

PERUVIAN ECONOMIC ASSOCIATION

High Dimensional Factor Models: An Empirical  
Bayes Approach

James Sampi

Working Paper No. 75, October 2016

The views expressed in this working paper are those of the author(s) and not those of the Peruvian Economic Association. The association itself takes no institutional policy positions.

# High Dimensional Factor Models: an Empirical Bayes Approach

*James Sampi\**

Universitat Pompeu Fabra and Barcelona GSE

September 19, 2016

---

\*I would like to thank Christian Brownlees, Domenico Giannone and Geert Mesters for helpful comments.  
James Sampi: james.sampi@upf.edu

## Abstract

We propose an empirical Bayesian implementation of principal components analysis for estimating high dimensional factor models. The method is evaluated in a large Monte Carlo study where we compare the traditional principal components estimator to the our proposed empirical Bayes version. We find that for increasingly weak factor specifications the mean squared error gain that is obtained from the empirical Bayes implementation increases. We further compare the standard and empirical Bayes principal components estimators to their maximum likelihood counterparts and document that in all cases the maximum likelihood estimates remain more accurate. The methodology is illustrated for two empirical applications. One for nowcasting macroeconomic time series and one for portfolio management. We find that the empirical Bayesian principal components estimates outperform the standard principal components estimates when compared the mean squared error for the inner product of the macroeconomic forecast estimates. Second, in the portfolio optimization problem the covariance matrix of the stock returns estimated by empirical Bayes methods achieve, in most cases, the highest Information Ratio and the highest expected return for the portfolio manager.

*JEL classification:* C32; C43

*Some keywords:* Shrinkage; Principal Component Analysis; Posterior modes; Nowcasting; Portfolio Management.

# 1 Introduction

Principal components analysis (PCA) is one of the main methods for estimating parameters in high dimensional approximate factor models. In macroeconomics these models are typically used for forecasting and structural analysis. Bai and Ng (2008) and Stock and Watson (2011) provide reviews of the factor model literature.

Bai and Ng (2002) and Bai (2003) show that under strong factor assumptions the principal components method yields consistent and asymptotically normal estimates for the factors. Unfortunately, for many empirical applications such strong factor assumptions are argued to be not realistic, see Onatski (2012). In fact, relaxing the strong factor assumption, Onatski (2012) shows that the principal components method no longer yields consistent estimates for the factors.

In this paper we explore an empirical Bayesian implementation of the PCA for estimating approximate factor models. In particular, using Gaussian prior densities we shrink the loadings and factors of the principal components estimates towards common means. We propose an easy algorithm to implement an estimator that carries out the procedure. Notably, the method under consideration requires only marginally more computational time when compared to the standard principal components estimator.

The Bayesian approach to PCA has already been considered in the machine learning literature for exact factor models with scalar variance, see Bishop (1999). Moreover, Bishop (2006) provides a textbook treatment. He argues that Bayesian approaches to PCA may give more accurate estimates in mean squared error sense for the inner product between the loadings and the factors when compared to the classical principal components method. However, no comparison between Bayesian and standard PCA exists neither for approximate factor models, where the error terms do not have a scalar variance structure, nor for weak factor settings. We expect that in exactly these settings where the noise is large relative to the factors, the Bayesian methods perform increasingly well, see Gelman, Carlin,

Stern, Dunson, Vehtari and Rubin (2012) for further discussion.

An alternative popular method for estimating parameters in high dimensional approximate factor models is the Maximum Likelihood Estimation (MLE), see Doz, Giannone and Reichlin (2012), Bai and Li (2012) and Bai and Li (2015). Both Doz et al. (2012) and Bai and Li (2015) show that maximum likelihood outperform principal components based methods in terms of accuracy. Nevertheless, the maximum likelihood method requires more computation time, which makes it less attractive for practitioners. An empirical Bayes implementation for the maximum likelihood method was recently proposed by Koopman and Mesters (2016). They show that the empirical Bayes implementation gives large gains in mean squared error for the inner product estimates.

The contributions of this paper can be summarized as follows. First, we propose an empirical Bayesian implementation of the PCA for estimating approximate factor models. Second, we compare this alternative implementation to the standard principal components method in a large Monte Carlo study. Here we study whether the Bayesian version of PCA can improve the estimation accuracy. We are especially interested in the weak factor settings. Third, we compare the empirical Bayes PCA to the Bayesian maximum likelihood method. Here we aim at documenting whether the empirical Bayesian implementation of the principal components method can give results that are close to the maximum likelihood estimates.

Finally, we illustrate one empirical application for macroeconomic nowcasting and another for portfolio optimization. We compare the different nowcasting methods among them upon macroeconomic time series. Nowcasting relies on different types of data based on real economic activities such as financial series and surveys. Because the time series are non-synchronized, the resulting panel is an unbalanced data set. This feature makes it a difficult task to efficiently use the information contained in various time series.

In the portfolio optimization problem, we study how these methods perform when estimating the covariances of different portfolios. Here the object of interest is not the forecast,

but the covariance matrix estimate is implied by the factor model.

Our paper yields three main results. First, the simulation exercise shows that the empirical Bayesian PCA is more accurate than the standard PCA when comparing the mean squared error of the inner product estimates. The gains can be as large as 98%, 64% and 15% depending on how strong the factors are. The weaker the factors the larger the gains implied by the empirical Bayes PCA. Also, under various disturbances' specification the gains stay the same. This conclusion complements Bai (2003) and Onatski (2012).

Second, within the Bayesian specifications, MLE outperform PCA. Despite that the empirical Bayesian MLE achieves the lower mean squared error for the inner product estimates, it requires much more computational time than the empirical Bayes PCA.

Third, the empirical illustrations show that the empirical Bayesian estimates outperform the standard estimates considered in each exercise. When dealing with macroeconomic forecasting, the empirical Bayesian methods outperform their standard counterparts. Similar results are obtained for the portfolio optimization problem, where largest Information Ratio is achieved by the empirical Bayesian estimates when compared to shrinkage covariance estimates, see Ledoit and Wolf (2004).

The remainder of the paper is organized as follow. The next section provides a brief literature overview of factor models in macroeconomics. Section 3 discusses the approximate factor model and explains why it is different from the exact factor model. The PCA method and the empirical Bayes implementation of it are discussed in Section 4. Their maximum likelihood counterparts are discussed in Section 5. The Monte Carlo study which compares all methods is shown in Section 6 and the empirical studies are shown in Section 7. Section 8 concludes.

## 2 Literature review

The seminar papers of Geweke (1989), Stock and Watson (1989) and, Bai and Ng (2002) have placed the dynamic factor model (DFM) as the predominant framework for research on macroeconomic forecasting using many predictors. Overall, this framework allows us to study large panels of time series through few common factors, especially, when the data series are strongly collinear.

The available methodologies for estimating DFMs can be divided into three groups. The first group of estimators entails nonparametric estimation with large  $N$  using cross-sectional averaging methods, primarily principal components. Principal components analysis (PCA) is the most popular factor extraction method in the treatment of dynamic factors models. PCA is appealing because of its computational advantages and asymptotic properties in large data sets, see Bai (2003). Moreover, few improvements have been introduced in order to enhance the estimator's efficiency. Stock and Watson (2006) presented a handy summary to novel PCA estimators, e.g. weighted principal components.

The second group consists of parametric models estimated in the time domain using maximum likelihood estimation (MLE) and the Kalman filter. MLE has been used successfully to estimate the parameters of low-dimensional DFMs. The likelihood function is computed by the Kalman filter and the factors are estimated by the Kalman smoother. However, there are significant computational requirements to maximize the likelihood function with many parameters. In order to deal with the dimensionality problem associated with the likelihood function, further estimators have been implemented. The main idea behind this methods is to use the consistent parameters implied by the first group for computing the factors implied by the second one, see Doz and Reichlin (2011) and Doz et al. (2012).

The third group relies on Bayesian methods to estimate DFMs. The main idea behind these methods is the Bayesian dynamic factor analysis introduced by Aguilar and

West (2000), Kaufman and Schumacher (2013) and, Nakajima and West (2013). Moreover, Stock and Watson (2012) provide an empirical comparison study using Bayesian methods. They suggest that further shrinkage estimators can potentially improve the estimation accuracy. Thereby, Bayesian methods appear to be a potential tool to solve the dimensionality problem associated with high dimensional DFMs.

### 3 Approximate factor model

Let  $y_{it}$  be the observed data for the  $i$ th variable at time  $t$ . In total we have  $N$  variables indexed by  $i = 1, \dots, N$ . Also, we have  $T$  time periods and  $t = 1, \dots, T$ . The approximate factor model decomposes  $N$  dimensional vectors  $y_t = (y_{1t}, \dots, y_{Nt})'$ , for  $t = 1, \dots, T$ , as follows

$$y_t = \Lambda f_t + \varepsilon_t \tag{1}$$

where  $\Lambda = (\lambda_1, \dots, \lambda_N)'$  is the  $N \times r$  matrix of factor loadings with  $r$  as the number of factors,  $f_t = (f_{1t}, \dots, f_{rt})'$  is the  $r \times 1$  vector of factors and  $\varepsilon_t$  is the  $N \times 1$  idiosyncratic disturbance term.

