Working paper
2021-03
Statistics and Econometrics
ISSN 2387-0303

Dynamic factor models: Does the specification matter?

Karen Miranda, Pilar Poncela, Esther Ruiz

uc3m | Universidad Carlos III de Madrid

Dynamic factor models: Does the specification matter?

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

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

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

20 March 2021

Abstract

Dynamic Factor Models (DFMs), which assume the existence of a small number of unobserved underlying factors capturing the comovements in large systems of variables, are very popular among empirical macroeconomists to reduce dimension and to extract factors with an economic interpretation. Factors can be extracted using either non-parametric Principal Components (PC) or parametric Kalman filter and smoothing (KFS) procedures, with the former being computationally simpler and robust against misspecification and the latter being efficient if the specification is correct and coping in a natural way with missing and mixed-frequency data, time-varying parameters, non-linearities and non-stationarity among many other stylized facts often observed in real systems of economic variables. This paper analyses the empirical consequences on factor estimation and forecasting of using alternative extraction procedures and estimators of the DFM parameters under various sources of potential misspecification. In particular, we consider factor extraction when assuming different number of factors and different factor dynamics. The factors are extracted from a popular data base of US macroeconomic variables that has been widely analyzed in the literature without consensus about the most appropriate model specification. We show that this lack of consensus is ony marginally cruzial when it comes to factor extraction but it matters when the objective is forecasting.

Keywords: EM algorithm, Kalman filter, Principal Components, State-space model.

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

†Financial support from the Spanish National Research Agency (Ministry of Science and Technology) Projects PID2019-108079GB-C21/AIE/10.13039/501100011033 and PID2019-108079GB-C22/AIE/10.13039/501100011033 is gratefully acknowledged by the first and third authors, and the second author, respectively.

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. DFMs generally assume the existence of a small number of unobserved factors capturing the comovements in the system.[1] Two main types of procedures for factor extraction are popular in the related literature. First, in many applications factors are extracted using non-parametric procedures based on Principal Components (PC), which are attractive because they are computational simple and have well-known theoretical properties. In particular, PC is consistent under mild conditions and, as far as the factors are pervasive and the idiosyncratic dependence is weak, it is robust to the underlying dependence of common factors and idiosyncratic components . As a consequence, PC procedures are very popular for factor estimation and several excellent surveys are available in the literature; see, among others, Bai and Ng (2008a) for a technical survey on the econometric theory for PC and Stock and Watson (2011) for an overview with a focus on applications. However, when the common factors and/or idiosyncratic components are serially dependent, PC procedures do not use this information and, consequently, they are not efficient.

Alternatively, after casting the DFM as a state-space model (SSM), factors can be extracted using Kalman Filter and Smoothing (KFS) procedures. One important feature of these procedures is that they open the door to Maximum Likelihood (ML) estimation of the model parameters. Furthermore, if the model specification is correct, KFS is efficient for factor extraction. KFS is also very flexible allowing to handle in a straightforward way data characteristics often observed in practice as, for example, missing data, mixed frequencies, seasonal dependencies, nonstationarity or regime-switching nonlinearity. Moreover, KFS procedures are also of interest in empirical applications because they allow incorporating restrictions on the factor loadings, as in multilevel DFMs, or on the idiosyncratic components and to perform counterfactual exercises; see, for example, Banbura, Giannone and Reichlin (2011), Forni and Reichlin (2001), Beck, Hubrich and Marcellino (2009, 2016), Camacho, Pacce and Perez-Quiros (2020) and Coroneo, Giannone and Modugno (2016) for multilevel models and Luciani (2015) for counterfactual analysis. However, KFS procedures also have drawbacks, the main one being that they require full specification of the dependence of the common and idiosyncratic components, opening the door to potential misspecification; see Poncela, Ruiz and Miranda (forthcoming) for a very recent survey on KFS for factor extraction in DFMs.

This paper analyses the empirical consequences on factor estimation and forecasting of using

[1] See Chen, Dolado and Gonzalo (2021) for a factor model with the factors being common in the quantiles.

alternative factor extraction procedures and estimators of the DFM parameters under various sources of potential misspecification. In particular, we consider factor extraction when assuming different number of factors and different factor dynamics. We focus our analysis on the factors extracted using PC and KFS from the ubiquitous data base of US macroeconomic variables described by McCracken and Ng (2016). Factor extraction procedures have been previously being compared using this data set; see, for example, Poncela and Ruiz (2015) and the references therein. However, as far as we know, the empirical properties of KFS extraction under potential sources of misspecification has not been analysed before when extracting factors from the same data set; see, Aruoba, Diebold and Scotti (2009) for the importance of comparing factor extraction procedures in the context of the same data set.

The rest of the paper is organized as follows. Section 2 briefly describes the representation of DFMs as SSMs and how factor extraction can be performed using KFS based on alternative estimates of the parameters. In Section 3, the factors are extracted from a system of US macroeconomic variables under the assumption of serially uncorrelated idiosyncratic components. We analyse the differences, both in terms of point and interval estimation of factors and forecasting, when factors are extracted using PC and the KFS under different assumptions on the number of factors and their dynamic dependence. Section 4 concludes the paper with our conclusions.

2 Dynamic Factor Models, KFS factor extraction and EM parameter estimation

In this section, we briefly describe how the DFM can be cast as an SSM, how to extract the factor using KFS (assuming that the parameters are known) and how the model parameters can be estimated.

2.1 DFMs as SSMs and KFS

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. 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 static DFM[2]

$$Y_t = \Lambda F_t + \varepsilon_t \tag{1}$$

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

where Λ is the $N \times r$ matrix of factor loadings and F_t, the $r \times 1$ vector of common factors, is assumed to evolve over time following a stationary VAR(p) model given by

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

where u_t is an $r \times 1$ white noise vector with covariance matrix Σ_u. Finally, $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$ are $N \times 1$ vectors representing the common idiosyncratic components, each of them specified as an AR(p_i^*) process as follows

$$\varepsilon_{it} = \theta_{1i}\varepsilon_{it-1} + \theta_{2i}\varepsilon_{it-2} + ... + \theta_{p_i^* i}\varepsilon_{it-p_i^*} + e_{it}. \tag{3}$$

Let $e_t = (e_{1t}, ..., e_{Nt})'$ be the vector of idiosyncratic errors, assumed to be white noise with covariance matrix Σ_e. 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*".

First, we consider the DFM with serially uncorrelated idiosyncratic components, i.e. $\theta_{ji} = 0$ for $i = 1, ..., N$ and $j = 1, ..., p_i^*$. In this case, it is straighforward to write the DFM as a SSM. Assuming that r and p as well as all DFM parameters are known, KFS can be implemented to extract the factors, regardless of the cross-sectional dimension, N; see Poncela, Ruiz and Miranda (forthcoming) for a detailed description.

