model are unknown and need to be estimated. However, there is an important difference between the identification restrictions in DFMs with non-stationary factors and those described above for the stationary case. If we are interested in extracting a particular set of nonstationary common factors, the identification restrictions need to ensure the superconsistency of the estimators of the factor loading matrix. Johansen and Tabor (2017) analyze an SSM with an unobserved multivariate random walk and check whether the extracted common trends and their estimators are cointegrated. They conclude that, although cointegration holds for the common components, $\chi_{i,t}$, $i = 1, .., N$, it is not always the case for the common trends. In the context of the ML estimator of the parameters of an autoregressive representation of the SSM model, they show that, after imposing the necessary identification restriction in order to separate the factor loading matrix from the common trends cointegration between the identified non-stationary factors and their estimator holds, if and only if the estimators of the factor loadings are consistent at a faster rate than $\sqrt{T}$.

Once the non-stationary factors are properly identified, the DFM parameters can be estimated as in the stationary case.[34] Most works dealing with KFS procedures for factor extraction in non-stationary DFMs obtain parameter estimates based on ML methods with the log-likelihood maximized either using Newton-type optimization algorithms as in Frale et al. (2011) or using the EM algorithm as in Zuur et al. (2003), Peña and Poncela (2004) and Seong et al. (2013) who consider systems with small cross-sectional dimension, in particular, $N = 12$ in the first work and $N = 5$ in the last two works. The EM algorithm is also able to handle large N systems; see, for example, Quah and Sargent (1993) and Barigozzi and Luciani (2017, 2019b). Barigozzi and Luciani (2019b) provide consistency results for the common components for the model estimated using the EM-algorithm and KFS when both N and T go infinity.

There are two important issues related with implementing the EM algorithm in non-stationary DFMs. First, it is important to point out that the estimation procedures described above, rely on the idiosyncratic noises being stationary. However, even if the factors were observed, if unit roots were allowed in the idiosyncratic components, it is no longer possible to "regress" the observed time series over the factors to estimate the factor loadings as in equation (43) of the EM algorithm. Alternatively, if the idiosyncratic component for the i-th series is non-stationary, Barigozzi and Lucinai (2019b) propose replacing $Y_{i,t}$ by $Y_{i,t} - \varepsilon_{i,t}$ in order to remove the sources of non-stationarity not coming from the common factors from $Y_{i,t}$.

The second issue relates to the parameters used to initialize the EM algorithm. Consider

[34] Obviously, the non-stationary nature of the data should be taken into account with the unit roots imposed before estimating the parameters. Recall that, in this subsection, we consider that the DFM specification is known.

first, the initial values for the loading matrix, Λ. Peña and Poncela (2006) assume that the non-stationarity is only present in the common factors and suggest as initial values of the columns of the factor loading matrix, Λ, the eigenvectors of the first-lag autocovariance or uncentered lagged second moments matrix of the observed series properly normalized by T^{2d} instead of T where d is the order of integrtion of the observed series. The idea is that, if there is no serial correlation in the idiosynratic noises or it is very weak, the eigenvectors of the lagged covariance matrices will be a better estimate of the factor loading matrix since they would not be contaminated by the zero lag correlation present in the noise. Alternatively, Barigozzi, Lippi and Luciano (2020b) and Barigozzi and Luciani (2019b) allow for $I(1)$ idiosyncratic noises and suggest to use the eigenvectors of the covariance matrix of the model in first differences in order to avoid spurious effects due to unit roots in the idiosyncratic noise.

As regards the initial values for the parameters of the state equation, needed for the ML estimator, consistent estimates of the common factors based on PC can be used to determine the dynamics of the common factors and give initial estimates for the transition matrix based on LS estimates of these initial estimates of the common factors. For instance, Barigozzi, Lippi and Luciani (2020b) use data on first differences to estimate the factor loading matrix but project the levels of the series over those estimates of the factor loading matrix to get pre-estimators of the common factors.

Alternatively, Johansen and Tabor (2017) consider a LS estimator based on using the cointegration implications of the non-stationary DFM.

7.3 Specification of non-stationary DFMs

As in the stationary case, we need to specify the model and write it in state space form prior to estimation of the parameters and factor extraction. The main issue in non-stationary DFMs is to determine the number of non-stationary and stationary common factors; see Barhoumi *et al.* (2013) who review procedures for the determination of the number of factors in non-stationary systems. It is important to note that, in empirical applications, it is often the case that the variance associated to the non-stationary common factors is much larger than that linked to the stationary factors and, consequently, the later ones may be difficult to detect; see Peña and Poncela (2006). In practice, if the number of stationary and non-stationary factors cannot be determined simultaneously, one should first determine the number of non-stationary factors and, second, determine the number of stationary common factors once the number of non-stationary common factors is determined. In order to determine the number of factors, it is important to consider three possible scenarios regarding the idiosyncratic components: (i) all of them are non-stationary, (ii) all of them are stationary and (iii) a mixture of both, some of them are non-stationary while some others are stationary. If all idiosyncratic components were non-stationary,

then the number of common factors could be determined after differentiation of the original systems of variables. In what follows we will review some of the procedures to set up the number of common factors when some of them are non-stationary.[35,36]