The observation equation of the model (1) can also be written as

$$y = \Lambda f + \varepsilon \tag{2}$$

where  $y = (y_1, \dots, y_N)'$  is the  $N \times T$  matrix of the observed data,  $f = (f_1, \dots, f_r)'$  is the  $r \times T$  matrix of factors, and  $\varepsilon$  is the  $N \times T$  disturbances matrix. This alternative representation is convenient for our exposition along this paper.

Typically, the *exact factor model* assumes idiosyncratic disturbances mutually uncorrelated for all  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . Unlike the exact factor model, we assume an *approximate factor structure* by allowing some serial and cross-sectional correlation among the idiosyncratic components, see Bai and Ng (2002).



In approximate factor settings, the consistency and asymptotic normality of the estimators when both  $N$  and  $T$  go to infinity have been recently shown by Bai (2003), Bai and Ng (2002) and Doz et al. (2012). In order to prove these properties, Bai (2003) makes a strong assumption related to the eigenvalues of the population covariance matrix of the data. Specifically, it requires that the ratio between the  $r - th$  largest and the  $r + 1 - th$  largest eigenvalues increase proportionately to  $N$ . Asymptotically, this implies that the cumulative effects of the normalized factors strongly dominate the idiosyncratic disturbances.

The strong factor assumption is sometimes called pervasive factor assumption, and it can be expressed as follows  $\Lambda'\Lambda/N = D_N$  with elements  $d_{1N} > \dots > d_{rN} > 0$  along the diagonal. Such that,  $D_N \xrightarrow{N \rightarrow \infty} D$  where  $D$  is a diagonal matrix with elements  $d_1 > \dots > d_r > 0$ . It requires that the cumulative explanatory power of the factors is measured by the diagonal elements of  $\Lambda'\Lambda$ , and it increases proportionally to  $N$ .

Recently, Onatski (2012) and Onatski (2015) show that the strong factor assumption requires one of the following two scenarios. Either, an overwhelming domination of the factors represented by higher values of  $d_r$  for all  $r$ , or  $\varepsilon\varepsilon'/T$  needs to be close to the identity matrix, where  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$  is the  $N \times T$  disturbances matrix. However, the former scenario is unwanted because we do not want to assume the overwhelming domination of factors over the idiosyncratic disturbances. The latest scenario does not hold as typically the expected covariance matrix of the disturbances is not the identity,  $E(\varepsilon_t\varepsilon_t') = \Omega \neq I_N$ .

Moreover, Onatski (2012) proposed the alternative weak factor assumptions. He assumes  $\Lambda'\Lambda = \Delta_N$  with elements  $\delta_{1N} > \dots > \delta_{rN} > 0$  along the diagonal. Such that,  $\Delta_N \xrightarrow{N \rightarrow \infty} \Delta$  where  $\Delta$  is a diagonal matrix with elements  $\delta_1 > \dots > \delta_r > 0$ . As explained in details in Onatski (2012), the asymptotic regime described by this assumption is meant to provide an adequate approximation to empirically relevant finite sample situations where a few of the largest eigenvalues of the sample covariance matrix are not overwhelmingly larger than the rest of the eigenvalues.

We notice that the pervasive factor assumption plays a key role in order to determine

the accuracy of the estimates. However, many of the macroeconomics applications do not verify this assumption. In fact, very few estimation procedures are dealing with the weak factor assumptions.

In this paper we address a Bayesian approach in order to deal with the weak factor assumptions. In particular, we rely on the following two arguments. The pervasive factor assumption only considers the deterministic loading matrix case. Nonetheless, the loadings may follow a stochastic process. This implies that the cumulative explanatory power of factors could be wrongly estimated. This shortcoming allows us to start considering the Bayesian approaches as the ideal framework to deal with high dimensional DFMs where the loading matrix is treated as latent variable.

Second, under the strong factor assumptions the disturbances are assumed to be uncorrelated and  $\varepsilon\varepsilon'/T$  is fixed as identity matrix. Therefore, the disturbances covariance matrix is not relevant to determine the factors and loading estimates. In our setting, we allow random loadings and factors in order to incorporate an approximate factor structure into the determinacy of our estimates by using the joint posterior density.

## 4 Principal components analysis

In this section, we provide a detailed discussion on the standard principal components analysis (PCA) and empirical Bayesian PCA estimators. PCA is often used for dynamic factor models because of its computational efficiency for practitioners. Moreover, it is a nonparametric procedure that relies on the strong factor assumption to be consistent. Notwithstanding, the strong factor assumption does not hold in many macroeconomics applications. Thereby, we implement an empirical Bayesian estimator in order to deal the with weak factor assumptions. Our proposed empirical Bayesian PCA is a nonparametric estimator that uses the joint posterior density to determine the estimates. This estimator requires marginally more computational time than the standard PCA estimator.

## 4.1 Standard principal component analysis

We estimate the common factor,  $\Lambda f_t$ , in large panels using the method of principal components. Consider an arbitrary number of factors  $r$  ( $r < \min\{N, T\}$ ), such that  $\text{rank} f = r$  and  $\text{rank} \Lambda = r$ , where  $f = (f_1, \dots, f_r)'$  is the  $r \times T$  matrix of factors.

The PCA estimates of  $\Lambda$  and  $f_t$  are obtained by solving the optimization problem

$$V = \min_{\Lambda, f} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \lambda_i f_t)^2 \quad (3)$$

subject to the normalization of either  $\Lambda' \Lambda / N = I_r$  with  $r(r+1)/2$  restrictions or  $f f' / T = I_r$  with  $r(r-1)/2$  restrictions. We used the notation  $\lambda_i$  as the  $i^{\text{th}}$  row of  $\Lambda$  for  $i = 1, \dots, N$ . The optimization problem is identical to maximizing  $\text{tr}(f(y'y)f')$  where  $y = (y_1, \dots, y_N)'$  is the  $N \times T$  matrix of the observed data. Here  $\text{tr}()$  denotes the trace operator. Let  $Q$  the  $r \times r$  diagonal matrix containing the  $r$  largest eigenvalues of sample covariance matrix  $S = \frac{1}{T} \sum_{t=1}^T y_t y_t'$ . The estimated factor matrix denoted by  $\hat{f}^{PCA}$ , is  $\sqrt{T}$  times the eigenvector,  $P$ , associated with  $Q$ . Hence,  $\hat{\Lambda}^{PCA} = y \hat{f}^{PCA'} / T$  is the corresponding matrix of factor loadings estimated.

The solution to the above minimization problem is not unique, even though the sum of squared residuals  $V$  is unique. Another equivalent solution explored by Doz and Reichlin (2011) and Bai and Ng (2002) is expressed as follow

$$\begin{aligned} \hat{f}_t^{PCA} &= Q^{-1/2} P' y_t \\ \hat{\Lambda}^{PCA} &= P Q^{1/2} \end{aligned} \quad (4)$$

Recent literature has shown that the principal components estimator of the common factors provides consistent estimates under the strong factor assumptions, see Bai and Ng (2000), Bai and Ng (2002) and Onatski (2012). However, the weak factor assumptions are more relevant for the empirical macroeconomics applications, see Boivin and Ng (2003).

## 4.2 Empirical Bayes principal components analysis

In this section we introduce our proposed empirical Bayesian implementation for the PCA. As stated in the standard PCA, the factors and disturbances follow a random distributions defined by  $f_t \sim NID(0, I_r)$  and  $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2 I_N)$ . Nevertheless in order to deal with the weak factor assumptions, we assume that the loading follows a random distribution of the form  $\lambda_i \sim NID(0, \Sigma)$ , where  $\Lambda = (\lambda_1, \dots, \lambda_N)'$  is the  $N \times r$  loading matrix and  $\Sigma$  is the loading covariance matrix. Hence, the asymptotic covariance matrix is defined by  $E(y_t y_t') = \text{tr}(\Sigma) I_N + \sigma_\varepsilon^2 I_N$ , where  $\text{tr}()$  denotes the trace operator and  $y_t = (y_{1t}, \dots, y_{Nt})'$  is the  $N \times 1$  vector of the observed data. In our setting  $\text{tr}(\Sigma)$  is not necessarily proportional to  $N$  and does not overwhelms to the disturbances. Thereby, weak factors are implied by smaller values of  $\text{tr}(\Sigma)$ .

Unlike the standard PCA, in our setting the disturbances play a key role in the maximization problem. The conditional joint posterior distribution of the loadings and factors,  $p(\Lambda, f|y)$ , allows us to consider a more general case in which the disturbances covariance matrix is not equal to the identity.