When the idiosyncratic components are serially correlated and, assuming for simplicity that $p_i^* = 1, \forall i$, the DFM can be reformulated as follows:

$$Y_t = \Theta Y_{t-1} + \begin{bmatrix} \Lambda & -\Theta\Lambda \end{bmatrix} \begin{bmatrix} F_t \\ F_{t-1} \end{bmatrix} + e_t \tag{4}$$

$$\begin{bmatrix} F_t \\ F_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 \\ I_r & 0 \end{bmatrix} \begin{bmatrix} F_{t-1} \\ F_{t-2} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \end{bmatrix}, \tag{5}$$

where I_r is the $r \times r$ identity matrix and Θ is a diagonal $N \times N$ matrix containing the autoregressive parameters θ_i in its main diagonal. Defining the observations as $Y_t - \Theta Y_{t-1}$, the model in (4) and (5) can be directly cast in state space form.[3]

2.2 Estimation of the parameters of DFMs

The KFS factor extraction requires not only a known specification of the DFM (r, p and p^*) but also knowledge of the model parameters in Λ, Φ_1,...,Φ_p, Σ_u, Θ_1,...,Θ_{p^*} and Σ_e. However,

[3]Alternatively, one can deal with the autocorrelation of the idiosyncratic noises by augmenting the state vector by ε_t. The main problem associated with this alternative is that the state vector dimension increases with N and can be unfeasible from a computational point of view for large cross-sectional dimensions. Additionally, the resulting measurement equation will lack measurement noise, which could be an issue when estimating the DFM parameters using the Expectation Maximization (EM) algorithm. If the parameters were known, both formulations lead to the same results when the initialization issues are properly accounted for.

in practice, both the specification and the parameters are unknown and need to be determined and estimated, respectively, before running the KFS algorithms. In this subsection, we describe estimation of the parameters when the model specification is assumed to be known; see the survey by Poncela, Ruiz and Miranda (forthcoming) for a discussion on how to deal with the DFM specification.[4]

First, we consider the DFM in (1) with $\Sigma_\varepsilon = \Sigma_e$ being a diagonal matrix, i.e. the idiosyncratic components are serial and cross-sectionally uncorrelated although they can be cross-sectionally heteroscedastic. In this case, after assuming normality, estimation of the parameters can be carried out by ML with the Kalman filter (KF) used to compute the innovation decomposition form of the Gaussian likelihood, which is given by

$$logL(Y;\Psi) = -\frac{NT}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{T} \nu_t'\Sigma_t^{-1}\nu_t, \tag{6}$$

where $Y = (Y_1, ..., Y_T)$ and Ψ is the vector of parameters to be estimated, namelly the loadings in Λ, the variances in the main diagonal of the covariance matrix of the idiosyncratic noises, $\sigma^2_{\varepsilon_1}, ..., \sigma^2_{\varepsilon_N}$, the autorregressive parameters of the VAR model for the factors in Φ_1,...,Φ_p and the parameters in the covariance matrix Σ_u. Finally, $\nu_t = Y_t - E(Y_t|Y_1, ..., Y_{t-1})$ is the innovation vector and Σ_t is its covariance matrix, and both can be obtained from the Kalman filter. Note that the likelihood decomposition in (6) requires inverting the covariance matrix of the innovations. Using the Woodbury identity, it is possible to see that

$$\Sigma_t^{-1} = \Sigma_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}\Lambda\left(P_{t|t-1}^{-1} + \Lambda'\Sigma_\varepsilon^{-1}\Lambda\right)^{-1}\Lambda'\Sigma_\varepsilon^{-1}, \tag{7}$$

where $P_{t|t-1}$ is the MSE of the one-step-ahead estimates of the underlying factors. Given that Σ_ε is diagonal, obtaining Σ_t^{-1} and, consequently, the expression of the log-likelihood in (6) is straightforward. After imposing the necessary identification restrictions, the log-likelihood can be maximized using, for instance, numerical optimization with, for example, Newton-Raphson algorithms.[5] The parameters are restricted before estimation in order to identify the model. In this paper we follow the suggestion of Harvey (1989) and restrict the parameters in such a way that $\lambda_{i,j} = 0$ for $j > i$ and $i = 1, ..., r$ and $\Sigma_u = I_r$. However, in this very simple DFM, the main hurdle found in the numerical maximization of the likelihood appears when N is extremely large, because the number of parameters to be estimated, $r^2 \times p + N \times (r + 1)$, increases with N.[6]

Alternatively, given that directly optimizing the log-likelihood can be unfeasible when N is large, the ML estimator of the DFM parameters can be obtained by the iterative expectation

[4] We focus on time domain frequentist methods. Bayesian estimators have shown to be very useful in the context of, for example, multilevel DFMs; see Camacho, Pacce and Perez-Quiros (2020) and the references therein.

[5] The optimization algorithm used in the empirical application of this paper is a Quasi-Newton-Raphson algorithm as implemented in the subroutine *Optim* in R.

[6] The proposal by Delle Monache and Petrella (2019) could be useful when the likelihood needs to be evaluated a very large number of times as when using Bayesian procedures based on simulations.

maximization (EM) algorithm proposed by Shumway and Stoffer (1982) and Watson and Engle (1983) for estimation in the context of state-space models. The EM algorithm works iteratively. To simplify the description of the EM algorithm, let us assume that $p = 1$, i.e., the factors are specified as a VAR(1) model.[7] First, starting values for the parameters, $\hat{\Lambda}^{(0)}, \hat{\Sigma}_{\varepsilon}^{(0)}, \hat{\Phi}^{(0)}$ and $\hat{\Sigma}_{u}^{(0)}$, should be obtained. These starting values are usually based on factors and loadings estimated by Principal Components. Denote by $\widetilde{f}^{PC}$, the $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of YY' arranged in decreasing order and $\widetilde{\Lambda}^{PC\prime} = \frac{1}{T}\widetilde{f}^{PC\prime}Y$. By construction, the estimators $\widetilde{f}^{PC}$ and $\widetilde{\Lambda}^{PC}$ are such that $\widetilde{f}^{PC\prime}\widetilde{f}^{PC}/T = I_r$ and $\widetilde{\Lambda}^{PC\prime}\widetilde{\Lambda}^{PC}/N$ is the $r \times r$ diagonal matrix consisting of the first r eigenvalues of the matrix $\frac{1}{TN}YY'$ arranged in decreasing order. Then, the starting parameters for the loadings are $\hat{\Lambda}^{(0)} = \tilde{\Lambda}^{PC}$ while the autoregressive parameters are estimated by the following Ordinary Least Squares (OLS) estimator

$$\hat{\Phi}^{(0)} = \left(\sum_{t=1}^{T} \tilde{f}_{t-1}^{PC} \tilde{f}_{t-1}^{PC\prime} \right)^{-1} \sum_{t=1}^{T} \tilde{f}_{t}^{PC} \tilde{f}_{t-1}^{PC\prime}, \tag{8}$$

and the covariance matrix of the idiosyncratic compoenents is estimated by

$$\hat{\Sigma}_{\varepsilon}^{(0)} = diag \left\{ \frac{1}{T} \sum_{t=1}^{T} \tilde{\varepsilon}_t \tilde{\varepsilon}_t' \right\} \tag{9}$$

where $\tilde{\varepsilon}_t = Y_t - \tilde{\Lambda}^{PC} \tilde{f}_t^{PC}$.

