Working paper
2022-11
Statistics and Econometrics
ISSN 2387-0303

Ignoring cross-correlated idiosyncratic components when extracting factors in dynamic factor models

Diego Fresoli, Pilar Poncela, Esther Ruiz

uc3m | Universidad Carlos III de Madrid

Ignoring cross-correlated idiosyncratic components when extracting factors in dynamic factor models*

Diego Fresoli

Dept. of Econ. Analysis: Quantitative Economics, Univ. Autónoma de Madrid, Spain

Pilar Poncela

Dept. of Econ. Analysis: Quantitative Economics, Univ. Autónoma de Madrid, Spain

and

Esther Ruiz†

Department of Statistics, Universidad Carlos III Madrid, Spain

December 24, 2022

Abstract

In economics, Principal Components, its generalized version that takes into account heteroscedasticity, and Kalman filter and smoothing procedures are among the most popular procedures for factor extraction in the context of Dynamic Factor Models. This paper analyses the consequences on point and interval factor estimation of using these procedures when the idiosyncratic components are wrongly assumed to be cross-sectionally uncorrelated. We show that not taking into account the presence of cross-sectional dependence increases the uncertainty of point estimates of the factors. Furthermore, the Mean Square Errors computed using the usual expressions based on asymptotic approximations, are underestimated and may lead to prediction intervals with extremely low coverages.

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

JEL Codes: C32, C38, C55

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

†Corresponding author: Dpto. de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain, email: ortega@est-econ.uc3m.es, Tel.: +34 916249851

1 Introduction

Dynamic Factor Models (DFMs) are popular to represent co-movements within large systems of economic variables; see Stock and Watson (2017) for their importance. Two main types of procedures are available for factor extraction. First, in many applications, factors are extracted using Principal Components (PC), with the Generalized Least Squares PC (GPC) estimator improving efficiency in the presence of cross-sectional heteroscedascity. However, in extracting the factors using GPC, there is not obvious way to improve efficiency by exploiting the covariances of the idiosyncratic components. Consequently, they are often wrongly assumed to be cross-sectionally uncorrelated, which may affect both the Mean Square Errors (MSE) of point estimates and the interval estimation of the factors, with the latter being of interest in situations in which the factors have a direct interpretation; see, for example, González-Rivera, Ruiz and Maldonado (2019) and González-Rivera, Rodríguez-Caballero and Ruiz (2021), who identify the factors with underlying economic conditions and use their uncertainty to construct economic scenarios. The factor uncertainty is also relevant when computing prediction intervals in the context of diffusion models in which the factors summarize the information contained in a large number of predictors; see, for example, Bai and Ng (2006).

Alternatively, one can use Kalman Filter and Smoothing (KFS) to extract the factors; see Poncela, Ruiz and Miranda (2021) for a survey. To be efficient, KFS procedures require full specification of the common and idiosyncratic components, which, when the cross-sectional dimension is large, depends on a large number of parameters. Consequently, it is common to assuming cross-sectionally uncorrelated idiosyncratic components, introducing potential misspecification that is not reflected in the model-based inference.

This paper analyses the consequences on point and interval factor estimation of using PC and KFS factor extraction procedures when the idiosyncratic components are wrongly assumed to be cross-sectionally uncorrelated. This analysis is carried out by running exhaustive Monte Carlo experiments, which cover situations of interest in the empirical applications.

In the following, Section 2 describes PC and KFS factor extraction, Section 3 summarizes the main conclusions from our simulations and Section 4 concludes.

2 Dynamic Factor Models: PC and KFS 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

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

where Λ is the $N \times r$ matrix of factor loadings, with r being the number of factors, assumed to be known. F_t is the $r \times 1$ vector of common factors, given by the following stationary VAR(1) model

$$F_t = \Phi F_{t-1} + u_t, \tag{2}$$

where u_t is an $r \times 1$ white noise vector with full rank covariance matrix Σ_u. Finally, $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$, the $N \times 1$ vector of idiosyncratic components, is a white noise with covariance matrix Σ_ε.