First, assuming that the idiosyncratic component is stationary, Peña and Poncela (2006) determine the number of common non-stationary factors as the number of nonzero eigenvalues of the following generalized covariance matrices

$$C_Y(k) = \frac{1}{T^\alpha} \sum_{t=k+1}^{T} (Y_{t-k} - \overline{Y})(Y_t - \overline{Y})' \tag{59}$$

where $\alpha = 2d$ if $d \geq 1$ while $\alpha = 1$ when $d = 0$ and $\overline{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$ is the time average of the vector of observed series. The procedures for determining the number of stationary common factors can then be implemented to $Y_t^* = Y_t - \Lambda_1(\Lambda_1'\Lambda_1)^{-1}\Lambda_1' Y_t$, where Λ_1 is the matrix of weights corresponding to the non-stationary factors. Although Peña and Poncela (2006) analyse the properties of this procedure in small cross-sectional dimension, Lam and Yao (2012) suggest that it could also be applied in models with large N. Peña and Poncela (2006) also argue that, instead of the generalized covariance matrices, the following generalized sample second moment matrices can be used to determine r_1

$$A_y(k) = \frac{1}{T^\alpha} \sum_{t=k+1}^{T} Y_{t-k} Y_t'. \tag{60}$$

This matrix is also used by Barigozzi and Trapani (2018) who propose a procedure based on randomization tests that allows for non-stationary idiosyncratic components and that can be implemented in settings with large cross-sectional dimension.

If the idiosyncratic component is assumed further to be not only stationary but also white noise, Peña and Poncela (2006) propose testing for the number of common factors, stationary or not, using the following test based on canonical correlations,

$$s = -(T-k) \sum_{j=0}^{N-r-1} log(1 - \lambda_{N-j}) \tag{61}$$

where $\lambda_1 > \cdots > \lambda_N$ are the ordered eigenvalues of the matrix $M(k) = C_Y'(k)C_Y^{-1}(0)C_Y(k)C_Y^{-1}(0)$. The statistic s is asymptotically a $\chi^2_{(N-r)^2}$ when T goes to ∞. The advantage of canonical corre-

[35] Notice that, in cases (ii) and (iii), there are cointegration relationships among the series in the system Y_t. For finite N, Escribano and Peña (1994) show that, if the idiosyncratic components were stationary and there were r_1 common trends or non-stationary factors, there must be $N \quad r_1$ cointegration relations among the observed series. So, the number of common trends, r_1, can be determined by estimating the number of cointegration relations $N \quad r_1$. Note that Bai and Wang (2016) show that the link between cointegration and common trends is broken in the setup of large N, T.

[36] Cointegration among the factors themselves is considered in Bariggozi, Lippi and Luciani (2020) and Barigozzi and Luciani (2019b). For simplicity, we assume that the non-stationary common factors are not cointegrated.

lation analysis is that you may detect the total number of common factors independently of their order of integration. The disadvantage of the procedure is that any remaining serial correlation present in the idiosyncratic noise is detected as an additional common factors; see also Pan and Yao (2008) for methods of detecting the number of common factors based on the same principle.

Apart from the previous methods that rely on analyzing the eigenstructure of certain matrices, Bai and Ng (2004), propose a procedure, PANIC, for the number of common non-stationary factors based on PC allowing for non-stationary idiosyncratic components as well. The procedure is as follows: first, apply PC to the data in first differences and extract the common factors. The number of common factors for stationary data can be determined applying the information criteria of Bai and Ng (2002). Denote them by f_t. Then recumulate to obtain the common factors in levels $\hat{F}_t = \sum_{s=2}^{t} f_s, t = 2, ..., T$. In this way, we have consistent estimates of the common factors even though there might be unit roots in the idiosyncratic components. If there is only one common factor, PANIC performs simple unit root test. If there are multiple factors, Bai and Ng (2004) suggests the tests for common trends proposed by Stock and Watson (1988) to determine the number of independent stochastic trends in the estimated r common factors; see Harvey (1989) who also suggests using Stock and Watson (1988) to determine the number of common trends. As regards the non-stationarity of the idiosyncratic noises, univariate unit root tests may have low power, so they proposed to use panel unit root tests. In particular, they propose the following test

$$S = \frac{-2\sum_{i=1}^{N} logs_i - 2N}{\sqrt{4N}}$$

where s_i is the p-value corresponding to the Dickey-Fuller test of the ith idiosyncratic residual. Pooled tests could not be used in the original data because of strong cross-correlation due to the common factors, but they can be used in the specific components since this strong cross-correlation has been removed after extracting the common factors.

Finally, Nyblom and Harvey (2000) propose a test for the number of common trends based on the Lagrange multiplier principle. The asymptotic distribution of their test statistics depends only on the rank r_1, the number of common non-stationary factors, of the covariance matrix of the disturbances driving the N random walks.

7.4 Empirical applications of non-stationary DFMs

The number of empirical applications of non-stationary DFMs using KFS for factor extraction is much more scarce than in the stationary case.

Many important applications of DFMs deal with real economic activity. Although the vast majority of papers using DFMs to track and forecast the real economy use stationary DFMs, there are a few papers that fit non-stationary DFMs using KFS procedures to extract the factors. In general, analysts are interested in: (i) detecting the turning points of the business cycle through