The expectation step consists in running the KFS algorithm with the parameters of the DFM substituted by the starting values above to obtain $f_{t|T}^{(0)}$, $P_{t|T}^{(0)}$ and $C_t^{(0)}$, where $f_{t|T}^{(0)}$ and $P_{t|T}^{(0)}$ are the smoothed estimate of F_t and its corresponding estimated MSE, given by the Kalman smoother, and $C_t^{(0)} = E\left[\left(F_t - f_{t|T}^{(0)} \right) \left(F_{t-1} - f_{t-1|T}^{(0)} \right)' |Y_1, ..., Y_T \right]$ can also be obtained by the Kalman smoother by augmenting the state vector to include F_{t-1}.[8] In the maximization step, the parameters of the DFM are estimated as follows

$$\hat{\Lambda}^{(1)} = \sum_{t=1}^{T} Y_t f_{t|T}^{(0)\prime} \left(\sum_{t=1}^{T} f_{t|T}^{(0)} f_{t|T}^{(0)\prime} + P_{t|T}^{(0)} \right)^{-1}, \tag{10}$$

$$\hat{\Phi}^{(1)} = \left(\sum_{t=1}^{T} f_{t|T}^{(0)} f_{t-1|T}^{(0)\prime} + C_t^{(0)} \right) \left(\sum_{t=1}^{T} f_{t-1|T}^{(0)} f_{t-1|T}^{(0)\prime} + P_{t-1|T}^{(0)} \right)^{-1}, \tag{11}$$

while $\Sigma_{(\varepsilon)}$ is estimated as in (9) with the PC residuals substituted by $\hat{\varepsilon}_t^{(1)} = Y_t - \hat{\Lambda}^{(0)} f_{t|T}^{(0)}$.[9] Recall that, for identification, the parameters of the DFM need to be restricted and, therefore, using

[7] Note that, if the VAR order is $p > 1$, the EM estimator can be easily modified.

[8] The estimates $f_{t|T}^{(0)}$ of this first iteration are usually known as two-step LS estimates of the factors; see Doz, Giannone and Reichlin (2011). In this case, the covariance matrix of u_t is estimated by

$$\hat{\Sigma}_u^{(0)} = \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t \tilde{u}_t'$$

where $\tilde{u}_t = \tilde{f}_t^{PC} - \hat{\Phi}^{(0)} \tilde{f}_{t-1}^{PC}$.

[9] Note that Bai and Li (2016) propose using $\Lambda^{(1)}$ instead of $\Lambda^{(0)}$ to calculate the idiosyncratic residuals.

the restrictions described above, $\Sigma_u = I_r$ does not need to be estimated. Furthermore, denoting by Y the $T \times N$ matrix of observations and by $F^{(S)}$ and $P^{(S)}$, the $T \times r$ matrix of smoothed factors and their corresponding MSE matrix in the steady state, respectively, the restrictions in the loadings can be imposed as follows[10]

$$vec\left(\hat{\Lambda}^{(1)*}\right) = vec\left(\hat{\Lambda}^{(1)}\right) - \left(Rvec\left(\hat{\Lambda}^{(1)}\right) - c\right)'\left[R\left(\left(F^{(S)\prime}F^{(S)} + P^{(S)}\right)^{-1} \otimes I_N\right)R'\right]^{-1} R\left(\left(F^{(S)\prime}F^{(S)} + P^{(S)}\right)^{-1} \otimes I_N\right) \quad (12)$$

where R is an $\frac{r(r-1)}{2} \times Nr$ matrix of zeros and ones of the coefficients of the parameters in the restrictions and c is a $\frac{r(r-1)}{2}$ vector of zeros for the restrictions considered in this case. Consider, for example, that $r = 3$, then the matrix of coefficients of the restrictions is given by the following $3 \times 3N$ matrix

$$R = \begin{bmatrix} 0 & 0 & ... & 0 & 1 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 & 1 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 & 0 & 1 & ... & 0 \end{bmatrix} \quad (13)$$

$$\underbrace{\qquad\qquad}_{N} \quad \underbrace{\qquad\qquad}_{N} \quad \underbrace{\qquad\qquad}_{N}$$

The expectation and maximization steps are iterated until convergence. The parameters of the DFM with serially and cross-sectionally uncorrelated idiosyncratic components can be estimated by ML using the EM algorithm regardless of N; see, among many others, Stock and Watson (1989, 1991) with $N = 4$, Quah and Sargent (1993) with $N = 60$ and Proietti (2011) with $N = 148$.

Several authors have shown the consistency of the Two-step LS and ML estimators of the parameters of the DFM and of the corresponding factors extracted by wrongly considering Σ_ε as diagonal when it is not. First, Doz, Giannone and Reichlin (2011) show that the smoothed factors extracted using the two-step OLS estimates of the parameters are consistent due to the misspecification error vanishing as N and T diverge to infinity. Later, Doz, Giannone and Reichlin (2012) extend the result to the EM estimator with the resulting estimator 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 (2019) who derive the conditions under which the asymptotic distribution can still be used for inference in case of miss-specification.[11]

[10] Note that, if the factors were known, the estimator of Λ can be written in a compact form as $\hat{\Lambda}^{(1)} = Y'F^{(S)}\left(F^{(S)\prime}F^{(S)} + P^{(S)}\right)^{-1}$ and vectorizing it, $vec\left(\hat{\Lambda}^{(1)}\right) = \left(\left(F^{(S)\prime}F^{(S)} + P^{(S)}\right)^{-1} \otimes I_N\right) vec\left(Y'F^{(S)}\right)$.

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

Finally, it is also relevant for empìrical factor extraction to consider DFM with serially correlated idiosyncratic errors, i.e. when $\theta_{ji} \neq 0$ for $i = 1, ..., N$ and $j = 1, ..., p_i^*$. In this case, regardless of N, the parameters of the DFM in equations (4) and (5), with serially correlated idiosyncratic errors specified as independent VAR models, 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).[12]

2.3 Forecasting with DFM

Once the factors have been extracted, it is very popular to obtain out-of-sample forecasts of the variables of interest using diffusion indexes (also known as factor augmented predictive regressions) proposed by Stock and Watson (2002), according to which the one-step-ahead forecast of the i-th variable in the system is given by

$$\hat{y}_{iT+1|T} = \mu + \sum_{j=1}^{q} \delta_j y_{T-j+1} + \sum_{j=1}^{s} B_j' F_{T-j+1} \tag{14}$$

where $B_j = (\beta_{1j}, ..., \beta_{rj})'$ are parameters and F_t are the underlying common factors of the system Y_t. In practice, the parameters of the diffusion indexes in (14) are estimated by LS after substituting the factors by the corresponding estimates. When the factors are extracted by PC, Stock and Watson (2002) show that $\hat{y}_{iT+1|T}$ is consistent for y_{iT+1}. Bai and Ng (2006) show that, if $\frac{\sqrt{T}}{N} \to 0$, the LS estimates if the parameters are $\sqrt{T}$ consistent and asymptotically normal. Furthermore, they show that the conditional mean predicted by the estimated factors is $min[\sqrt{T}, \sqrt{N}]$ consistent and asymptotically normal.[13] Finally, Bai and Ng (2006) also derive the asymptotic distribution of the forecasts of y_{iT+1}, which can be used to construct forecast intervals.[14]