2.1 Principal Components

Define $Y = (Y_1, Y_2, ..., Y_T)'$ as the $T \times N$ matrix of observations and $F = (F_1, F_2, ..., F_T)'$ as the $T \times r$ matrix of factors. From (1) and using the normalization $\frac{F'F}{T} = I_r$, the estimated PC factors, $\widetilde{f}^{PC}$, are $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of YY' arranged in decreasing order, and $\widetilde{\Lambda}^{PC\prime} = \frac{1}{T}\widetilde{f}^{PC\prime}Y$. If the idiosyncratic cross-sectional correlations are weak and the factors are pervasive, Bai (2003) establishes consistency of the space spanned by factors when both N and T tend simultaneously to infinity. Furthermore, if $\frac{F'F}{T} = I$, and further $\frac{\sqrt{N}}{T} \to 0$, Bai (2003) derives the following asymptotic distribution

$$\sqrt{N}\left(\tilde{f}_t^{PC} - F_t\right) \xrightarrow{d} N\left(0, \Sigma_\Lambda^{-1}\Gamma_t\Sigma_\Lambda^{-1}\right) \tag{3}$$

where $\Sigma_\Lambda = lim_{N\to\infty}\frac{1}{N}\Lambda'\Lambda$ is a positive definite matrix and $\Gamma_t = lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \lambda_i\lambda_j' E(\varepsilon_{it}\varepsilon_{jt})$, with λ_i' being the i'th row of Λ. If the idiosyncratic noises are serially uncorrelated, the limiting distributions in (3) are asymptotically independent across t. From (3), the asymptotic approximation of the MSE of $\tilde{f}_t^{PC}$ can be estimated as follows

$$\widetilde{MSE}_t^{PC} = \left(\frac{\tilde{\Lambda}^{PC\prime}\tilde{\Lambda}^{PC}}{N}\right)^{-1}\frac{\tilde{\Gamma}_t}{N}\left(\frac{\tilde{\Lambda}^{PC\prime}\tilde{\Lambda}^{PC}}{N}\right)^{-1} \tag{4}$$

where $\tilde{\Gamma}_t$ is an estimate of Γ_t. According to Bai and Ng (2006), $\tilde{\Gamma}_t$ can be obtained using the following heteroscedasticity robust (HR) matrix

$$\widetilde{\Gamma}_t^{HR} = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i^{PC}\tilde{\lambda}_i^{PC\prime}\tilde{\varepsilon}_{it}^{PC2} \tag{5}$$

with $\tilde{\varepsilon}_{it}^{PC} = Y_{it} - \tilde{\lambda}_i^{PC\prime}\widetilde{f}_t^{PC}$. At each moment of time t, prediction regions for the factors can be constructed based on the asymptotic normality and using $\widetilde{MSE}_t^{PC}$. In the presence of cross-sectional dependence, $\widetilde{\Gamma}_t^{HR}$ is inconsistent, so that the associated prediction intervals are invalid; see, for example, Kim (2022). In this case, Bai and Ng (2006) also propose the following estimator of Γ_t, which is robust not only to heteroscedasticity but also to cross-sectional dependence

$$\widetilde{\Gamma}_t^{BN} = \frac{1}{N_{sub}}\sum_{i=1}^{N_{sub}}\sum_{j=1}^{N_{sub}}\tilde{\lambda}_i^{PC}\tilde{\lambda}_j^{PC\prime}\frac{1}{T}\sum_{t=1}^T\tilde{\varepsilon}_{it}^{PC}\tilde{\varepsilon}_{jt}^{PC}, \tag{6}$$

where $\frac{N_{sub}}{min[N,T]} \to 0$ and N_{sub} is the cardinality of a subset of the cross-sectional units. In practice, researchers should select N_{sub} variables in such a way that $\widetilde{\Gamma}_t^{BN}$ replicates the overall dependence structure of the idiosyncratic components. The simulation results in Kim (2022) show that the performance of $\widetilde{\Gamma}_t^{BN}$ depends strongly on this selection. As a consequence, Bai and Ng (2006) recommend using $\widetilde{\Gamma}_t^{HR}$ even in the presence of cross-sectional dependence.[1]