Hence, instead of the optimization problem (3), we propose one associated with the posterior density given by

$$\left\{ \hat{\Lambda}^{EB/PCA}, \hat{f}^{EB/PCA} \right\} = \arg \max_{\Lambda, f} \log p(\Lambda, f|y) \quad (5)$$

where

$$\log p(\Lambda, f|y) = \log p(y|f, \Lambda) + \log p(f) + \log p(\Lambda) - \log p(y), \quad (6)$$

$f = (f_1, \dots, f_r)'$  is the  $r \times T$  matrix of factors, and  $y = (y_1, \dots, y_N)'$  is the  $N \times T$  matrix of the observed data. An analytical solution of (5) is complicated as the first order conditions to  $\Lambda$  and  $f$  depend on each other. However, we can obtain the modes  $\hat{\Lambda}^{EB/PCA}$  and  $\hat{f}^{EB/PCA}$  under the iterative conditional mode algorithm described as follows

- (i)  $f^{(s)} = \Lambda^{(s-1)'} \text{inv}\left(\Lambda^{(s-1)}\Lambda^{(s-1)'} + \sigma_\varepsilon^{2(s-1)}I_N\right) y$
- (ii)  $\Lambda^{(s)} = y \text{inv}\left(f^{(s)'}\Sigma^{(s-1)}f^{(s)} + \sigma_\varepsilon^{2(s-1)}I_T\right) f^{(s)'}\Sigma^{(s-1)}$
- (iii)  $\sigma_\varepsilon^{2(s)} = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \left(y_{it} - \lambda_i^{(s)} f_t^{(s)}\right)^2$ , and  
 $\Sigma^{(s)} = \text{diag}\left(\sigma_{\lambda,1}^{2(s)}, \dots, \sigma_{\lambda,r}^{2(s)}\right)$  with  $\sigma_{\lambda,j}^{2(s)} = \frac{1}{N} \sum_{i=1}^N \lambda_{ij}^{2(s)}$  for all  $j = 1, \dots, r$ .
- (iv)  $s = s + 1$ , until they converge

We use the convergence criteria proposed by Onatski (2015) described as follows,  $\text{tr}\left[\left(\Lambda^{(s)}f^{(s)} - \Lambda^{(s-1)}f^{(s-1)}\right)\left(\Lambda^{(s)}f^{(s)} - \Lambda^{(s-1)}f^{(s-1)}\right)'\right] < 10^{-6}$ . As well,  $\lambda^{(0)}$  and  $f^{(0)}$  are chosen as the standard principal components estimates.

Since our setting does not involve any analytical solution to the optimization problem (5), an evaluation of its consistency properties would be interesting to develop. In fact, it is an important next step in our research agenda.

## 5 Maximum likelihood based approaches

In this section we discuss the standard and the empirical Bayesian maximum likelihood methods. For this methods, we need to make an assumption about the stochastic process of the factors such that the model can be written in state space form. In particular, we assume that the factors follow a vector autoregressive model of order one. We have

$$f_t = Hf_{t-1} + \eta_t \quad \eta_t \sim IID(0, \Sigma_\eta) \quad (7)$$

where  $H$  is the  $r \times r$  transition matrix and  $\eta_t$  is the  $r \times 1$  factor error term that has mean zero and variance  $\Sigma_\eta$ . This specification can easily be extended to allow for higher order vector autoregressions. Together with the observation equation (1) the model can be viewed as a state space model, see Durbin and Koopman (2012). Furthermore, we assume a more general case for the disturbances covariance matrix, such that  $\varepsilon_t \sim NID(0, \Omega)$ .

The reason for introducing the Bayesian estimation is the same as stated in Section 4.2. However, the factors follow an autoregressive process defined by the equation (7). See Koopman and Mesters (2016) for a detailed discussion.

## 5.1 Standard maximum likelihood estimation

We describe the parametric MLE method documented upon Durbin and Koopman (2012) and Ghahramani and Hinton (1996). The method relies on the Kalman filter. We define the conditional moments  $a_{t|s} = E(f_t|y_1, \dots, y_s; \psi^{MLE})$  and  $P_{t|s} = Var(a_{t|s} - f_t|y_1, \dots, y_s; \psi^{MLE})$  for  $t, s = 1, \dots, T$ , where  $\psi^{MLE} = \{\Lambda, \Omega, H, \Sigma_\eta\}$  contains the parameters that pertain to the distribution of the factors as well as to the disturbances in the equations (1) and (7). Moreover, we assume that the initial factor has density  $N(0, P_1)$  where  $P_1 = \text{inv}(I_r - HH')$  and  $\varepsilon_t \sim NID(0, \Omega)$  is the  $N \times 1$  disturbance term.

The Kalman filter is a recursive procedure through the time index  $t$  and is given by

$$\begin{aligned} F_t &= \Lambda P_{t|t-1} \Lambda' + \Omega, & K_t &= H P_{t|t-1} \Lambda', \\ a_{t+1|t} &= H a_{t|t-1} + K_t F_t^{-1} v_t, & P_{t+1|t} &= H P_{t|t-1} H' - K_t F_t^{-1} + \Sigma_\eta, & v_t &= y_t - \Lambda a_{t|t-1} \end{aligned} \quad (8)$$

The conditional mean of the state  $f_t$  given the observations  $y$ , where  $y = (y_1, \dots, y_N)'$  is the  $N \times T$  matrix of the observed data, that is  $\hat{f}_t^{MLE} = E(f_t|y)$ , together with its mean squared error matrix  $V_t$  can be therefore calculated using the backwards recursion

$$\begin{aligned} r_{t-1} &= \Lambda' F_t^{-1} v_t + L_t' r_t, & N_{t-1} &= \Lambda' F_t^{-1} \Lambda + L_t' N_t L_t, \\ \hat{f}_t^{MLE} &= a_t + P_t r_{t-1}, & V_t &= P_t - P_t N_{t-1} P_t, & L_t &= H - H P_t \Lambda' F_t^{-1} \Lambda \end{aligned} \quad (9)$$

for  $t = T, T-1, \dots, 1$ . Moreover, we set  $r_T = 0$  and  $N_T = 0$ . The relations in (9) are called the state smoothing recursions.

Defined the variables by the equations (1) and (7), the log-likelihood function associated

to the Gaussian density is given by

$$\log L(y; \psi^{MLE}) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \left( \log |F_t| + v_t' F_t^{-1} v_t \right) \quad (10)$$

where the quantities  $v_t$  and  $F_t$  are evaluated by the Kalman filter. As well, the vector  $\psi^{MLE} = \{\Lambda, \Omega, H, \Sigma_\eta\}$  of unknown parameters, contains the parameters that pertain to the distribution of the factors and disturbances.

The estimation of the parameter vector  $\psi^{MLE}$  is based upon maximizing the log-likelihood function  $L(y; \psi^{MLE})$  in (10) with respect to  $\psi^{MLE}$ . In large data sets, the number of parameters is typically large (i.e. 500). Therefore, numerical optimization procedures are used for the maximization of the log-likelihood function. Within these procedures we chose the Expectation-Maximization (EM) algorithm due to its relative computational efficiency advantage over other methods such as the quasi-Newton and Broyden-Fletcher-Goldfarb-Shanno (BFGS), see appendix A for more details.

Moreover, Doz et al. (2012) show that MLE estimates are consistent under the strong factor assumption. In fact, the effects of misspecification on the estimation of the common factors is negligible when  $N, T \rightarrow \infty$ .

## 5.2 Empirical Bayes implementation of maximum likelihood estimation

The parametric empirical Bayesian estimator was proposed by Koopman and Mesters (2016). As proved by the authors, this method is appealing to perform in order to deal with the weak factor assumptions in the context of the state space model described by the equations (1) and (7).

We allow the loading matrix to follow a random distribution  $\lambda_i \sim NID(\delta, \Sigma)$  as in section 4.2 where we set  $\delta = 0$ , and  $\Sigma$  is the loading covariance matrix. The factors follow the autoregressive process described by the equation (7) and  $\varepsilon_t \sim NID(0, \Omega)$  is the  $N \times 1$

disturbance term.

Koopman and Mesters (2016) described two general steps for performing the likelihood evaluation and the pertinent parameters. They redefined the parameter vector  $\psi^{EB/MLE} = \{\psi_1^{EB/MLE}, \psi_2^{EB/MLE}\}$ , as  $\psi_1^{EB/MLE} = \{\Omega, H, \Sigma_\eta\}$  and  $\psi_2^{EB/MLE} = \{\delta, \Sigma\}$ . The first step is concerning to obtain the  $\psi_1^{EB/MLE}$  estimates by similar procedures described in the standard maximum likelihood, solving

$$\left\{ \hat{\lambda}^{EB/MLE}, \hat{\psi}_1^{EB/MLE} \right\} = \arg \max_{\lambda, \psi_1} \log L\left(y; \psi_1^{EB/MLE}\right) \quad (11)$$

where  $\log L\left(y; \psi_1^{EB/MLE}\right)$  is given by the equation (10),  $y = (y_1, \dots, y_N)'$  is the  $N \times T$  matrix of the observed data, and  $\lambda = \text{vec}(\Lambda) = (\lambda'_1, \dots, \lambda'_r)'$  is treated as a deterministic variable. The problem (11) can be optimized with respect to  $\lambda$  and  $\psi_1^{EB/MLE}$  using the standard numerical methods, e.g. the EM algorithm as used in the Section 5.1, see the appendix A.

The second step aims at obtaining an estimate for  $\psi_2^{EB/MLE}$ , which is achieved by solving

$$\hat{\psi}_2 = \arg \max_{\psi_2^{EB/MLE}} \log p\left(\hat{\lambda}^{EB/MLE}; \psi_2^{EB/MLE}\right) \quad (12)$$

It is easy to verify that the procedure yields consistent estimates for the parameter vector because  $\hat{\lambda}^{EB/MLE}$  is obtained in the first step and the density  $p\left(\hat{\lambda}^{EB/MLE}; \psi_2^{EB/MLE}\right)$  is implied by the normality assumption of the loadings, see appendix B.