indexes of economic activity and (ii) forecasting real GDP or other variables related to real economic activity. One of the pioneers paper is that by Quah and Sargent (1993) who use information from a large set of cross-sectional data to track the temporal evolution of aggregate fluctuations. They use a non-stationary DFM with annual data ($T = 48$) over a cross section of $N = 60$ of sectoral US unemployment. This seminal paper has spurred a vast literature regarding DFMs, stationary or not, for the analysis of the business cycle. García-Ferrer and Poncela (2002) use a small scale factor model with stationary and nonstationary common factors to forecast real GNP of a set of euro area countries that represent about 80% of the euro area GDP. They find the DFMs outperform other forecasting alternatives. The same data set is used in Peña and Poncela (2004) to illustrate the smaller MSEs obained with DFMs over the pooled term in García-Ferrer et al. (1987) and univariate alternatives. They also show that the common factor estimated as the pooling term exhibits minimum MSE if the common factor follows a random walk. Barigozzi and Luciani (2017) use nonstationary DFMs to separate long-run from short-run comovements in real economic activity using a large data set of US macroeconomic indicators. Frale et al. (2011) are interested in an index of economic activity and use mixed frequency non-stationary DFMs to build EURO-MIND, an index to track the business cycle in the euro area and Seong et al. (2013) use the data of the four big US monthly indicators (industrial production, employment, income and sales) and quarterly GDP to compute in-sample monthly smoothed estimates and out-of-sample monthly forecasts of GDP. Bujosa et al. (2013, 2020) determine non-stationary common trends based on the identification procedure in Peña and Poncela (2006) to build monthly coincident and leading indicators for the Spanish economy. They keep only the lower frequency component in order to obtain smooth indexes with the aim of reducing the uncertainty of the state of the economy. When building indexes of economic activity, usually one factor models are estimated. However, Martínez, Nieto and Poncela (2016) ask the question of which common factor should be used as the index of interest in multifactor models. They complement the analysis comparing the peaks of the estimated common factors with those of a reference series using Fisher's randomization test for matching pairs. They apply it to real economic activity data in Colombia and to financial series from stock markets also in Colombia. A different application related with GDP can be found in Corona, Poncela and Ruiz (2020) where non-stationary DFMs are used to detect risk sharing or the amount of GDP shocks that are smoothed, and therefore not passed into consumption, through cross-border international flows.

In the nonstationary case there are also DFMs to monitor inflation; see, for instance, Delle Monache, Petrella and Venditti (2016).

Another set of applications is more related to finance. Harvey, Ruiz and Shephard (1994) find that two common trends explain the evoluation of volatilities of four exchange rates in the context of multivariate stochastic volatility models. Peña and Poncela (2006) use stationary

and non-stationary common factors to model the interest rates at different maturities. Broto and Perez-Quiros (2015) work with sovereign credit default swaps (CDS) spreads of ten OECD economies and decompose them into a global common random walk, a specific random walk linked to European peripheral countries and an idiosyncratic component in order to study contagion.

Other applications include hourly electricity prices as in Alonso et al. (2011) and Carpio, Juan and López (2014). In demography, Ortega and Poncela (2005) use nonstationary DFMs for $I(2)$ time series in order to forecast fertitlity rates of Southern European countries. In environmetrics, Zuur et al. (2003) apply nonstationary DFMs to model biomass of marine species and Nieto, Peña and Saboyá (2016) extend the non-stationary DFM to seasonal data and apply it to monthly measures of rainfall measured at six airports in Colombia.

8 Conclusions

Using KFS procedures for factor extraction is a powerful instrument that generates efficient estimates of the underlying latent factors, being able to deal in a direct way with different types of data anomalities and non-stationarities. Through the last years, some issues dealing with the computational complexity of KFS factor extraction have been solved and the EM algorithm is now predominant in the estimation of the parameters of DFMs even when the cross-sectional dimension is very large. When fitting DFMs written in SSF to large systems of time series, it is crucial to take special care of the identification of the parameters and the specification of the model (mainly in cases of dynamic and non-stationary factors). Due to the importance of a correct specification of the factors, there is a need for more evaluation and testing of the model specification.

References

[1] Aastveit, K.A. and Trovik, T. (2012), Nowcasting Norwegian GDP: the role of asset prices in a small open economy, *Empirical Economics*, 42(1), 95-119.

[2] Ahn, S. and Horenstein, A. (2013), Eigenvalue ratio test for the number of factors, *Econometrica*, 81(3), 1203-1227.

[3] Alessi, L., Barigozzi, M. and Capasso, M. (2010), Improved penalization for determining the number of factors in approximate factor models, *Statistics and Probability Letters*, 80(1), 1806-1813.

[4] Alonso, A.M., C. García-Martos, J. Rodríguez and M.J. Sánchez (2011), Seasonal dynamic factor analysis and bootstrap inference: application to electricity market forecasting, *Technometrics*, 53(2), 137-151.

[5] Amengual, D. and Watson, M.W. (2007), Consistent estimation of the number of dynamic factors in a large N and T panel, *Journal of Business & Economic Statistics*, 25(1), 91-96.

[6] Anderson, T.W. and Rubin, H. (1956), Statistical inference in factor analysis, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and probability: Contributions to the Theory of Statistics*, University of California Press, Berkeley.

[7] Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin and G. Rünstler (2011), Short-term forecasts of euro area GDP growth, *The Econometrics Journal*, 14(1), 25-44.

[8] Ansley, C.F. and R. Kohn (1985), Estimation, filtering and smoothing in state space models with incomplete specified initial conditions, *Annals of Statistics*, 13(4), 1286-1316.

[9] Aruoba, S.B. and Diebold, F.X. (2010), Real-time macroeconomic monitoring: Real activity, inflation and interactions, *American Economic Review*, 100(2), 10-24.

[10] Aruoba, S.B., Diebold, F.X. and Scotti, C. (2009), Real-time measurement of business conditions, *Journal of Business & Economic Statistics*, 27(4), 417-427.

[11] Bai, J. (2004), Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics*, 122(1), 137-183.

[12] Bai, J. and Li, K. (2012), Statistical analysis of factor models of high dimension, *Annals of Statistics*, 40(1), 436-465.