2.2 Generalized Least Squares Principal Components

In practice, the idiosyncratic components can be cross-sectionally heteroscedastic and/or correlated. In this case, one can estimate the factors using the GPC estimator, $\tilde{f}^{GPC}$, which is given by $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of $Y\Sigma_\varepsilon^{-1}Y'$ arranged in decreasing order and $\tilde{\Lambda}^{GPC\prime} = \frac{1}{T}\tilde{f}^{GPC\prime}Y$. The GPC estimator of the factors is $min\left(\sqrt{N},T\right)$-consistent if the idiosyncratic cross-correlations are weak; see Bai and Li (2012, 2016). Furthermore, Choi (2012) derives the following asymptotic distribution of the GPC factors

$$\sqrt{N}\left(\tilde{f}_t^{GPC} - F_t\right) \xrightarrow{d} N\left(0, \Sigma_{\Lambda^*}^{-1}\right), \tag{7}$$

where $\Sigma_{\Lambda^*} = lim_{N\to\infty}\frac{1}{N}\Lambda'\Sigma_\varepsilon^{-1}\Lambda$; see also Bai and Li (2012, 2016) for the asymptotic properties. If Σ_ε^{-1} were known, the asymptotic approximation of the MSE can be estimated by

$$\widetilde{MSE}_t^{GPC} = \frac{\tilde{\Lambda}^{GPC\prime}\Sigma_\varepsilon^{-1}\tilde{\Lambda}^{GPC}}{N}. \tag{8}$$

In practice, Σ_ε should be substituted by an estimate, $\widetilde{\Sigma}_\varepsilon$, both when calculating $\tilde{f}_t^{GPC}$ and its MSE in (8). There is no obvious way to exploit the entire Σ_ε matrix to improve efficiency and, consequently, Boivin and Ng (2006) suggest assuming that it is diagonal and estimating Σ_ε as follows

$$\widetilde{\Sigma}_\varepsilon = diag\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'\right). \tag{9}$$

However, note that, in the presence of cross-sectional dependence, the asymptotic variance of the feasible GPC estimator should be modified to take into account that the matrix $\widetilde{\Sigma}_\varepsilon$ used to extract the factors is diagonal while Σ_ε is full. The true approximation of the asymptotic MSE of $\tilde{f}_t^{GPC}$ is given by

$$MSE_t^{GPC} = \left(\frac{\Lambda'\Sigma_\varepsilon^{*-1}\Lambda}{N}\right)^{-1}\frac{\Lambda'\Sigma_\varepsilon^{*-1}\Sigma_\varepsilon\Sigma_\varepsilon^{*-1}\Lambda}{N}\left(\frac{\Lambda'\Sigma_\varepsilon^{*-1}\Lambda}{N}\right)^{-1}, \tag{10}$$

where $\Sigma_\varepsilon^* = diag\left(\Sigma_\varepsilon\right)$.

2.3 Kalman filter and smoothing algorithms

Model (1)-(2) can be written as a State Space Model and, if the model parameters were known, the KFS can be implemented to extract the factors and to obtain their MSEs and their corresponding prediction intervals.

[1]Assuming that the process $\lambda_i\varepsilon_i$ follows a linear-array process, Kim (2022) proposes estimating the MSEs of the PC factors by a time-series average of spatial HAC estimators. Analysing the performance of this estimator is beyond our objectives, as we focus on the most popular estimators often used in empirical applications.

In practice, the parameters are unknown and need to be estimated before running the KFS algorithms. Given that numerical maximization of the likelihood can be a difficult task when N is large, the Gaussian log-likelihood is often maximized using the Expectation Maximization (EM) algorithm; see Miranda, Poncela and Ruiz (2022) for a description of the EM algorithm in the context of DFMs. The EM and/or KFS algorithms are often run by assuming that Σ_ε is diagonal. For example, Doz, Giannone and Reichlin (2011) propose a Two-Step Least Squares (TS-LS) estimator based on fitting a VAR(1) model to the PC factors and running the KFS using the corresponding estimates. They show that the smoothed factors extracted using the TS-LS estimated parameters are consistent due to the misspecification error vanishing as N and T diverge to infinity. Alternatively, Doz, Giannone and Reichlin (2012) propose estimating the parameters by the EM algorithm assuming that Σ_ε is diagonal even if it is not. The resulting estimator is known in the related literature as Quasi-ML (QML) and will be denoted as EM-QML. The $min\left(\sqrt{N}, \sqrt{T}\right)$- consistency and asymptotic normality of the estimates of the factors is derived by Barigozzi and Luciani (2019), who show under which conditions the asymptotic distribution can still be used for inference in case of miss-specification.