Then, given the estimated parameters described by  $\hat{\psi}^{EB/MLE}$  we compute the posterior modes for  $\Lambda$  and  $f$  as follow

$$\left\{ \hat{\lambda}^{EB/MLE}, \hat{f}^{EB/MLE} \right\} = \arg \max_{\lambda, f} \log p\left(\lambda, f | y; \hat{\psi}^{EB/MLE}\right), \quad (13)$$

where  $f = (f_1, \dots, f_r)'$  is the  $r \times T$  matrix of factors. However, the direct optimization of (13) with respect to  $\lambda$  and  $f$  is complicated as the first order conditions for  $\lambda$  and  $f$  depend



on each other and solving analytically, or numerically for either one is infeasible when  $N$  and  $T$  are large. Instead, Koopman and Mesters (2016) proposed a simple algorithm as follows

$$(i) \quad f^{(s)} = E\left(f|y; \lambda = \lambda^{(s-1)}, \hat{\psi}^{EB/MLE}\right);$$

$$(ii) \quad \lambda^{(s)} = E\left(\lambda|y; f = f^{(s)}, \hat{\psi}^{EB/MLE}\right);$$

(iii)  $s = s + 1$ , until they converge

We notice that the algorithm presented here is similar to the one we set in Section 4.2. Nonetheless, they rely on the Kalman filter in order to compute the estimates.

In practice, we use the convergence criteria described by  $\|\lambda_{i,j}^{(s)}/\lambda_{i,j}^{(s-1)} - 1\| < 10^{-5}$  and  $\|f_{i,j}^{(s)}/f_{i,j}^{(s-1)} - 1\| < 10^{-5}$ , for all  $i = 1, \dots, N$ ,  $j = 1, \dots, r$  and  $t = 1, \dots, T$ . The initial points  $\lambda^{(0)}$  and  $f^{(0)}$  are chosen as the standard maximum likelihood estimates.

## 6 Simulation Study

We present results from a simulation study under different data generating processes. The main interest is to compare the empirical Bayesian principal components to the standard principal components estimates. Moreover, we want to compare the empirical Bayesian maximum likelihood estimates with their standard counterpart. Finally, we compare the estimates between both Bayesian approaches, MLE and PCA. The estimates are computed given the estimated parameters vector  $\hat{\psi}^m$  for  $m = EB/PCA, PCA, EB/MLE$  or  $MLE$  which are obtained by the methods discussed in Sections 4 and 5.

### 6.1 Simulation design

We design two simulation exercises that rely upon specific factor assumptions which allow us to perform a clear comparison between methods. In the first exercise, we simulate the

approximate factor structure described in Section 4. The second exercise is to simulate the state space model described in Section 5.

In both simulation designs the dynamic factor model is simulated for different cross-sections and time series dimensions. We include combinations for  $N, T = 50$  and  $100$ . The number of common factors is chosen to be equal to  $r = 3$  and  $5$ . During the study is assumed that the true number of factors is known.

### 6.1.1 Simulation design 1

We draw  $J = 1000$  different panels of observations for the general model (1) and for each combination of panel size. Here, each observation has its own “true” loadings matrix and factors:  $\lambda(l)$  and  $f(l)$  where  $\lambda(l) = \text{vec}(\Lambda(l)) = (\lambda_1(l)', \dots, \lambda_r(l)')$  and  $f(l) = (f_1(l), \dots, f_r(l))'$  is the  $r \times T$  matrix of the true factors. The loadings  $\lambda_i$  are drawn from a distribution described by  $NID(0, \Sigma)$  where  $\Sigma = \sigma_\lambda^2 I_r$  and  $f_t$  by  $NID(0, I_r)$ . The disturbance  $\varepsilon_t$  incorporates cross-sectional and serial correlation, which is generated by

$$\begin{aligned} D(L)\varepsilon_t &= \nu_t && \text{with } \nu_t \sim NID(0, \Gamma) \\ d_{ij}(L) &= (1 - \rho L) && \text{if } i = j \\ \Gamma_{ij} &= \tau^{|i-j|}(1 - \rho^2) && \forall i, j = 1, \dots, n \end{aligned} \tag{14}$$

Notice that we defined the cross-correlation among the idiosyncratic elements by the Toeplitz matrix  $\Gamma$ . Where  $\rho$  and  $\tau$  are defined as the serial correlation and cross correlation coefficient, respectively.

We control the signal-noise ratio of the model (1) such that

$$y_t = \Lambda f_t + \sqrt{\theta} \varepsilon_t \quad t = 1, \dots, T \tag{15}$$

where the common factor  $\Lambda f_t$  and the idiosyncratic disturbance component  $\varepsilon_t$  are independent. The parameter  $\theta$  measures the inverse of the signal-to-noise ratio, see Onatski

(2015).

Moreover the loadings' variance,  $\sigma_\lambda^2$ , is chosen such that it mimics the empirical distribution of the loadings that is found in many macroeconomics applications as explained in Onatski (2012). In particular,  $\sigma_\lambda^2 = \{0.01, 0.1, 0.25, 0.5, 1, 2\}$ . Since  $E(\lambda_i' \lambda_i)$  is equivalent to  $\Sigma$  for all  $i = 1, \dots, N$ , we represent the strong factor assumptions with high values of  $\sigma_\lambda^2$ . Therefore the weak factors are represented by smaller values of  $\sigma_\lambda^2$ .

Along this exercise, we set  $\rho = 0.5$  and  $\tau = 0.5$  in order to represent serial and cross-sectional correlation among the idiosyncratic disturbances, or approximate factor model setting. Therefore,  $\rho = 0$  and  $\tau = 0$  represent idiosyncratic disturbances mutually uncorrelated, or exact factor model setting. Finally, we set  $\theta = 0.5$ , which sets the ratio between the variances to 0.5 on average.

### 6.1.2 Simulation design 2

There are  $J = 500$  different panels of observations for the general model (1) and (7) for each combination of panel size. Here, each observation has its own "true" loadings matrix and factors:  $\lambda(l)$  and  $f(l)$  where  $\lambda(l) = \text{vec}(\Lambda(l)) = (\lambda_1(l)', \dots, \lambda_r(l)')$  and  $f(l) = (f_1(l), \dots, f_r(l))'$  is the  $r \times T$  matrix of the true factors.

The loadings vectors  $\lambda_i(l)$  are drawn from two different mixture distributions. The distributions are chosen such that they mimic the empirical distribution of the loadings that is found in many macroeconomic applications. In particular, we draw the elements of the loading vector from

$$\pi(\lambda_{i,j}) = \kappa_1 N(\mu_1, \sigma_1^2) + \dots + \kappa_s N(\mu_s, \sigma_s^2) \quad (16)$$

where the values for  $\mu_n$ ,  $\sigma_n^2$ ,  $\kappa_n$  for  $n = 1, \dots, s$  and  $s$  are taken such that the loadings have normal and bimodal distributions. The values for the loading setting are given in Marron and Wand (1992).

The dynamics factors are simulated from the autoregressive process given by the equation (7) with autoregressive polynomial matrix  $H$  and covariance matrix  $\Sigma_\eta = I_r$ , such that  $\text{Var}(f_t) = \text{inv}(I_r - HH')$ . The elements for the diagonal of  $H$  are drawn from the uniform distribution  $U(0.5, 0.9)$ . The off-diagonal elements are set to zero. For both the standard maximum likelihood and the empirical Bayesian maximum likelihood methods, the factor process is correctly specified.

The error term  $\epsilon_t$  incorporates cross-sectional and serial correlation. Given draws  $\lambda_{i,j}$ ,  $\epsilon_t$  is generated by

$$\begin{aligned}
D(L)\epsilon_t &= \nu_t && \text{with } \nu_t \sim NID(0, \Gamma) \\
d_{ij}(L) &= (1 - \rho L) && \text{if } i = j \\
\alpha_i &= \frac{\beta_i}{1 - \beta_i} \lambda_i' \text{inv}(I_r - HH') \lambda_i && \text{with } \beta_i \sim U(\mu, 1 - \mu) \\
\Gamma_{ij} &= \sqrt{\alpha_i \alpha_j} \tau^{|i-j|} (1 - \rho^2) && \forall i, j = 1, \dots, N
\end{aligned} \tag{17}$$

We defined the cross-correlation among the idiosyncratic disturbance elements as the Toeplitz matrix  $\Gamma$ . Where  $\rho$  and  $\tau$  are defined like the serial correlation and cross correlation coefficient, respectively. The coefficient  $\beta_i$  captures the ratio between the variance of the idiosyncratic component,  $\epsilon_{it}$ , and the total variance of  $y_{it}$ . We draw  $\beta_i$  from the uniform distribution  $U(\mu, 1 - \mu)$  with  $\mu = 0.1$ , which sets the ratio between the variances to 0.5 on average. Finally, we set  $\rho = 0.5$ , and  $\tau = 0.5$  in order to represent serial and cross-sectional correlation among the idiosyncratic disturbances, or approximate factor model. Therefore,  $\rho = 0$  and  $\tau = 0$  represent idiosyncratic disturbances mutually uncorrelated, or exact factor model.