3 Empirical extraction of factors

The forecasting performance of KFS procedures for factor extraction are illustrated, both in-sample and out-of-sample, in the context of the ubiquitous database described in McCracken and

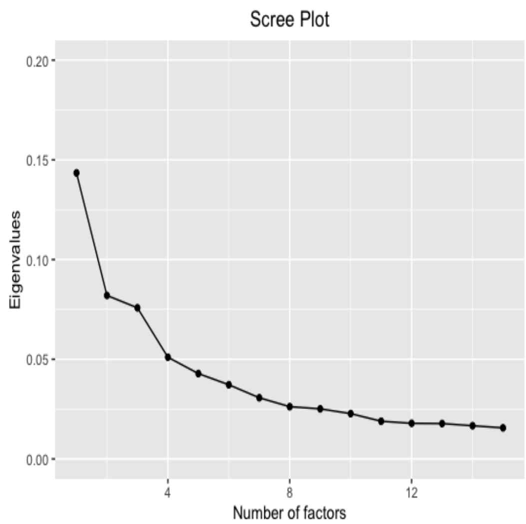
[12] Alternatively, one can deal with serially dependent idiosyncratic noises by considering the SSM with the state vector augmented with the idiosyncratic noises. Banbura and Modugno (2014) modify the steps of the EM algorithm for a general pattern of missing data and show how to model serial correlation by augmenting the state; Alvarez, Camacho and Perez-Quiros (2016) for an empirical application. Augmenting the state may be computationaly expensive when N is very large. Consequently, Jungbacker et al. (2011) show how to reduce the computational burden by alternating between the representation of Reiss and Watson (2010) and the augmented SSM; see Pinheiro, Rua and Dias (2013) for an application. Alternatively, one can also use Bai and Li (2016) to estimate the DFM with serially correlated errors.

[13] As far as we know there are not corresponding results when the factors are extracted using the Kalman filter.

[14] The asymptotic variance is obtained using the usual formulae for regression models, adding an additional term due to the estimation of the factors. However, if N is large, this additional term dissapears.

Ng (2016) that consists of $N = 128$ variables observed monthly from January 1983 up to and including December 2020, with a total of 444 observations *per series*.[15] We consider forecasting in the context of stationary DFM. With this purpose, previous to its analysis, the data are transformed to stationarity and outliers and missing observations are dealt with as described in McCracken and Ng (2016). Then, all variables in the system are centered and standardized. The sample period is split into an estimation period from January 1983 to December 2016 ($T = 396$) and an out-of-sample forecast period, from January 2017 to December 2020 ($H = 48$). The focus of prediction are Industrial Production (IP) Index and inflation first differences; see applications in Quah and Sargent (1993), Bai and Ng (2008), Alvarez, Camacho and Perez-Quiros (2016) and McCracken and Ng (2016), for forecasts of these same variables.

Figure 1: Scree plot of the eigenvalues of the covariance matrix of the US macroeconomic data set.



To determine the number of static factors, we inspect the scree plot proposed by Cattell (1966), which appears in Figure 1; see, for example, Hindrayanto, Koopman and de Winter (2016) who also look at the scree plot to determine r. The message from the scree plot is not clear with the presence of one factor being obvious but not other factors appearing in a neat way. Alternatively, by using the second criteria proposed by Alessi, Barigozzi and Capasso (2010), the number of factors is determined to be $r = 7$, which is also the number determined using the criteria proposed by Bai and Ng (2002) by many authors, as, for example, Stock and Watson

[15]This data base is an update version of the data base used by Stock and Watson (2002, 2012). Although data are available from January 1959, there is a generalized consensus about the presence of a structural break in 1982 due to the change of policy rule of the Federal Reserve which swichted from targeting non-borrowed reserve to targeting the federal funds rate; see, for example, Hallin and Liska (2007) and Luciani (2015). To avoid the problems associated to the presence of structural breaks when determining the number of factors, we analyse the data set observed from 1983; see Breitung and Eickmeier (2011) who show that, in the presence of a structural break, the number of factors is overestimated and the factor loadings are inconsistently estimated, and Chen, Dolado and Gonzalo (2014) for a test for breaks in DFMs.

(2005), Bai and Ng (2007), Pinheiro, Rua and Dias (2013), McCracken and Ng (2016), Kristensen (2017) and Despuis and Doz (2020), analysing the same data set over a different period of time. Furthermore, the first criteria of Alessi, Barigozzi and Capasso (2010) determines, $r = 5$; see also Poncela and Ruiz (2015) who chose $r = 4$. Moreover, the criteria proposed by Onatski (2010) determines $r = 1$; see, for example, Alvarez, Camacho and Perez-Quiros (2016) who consider the case of $r = 1$ factors in this data set. Therefore, there is no agreement about the number of factors. Is this important? In order to analyse the effect of the number of factors on the conclusions, we carry out the analysis by assuming the three most likely scenarios given the scree plot, namely, $r = 1$ and $r = 3$, and analyse the implications in the estimation of the first factor and on the forecasts of industrial production and inflation.

3.1 Extracting one single factor

After extracting a single factor by PC, its correlogram and partial correlogram suggest that the factor could be represented by an AR(3) model. Should we worry about an adequate specification of the dynamic dependence of the factors? In this subsection, in order to analyse the effect of the dynamic dependence assumed for the factors on the estimated factors and on the corresponding forecasts, we estimate the exact DFM with $r = 1$ assuming either that $p = 1$ or $p = 3$. For each of these cases, the parameters of the DFM are estimated by TS-LS and ML, the latter maximized either using numerical optimization or by the EM algorithm. Table 1 reports some summary of the results. In particular, it reports $\sum_{i=1}^{N} \hat{\lambda}_i^2$, $\sum_{i=1}^{N} \hat{\sigma}_{ei}^2$ as well as the estimated autoregressive parameters and the MSE of the smoothed factor.[16] First of all, we can observe that the sums of squared loadings and of idiosyncratic variances and MSE($\hat{f}_{t|T}$) are the same regardless of whether the factor is assumed to be AR(1) or AR(3) or whether we estimate the model parameters by ML either using EM or numerically maximizing the log-likelihood. When the Kalman filter is run with the parameters estimated by TS-LS, we can observe that the sum of squared loadings is slightly larger and the sum of idiosyncratic variances is slightly smaller. As a consequence, the Kalman (steady) MSE of the smoothed factor, $\hat{f}_{t|T}$, is smaller with an apparent increase in precision as compared with the steady MSE obtained when the parameters are estimated by ML. In any case, it is remarkable that the MSE of the PC extracted factor estimated as proposed in Bai (2003) is 0.01, approximately 5 times smaller than that obtained when the factors are extracted using the Kalman filter with ML estimates of the parameters.