The MSEs used to construct prediction intervals for the KFS factors are based on the steady state MSEs. Note that they are not the true ones but are obtained by substituting the true parameters by estimated ones when running the KFS algorithms. Moreover, they do not incorporate parameter uncertainty. Furthermore, when the filter is run assuming that the covariance matrix of the idiosyncratic components is diagonal, the resulting MSEs are not reflecting the true uncertainty of the factors as they are not taking into account the uncertainty associated with the model misspecification. In order to obtain the true MSEs in the presence of misspecification, one can use the results in Harvey and Delle Monache (2009).

3 Simulation results

In this section, we carry out a simulation study about the performance of the asymptotic distribution of the factors extracted either by PC, GPC, TS-LS and EM-QML, in the presence of weakly cross-correlated idiosyncratic components when they are wrongly assumed to be uncorrelated.

Systems of Y_t with $N = 30, 100$ and 200 and $T = 50, 100$ and 500, are generated by the DFM in (1)-(2) with $r = 1$ and the loadings independently generated by $\lambda_i \sim U(0,1)$. The factor is generated by an AR(1) model with autoregressive parameter $\phi = 0.7$ and $\sigma_u^2 = 1 - \phi^2$ and standardized so that $\frac{1}{T}\sum_{t=1}^{T} F_t^2 = 1$. Finally, ε_t, are generated by independent Gaussian white noise vectors with the variances in the main diagonal of Σ_ε, $\sigma_i^2, i = 1, ..., N$, being independently generated by $\sigma_i^2 \sim U(0.5, 10)$ while the covariances are generated by

$$\sigma_{ij} = \sigma_i \sigma_j \tau^{|i-j|}. \tag{11}$$

The values of τ considered are $0, 0.1, 0.2, ..., 0.9$. The number of replicates is $M = 1000$.

3.1 Point estimates

For each replicate, $m = 1, ..., M$, and factor extraction procedure, the MSE of the factor is calculated as

$$MSE^{(m)} = \frac{1}{T}\sum_{t=1}^{T}\left(F_t^{(m)} - \hat{F}_t^{(m)}\right)^2. \tag{12}$$

Given that the empirical variances of both the simulated and estimated factors are one, the MSEs are given by

$$MSE^{(m)} = 2\left(1 - \frac{1}{T}\sum_{t=1}^{T} F_t^{(m)}\hat{F}_t^{(m)}\right) = 2\left(1 - Corr\left(F_t^{(m)}, \hat{F}_t^{(m)}\right)\right), \tag{13}$$

where $Corr\left(F_t^{(m)}, \hat{F}_t^{(m)}\right)$ is the sample correlation between the true factor $F_t^{(m)}$ and the estimated factor, $\hat{F}_t^{(m)}$.

Table 1 reports the average MSEs through Monte Carlo replicates, denoted by MSE^T, for $\tau = 0$ and 0.5.[2] We can observe that, when $\tau = 0.5$, PC has the largest MSE followed by GPC, TS-LS and EM-QML. The difference between the procedures is more pronounced if N is moderate and T is large. For example, when $N = 100$ and $T = 500$, the average MSE of PC, GPC, TS-LS and EM-QML are 0.239, 0.213, 0.186, and 0.178, respectively. These MSEs imply average correlations between the true and estimated factors of 0.88, 0.89, 0.90 and 0.91, respectively. As expected, the performance of all procedures deteriorates with the presence of cross-dependence of the idiosyncratic components.