## 6.2 Results

We study the accuracy of the empirical Bayesian principal components and the empirical Bayesian maximum likelihood estimates comparing these with their standard counterparts. We define the accuracy of the estimates for the common factors by computing the average

mean squared error statistics. In particular, we compare

$$MSE = J^{-1} \sum_{l=1}^J (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left( \hat{\lambda}_i(l) \hat{f}_t(l) - \lambda_i(l) f_t \right)^2 \quad (18)$$

where the loading  $\lambda_i$  is the  $i^{th}$  row of  $\Lambda$  for  $i = 1, \dots, N$  and the averaging is over the  $J$  samples. Here,  $\hat{\lambda}$  and  $\hat{f}$  may represent the estimates obtained by the empirical Bayesian principal components ( $\hat{\lambda}^{EB/PCA}$ ,  $\hat{f}^{EB/PCA}$ ), standard principal components ( $\hat{\lambda}^{PCA}$ ,  $\hat{f}^{PCA}$ ), empirical Bayesian maximum likelihood ( $\hat{\lambda}^{EB/MLE}$ ,  $\hat{f}^{EB/MLE}$ ) or standard maximum likelihood ( $\hat{\lambda}^{MLE}$ ,  $\hat{f}^{MLE}$ ). For each data generating process described above we present the relative mean squared error statistics  $MSE^{EB/q}/MSE^q$  where  $q = PCA$  or  $MLE$ .

### 6.2.1 Results 1

In Table 1 we show the relative statistics between the empirical Bayesian principal components estimates and the standard principal components estimates. These results are simulated as described in Section 6.1.1.

We deduct the following five conclusions. First, the relative statistics are smaller than one in most cases. It shows that the empirical Bayesian principal components estimates are on average more accurate when compared to the standard principal component estimates. Second, the relative statistics tend to one when the variance of the loadings,  $\sigma_\lambda^2$ , is larger. The intuition behind this result is implied by the strong factor assumptions. As is stated in Section 4.2, high values of  $\sigma_\lambda^2$  leads to consistent estimates, see Bai (2003). Thereby, the relative mean squared error ratio between empirical Bayesian principal components and standard principal components estimates converge to one, see Figure 1. We notice that under weak factor settings, represented by smaller values of  $\sigma_\lambda^2$ , our empirical Bayesian principal components estimator is more accurate than its standard counterpart.

Third, the relative statistics depend on the panel dimensions. The relative performance of the empirical Bayesian principal component estimates is better when  $N \geq T$ . Fourth,

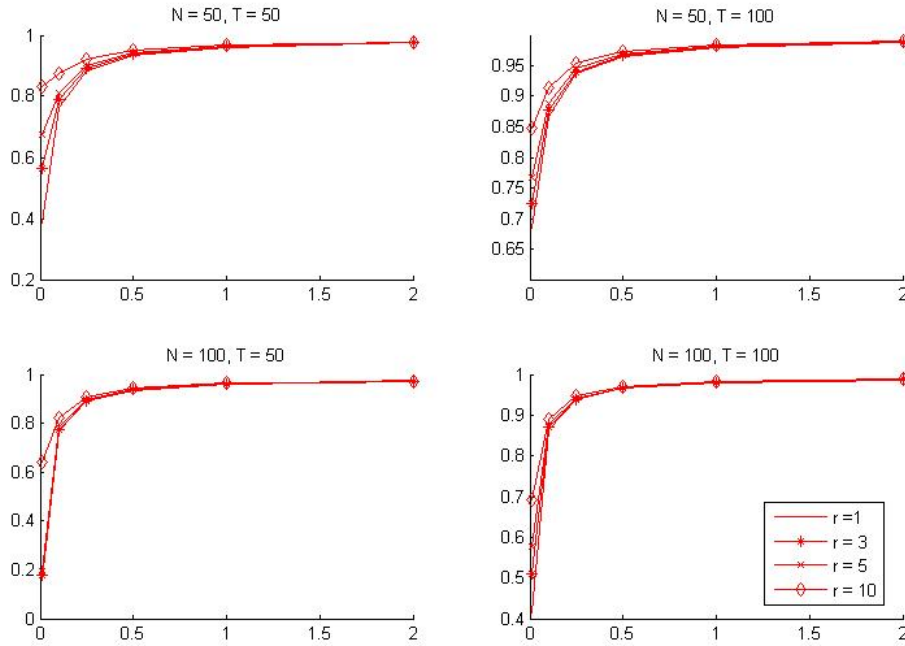


Figure 1:  $MSE^{EB/PCA}/MSE^{PCA}$  for different values of  $\sigma_\lambda^2$ . The DGP and parameters are chosen as discussed in Section 6.1.1. We fixed  $\rho = 0$  and  $\tau = 0$

the relative performances increases (declines) for all the cases when the number of factors  $r$  decreases (increases). Finally, the overall performance of the empirical Bayesian principal components estimates is not affected by different disturbance term's sampling schemes.

The magnitude of the gain depend on the data generating process. The gains of empirical Bayesian principal component estimates can be as large as 98% while the largest loss has only been 1%. For the case  $N \geq T$  the relative gains (losses) are around 52% (0.6%), when  $N \leq T$  the relative gains (losses) are reduced to 28% (4%). As well, for  $r = 3$  the relative gains (losses) are around 64% (0.5%) and, 15% (0.5%) for  $r = 5$ .

Overall, we may conclude that the empirical Bayesian principal component estimates are more accurate than standard principal component estimates under strong and weak factor settings. The relative gains depend on the loadings' sampling scheme and the panel dimensions.

$N$	$T$	$r$	$E$	$\sigma_\lambda^2 = 0.01$	$\sigma_\lambda^2 = 0.1$	$\sigma_\lambda^2 = 0.25$	$\sigma_\lambda^2 = 0.5$	$\sigma_\lambda^2 = 1$	$\sigma_\lambda^2 = 2$
50	50	3	$n$	0.8298	0.9275	0.9598	0.9739	0.9818	0.9861
50	50	3	$s$	0.9251	0.9739	0.9897	0.9986	1.0024	1.0060
50	50	3	$c$	0.8796	0.9302	0.9603	0.9743	0.9821	0.9858
50	50	3	$sc$	0.4585	0.4801	0.8168	0.8796	0.9376	0.9737
50	50	5	$n$	0.8612	0.9346	0.9644	0.9754	0.9820	0.9855
50	50	5	$s$	0.9514	0.9791	0.9901	0.9982	1.0034	1.0059
50	50	5	$c$	0.9093	0.9365	0.9635	0.9748	0.9821	0.9855
50	50	5	$sc$	0.9677	0.9878	0.9947	1.0009	1.0037	1.0061
50	100	3	$n$	0.8818	0.9600	0.9800	0.9881	0.9923	0.9943
50	100	3	$s$	0.9300	0.9743	0.9890	0.9959	1.000	1.0011
50	100	3	$c$	0.9194	0.9622	0.9806	0.9878	0.9919	0.9943
50	100	3	$sc$	0.6789	0.7440	0.8459	0.9176	0.9575	0.9799
50	100	5	$n$	0.8966	0.9633	0.9819	0.9888	0.9927	0.9944
50	100	5	$s$	0.9420	0.9769	0.9908	0.9965	0.9996	1.0010
50	100	5	$c$	0.9350	0.9647	0.9823	0.9887	0.9927	0.9947
50	100	5	$sc$	0.9603	0.9863	0.9928	0.9975	0.9999	1.0015
100	50	3	$n$	0.7515	0.9205	0.9570	0.9706	0.9780	0.9824
100	50	3	$s$	0.8904	0.9658	0.9904	1.0001	1.0071	1.0087
100	50	3	$c$	0.8141	0.9215	0.9573	0.9699	0.9785	0.9810
100	50	3	$sc$	0.0564	0.5198	1.6762	0.9588	0.9580	0.9863
100	50	5	$n$	0.7844	0.9253	0.9584	0.9720	0.9777	0.9818
100	50	5	$s$	0.9198	0.9711	0.9926	1.0010	1.0039	1.0081
100	50	5	$c$	0.8421	0.9268	0.9597	0.9703	0.9784	0.9810
100	50	5	$sc$	0.9373	0.9764	0.9927	1.0014	1.0045	1.0081
100	100	3	$n$	0.8258	0.9576	0.9787	0.9867	0.9904	0.9930
100	100	3	$s$	0.8980	0.9733	0.9912	0.9976	1.0012	1.0031
100	100	3	$c$	0.8746	0.9585	0.9789	0.9872	0.9909	0.9927
100	100	3	$sc$	0.2469	0.6656	0.8547	0.9288	0.9676	0.9858
100	100	5	$n$	0.8398	0.9591	0.9795	0.9867	0.9905	0.9927
100	100	5	$s$	0.9099	0.9752	0.9913	0.9987	1.0013	1.0027
100	100	5	$c$	0.8869	0.9593	0.9798	0.9865	0.9910	0.9930
100	100	5	$sc$	0.9320	0.9769	0.9930	0.9980	1.0014	1.0028

Table 1. Simulations results for  $\frac{MSE^{EB/PCA}}{MSE^{PCA}}$ . The DGP and parameters are chosen as discussed in Section 6.1.1. We fixed  $\theta = 1/2$ . The codes  $n, c, s$  and  $sc$  indicate: uncorrelated disturbances, serial correlated disturbances, cross-section correlated disturbances and serial and cross-section correlated disturbances, respectively.