[13] Bai, J. and Li, K. (2016), Maximum likelihood estimation and inference for approximate factor models of high dimension, *Review of Economics and Statistics*, 98(2), 298-309.

[14] Bai, J. and Ng, S. (2002), Determining the number of factors in approximate factor models, *Econometrica*, 70(1), 191-221.

[15] Bai, J. and Ng, S. (2004), A PANIC attack on unit roots and cointegration, *Econometrica*, 72(4), 1127-1177.

[16] Bai, J. and Ng, S. (2007), Determining the number of primitive shocks in factor models, *Journal of Business & Economic Statistics*, 25(1), 52-60.

[17] Bai, J. and Ng, S. (2008), Large dimensional factor analysis, *Foundations and Trends in Econometrics*, 3, 89-163.

[18] Bai, J. and Ng, S. (2013), Principal Components estimation and identification of static factors, *Journal of Econometrics*, 176(1), 18-29.

[19] Bai, J. and Wang, P. (2016), Econometric analysis of large factor models, *Annual Review of Economics*, 8, 53-80.

[20] Bańbura, M. and Modugno, M. (2014), Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data, *Journal of Applied Econometrics* 29(1), 133-160.

[21] Bańbura, M. and Rünstler, G. (2011), A look into the factor model black box. Publication lags and the role of hard and soft data in forecasting, *International Journal of Forecasting*, 27(2), 333-346.

[22] Bańbura, M., Giannone, D. and Reichlin, L. (2011), Nowcasting, in Clements, M.P. and Hendry, D.F. (eds.), *Oxford Handbook on Economic Forecasting*, Oxford University Press, New York.

[23] Bańbura, M., Giannone, D., Modugno, M. and Reichlin, L. (2013), Nowcasting and the real-time data flow, in Elliot, G. and A. Timmermann (eds.), *Handbook on Economic Forecasting*, Vol. 2, 195-237, Elsevier.

[24] Banerjee, A., M. Macellino and I. Masten (2017), Structural FECM: Cointegration in large-scale structural FAVAR models, *Journal of Applied Econometrics*, 32(6), 1069-1086.

[25] Barhoumi, K., Darné, O., Ferrara, L. (2013), Dynamic factor models: A review of the literature, *Journal of Business Cycle Measurement and Analysis* 8(2), 73-107.

[26] Barigozzi, M. and M. Luciani (2017), Common factors, trends, and cycles in large datasets, A Finance and Economics Discussion Series 2017-111, Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2017.111

[27] Barigozzi, M. and Luciani, M. (2019a), Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm, *Manuscript*, arXiv:1910.03821v1[math.ST].

[28] Barigozzi, M. and Luciani, M. (2019b), Quasi maximum likelihood estimation of non-stationary large approximate dynamic factor models, *ManuscriptManuscript*, arXiv:1910.09841v1[econ.EM]

[29] Barigozzi, M. and Trapani (2018), Determining the dimension of factor structures in large datasets, *Manuscript*, arXiv:1806.03647v1[stat.ME].

[30] Barigozzi, M., Lippi, M. and Luciani, M. (2020a), Cointegration, and error correction mechanisms for singular stochastic vectors, *Econometrics*, 8,3.

[31] Barigozzi, M., Lippi, M. and Luciani, M. (2020b), Large-dimensional dynamic factor models: estimation of impulse-response functions with I(1) cointegrated factors, *Journal of Econometrics*, forthcoming.

[32] Bräuning, F. and S.J. Koopman (2014), Forecasting macroeconomic variables using collapsed dynamic factor analysis, *International Journal of Forecasting*, 30(3), 572-584.

[33] Breitung, J. and Eickmeier, S. (2006), Dynamic Factor Models, in *Modern Econometric Analysis*, Heidelberg: Springer, 25-40.

[34] Breitung, J. and Pigorsch, U. (2013), A canonical correlation approach for selecting the number of dynamic factors, *Oxford Bulletin of Economics and Statistics*, 75(1), 23-26.

[35] Broto, C. and G. Perez-Quiros (2015), Disentangling contagion among sovereign CDS spreads during the European debt crisis, *Journal of Empirical Finance*, 32, 165-179.

[36] Bujosa, M., A. García-Ferrer, A. de Juan, A. Martín-Arroyo (2020), Evaluating early warning and coincident indicators of business cycles using smooth trends, *Journal of Forecasting*, 39(1), 1-17.

[37] Bujosa, M., A. García-Ferrer, A. de Juan (2013), Predicting recessions with factor linear dynamic harmonic regressions, *Journal of Forecasting*, 32(6), 481-499.

[38] Camacho, M. and Martinez-Martin, J. (2014), Real-time forecasting US GDP from small-scale factor models, *Empirical Economics*, 41, 347-364.

[39] Camacho, M. and Perez-Quiros, G. (2010), Introducing the Euro-Sting: Short term indicators of the euro area growth, *Journal of Applied Econometrics*, 25(4), 663-694.

[40] Camacho, M., dal Bianco, M. and Martinez-Martin, J. (2015a), Toward a more reliable picture of the economic activity: An application to Argentina, *Economic Letters*, 132, 129-132.

[41] Camacho, M., dal Bianco, M. and Martinez-Martin, J. (2015b), Short-run forecasting of Argentine GDP growth, *Emerging Markets Finance and Trade*, 51, 473-485.

[42] Camacho, M., Leiva-Leon, D. and Perez-Quiros, G. (2016), Country shocks, monetary policy expectations and ECB decisions: A dynamic non-linear approach, *Advances in Econometrics*, 35, 283-316.