3.2 Prediction intervals

For each Monte Carlo replicate and period of time, we obtain 95% prediction intervals for the factors estimated by each of the procedures considered. The intervals are obtained based on the asymptotic normality with the MSEs calculated by (4) with $\widetilde{\Gamma}_t$ estimated as in (5), (8) with Σ_ε estimated by (9), and from the steady-state Kalman filter, as appropriate. For each factor extraction procedure, the coverage is computed by counting how many times in a given replicate the simulated factors are contained in their corresponding intervals. Figure 1 plots averages of the empirical coverages obtained through the replicates for different values of the cross-sectional dependence, τ. We can observe that, when T is small, the coverages are well below the 95% nominal level regardless of the cross-sectional dimension even if there is not idiosyncratic cross-sectional dependence; see Maldonado and Ruiz (2021), who show that, in the case of PC, parameter uncertainty is not included in the asymptotic MSE and, consequently, the coverage is below nominal. The coverages are close to the nominal level when T is large

[2]Results for other values of τ are available upon request.

and $\tau = 0$. In this latter case, if $N = 30$, the best performance is observed for the EM-QML estimator, with GPC and TS-LS being similar and slightly worse, and finally PC having a slightly smaller coverage. The presence of cross-sectional dependence seriously affects the coverages of the prediction intervals of the factors even when N and T are large. For example, if $\tau = 0.5$, $N = 100$ and $T = 500$, the coverages of PC, GPC, TS-LS and EM-QML are 0.735, 0.761, 0.765 and 0.783, respectively.

To understand whether this deterioration is due to parameter estimation uncertainty or to a poor approximation of the asymptotic MSE used to obtain the prediction intervals, for each of the procedures considered to extract the factors, we consider the following decomposition of the difference between the true empirical MSE, MSE_t^T, and the MSE used in practice to construct the intervals, $\widetilde{MSE}_t^M$,

$$MSE_t^T - \widetilde{MSE}_t^M = \left(MSE_t^T - MSE_t^A\right) + \left(MSE_t^A - MSE_t^M\right) + \left(MSE_t^M - \widetilde{MSE}_t^M\right), \quad (14)$$

where MSE_t^A is the asymptotic MSE obtained when the model parameters are known (which does not incorporate parameter uncertainty), which is computed by using (4) with known parameters for PC, by (10) for GPC and by the Kalman filter steady-state MSE obtained using Harvey and Delle Monache (2009) with known parameters for TS-LS and EM-QML. Finally, MSE_t^M is the misspecified asymptotic MSE obtained with known parameters and assuming that Σ_ε is diagonal. In particular, MSE_t^M is obtained using (4) with Γ_t estimated by (6) for PC, using (8) with known parameters for GPC and using the steady-state MSE delivered by the Kalman filter in the misspecified model.

Table 1 reports MSE_t^T and $\widetilde{MSE}_t^M$, together with the three terms in the decomposition in (14), denoted as P, M and E, respectively, when $\tau = 0$ and 0.5. First, comparing MSE_t^T and $\widetilde{MSE}_t^M$, we can observe that, regardless, of N, T and τ, and of the procedure used to extract the factors, the former MSE is clearly larger than the latter, explaining the large undercoverage of the prediction intervals for the factors observed in Figure 1. Second, underestimation of the MSE has different causes depending on whether there is idiosyncratic cross-correlation or not. Consider first the results when there is not idiosyncratic cross-correlation and, therefore, there is not misspecification when Σ_ε is assumed to be diagonal. In this case, the difference between MSE_t^T and $\widetilde{MSE}_t^M$ is mainly due to the parameter estimation uncertainty not being considered when computing $\widetilde{MSE}_t^M$; observe that M is always close to zero. The parameter estimation uncertainty decreases with T and can be incorporated into the MSE by using resampling procedures as those proposed by Maldonado and Ruiz (2021) and Rodríguez and Ruiz (2012) for PC-based and KFS procedures, respectively.