### 6.2.2 Results 2

In Table 2 we compare the empirical Bayesian maximum likelihood estimates with its standard counterpart, which are simulated as described in the Section 6.1.2. Moreover, we compare the empirical Bayesian maximum likelihood and empirical Bayesian principal component estimates.

We reach the following conclusion. The relative statistics are smaller than one in most of the cases, mainly compared to the empirical Bayesian principal component estimates. It shows that the empirical Bayesian maximum likelihood estimates are on average more accurate when compared to standard maximum likelihood and empirical Bayesian principal component estimates.

The magnitude of the gain depends on the data generating process. When we compare the empirical Bayesian maximum likelihood to the standard maximum likelihood estimates, the gains can be as large as 22%. Moreover, the different panel size specifications leads to different gains. For the case  $N \geq T$  the relative gains are around 15% and for  $N \leq T$  are around to 13%.

On the other hand, when comparig the empirical Bayesian maximum likelihood to the empirical Bayesian principal component estimates the gains can be as large as 48 %. As well, for the case  $N \geq T$  the relative gains are around 42% and for the case  $N \leq T$  are around to 44%.

Overall we may conclude that the empirical Bayesian maximum likelihodd estimates are more accurate than the standard maximum likelihood and empirical Bayesian principal component estimates. The relative gains depend on the sampling scheme for the loadings and the panel dimensions. However, the computational time required by the empirical Bayes PCA is not as sensitive as the empirical Bayes MLE when  $N$  and  $T$  tend to infinity (e.g.  $N = 1000$  and  $T = 1000$ ). For all computations in this study, we have written the codes in MATLAB programming language.



$N$	$T$	$r$	$E$	Panel (i)		Panel (ii)	
				$\frac{MSE^{EB/MLE}}{MSE^{MLE}}$	$\frac{MSE^{EB/MLE}}{MSE^{EB/PCA}}$	$\frac{MSE^{EB/MLE}}{MSE^{MLE}}$	$\frac{MSE^{EB/MLE}}{MSE^{EB/PCA}}$
50	50	3	$n$	0.8967	0.7585	0.8666	0.6530
50	50	3	$s$	0.8268	0.5805	0.7848	0.5570
50	50	3	$c$	0.8662	0.6575	0.8652	0.5955
50	50	3	$sc$	0.8067	0.5891	0.8202	0.6294
50	50	5	$n$	0.8396	0.6216	0.8133	0.6305
50	50	5	$s$	0.7723	0.6069	0.7515	0.5598
50	50	5	$c$	0.7401	0.5166	0.7524	0.5112
50	50	5	$sc$	0.7560	0.6081	0.7613	0.5957
50	100	3	$n$	0.9856	0.7255	0.9590	0.7769
50	100	3	$s$	0.8937	0.6870	0.8750	0.6177
50	100	3	$c$	0.9830	0.6881	0.9490	0.6465
50	100	3	$sc$	0.8803	0.6330	0.8716	0.6204
50	100	5	$n$	0.9615	0.6664	0.9335	0.6940
50	100	5	$s$	0.8493	0.6421	0.8356	0.6199
50	100	5	$c$	0.9092	0.5875	0.9120	0.5426
50	100	5	$sc$	0.8389	0.5996	0.8389	0.5879
100	50	3	$n$	0.8790	0.6898	0.9122	0.6895
100	50	3	$s$	0.7671	0.5595	0.7844	0.5458
100	50	3	$c$	0.8932	0.6787	0.9105	0.6869
100	50	3	$sc$	0.7990	0.5579	0.7767	0.5740
100	50	5	$n$	0.8249	0.6514	0.7825	0.6784
100	50	5	$s$	0.7529	0.5801	0.7318	0.5564
100	50	5	$c$	0.7557	0.6213	0.7568	0.5616
100	50	5	$sc$	0.7335	0.5776	0.7247	0.5931
100	100	3	$n$	0.9872	0.7386	0.9660	0.7663
100	100	3	$s$	0.9007	0.5970	0.8674	0.6318
100	100	3	$c$	0.9723	0.8208	0.9378	0.7155
100	100	3	$sc$	0.9040	0.6154	0.8655	0.6468
100	100	5	$n$	0.8900	0.7455	0.9043	0.7390
100	100	5	$s$	0.8493	0.6380	0.8234	0.6218
100	100	5	$c$	0.8985	0.7135	0.8661	0.6219
100	100	5	$sc$	0.8226	0.6204	0.8191	0.6267

Table 2. Simulations results for the empirical Bayesian maximum likelihood, empirical Bayesian principal component and standard maximum likelihood estimates. The DGP and parameters are chosen as discussed in Section 6.1.2. The Panel (i) shows the results for normal distribution  $N(0, 1)$  and Panel (ii), bimodal distribution for the true-loadings. The codes  $n, c, s$  and  $sc$  indicate: uncorrelated disturbances, serial correlated disturbances, cross-section correlated disturbances and serial and cross-section correlated disturbances, respectively.

## 7 Empirical studies

In this section we discuss the results from two empirical applications.

### 7.1 Nowcasting macroeconomic variables

This study is concerned with the comparison between the empirical Bayesian principal component estimates and the standard principal component and standard maximum likelihood estimates for macroeconomic forecasting based on monthly U.S dataset. Moreover, we compare the empirical Bayesian maximum likelihood estimates to the standard principal component and standard maximum likelihood estimates. The key question is whether and to what extent the empirical Bayes methods improve out-of-sample forecasts when compared with the standard methods.

We consider a similar data set as in Giannone, Reichlin and Small (2008), which includes  $N = 139$  macro economic and financial time series. Table 3 summarizes the categories for which the time series are included. They are constructed as stationary time series following the guidelines in the appendix of Giannone et al. (2008). The resulting panel ranges from January 2000 until December 2015, with  $T = 179$ .

According to Stock and Watson (2012), Giannone et al. (2008) and Koopman and Mesters (2016) we consider the dynamic factor model with  $r = 5$  factors and a diagonal variance matrix for the disturbances. Similar results can be obtained for models with  $r = 3$  factors. Moreover, the factors are assumed to follow an autoregressive process described by equation (7).

The out-of-sample forecasting study for the panel of macroeconomic time series is designed as follows. We forecast each time series 1 and 2 months ahead for 2013-06 until 2015-12. All regressions start in 2000-01, with earlier observations used for initial conditions. In total we compute  $k = 60$  out-of-sample predictions for each horizon. In particular, let the integer  $n$  denote the sample split point (2013-05 for  $h = 1$  and 2013-04 for  $h = 2$ ).

Category	number of series
Employment Situation	38
Foreign Exchange Rates	5
Industrial Production and Capacity Utilization	29
Inventories	11
Money, Loans and Reserves	11
Personal Income/GDP Componets	10
Prices	21
Selected Interest Rates	9
Stock Market	5

Table 3. Summary of the time series that are included in the empirical application

The forecasts are computed for  $n + 1, \dots, n + k$  based on subsamples of the observations  $Y_1, \dots, Y_{n+j-h}$  for  $j = 1, \dots, k$ , where  $Y_j = (y_{j,1}, \dots, y_{j,N})'$ . We estimate the parameter vectors  $\psi^{PCA}, \psi^{EB/PCA}, \psi^{MLE}$  and  $\psi^{EB/MLE}$  for each subsample using the methods explained in Sections 4 and 5. We compute the forecast defined by

$$\hat{Y}_{n+j}^m = \hat{\Lambda}^m \hat{f}_{n+j}^m \quad (19)$$

where  $m = EB/PCA, PCA, EB/MLE$  or  $MLE$  and  $\hat{f}_{n+j}^m = (\hat{f}_{n+j,1}^m, \dots, \hat{f}_{n+j,r}^m)'$  is the forecast for the factors based on the principal component, empirical Bayes principal component, maximum likelihood or empirical Bayes maximum likelihood procedures. These forecasts are computed for all horizons  $h = 1, 2$ .

As a measure of accuracy we consider the mean squared error (MSE) of the out-sample forecasts. In particular we compute for each time series

$$MSE_i = k^{-1} \sum_{j=1}^k (y_{i,n+j} - \hat{y}_{i,n+j}^m)^2 \quad (20)$$

where  $y_{i,n+j}$  and  $\hat{y}_{i,n+j}^m$  are elements of  $Y_{n+j}$  and  $\hat{Y}_{n+j}^m$ , respectively. In this way we compute the mean squared error statistics for the 139 time series for all forecastings horizons.

At the top of table 4 we present the summary statistics for the relative mean squared

error statistics to  $MSE_i^{EB/PCA}/MSE_i^{PCA}$  and  $MSE_i^{EB/PCA}/MSE_i^{MLE}$ . Moreover, at the bottom of the table we present the summary statistics to  $MSE_i^{EB/MLE}/MSE_i^{PCA}$  and  $MSE_i^{EB/MLE}/MSE_i^{MLE}$ . In the last row, the average statistics over all series are displayed.