[43] Camacho, M., Lovcha, Y. and Perez-Quiros, G. (2015), Can we use seasonally adjusted indicators in dynamic factor models?, *Studies in Nonlinear Dynamics and Econometrics*, 19(3), 377-391.

[44] Camacho, M., Perez-Quiros, G. and Poncela, P. (2013), Short-term forecasting for empirical economists: A survey of the recently proposed algorithms, *Foundations and Trends in Econometrics*, 6(2), 101-161.

[45] Camacho, M., Perez-Quiros, G. and Poncela, P. (2015), Extracting nonlinear signals from several economic indicators, *Journal of Applied Econometrics*, 30(7), 1073-1089.

[46] Camacho, M., Perez-Quiros, G. and Poncela, P. (2018), Markov-switching dynamic factor models in real time, *International Journal of Forecasting*, 34(4), 598-611.

[47] Carpio, J., J. Juan and D. López (2014), Multivariate exponential smoothing and dynamic factor model applied to hourly electricity price analysis, *Technometrics*, 56(4), 494-503.

[48] Carriero, A. and Marcellino, M. (2007), A comparison of methods for the construction of composite coincident and leading indexes for the UK, *International Journal of Forecasting*, 23(2), 219-236.

[49] Cattell, R.B. (1966), The scree test for the number of factors, *Multivariate Behavioral Research*, 1, 245-276.

[50] Cavanaugh, J.E. and R.H. Shumway (1996), On computing the expected Fisher information matrix for state-space model parameters, *Statistics Probability Letters*, 26(4), 347-355.

[51] Chan, J.C., R. Leon-Gonzalez and R.W. Strachan (2018), Invariant inference and efficient computation in the static factor model, *Journal of the American Statistical Association*, 113(522), 819-828.

[52] Chauvet, M. and Z. Senyuz (2016), A dynamic factor model of the yield curve components as a predictor of the economy, *International Journal of Forecasting*, 32(2), 324-343.

[53] Choi, I. and H. Jeong (2019), Model selection for factor analysis: some new criteria and performance comparisons, *Econometric Reviews*, 38(6), 577-596.

[54] Corona, F., P. Poncela and E. Ruiz (2017), Determining the number of factors after stationary univariate transformations, *Empirical Economics*, 53, 351-372.

[55] Corona, F., P. Poncela and E. Ruiz (2020), Estimating non-stationary common factors: implications for risk-sharing, *Computational Economics*, 55, 37-60.

[56] Coroneo, L., Giannone, D. and Modugno, M. (2016), Unspanned macroeconomic factors in the yield curve, *Journal of Business Economic Statistics*, 34(3), 472-485.

[57] Coulombe,G., F. Diebold, M. Gobel, G. Rudebush and B. Zhang (this issue), Optimal combination of Artic Sea ice extent measures: A dynamic factor modelling approach, *International Journal of Forecasting*.

[58] D'Agostino, A. and Giannone, D. (2012), Comparing alternative predictors based on large-panel factor models, *Oxford Bulletin of Economics and Statistics*, 74(2), 306-326.

[59] D'Agostino, A., McQuinn, K. and O'Brien, D. (2012), Now-casting Irish GDP, *OECD: Journal of Business Cycle Measurement and Analysis*, 2012/2.

[60] de Jong, P. (1991), The diffuse Kalman filter, *Annals of Statistics*, 19(2), 1073-1083.

[61] Delle Chiaie, S., L. Ferrara and D. Giannone (2017). Common factors of commodity prices. Banque de France Working Paper No. 645.

[62] Delle Monache, D. and I. Petrella (2019). Efficient matrix approach for classical inference in state space models. *Economics Letters*, 181, 22-27.

[63] Delle Monache, D., I. Petrella and F. Venditti (2016). Common faith or parting ways?, in Hillebrand, E. and Koopman, S.J. (eds.), *Dynamic Factor Models*, Advanced Econometrics, 35, 539-565.

[64] Dempster, A.P., N.M. Laird and D.B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, 39(1), 1-38.

[65] den Reijer, A. and Johansson, A. (2019). Nowcasting Swedish GDP with a large and unbalanced data set. *Empirical Economics*, 57(4), 1351-1373.

[66] Diebold, F.X. (2003), Big data dynamic factor models for macroeconomic measurement and forecasting, in Dewatripoint, M. and Turnovsky, S. (eds.), *Advances in Ecvonomics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, 115-122.

[67] Diebold, F.X. and C. Li (2006), Forecasting the term structure of Government Bond yields, *Journal of Econometrics*, 130(2), 337-364.

[68] Diebold, F.X., S. Rudebusch and S. Aruoba (2006), The macroeconomy and the yield curve, *Journal of Econometrics*, 131(1), 309-338.

[69] Doz,C. and P. Fuleky (2020), Dynamic factor models, in Fuleky, P. (ed.), *Macroeconomic Forecasting in the Era of Big Data*, Advanced Studies in Theoretical and Applied Econometrics Book Series, Vol. 52, Springer Nature.

[70] Doz, C., D. Giannone and L. Reichlin (2011), A two-step estimator for large approximate dynamic factor models based on Kalman filtering, *Journal of Econometrics*, 164(1), 188-205.

[71] Doz, C., D. Giannone and L. Reichlin (2012), A quasi-maximum likelihood approach for large, approximate dynamic factor models, *Review of Economics and Statistics*, 94(4), 1014-1024.

[72] Eickmeier, S. and C. Ziegler (2008), How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach, *Journal of Forecasting*, 27(3), 237-265.

[73] Eickmeier, S., W. Lemke and M. Marcellino (2015), Classical time varying factor-augmented vector auto-regressive models: estimation, forecasting and structural analysis, *Journal of Royal Statistical Society, A*, 178(3), 493-533.