However, when $\tau = 0.5$, the percentage of the difference between MSE_t^T and $\widetilde{MSE}_t^M$ due to misspecification, M, is different from zero and its magnitude decreases with the cross-sectional dimension, N, but increases with T. For example, if $T = 50$, this percentage is 18% when $N = 30$

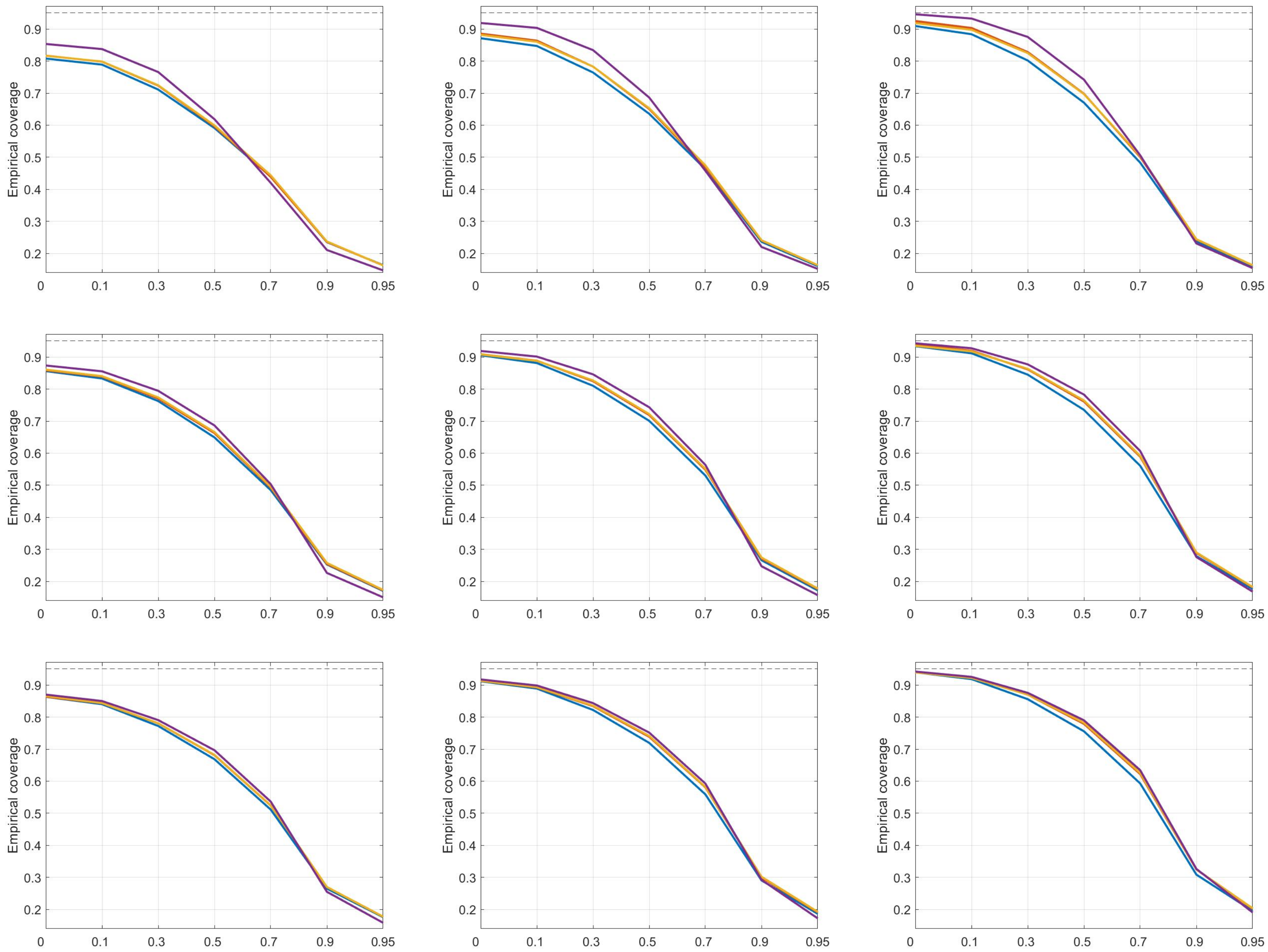

Figure 1: Monte Carlo 95% empirical coverages of prediction intervals for the factors constructed using PC (blue), GPC (red), TS-LS (yellow) and EM-QML (purple) with $N = 30$ (first row), $N = 100$ (second row) and $N = 200$ (third row) and $T = 50$ (first column), $T = 100$ (second column) and $T = 500$ (third column).

and decreases to 14% for $N = 200$. However, if $T = 500$, the percentages are 23% and 21%, respectively. It is important to note that, as far as we are concerned, this underestimation of the MSE of the factors cannot be corrected by using available resampling procedures; see, for example, the results in Maldonado and Ruiz (2021), who use subsampling to incorporate parameter uncertainty but still have large biases in the MSEs as they do not consider the misspecification of Σ_ε. Constructing prediction intervals for the factors with coverages close to the nominal is still an open challenge in the presence of idiosyncratic cross-sectional dependence.