When we compare the empirical Bayesian principal components estimates to the standard principal components estimates, the gains are 4% and 0.4% for the one month and two months ahead forecasts, respectively. However, standard maximum likelihood estimates are more accurate than the empirical Bayesian principal component estimates.

Furthermore, when we compare the empirical Bayesian maximum likelihood estimates to the standard maximum likelihood ones, the gains are 5% and 4% for the one and two months ahead forecasts. Also, when we compare the empirical Bayesian maximum likelihood estimates to the standard principal components estimates the gains are 11% and 5% for the one and two months ahead forecasts, respectively.

The relative improvements in accuracy of empirical Bayes principal components forecasts declines as we forecast further in the future. As well, the relative accuracy of empirical Bayes maximum likelihood declines as we forecast further into the future. Similar results are obtained by Koopman and Mesters (2016) to empirical Bayes maximum likelihood and standard maximum likelihood because both estimates are based on the same vector autoregressive process for the factors.

Finally, we summarize the relative mean squared error statistics per category. We find largest gains when using the empirical Bayes principal components for Employment Situation, Foreign Exchange Rates, Money, Personal Income, Prices, Interest Rates and Stock Market with respect to the standard principal components estimates. For the other categories we do not obtain much improvements. For the comparison between the empirical Bayes maximum likelihood estimates and their standard counterpart, the largest gains are obtained for Employment Situation, Foreign Exchange Rates, Industrial Production, Prices and Interest Rates. For the other categories we do not obtain much improvements.

Components (mean)	$h = 1$	$h = 1$	$h = 2$	$h = 2$
(i) EB/PCA estimates	$\frac{MSE^{EB/PCA}}{MSE^{MLE}}$	$\frac{MSE^{EB/PCA}}{MSE^{PCA}}$	$\frac{MSE^{EB/PCA}}{MSE^{MLE}}$	$\frac{MSE^{EB/PCA}}{MSE^{PCA}}$
Employment Situation	1.0732	0.9506	1.0631	0.9918
Foreign Exchange Rates	1.0110	0.9531	1.0151	0.9991
Industrial Production	1.0659	0.9957	1.0446	0.9972
Inventories	1.0168	0.9991	1.0248	0.9990
Money, Loans and Reserves	1.1669	0.9518	1.0970	0.9937
Personal Income/GDP Componets	1.0300	0.9609	1.0210	0.9971
Prices	1.1514	0.9583	1.0849	0.9940
Selected Interest Rates	0.9833	0.9651	0.9753	0.9936
Stock Market	0.9321	0.9657	0.9123	0.9959
Global Mean	1.1127	0.9667	1.0840	0.9957
(ii) EB/MLE estimates	$\frac{MSE^{EB/MLE}}{MSE^{MLE}}$	$\frac{MSE^{EB/MLE}}{MSE^{PCA}}$	$\frac{MSE^{EB/MLE}}{MSE^{MLE}}$	$\frac{MSE^{EB/MLE}}{MSE^{PCA}}$
Employment Situation	0.9764	0.8649	0.9902	0.9238
Foreign Exchange Rates	0.8789	0.8286	0.9153	0.9009
Industrial Production	0.9635	0.9000	0.9739	0.9297
Inventories	1.0622	1.0437	1.0402	1.0140
Money, Loans and Reserves	1.0148	0.8277	1.0132	0.9178
Personal Income/GDP Componets	1.0150	0.9469	1.0116	0.9879
Prices	0.8929	0.7432	0.9239	0.8465
Selected Interest Rates	0.8125	0.7975	0.9675	0.9857
Stock Market	1.0054	1.0416	0.9956	10868.
Global Mean	0.9526	0.8882	0.9557	0.9548

Table 4. Relative mean squared error statistics for out-sample forecasting using empirical Bayes principal component, empirical Bayes maximum likelihood, standard principal component and standard maximum likelihood methods. The results summarize the statistics  $MSE_i^{EB/PCA}/MSE_i^{PCA}$ ,  $MSE_i^{EB/PCA}/MSE_i^{MLE}$ ,  $MSE_i^{EB/MLE}/MSE_i^{PCA}$  and  $MSE_i^{EB/MLE}/MSE_i^{MLE}$  for  $i = 1, \dots, 139$  and forecasts horizons  $h = 1, 2$ .

## 7.2 Portfolio optimization

The mean-variance optimization approach has been the most rigorous way to pick stocks where to invest. The two fundamental ingredients are the expected return for each stock, which represents the portfolio manager's ability to forecast future price movements, and the covariance matrix of stock returns, which represents risk control.

Estimating the covariance matrix of stock returns has always been one of the toughest points. Even if the sample covariance matrix is easy to compute, it contains a lot of estimation error when the number of data points is of comparable or even smaller order than the number of individual stocks; this is the common situation in financial applications. Ledoit and Wolf (2004) proposed a shrinkage estimator of the covariance matrix that combine the sample covariance matrix  $S$  and a highly structured estimator, denoted by  $\Pi$ , in a linear combination described as follows,  $\delta\Pi + (1 - \delta)S$  where  $\delta$  is a shrinkage target. See Ledoit and Wolf (2004) for a detailed discussion to the estimation procedure.

Another approach to estimate the covariance matrix of stock returns is to use statistical factors, such as principal components or maximum likelihood approaches, with the total number of factors being on the order of 5. In particular, in this study we consider the empirical Bayesian principal components and empirical Bayesian maximum likelihood estimates in order to compute the covariance matrix of stock returns.

The key question is whether the empirical Bayes methods improve the covariance matrix estimated in a portfolio optimization problem.

Formally, the optimization problem of the manager is defined as follow

$$\begin{aligned} \text{Minimize: } & \sigma_P^2 = \frac{1}{2}x'\Sigma x \\ \text{such that: } & x'\alpha \geq \alpha_P \\ & x'1_N = 1 \end{aligned} \tag{21}$$

where  $\alpha$  is the vector expected stock returns,  $\alpha_P = x'\alpha$  is the expected return on portfolio

and  $\Sigma$  the covariance matrix of stock returns.

As far as the equity size  $N$  is concerned, we employ  $N = 30, 50, 100$  and  $150$ . The stocks are provided by monthly U.S. stock data (NYSE), starting in January 2005 until December 2015. We present the solutions to the expected portfolio returns and variance portfolio of the problem (21), where the covariance matrix is performed using the empirical Bayesian principal component, empirical Bayesian maximum likelihood and shrinkage estimates. Moreover, we introduced the Information Ratio (IR)<sup>1</sup> to measure the portfolio performances for each estimate with respect to the benchmark<sup>2</sup>. The higher the Information Ratio the better performance implied by the estimators when compared to the standard benchmark.

In table 5, we present the results. First, in most cases, the empirical Bayesian principal componentss estimates yield the highest portfolio returns. However, the empirical Bayesian maximum likelihood achieves the lower standard deviation of the portfolio returns. Second, for  $N \leq 50$  the empirical Bayesian principal components achieves the higher Information Ratios. For the case  $N > 50$  the results are not clear. Finally, in most cases, the empirical Bayesian principal component and empirical Bayesian maximum likelihood estimates yield the highest portfolio expected returns.

Overall, we may conclude that the empirical Bayes estimates in most of the cases increase substantially the realized Information Ratio of the portfolio manager. Mainly with a small stock portfolio.

---

<sup>1</sup>We define the Information Ratio as follow,  $IR = \frac{\alpha_P - \alpha_B}{\sigma_{P,B}}$ , where  $\alpha_P$  is the expected return of the portfolio,  $\alpha_B$  is the expected return of the benchmark and,  $\sigma_{P,B}$  is the standard deviation of the difference between returns of the portfolio and the returns of the benchmark.

<sup>2</sup>The benchmark is implied by the standard solution obtained in Markowitz (1952).

	<i>EB/PCA</i> <i>r = 3</i>	<i>EB/PCA</i> <i>r = 5</i>	<i>EB/MLE</i> <i>r = 3</i>	<i>EB/MLE</i> <i>r = 5</i>	<i>Shrink</i>	<i>Benchmark</i>
<i>N = 30</i>						
Portfolio Return	0.1332	0.1330	0.1208	0.1184	0.0400	0.0324
S.D Portfolio	0.6029	0.7483	0.0378	0.0338	0.0612	0.0403
IR	0.4390	0.4390	0.4231	0.4052	0.3273	
<i>N = 50</i>						
Portfolio Return	0.1430	0.1430	0.1223	0.1157	0.0359	0.0483
S.D Portfolio	0.4643	0.5983	0.0279	0.0226	0.0478	0.0249
IR	0.4145	0.4145	0.3760	0.3632	-0.5166	
<i>N = 100</i>						
Portfolio Return	0.1383	0.1383	0.1341	0.1290	0.0557	0.0353
S.D Portfolio	0.3420	0.4263	0.0793	0.0267	0.0377	0.0082
IR	0.4579	0.4579	0.4595	0.4692	0.8191	
<i>N = 150</i>						
Portfolio Return	0.1298	0.1298	0.1268	0.1199	0.0524	0.1734
S.D Portfolio	0.2797	0.3549	0.0569	0.0186	0.0315	0.1792
IR	-0.1645	-0.1645	-0.1807	-0.2189	-0.6811	

Table 5. Mean-Summary Statistics for expected portfolio return and variance portfolio. This table presents means, and standard deviations of realized expected portfolio returns. The out-of-sample period is 01.2005 until 12.2015, yielding 120 monthly data. The size of the benchmark is denoted by  $N$ . "Benchmark" denotes the sample covariance matrix; "Shrink" denotes the shrinkage estimator proposed by Ledoit and Wolf(2003), and "EB/PCA" and "EB/MLE" the empirical Bayes estimators. All numbers are annualized



## 8 Conclusion

In this paper, we have developed an empirical Bayes principal components estimator for dynamic factor models in order to deal with the weak factor assumptions. In particular, we designed a simulation study in order to evaluate our estimator with its standard principal components counterpart. Moreover, we provide a comparison between the empirical Bayes maximum likelihood and standard maximum likelihood methods.