[74] Engle, R. and M. Watson (1981), A one-factor multivariate time series model for metropolitan wage rates, *Journal of the American Statistical Association*, 76(376), 774-781.

[75] Escribano, A. and Peña, D. (1994), Cointegration and common factors, *Journal of Time Series Analysis*, 15(6), 577-586.

[76] Evans, M. (2005), Where are we now? Real time estimates of the macroeconomy, *International Journal of Central Banking*, 1(2), 127-175.

[77] Fiorentini, G. and E. Sentana (2019), Dynamic specification tests for dynamic factor models, *Journal of Applied Econometrics*, 34(3), 325-346.

[78] Fiorentini, G., A. Galesi and E. Sentana (2018), A spectral EM algorithm for dynamic factor models, *Journal of Econometrics*, 205(1), 249-279.

[79] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), The generalized dynamic factor model: identification and estimation, *Review of Economics and Statistics*, 82(4), 540-554.

[80] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004), The generalized dynamic factor model: consistency and rates, *Journal of Econometrics*, 119(2), 231-255.

[81] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), The generalized dynamic factor model: one-sided estimation and forecasting, *Journal of the American Statistical Association*, 100(471), 830-839.

[82] Foroni, C. and Marcellino, M. (2014), A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates, *International Journal of Forecasting*, 30(3), 554-568.

[83] Frale, C., M., Marcellino, G.L. Mazzi and T. Proietti (2011), EUROMIND: A monthly indicator of the euro area economic conditions, *Journal of the Royal Statistical Society, A*, 174(2), 439-470.

[84] García-Ferrer, A. and Poncela, P. (2002), Forecasting European GNP data through common factor models and other procedures, *Journal of Forecasting*, 21(4), 225-244.

[85] García-Ferrer, A., R.A. Highfield, F. Palm and A, Zellner (1987), Macroeconomic forecasting using pooled international data, *Journal of Business Economic Statistics*, 5(1), 53-67.

[86] Giannone, D., Reichlin, L. and Sala, L. (2005), Monetary policy in real time, *Macroeconomic Annual*, 19, 161-200.

[87] Giannone, D., Reichlin, L. and Small, D. (2008), Nowcasting: the real-time informational content of macroeconomic data, *Journal of Monetary Economics*, 55(4), 665-676.

[88] Grassi, S., Proietti, T., Frale, C., Marcellino, M. and Mazzi, G. (2015), EuroMInd-C: A disaggregate monthly indicator of economic activity for the Euro area and member countries, *International Journal of Forecasting*, 31(3), 712-738.

[89] Hallin, M. and Liska, R. (2007), Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association*, 102(478), 603-617.

[90] Harvey, A. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, 2nd ed., Cambridge University Press, Cambridge.

[91] Harvey, A.C. and S.J. Koopman (1992), Diagnostic checking of unobserved components time series models, *Journal of Business & Economic Statistics*, 3, 84-107.

[92] Harvey, A.C., E. Ruiz and N. Shephard (1994), Multivariate stochastic variance models, *The Review of Economic Studies*, 61(2),247-264.

[93] Hindrayanto, I., S.J. Koopman and J. de Winter (2016), Forecasting and nowcasting economic growth in the euro area using factor models, *International Journal of Forecasting*, 32(4), 1284-1305.

[94] Jansen, W.J. and J.M. de Winter (2018), Combining model-based near term GDP forecasts and judgmental forecasts: A real-time exercise for G-7 countries, *Oxford Bulletin of Economics and Statistics*, 80(6), 1213-1242.

[95] Jansen, W.J., X. Jiu and J.M. de Winter (2016), Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts, *International Journal of Forecasting*, 32(2), 411-436.

[96] Jackson, L.E., M.A. Kose, C. Otrok and M.T. Owyang (2016), Specification and estimation of Bayesain dynamic factor models: A Monte Carlo analysis with an application to global house price comovement, in *Advances in Econometrics*, Emerald Group Publishing Limited: Bingley, 361-400.

[97] Johansen, S. and M.N. Tabor (2017), Cointegration between trends and their estimators in state space models and cointegrated vector autoregressive models, *Econometrics*, 5(3), 36.

[98] Joseph, A., E. Kalamara, G. Potjagailo and G. Kapetanios (this issue), Forecasting UK inflation bottom up, *International Journal of Forecasting*.

[99] Jungbacker, B. and S.J. Koopman (2015), Likelihood-based dynamic factor analysis for measurement and forecasting estimation, *Econometrics Journal*, 18(2), C1-C21.

[100] Jungbacker, B., S.J. Koopman and M. van der Wel (2011), Maximum likelihood estimation for dynamic factor models with missing data, *Journal of Economic Dynamics and Control*, 35(8), 135-168.

[101] Jungbacker, B., S.J. Koopman and M. van der Wel (2014), Smooth dynamic factor analysis with application to the US term structure of interest rates, *Journal of Applied Econometrics*, 29(1), 65-90.

[102] Kapetanios, G. and Marcellino, M. (2009), A parametric estimation method for dynamic factor models of large dimensions, *Journal of Time Series Analysis*, 30(2), 208-238.

[103] Kaufmann, S. and C. Schumaker (2019), Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification, *Journal of Econometrics*, 210(1), 116-134.

[104] Koop, G. and D. Korobilis (2014), A new index of financial conditions, *European Economic Review*, 71, 101-116.

[105] Koopman, S.J. and M. Van der Wel (2013), Forecasting the US term structure of interest rates using a macroeconomic smooth dynamic factor model, *International Journal of Forecasting*, 29(4), 676-694.