[16]Note that for the results to be comparable, all estimated loadings and factors have been rotated to the same base as those estimated by PC; see Poncela and Ruiz (2015) for the rotation. This rotation is need because of the different restrictions used for identification by PC and ML. While under PC the sample variance of the factors is assumed to be one, the restriction imposed when estimating by ML is that $\sigma_u^2 = 1$. The ML estimate of the autoregressive parameter, $\hat{\phi}_1 = 0.87$ implies that the factor has variance 4.11. Therefore, the ML weights should be divided by 2.03 for the variance of the common component to be the same.

Furthermore, the implications of the estimation method and specification assumed for the factor are also clear when estimating its dynamic dependence. Consider first the estimated parameters in the model with $p = 1$. The ML estimate of the autoregressive parameter (regardless of whether it is estimated maximizing numerically the likelihood, 0.87, or using the EM algorithm, 0.85) is larger than that based on PC, 0.78. Furthermore, note that the slight differences between the ML results obtained when the likelihood is maximized numerically or when using the EM algorithm dissappear when $p = 3$ is assumed. It seems that when the "true" log-likelihood is maximized its value at the maximum is the same regardless of the procedure used for its maximization. Finally, it is important to look at the roots implied by the estimated parameters of the AR(3) model. When the parameters are estimated based on the PC factors, the roots are 0.94 and $-0.30 \pm 0.35i$ while, if they are estimated by ML, the roots are 0.95 and $-0.27 \pm 0.28i$. In both cases, there is a cyclical behaviour of the factor which has a largest real root when the parameters are estimated by ML. In any case, the persistence of this real root is clearly larger than when an AR(1) model is assumed for the factor. These differences in the estimated persistence and number of lags of the factor may have implications for forecasting, mainly in periodos of changing points because the forecasts adapt quicker if the number of lags is smaller. Finally, Table 1, which also reports the value of the log-likelihood at the maximum for the ML estimates, shows that, although there are not significant differences between the log-likelihood values obtained when the maximization is based on EM or numerical optimization, the difference between the log-likelihood when $p = 1$ and $p = 3$ is significant, according to the log-likelihood ratio test.

Figure 2 plots the loadings estimated by PC and ML, in the latter case by using both numerical optimization and EM. As an illustration, Figure 3 plots the factors together with their corresponding 95% confidence intervals obtained by the KFS based on the PC and EM parameter estimates reported in Table 1, together with their 95% confidence intervals.[17] As before, the EM estimated factors have been rotated to be in the same space as those estimated by PC. More importantly, the intervals constructed using ML parameter estimates are clearly larger than those obtained using PC parameters; see also Poncela and Ruiz (2005) who conclude that the asymptotic RMSEs obtained from the asymptotic distribution of the PC factors are unreallistically small.[18]

To analyse whether the differences in the estimation of the parameters above have implications in forecasting, we obtain out-of-sample one-step-ahead forecasts of IP and inflation first differences from January 2017 up to December 2020 using the factor-augmented predictive regressions in equation (14) with $q = s = 4$.[19] Table 2 reports the estimates of the parameters of

[17]The confidence intervals of the PC factors are based on the asymptotic distribution derived by Bai (2003).

[18]The intervals for the factors could be more realistic if the asymptotic MSE is modified by subsampling as proposed by Maldonado and Ruiz (forthcoming).

[19]Residual diagnosis analysis of the factor augmented predictive regressions is available upon request.

Table 1: Parameter estimates of Static-DFMs: Two-step Least Squares (TS-LS); Maximum Likelihood with numerical optimization (ML-NO); Maximum Likelihood with EM (ML-EM).

	TS-LS	ML-NO	ML-EM	TS-LS	ML-NO	ML-EM
		$p=1$			$p=3$	
$r=1$						
$\sum_{i=1}^{N}\hat{\lambda}_i^2$	18.09	17.72	17.75	18.09	17.75	17.75
$\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2$	109.87	110.97	110.87	109.86	110.88	110.86
$\hat{\phi}_1$	0.78	0.87	0.85	0.34	0.42	0.41
$\hat{\phi}_2$	-	-	-	0.35	0.36	0.36
$\hat{\phi}_3$	-	-	-	0.20	0.14	0.15
MSE($\hat{F}_t$)	0.030	0.038	0.044	0.03	0.05	0.05
log-Lik	-	-69475.1	-69485.94	-	-69437.7	-69438.5
$r=3$						
$\sum_{i=1}^{N}\hat{\lambda}_i^2$	18.09	-	17.90	18.09	-	17.90
$\sum_{i=1}^{N}\hat{\sigma}_{\varepsilon i}^2$	89.52		92.67	89.52		92.65
$\hat{\phi}_1$	0.78	-	0.86	0.24	-	0.85
$\hat{\phi}_2$	-	-	-	0.49	-	0.31
$\hat{\phi}_3$	-	-	-	0.15	-	-0.27
MSE($\hat{F}_t$)	0.022		0.022	0.024	-	0.027
log-Lik	-	-	-61211	-	-	-61123.9

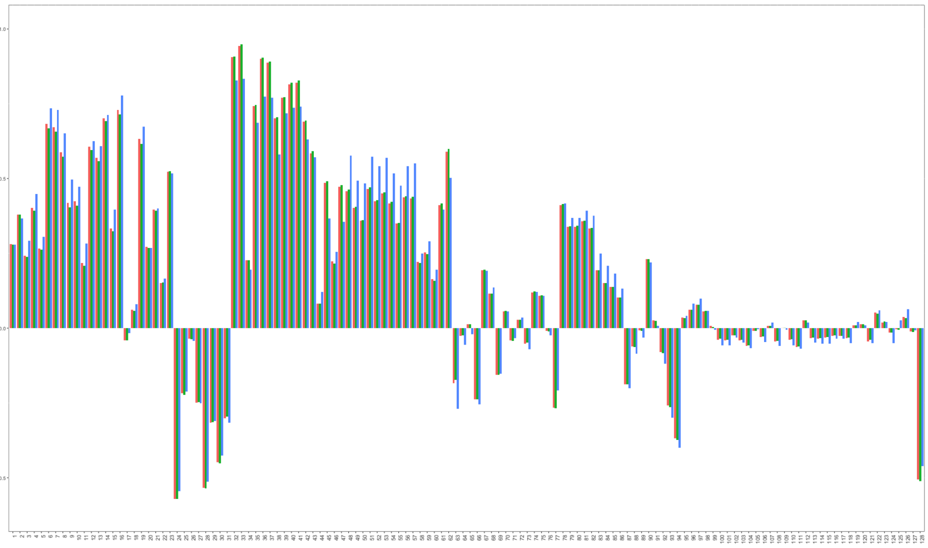
these regressions obtained in-sample for the IP growth together with their corresponding p-values obtained under the assumption of homoscedastic forecast errors, $e_{it} = y_{it} - \hat{y}_{it|t-1}$.[20]

Finally, using the estimated factor-augmented regressions reported in Table 2 and the filtered factors obtained from the Kalman filter run in the out-of-sample period, we obtain one-step-ahead forecasts of IP variations from January 2017 to December 2020. Table 3 reports the variance of the one-step-ahead forecasts of the IP variations. It also reports the empirical Mean Square Forecast Errors (MSFEs) and the empirical coverages of the 70% forecast intervals computed both with the forecasts obtained until December 2019 and until December 2020.[21] Note that in the latter case, we are incorporating in the analysis the forecasts obtained during the turbulent times due to the recession induced by the COVID-19 pandemic. However, in the former case,

[20]The results for inflation are not reported as, regardless of the particular procedure used to estimate the factor, the parameters associated with the factor and its lags in the regressions are not significant.