The methods are evaluated in a Monte Carlo study for dynamic factor models with different dimensions and number of factors. Our empirical Bayesian principal component estimates for common factors outperform the standard principal component estimates for the data generating process considered. Moreover, the empirical Bayesian maximum likelihood estimates for common factors outperform the standard maximum likelihood estimates. The results hold for different consideration of non-Gaussian process and weak factor settings. We have further illustrated this methods in an empirical application for forecasting macroeconomic time series and portfolio optimization problem. Our empirical Bayesian principal component estimates outperform the standard principal component estimates when compared to the mean squared error for the inner product of the macroeconomic forecast estimates. Second, in the portfolio optimization problem the covariance matrix of the stock returns estimated by the empirical Bayes methods achieve, in most cases, the highest Information Ratio and the highest expected return for the portfolio manager.

## References

- Aguilar, O. and West, M.: 2000, Bayesian Dynamic Factor Models and Portfolio Allocation, *Journal of Business and Economic Statistics* **18**, 338–357.
- Bai, J.: 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71**, 135–171.
- Bai, J. and Li, K.: 2012, Statistical Analysis of Factor Models of High Dimension, *The Annals of Statistics* **40**, 436–465.
- Bai, J. and Li, K.: 2015, Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension, *Review of Economics and Statistics* . forthcoming.
- Bai, J. and Ng, S.: 2000, Determining the Number of Primitive Shocks in Factor Models, *Journal of Business and Economic Statistics* **25**, 52–60.
- Bai, J. and Ng, S.: 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70**, 191–221.
- Bai, J. and Ng, S.: 2008, Large Dimensional Factor Analysis, *Foundations and Trends in Econometrics* **3**, 89–163.
- Bishop, C. M.: 1999, Bayesian PCA, in M. S. Kearns, S. A. Solla and D. A. Cohn (eds), *Advances in Neural Information Processing Systems*, MIT Press, Cambridge.
- Bishop, C. M.: 2006, *Pattern Recognition and Machine Learning*, Springer, New York.
- Boivin, J. and Ng, S.: 2003, Are More Data Always Better for Factor Analysis, *NBER* **11285**. Working Paper.
- Doz, C., Giannone, D. and Reichlin, L.: 2012, A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models, *Review of Economics and Statistics* **94**, 1014–1024.

- Doz, C. and Reichlin, L.: 2011, A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering, *Journal of Econometrics* **164**, 188–205.
- Durbin, J. and Koopman, S. J.: 2012, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D.: 2012, *Bayesian Data Analysis, third edition*, Chapman & Hall, New York.
- Geweke, J. F.: 1989, Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica* **57**, 1317–1339.
- Ghahramani, Z. and Hinton, G.: 1996, Parameter Estimation for Linear Dynamical Systems, *University of Toronto* . Technical Report CRG-TR-96-2.
- Giannone, D., Reichlin, L. and Small, D.: 2008, Nowcasting Gdp: the Real-time Informational Content of Macroeconomic Data, *Journal of Monetary Economics* **55**, 665–676.
- Kaufman, S. and Schumacher, C.: 2013, Bayesian Estimation of Sparse Dynamic Factor Models with Order-independent Identification, *Swiss National Bank* **13**. Working Paper.
- Koopman, S. J. and Mesters, G.: 2016, Empirical bayes methods for dynamic factor models, *Review of Economics and Statistics* . forthcoming.
- Ledoit, O. and Wolf, M.: 2004, Honey, I Shrunk the Sample Covariance Matrix, *Journal of Portfolio Management* **4**, 110–119.
- Markowitz, H.: 1952, Portfolio Selection, *Journal of Finance* **7**, 77–91.
- Marron, J. S. and Wand, M. P.: 1992, Exact Mean Integrated Squared Error, *The Annals of Statistics* **20**, 712–736.

- Nakajima, J. and West, M.: 2013, Bayesian Analysis of Latent Threshold Dynamic Models, *Journal of Business and Economic Statistics* **31**, 151–164.
- Onatski, A.: 2012, Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors, *Journal of Econometrics* **168**, 244–258.
- Onatski, A.: 2015, Asymptotic Analysis of the Squared Estimation Error in Misspecified Factor Models, *Journal of Econometrics* **186**, 388–406.
- Stock, J. H. and Watson, M. W.: 2006, Forecasting with Many Predictors, in G. C. Elliot, Graham and A. Timmermann (eds), *Handbook of Economic Forecasting*, North Holland, Michigan.
- Stock, J. H. and Watson, M. W.: 2011, Dynamic Factor Models, in M. P. Clements and D. F. Hendry (eds), *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford.
- Stock, J. and Watson, M.: 1989, New Indexes of Coincident and Leading Economic Indicators, in O. Blanchard and S. Fischer (eds), *NBER Macroeconomics Annual*, MIT Press, Cambridge.
- Stock, J. and Watson, M.: 2012, Generalized Shrinkage Methods for Forecasting Using Many Predictors, *Journal of Business and Economic Statistics* **30**, 481–493.

# Appendices

## A EM Algorithm

The EM algorithm is an iterative maximum likelihood estimation for many state space models, in which the direct maximization of the likelihood function is intractable or computationally difficult. For a given parameter vector  $\psi_0$ . The procedure relies on two steps

E-step: the expectation of the log-likelihood conditional on the data is calculated using the estimates from the previous iteration,  $\psi_j$ , for  $j = 0, 1, \dots$

M-step: the new parameter,  $\psi_{j+1}$ , is estimated through the maximization of the expected log-likelihood expressed by the equation (10) with respect to  $\psi$ , for  $j = 0, 1, \dots$ . In particular, consider the Kalman filter expressed by the equations in (8), the parameters are estimated as follow

(i) Loading matrix:  $\Lambda^{new} = \left( \sum_{t=1}^T y_t \hat{f}_t' \right) \text{inv} \left( \sum_{t=1}^T \hat{f}_t \hat{f}_t' \right)$

(ii) Noise covariance:  $\Omega^{new} = \frac{1}{N} \sum_{t=1}^T \left( y_t y_t' - \Lambda^{new} \hat{f}_t y_t' \right)$

(iii) State dynamic matrix:  $H^{new} = \left( \sum_{t=2}^T \hat{f}_t \hat{f}_{t-1}' \right) \text{inv} \left( \sum_{t=1}^T \hat{f}_{t-1} \hat{f}_{t-1}' \right)$

(iv) State noise covariance:  $\Sigma_\eta^{new} = \frac{1}{T-1} \left( \sum_{t=2}^T \hat{f}_t \hat{f}_t' - H^{new} \sum_{t=2}^T \hat{f}_{t-1} \hat{f}_{t-1}' \right)$

where  $\hat{f}_t$  is the  $r \times 1$  factor vector estimated by the state smoothing recursion given by the equations in (9).

The evaluation of  $L(y; \psi_{j+1})$  relies on the Kalman filter and smoother. The two steps are repeated until the convergence criteria is satisfied or reaches a local maximum. In particular, we control the EM's convergence criteria by looking at  $J_m = \frac{\log L(y; \psi_{j+1}) - \log L(y; \psi_j)}{|\log L(y; \psi_j)|}$ . And, we stop after  $M$  iterations if  $J_m < 10^{-5}$ .

## B Assumptions on Maximum Likelihood - Bayes estimation

Consider the dynamic factor model expressed in (1) and (4). We assume that,

- (i) **Common Factors:** The  $r \times 1$  vector of common factors  $f_t$  is stationary and restricted such that  $\text{Var}(f_t) = \text{inv}(I_r - HH')$ . The initial state is given by  $f_1 \sim N(0, \text{inv}(I_r - HH'))$ , that is  $f_1 = 0$  and  $P_1 = \text{inv}(I_r - HH')$ . The common innovations  $\eta_t$  and the initial state  $f_1$  are mutually independent and distributed independent of the loadings vectors  $\lambda_i$ , and disturbances  $\epsilon_{i,s}$  for all  $i = 1, \dots, N$  and  $s, t = 1, \dots, T$ .
- (ii) **Loading vectors:** The loading vectors  $\lambda_i$  are distributed normally and independently with mean  $\delta$  and positive homoscedasticity variance  $\Sigma$ . Moreover, loadings and disturbances are mutually uncorrelated.
- (iii) **Disturbances:** The disturbance vector  $\epsilon_t$  is normally and independently distributed with mean zero and diagonal matrix  $\Omega = \text{diag}\{\sigma_{\epsilon,i}^2, \dots, \sigma_{\epsilon,N}^2\}$