[106] Koopman, S.J., M.I.P. Mallee and M. Van der Wel (2010), Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters, *Journal of Business & Economic Statistics*, 28, 329-343.

[107] Kose, M.A., C. Otrok and C.H. Whiteman (2003), International business cycles: World, region, and country-specific factors, *American Economic Review*, 93(4), 1216-1239.

[108] Kuzin, V., M. Marcellino and C. Schumacher (2013), Pooling versus model selection for nowcasting GDP with many predictors: empirical evidence for six industrialized countries, *Journal of Applied Econometrics*, 28(3), 392-411.

[109] Lam, C. and Yao, Q. (2012), Factor modeling for high-dimensional time series: inference for the number of factors, *The Annals of Statistics*, 40(2), 694-726.

[110] Lopes, H.F. and M. West (2004), Bayesian model assessment in factor analysis, *Statistica Sinica*, 14(1), 41-67.

[111] Luciani, M. (2014), Forecasting with approximate dynamic factor models: the role of non-pervasive shocks, *International Journal of Forecasting*, 30(1), 20-29.

[112] Luciani, M. (2015), Monetary policy and the housing market: A structural factor analysis, *Journal of Applied Econometrics*, 30(2), 199-208.

[113] Luciani, M. (2017), Large-dimensional dynamic factor models in real-time: A Survey, in Mazzi, G.L. and Ozyildirim (eds.), *Handbook on Cyclical Composite Indicators*, Eurostat.

[114] Maciejowska, K. (2010), Common factors in nonstationary panel data with a deterministic trend-Estimation and distribution theory, EUI Working papers, ECO2010/28.

[115] Marcellino, M. and Schumacher, C. (2010), Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP, *Oxford Bulletin for Economics and Statistics*, 72(4), 518-550.

[116] Marcellino, M. and Sivec, V. (2016), Monetary, fiscal and oil shocks: Evidence based on mixed frequency structural FAVARs, *Journal of Econometrics*, 193(2), 335-348.

[117] Mariano, R.S. and Y. Murasawa (2003), A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics*, 18(4), 427-443.

[118] Mariano, R.S. and Y. Murasawa (2010), A coincident index, common factors, and monthly real GDP, *Oxford Bulletin of Economics and Statistics*, 72(1), 27-46.

[119] Martínez, W., F.H. Nieto and P. Poncela (2016), Choosing a dynamic common factor as a coincident index, *Statistics Probability Letters*, 109, 89-98.

[120] Martinez-Martin, J. and E. Rusticelli (in press), Keeping track of global trade in real time, *International Journal of Forecasting*.

[121] Matheson, T.D. (2010), An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys, *Economic Modelling*, 27(1), 304-314.

[122] Meng, X.-L. and D.B. Rubin (1991), Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm, *Journal of the American Statistical Association*, 86(416), 899-909.

[123] Molenaar, P.C.M., De Gooijer, J. and Schmitz, B. (1992), A dynamic factor analysis of nonstationary multivariate time series, *Psycometrika*, 57, 333-349.

[124] Nieto, F.H., D. Peña and D. Saboyá (2016), Common seasonality in multivariate time series, *Statistica Sinica*, 26(4), 1389-1410.

[125] Nyblom, J. and Harvey, A.C. (2000), Tests of common stochastic trends, *Econometric Theory*, 16(2), 176-199.

[126] Onatski, A. (2009), Hypothesis testing about the number of factors in large factor models, *Econometrica*, 77(5), 1447-1479.

[127] Onatski, A. (2010), Determining the number of factors from empirical distribution of eigenvalues, *The Review of Economics and Statistics*, 92(4), 1004-1016.

[128] Onatski, A. and C. Wang (2020), Spurius factor analysis, CWPE2003.

[129] Ortega, J.A. and P. Poncela (2005), Joint forecasts of Southern European fertility rates with non-stationary dynamic factor models, *International Journal of Forecasting*, 21(3), 539-550.

[130] Oakes, D. (1999), Direct calculation of the information matrix via the em algorithm, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 61(2), 479-482.

[131] Otrok, C. and C.H. Whiteman (1998), Bayesian leading indicators: measuring and predicting economic conditions in Iowa, *International Economic Review*, 39(4), 997-1014.

[132] Pan, J. and Yao, Q. (2008), Modelling multiple time series via common factors, *Biometrika*, 95(2), 365-379.

[133] Peña, D. and P. Poncela (2004). Forecasting with nonstationary dynamic factor models, *Journal of Econometrics*, 119(2), 291-321.

[134] Peña, D. and Poncela, P. (2006), Nonstationary dynamic factor analysis, *Journal of Statistical Planning and Inference*, 136(4), 1237-1257.

[135] Peña, D. and R. Tsay (2020), *Statistical Learning for Big Dependent Data*, Wiley.

[136] Peña, D., E. Smucler and V. Yohai (2019), Forecasting multiple time series with one-sided dynamic principal components, *Journal of the American Statistical Association*, 114(528), 1683-1694.

[137] Pinheiro, M., A. Rua and F. Dias (2013), Dynamic factor models with jagged edge panel data: taking on board the dynamics of the idiosyncratic components, *Oxford Bulletin of Economics and Statistics*, 75(1), 0305-9049.

[138] Poncela, P. (2013), Comments on "Forecasting the US term structure of interest rates using a macroeconomic smooth dynamic factor model", *International Journal of Forecasting*, 29(4), 695-697.