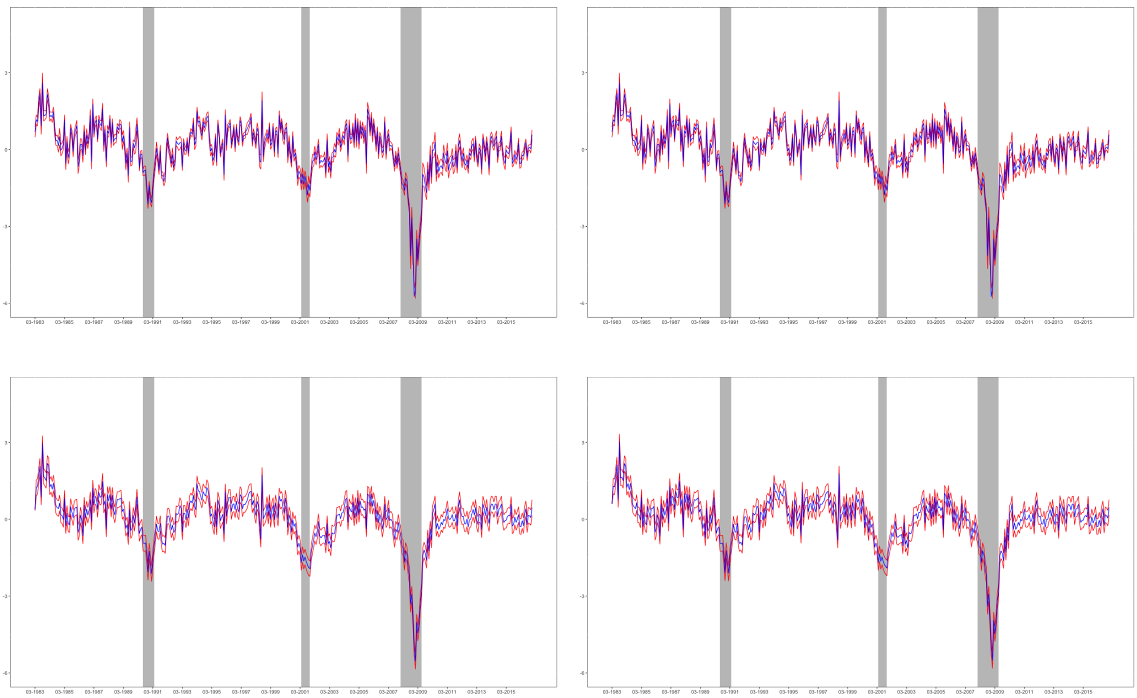
[21]Note that these quantities are merely descriptive as they are based on just 36 forecasts in the former case and 48 in the latter. Results for intervals with 95% nominal coverage are also available upon request.

Figure 2: Factor loadings estimated for the set of macroeconomic variables using: i) PC (blue bars) and ii) ML with numerical optimization, ML-NO (green bars) and with EM, ML-EM (orange bars).



the forecasts are obtained in a "normal" time in the evolution of the variables. First of all, Table 3 shows that, even if the differences between the in-sample estimated factors are very minor, the performance of one-step-forecasts can be quite different. The procedure used to extract the factors and the estimator of the DFM parameters when the factors are extracted using the Kalman filter and smoother is relevant for the out-of-sample one-step-ahead forecasts performance. Note that the differences are more obvious when there are extraordinary movements in the series, as those observed during the COVID-19 crisis; see Figure 4. This is so for industrial production. Notice that this first factor captures quite well the business cycle. When we take into account 2020, the differences in the MSFE obtained with PC are striking (for instance, the out of sample MSE of the PC extracted factors is more than twice that of the ML extracted factors). However, removing the year 2020 gives very different numerical results. First, the magnitude of the MSFEs is considerably reduced. In particular, PC induced MSE is around 10 times smaller when excluding 2020. Nevertheless, the PC extracted factors still renders out of sample MSFEs around 20% larger than those of ML methods. Regarding the length of the AR polynomial of the common factor for IP, notice that we always obtain smaller MSEs for $p = 1$, that is, the shorter the memory of the common factor, the smaller the MSFE. The picture for inflation is a little bit different. The MSEs without the year 2020 are also smaller for all methods, as expected, but those induced by PC with 1 factor are still the largest ones. However, figures are closer among them for the three methods. Finally, for inflation, the MSFE is not always smaller for $p = 1$ than for $p = 3$, as expected, given the different induced fluctuations of the pandemic crisis in prices than in real economic activity.

Figure 3: A single factor (blue) extracted from the set of macroeconomic variables using PC (first row) and KFS with EM estimates of the parameters (second row). The first column plots the smoothed factor extracted assuming an AR(1) dependence while in the second column the factor is assumed to be an AR(3) process. The red lines represent the corresponding 95% confidence intervals.



3.2 Extracting three factors

As mentioned above, the number of common factors in the system is unknown and should be determined beforehand. In the case of the data set considered in this paper, r is not clearly determined with a variety of possibilities depending on the criteria used. In this subsection, we analyse whether the potential misspecification about r has implications for factor extraction and forecasting. We assume that $r = 3$ and analyse the differences observed in the first factor with respect to that extracted above and in forecasting industrial production and inflation.

Table 1 reports a summary of the estimation results.[22] Consider first the case with $p = 1$. Comparing the results reported in Table 1 with those obtained when r was assumed to be one, we can observe that the only difference is that, obviously, the sum of idiosyncatic variances is now smaller and, consequently, the MSE of the extracted factors is reduced to half.[23] It is also remarkable that the maximum of the log-likelihood reported in Table 1 is significantly larger when $r = 3$ than when $r = 1$. Similarly, when we assume that $r = 3$ and $p = 3$ and the parameters are estimated by ST-LS, we can observe that the estimation results are very similar to those obtained

[22]In this case, it is unfeasible to maximize numerically the likelihood. Consequently, only the ML results for EM are reported.

[23]Note, however, that the MSE still doubles that obtained for the PC factors, which is 0.009.

Table 2: Parameter estimates of factor-augmented predictive regressions for IP based on factors estimated using: PC; Two-step Least Squares (TS-LS); Maximum Likelihood with EM (ML-EM). p-values in parenthesis.