Finally, Table 1 shows that the advantage of KFS over PC-based procedures when extracting factors, is larger when the cross-sectional dimension, N, is small and disappears when N is large; see also the results in Ruiz and Poncela (2022).

4 Main conclusions

Factor extraction is an important tool in the analysis of large stes of economic and financial variables. We analyse the performance of point and interval estimates of the factors when they are extracted wrongly assuming cross-sectionally uncorrelated idiosyncratic components. We show that point and interval estimates of the factors deteriorate with the presence of cross-sectional dependence if it is not taken into account. More importantly, a large part of the MSE of the factors can be attributed to misspecification and cannot be corrected using procedures designed to incorporate parameter uncertainty. Procedures designed to take into account the misspecification of the covariance matrix of the idiosyncratic component should be developed if appropriate prediction intervals for the factors are required.

We also show that, regardless of whether there is cross-sectional dependence, the advantage of KFS over PC-based procedures is relevant when the cross-sectional dimension is small.

References

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

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

[3] 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.

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

N	30					100					200				
	MSE^T	$\widetilde{MSE}^M$	P	M	E	MSE^T	$\widetilde{MSE}^M$	P	M	E	MSE^T	$\widetilde{MSE}^M$	P	M	E
$T=50$															
$\tau=0$															
PC	0.467	0.205	0.373	-0.015	-0.096	0.149	0.081	0.121	-0.004	-0.049	0.076	0.043	0.062	-0.002	-0.027
GPC	0.462	0.195	0.351	0.006	-0.090	0.145	0.077	0.112	0.002	-0.046	0.074	0.041	0.058	0.001	-0.026
TS-LS	0.374	0.156	0.298	-0.005	-0.075	0.127	0.068	0.102	-0.003	-0.040	0.067	0.038	0.054	-0.001	-0.024
EM-QML	0.360	0.314	0.284	-0.005	-0.233	0.124	0.072	0.099	-0.003	-0.044	0.067	0.039	0.054	-0.001	-0.025
$\tau=0.5$															
PC	0.789	0.142	0.563	0.117	-0.033	0.317	0.070	0.249	0.036	-0.038	0.166	0.040	0.132	0.018	-0.024
GPC	0.781	0.135	0.257	0.152	-0.030	0.301	0.067	0.223	0.047	-0.036	0.154	0.038	0.115	0.024	-0.023
TS-LS	0.676	0.118	0.511	0.084	-0.037	0.271	0.061	0.213	0.030	-0.033	0.144	0.036	0.114	0.016	-0.022
EM-QML	0.705	0.129	0.540	0.084	-0.048	0.262	0.064	0.204	0.030	-0.036	0.139	0.036	0.109	0.016	-0.022
$T=100$															
$\tau=0$															
PC	0.353	0.228	0.261	-0.017	-0.119	0.118	0.086	0.091	-0.005	0.054	0.060	0.046	0.047	-0.003	-0.030
GPC	0.345	0.217	0.244	-0.003	0.113	0.114	0.083	0.084	-0.001	-0.052	0.058	0.044	0.043	0.000	-0.029
TS-LS	0.271	0.166	0.195	-0.005	-0.085	0.098	0.071	0.073	-0.003	-0.043	0.053	0.040	0.04	-0.001	-0.004
EM-QML	0.242	0.190	0.166	-0.005	-0.109	0.096	0.075	0.071	-0.003	-0.047	0.052	0.041	0.039	-0.001	-0.027
$\tau=0.5$															
PC	0.703	0.157	0.483	0.111	-0.048	0.268	0.075	0.201	0.035	-0.043	0.140	0.042	0.107	0.017	-0.026
GPC	0.681	0.151	0.440	0.137	-0.047	0.246	0.073	0.173	0.042	-0.042	0.126	0.041	0.089	0.022	-0.026
TS-LS	0.578	0.129	0.415	0.082	-0.048	0.218	0.065	0.160	0.030	-0.037	0.116	0.038	0.086	0.015	-0.023
EM-QML	0.578	0.145	0.415	0.082	-0.064	0.208	0.068	0.150	0.030	-0.040	0.113	0.038	0.083	0.015	-0.023
$T=500$															
$\tau=0$															
PC	0.295	0.244	0.205	-0.018	-0.136	0.099	0.090	0.073	-0.006	-0.058	0.051	0.047	0.038	-0.003	-0.031
GPC	0.284	0.233	0.189	-0.009	-0.129	0.095	0.086	0.067	-0.003	-0.055	0.049	0.045	0.035	-0.001	-0.030
TS-LS	0.220	0.172	0.145	-0.006	-0.091	0.083	0.073	0.059	-0.004	-0.045	0.045	0.041	0.033	-0.003	-0.026
EM-QML	0.200	0.193	0.125	-0.006	-0.112	0.081	0.076	0.057	-0.004	-0.048	0.044	0.041	0.032	-0.003	-0.029
$\tau=0.5$															
PC	0.643	0.168	0.427	0.108	-0.060	0.239	0.079	0.174	0.033	-0.047	0.124	0.044	0.091	0.017	-0.028
GPC	0.597	0.164	0.367	0.126	-0.060	0.213	0.076	0.143	0.039	-0.045	0.109	0.043	0.074	0.040	-0.028
TS-LS	0.503	0.137	0.341	0.081	-0.056	0.186	0.068	0.129	0.029	-0.040	0.099	0.039	0.069	0.015	-0.024
EM-QML	0.477	0.157	0.315	0.081	-0.076	0.178	0.070	0.121	0.029	-0.042	0.097	0.040	0.067	0.015	-0.025