[139] Poncela, P. and García-Ferrer, A. (2014), The effects of disaggregation on forecasting non-stationary time series, *Journal of Forecasting*, 33(4), 300-314.

[140] Poncela, P. and Ruiz, E. (2015), More is not always better: back to the Kalman filter in Dynamic Factor Models, in Koopman, S.J. and Shephard, N.G. (eds.), *Unobserved Components & Time Series Econometrics*, Oxford University Press.

[141] Poncela, P. and Ruiz, E. (2016), Small versus big-data factor extraction in Dynamic Factor Models, in Hillebrand, E. and Koopman, S.J. (eds.), *Dynamic Factor Models*, Advanced Econometrics, 35, 401-434.

[142] Poncela, P. and Ruiz, E. (2020), A comment on the dynamic factor model with dynamic factors, *Economics*, Discussion Paper 2020-7.

[143] Poncela, P., E. Senra and L.P. Sierra (2020), Global vs sectoral factors and the impact of the financialization in commodity price changes, *Open Economic Review*.

[144] Proietti, T. (2011), Estimation of common factors under cross-sectional and temporal aggregation constrains, *International Statistical Review*, 79(3), 455-476.

[145] Quah, D. and T.J. Sargent (1993), A dynamic index model for large cross sections, in Stock, J. and M. Watson (eds.), *Business Cycles, Indicators, and Forecasting*, NBER and University of Chicago Press.

[146] Reis, R. and Watson, M.W. (2010), Relative goods' prices, pure inflation and the Phillips correction, *American Economic Review: Macroeconomics*, 2(3), 128-137.

[147] Rünstler, G. (2016), On the design of data sets for forecasting with dynamic factor models, in Hillebrand, E. and Koopman, S.J. (eds.), *Dynamic Factor Models*, Advanced Econometrics, 35.

[148] Rünstler, G., Barhoumi, K., Cristadoro, R., Reijer, A.D., Jakaitiene, A. Jelonek, P. Rua, A., Ruth, K., Benk, S. and Nieuwenhuyze, C.V. (2009), Short-term forecasting of GDP using large monthly data sets: a pseudo real-time forecast evaluation, *Journal of Forecasting*, 28(7), 595-611.

[149] Schiavoni, C., Palm, F., Smeekes, S. and van der Brakel, J. (2019), A dynamic factor model approach to incorporate Big Data in state space models for official statistics, *Working paper*.

[150] Scotti, C. (2016), Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises, *Journal of Monetary Economics*, 82, 1-19.

[150] Seong, B., S.K. Ahn and P.A. Zadrozny (2013), Estimation of vector error correction models with mixed-frequency data, *Journal of Time Series Analysis*, 34(2), 194-205.

[151] Shumway, R.H. and D.S. Stoffer (1982), An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis*, 3(4), 253-264.

[152] Solberger, M. and Spanberg, E. (2020), Estimating a dynamic factor model in Eviews using the Kalman filter and smoother, *Computational Economics*, 55, 875-900.

[153] Stock, J.H. and Watson, M.W. (1988), Testing for common trends, *Journal of the American Statistical Association*, 83(404), 1097-1107.

[154] Stock, J.H. and Watson, M.W. (1989), New indexes of coincident and leading economic indicators, *Macroeconomics Annual 1989*, National Bureau of Economic Research, 4, 351-409.

[155] Stock, J.H. and Watson, M.W. (1991), A probability model of the coincident economic indicators, in Moore, G. and Lahiri, K. (eds.), *The Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press, Cambridge.

[156] Stock, J.H. and Watson, M.W. (2002), Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association*, 97(460), 1167-1179.

[157] Stock, J.H. and Watson, M.W. (2005), Implications of dynamic factor models for the VAR analysis, Working paper 11467, National Bureau of Economic Research.

[158] Stock, J.H. and Watson, M.W. (2011), Dynamic factor models, in Clements, M.P. and D.F. Hendry (eds.), *Oxford Handbook of Economic Forecasting*, Oxford University Press.

[159] Stock, J.H. and Watson, M.W. (2016a), Dynamic factor models: A brief retrospective, in Hillebrand, E. and S.J. Koopman (eds.), *Dynamic Factor Models, (Advances in Econometrics)*, Vol. 35, Emerald Group Publishing Limited.

[160] Stock, J.H. and Watson, M.W. (2016b), Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics, in Taylor, J.B. and H. Uhlig (eds.), *Handbook of Macroeconomics*, Vol. 2A, Chapter 8, 415-526, Elsevier.

[161] Stock, J.H. and Watson, M.W. (2017), Twenty years of time series econometrics in ten pictures, *Journal of Economic Perspectives*, 31(2), 59-86.

[162] Stoffer, D.S. and K.D. Wall (1991), Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter, *Journal of the American Statistical Association*, 86(416), 1024-1033.

[163] Trenkler, C. and Weber, E. (2016), On the identification of multivariate uncorrelated unobserved components models, *Economics Letters*, 138, 15-18.

[164] Watson, M.W. and Engle, R.F. (1983), Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models, *Journal of Econometrics*, 23(3), 385-400.

[165] Wu, C.-F. J. (1983), On the convergence properties of the EM algorithm, *Annals of Statistics*, 11(1), 95-103.

[166] Zhao, Z., Cui, G. and Wang, S. (2017), A Monte Carlo comparison of estimating the number of dynamic factors, *Empirical Economics*, 53(3), 1217-1241.

[167] Zuur, A.F., R.J. Fryer, I.T. Jolliffe, R. Dekker and J.J. Beukema (2003), Estimating common trends in multivariate time series using dynamic factor analysis, *Environmetrics*, 14(7), 665-685.