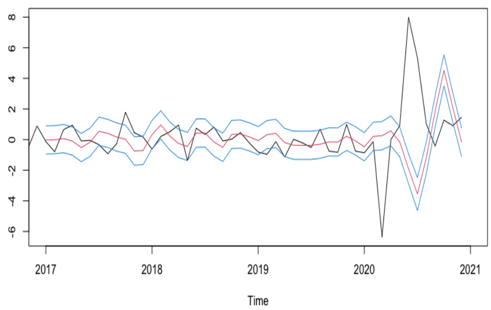
	$r=1$					$r=3$				
	PC	$p=1$		$p=3$		PC	$p=1$		$p=3$	
		TS-LS	ML-EM	TS-LS	ML-EM		TS-LS	ML-EM	TS-LS	ML-EM
μ	-0.006	-0.006	-0.006	-0.005	-0.005	-0.006	-0.005	-0.002	-0.005	-0.002
	(0.88)	(0.90)	(0.89)	(0.91)	(0.90)	(0.89)	(0.91)	(0.95)	(0.91)	(0.97)
δ_1	-0.216	-0.256	-0.224	-0.224	-0.191	-0.179	-0.262	-0.128	-0.231	-0.102
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.03)	(0.01)	(0.07)	(0.01)	(0.15)
δ_2	-0.094	-0.096	-0.032	-0.113	-0.060	-0.083	-0.125	-0.030	-0.131	-0.046
	(0.23)	(0.23)	(0.64)	(0.16)	(0.39)	(0.32)	(0.17)	(0.66)	(0.15)	(0.50)
δ_3	0.300	0.306	0.271	0.298	0.262	0.238	0.252	0.254	0.220	0.243
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)
δ_4	0.408	0.382	0.288	0.381	0.293	0.370	0.427	0.283	0.437	0.287
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
β_{11}	0.486	0.584	0.671	0.521	0.580	0.322	0.526	0.783	0.421	0.817
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)
β_{12}	0.486	0.472	0.359	0.519	0.471	0.467	0.572	0.607	0.705	0.643
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
β_{13}	-0.220	-0.282	-0.321	-0.267	-0.305	-0.177	-0.433	-1.003	-0.480	-1.097
	(0.10)	(0.05)	(0.03)	(0.05)	(0.03)	(0.24)	(0.02)	(0.00)	(0.01)	(0.00)
β_{14}	-0.603	-0.586	-0.508	-0.588	-0.544	-0.429	-0.441	-0.270	-0.419	-0.239
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.01)	(0.17)	(0.01)	(0.23)
σ_ε^2	0.723	0.725	0.726	0.731	0.698	0.719	0.718	0.708	0.716	0.698
R_A^2	0.27	0.27	0.27	0.27	0.30	0.28	0.28	0.29	0.28	0.30

when we assumed that $r = 1$ and $p = 3$. Looking at the estimated dynamics of the first factor, we can observe that, they are very similar to those estimated when assuming that $r = 1$.[24] In particular, when the parameters are estimated by TS-LS, the roots of the characteristic equation are 0.935 and $0.348 \pm 0.20i$, very close to those estimated above. However, the results are rather different when the parameters are estimated by ML. In this case, the roots are 0.807, -0.557 and 0.6, rather different from those obtained when the parameters are estimated by TS-LS and when assuming that $r = 1$.

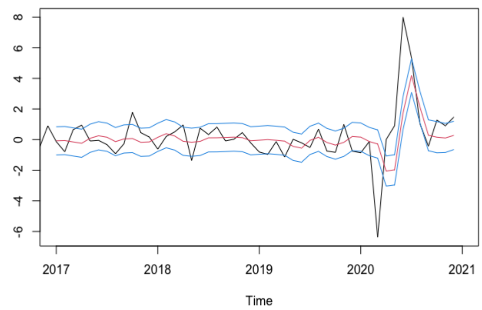
Finally, it is important to mention that the sample pairwise correlations between the first factor estimated in the different specifications and estimators considered range from 0.96 to 1.00. The minimum correlation, 0.96, is obtained when the factor is extracted assuming that $r = 1$ and $p = 1$ and estimating the parameters by ML and when it is assumed that $r = 3$ and $p = 3$ and the parameters of the DFM used to extract the factors are estimated by TS-LS. On the other hand, the maximum correlation, 1.00, is obtained when it is assumed that $r = 1$ and $p = 3$ and the parameters are estimated by ML either maximizing numerically the likelihood or using

[24]The off-diagonal elements of the estimated autoregressive matrices are fixed to zero as, in practice, they are not far from zero.

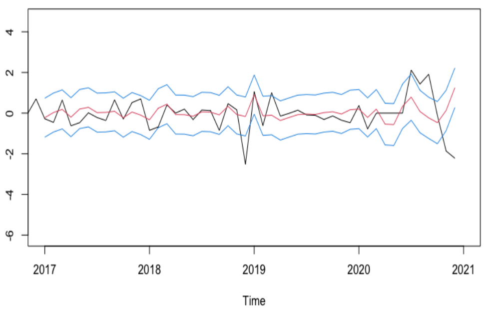
Figure 4: Out-of-sample forecasts of IP (first row) and inflation (second row) together with the corresponding 70% confidence intervals.



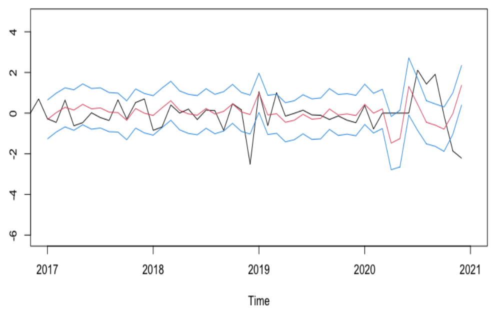
(a) IP forecasts with one PC factor and p=1



(b) IP forecasts with one EM factor and p=1



(c) Inflation rate forecasts with one EM factor and p=1



(d) Inflation rate forecasts with three EM factors and p=1

the EM algorithm; see also Lewis *et al.* (2020), who conclude that the factors are robust to whether PC or KFS is implemented for factor extraction when constructing a weekly index of real activity (EWI) based on $N = 10$ variables for USA.

When instead of one factor, we estimate the predictive regressions using three factors, the results are similar for IP forecasts; see Table 2, which only reports the parameter estimates for the first factor, as the second and third factors are not significant in the predictive regression of IP. Note that McCracken and Ng (2016) interpret the first common factor (extracted using PC) as a real activity/employment factor. They find that although in initial forecasting samples more than one factor seems to have predictive information over IP, when we move to latter samples only the first common factors seems to retain its predictive information. However, when forecasting inflation, the second lag of the three factors is significant. McCracken and Ng (2016) interpret the third common factor as an inflation factor while the second common factor was dominated by forward-looking variables such as term interest rate spreads and inventories. They show that, in the sample period they consider, the first common factor does not have any predictive content for forecasting inflation in later times. According to our results, forecasts of inflation based on models with $r = 1$ or $r = 3$ are different; see Figure 4. On top of the noticeable differences

Table 3: One-step-ahead out-of-sample forecasts of first differences of industrial production and inflation. Factor-augmented predictive regressions based on factors estimated by: PC; Two-step Least Squares (TS-LS); Maximum Likelihood with EM (ML-EM).