Table 1: Averages of Monte Carlo MSEs (MSE^T) and of estimated MSEs ($\widetilde{MSE}^M$), together with the decomposition of their differences into the part due to parameter uncertainty (P), to misspecification (M) and to using parameter estimates (E), for factors extracted using PC, GPC, TS-LS and EM-QML.

[5] 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].

[6] Boivin, J. and S. Ng (2006), Are more data always better for factor analysis?, *Journal of Econometrics*, 132, 169-194.

[7] Choi, I. (2012), Efficient estimation of factor models, *Econometric Theory*, 28(2), 274-308.

[8] 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.

[9] 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.

[10] González-Rivera, G., C.V. Rodríguez-Caballero and E. Ruiz (2021), Expecting the unexpected: Economic growth under stress, WP21-02 (Statistics and Econometrics), Universidad Carlos III de Madrid.

[11] González-Rivera, G., E. Ruiz and J.V. Maldonado (2019), Growth in stress, *International Journal of Forecasting*, 35(3), 948-966.

[12] Harvey, A.C. and D. Delle Monache (2009), Computing the mean square error of unobserved components extracted by misspecified time series models, *Journal of Economic Dynamics & Control*, 33, 283-295.

[13] Kim, M.S. (2022), Robust inference for diffusion-index forecasts with cross-sectionally dependent data, *Journal of Business & Economic Statistics*, 40(3), 1153-1167.

[14] Maldonado, J.V. and E. Ruiz (2021), Accurate confidence regions for principal components factors, *Oxford Bulletin of Economics and Statistics*, 83(6), 1432-1453.

[15] Miranda, K., P. Poncela and E. Ruiz (2022), Dynamic factor models: Does the specification matter?, *SERIeS*, 13, 397-428.

[16] Poncela, P., E. Ruiz and K. Miranda (2021), Factor extraction using Kalman filter and smoothing: This is not just another survey, *International Journal of Forecasting*, 37(4), 1399-1425.

[17] Rodríguez, A. and E. Ruiz (2012), Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters, *Computational Statistics & Data Analysis*, 56, 62-74.

[18] Ruiz, E. and P. Poncela (2022), Factor extraction in dynamic factor models: Kalman filter versus Principal Components, *Foundations and Trends in Econometrics*, 12(3), 1-111.

[19] 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.