	St. error	MSE (2019)	MSE (2020)	Cov. 70% (2019)	Cov. 70% (2020)	St. error	MSE (2019)	MSE (2020)	Cov. 70% (2019)	Cov. 75% (2020)
			$p = 1$				$p = 3$			
Industrial production growth										
$r = 1$										
PC	0.88	0.61	5.62	72.22%	60.42%					
TS-LS	0.87	0.51	3.16	80.56%	64.58%	0.88	0.51	3.24	80.56%	64.58%
ML-EM	0.88	0.50	2.37	77.78%	64.58%	0.89	0.51	2.51	77.78%	66.67%
$r = 3$										
PC	0.87	0.59	5.09	75%	62.5%					
TS-LS	0.87	0.51	3.17	77.78%	64.58%	0.87	0.52	3.25	72.22%	60.42%
ML-EM	0.94	0.55	2.95	75%	64.58%	0.93	0.55	3.16	75%	62.5%
Inflation rate										
$r = 1$										
PC	0.92	0.34	0.91	94.44%	85.42%					
TS-LS	0.94	0.35	0.84	91.67%	83.33%	0.94	0.35	0.84	91.67%	83.33%
ML-EM	0.95	0.34	0.78	94.44%	85.42%	0.95	0.34	0.77	94.44%	85.42%
$r = 3$										
PC	0.90	0.33	0.89	94.44%	85.42%					
TS-LS	0.93	0.35	0.90	94.44%	85.42	0.93	0.36	0.89	94.44%	85.42%
ML-EM	1.034	0.38	1.028	94.44%	83.33	0.93	0.38	0.89	94.44%	85.42%

between results including pre-COVID times and those that do not include them that we can also observe with three factor models, notice that both for industrial production and inflation including more factors does not necessarily translate into smaller out of sample MSEs. Indeed, in occasions those are larger than the corresponding ones from one factor predictive regressions.

4 Conclusions

Given the model specification, the main difference between factor estimates obtained using PC or two-step LS versus ML (either numerical optimization or EM) appears in the factor dynamics. In the particular US macroeconomic data set analysed in this paper, the largest autoregressive root is closer to one when the model parameters are estimated by ML. This stronger persistence has implications in forecasting. Furthermore, the sum of squared loadings (idiosyncratic variances) is larger (smaller) when the parameters are estimated by TS-LS than when estimated by ML and, consequently, the confidence intervals for the factors are larger (and more realistic?) when the parameters are estimated by ML. Assuming a larger number of factors imply reducing the sum of idiosyncratic variances and, consequently, decreasing the MSE of the factors extracted using KF. While the likelihood-ratio tests favour specifications with more factors and more lags, the forecasting results point towards the importance of the factor extraction procedure and the parameter estimator. Therefore, it seems that, in practice, one should analyse the sensitivity of alternative specifications and factor extraction procedures case by case. Although in sample results might render the conclusion that the differences in the estimation method are not so important, the picture changes when we focus on out of sample results. On the one hand, more factors does not necessarily mean smaller out of sample MSFEs. On the other hand, small differences in the in-sample results, might render noticeable differences in the out-of-sample

MSFEs. However, this is case dependent, specially in the vicinity of the turning points. Moreover, we do not obtain the same results for industrial production and inflation, leaving the issue as an empirical matter. In any case, answering the question in the title of this work, a careful specification of the DFM before factor extraction could be important in terms of forecasting.

References

[1] 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.

[2] Alvarez, A., M. Camacho and G. Perez-Quiros (2016), Aggregate versus disaggregate informatio in dynamic factor models, *International Journal of Forecasting*, 32, 680-694.

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

[4] Bai, J. (2003), Inferential theory for factor models of large dimensions, *Econometrica*, 71(1), 135-171.

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

[6] Bai, J. and S. Ng (2006), Confidence intervals for diffusion index forecasts and inference for factor augmented regressions, *Econometrica*, 74(4), 1133-1150.

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

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

[9] Bai, J. and S. Ng (2008b), Forecasting economic time series with targeted predictors, *Journal of Econometrics*, 146, 304-317.

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

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

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

[13] Beck, G., K. Hubrich and M. Marcellino (2009), Regional inflation dynamics within and across euro area countries and comparison with the US, *Economic Policy*, 24, 141-184.

[14] Beck, G., K. Hubrich and M. Marcellino (2016), On the importance of sectoral and regional shocks for price setting, *Journal of Applied Econometrics*, 31(7), 1234-1253.

[15] Breitung, J. and Eickmeier, S. (2011), Testing for structural breaks in dynamic factor models, *Journal of Econometrics*, 163, 71-74.

[16] Camacho, M., M. Pacce and G. Perez-Quiros (2020), Spillover effects in international business cycles, Documentos de Trabajo no. 2034, Banco de España.

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

[18] Chen, L., J.J. Dolado and J. Gonzalo (2014), Detecting big structural breaks in large factor models, *Journal of Econometrics*, 180(1), 30-48.

[19] Chen, L., J.J. Dolado and J. Gonzalo (2021), Quantile factor models, *Econometrica*, 89(2), 875-910.

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

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

[22] Despuis, T. and C. Doz (2020), Another look into the factor blackbox: factors interpretation and structural (in)stability, halshs-02235543v2.

[23] 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, 188-205.

[24] 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.

[25] Forni, M. and L. Reichlin (2001), Federal policies and local economies: Europe and the US, *European Economic Review*, 45, 109-134.

[26] 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.

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

[28] 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.

[29] 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.

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

[31] Kristensen, J.T. (2017), Diffusion indexes with sparse loadings, *Journal of Business & Economic Statistics*, 35(3), 434-451.

[32] Lewis, D.J., K. Mertens, J.H. Stock and M. Trivedi (2020), Measuring real activity using a weekly economic index, Federal Reserve Bank of New York Staff reports, no. 920.

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

[34] Maldonado, J.V. and E. Ruiz (forthcoming), Accurate confidence regions for principal components factors, *Oxford Bulletin of Economics and Statistics.*

[35] McCracken, M.W. and S. Ng (2016), FRED-MD: A monthly database for macroeconomic research, *Journal of Business & Economic Statistics*, 34(4), 574-589.

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

[37] 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.

[38] Poncela, P. and E. Ruiz (2015), Small versus big-data factor extraction in Dynamic factor models: An Empirical assessment, in E. Hillebrand and S.J. Koopman (eds.), Advanced Econometrics, Vol. 35, Dynamic Factor Models, Emerald Publishing Group.

[39] Poncela, P., E. Ruiz and K. Miranda (forthcoming), Factor extraction using Kalman filter and smoothing: This is not just another survey, *International Journal of Forecasting.*

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

[41] Quah, D. and T.J. Sargent (1993), A dynamic index model for large cross sections, *Business Cycles, Indicators, and Forecasting*, NBER.

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

[43] 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.

[44] 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.

[45] 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.

[46] Stock, J.H. and Watson, M.W. (2002), Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics*, 20, 147-162.

[47] 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.

[48] 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.

[49] Stock, J.H. and Watson, M.W. (2012), Disentangling the channels of the 2007-09 recession, *Brooking Papers on Economic Activity*, Spring, 81-135.

[50] 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.

[51] 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, 385-400.