

Estimation Methods for Dynamic Factor Models of Large Dimension

Metodi di Stima per Modelli a Fattori Dinamici di Grande Dimensione

Massimiliano Marcellino
IEP-Università Bocconi, IGIER and CEPR,
e-mail: massimiliano.marcellino@uni-bocconi.it

Riassunto: In questo lavoro prendiamo in esame tre metodi alternativi di stima di modelli a fattori dinamici con elevato numero di osservazioni: componenti principali statiche (SW), componenti principali dinamiche (FHLR) e *subspace algorithms* (KM). Le nostre simulazioni evidenziano come SW e KM ottengano risultati molto simili e si dimostrino sufficientemente attendibili anche nel caso di campioni ristretti; al contrario, la resa di FHLR è di poco peggiore, in una serie di esperimenti diversamente costruiti. Una seconda serie di simulazioni valuta la capacità previsiva dei fattori SW, a confronto con i metodi tradizionali di serie storiche. I fattori SW danno in genere buoni risultati, anche se è necessaria una dimensione temporale ampia perché i vantaggi in termini di previsione siano sistematici e considerevoli.

Keywords: Dynamic Factor Models, Principal Components, Dynamic Principal Components, Subspace algorithms on state space form.

1. Introduction

The factor model has attracted considerable attention in the statistical and econometric literature because it provides a good representation for several types of variables. The main underlying hypothesis is that the variables are driven by a set of common forces, the factors, and by idiosyncratic components. In the statistical literature, typically the factors are assumed to be i.i.d., the idiosyncratic components i.i.d. and uncorrelated across variables, and the number of variables under analysis rather limited. Recent developments in factor analysis have tried to relax these restrictions.

In particular, Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1993) allowed for some correlation across the idiosyncratic components and developed estimation methods that are suited also for large datasets, maintaining though the hypothesis of i.i.d. variables.

Geweke (1977) and Sargent and Sims (1977) studied a dynamic factor model for a limited number of series, with further developments due to Stock and Watson (1989, 1991), Quah and Sargent (1993) and Camba-Mendez et al (2001). Yet, all these methods are not suited when the number of variables is very large due to the computational cost, even when a sophisticated EM algorithm is used for optimization, as in Quah and Sargent (1993).

Stock and Watson (2002a, 2002b) (SW), Forni, Hallin, Lippi and Reichlin (1999, 2000) (FHLR), and Kapetanios and Marcellino (2003, KM) have developed methods that are suited to analyze dynamic factor models even in the presence of a very large number of variables, possibly tending to infinity.

More specifically, SW have suggested a non-parametric principal component based estimation approach in the time domain, and shown that principal components can estimate consistently the factor space asymptotically. FHLR have developed an alternative non-parametric procedure in the frequency domain, based on dynamic principal components (see Chapter 9 of Brillinger (1981)), that incorporates an explicitly dynamic element in the construction of the factors. KM have suggested a third approach for factor estimation that retains the attractive framework of a parametric state space model but is computationally feasible for very large datasets because it does not use maximum likelihood but linear algebra methods, based on subspace algorithms used extensively in engineering, to estimate the state.

In this paper we review these three approaches to factor model estimation (Section 2); compare their small sample performance through simulation experiments (Section 3); and provide directions for further research in this field (Section 4).

2. The alternative factor estimation methods

In this section we review the three alternative estimation methods for dynamic factor models of large dimension, namely, dynamic principal components (subsection 1), static principal components (subsection 2), and subspace algorithm (subsection 3). Additional details and proofs can be found in the original papers.

2.1 Dynamic principal components (FHLR)

Frequency domain analysis of the dynamic factor model was recently proposed by Forni and Reichlin (1996, 1997, 1998), Forni, Hallin Lippi and Reichlin (2000). The model they adopt is

$$x_{it} = b_i'(L)u_t + \xi_{it}, i = 1, \dots, N, t = 1, \dots, T, \quad (1)$$

where x_{it} is a stationary univariate random variable, u_t is a $q \times 1$ vector of common factors, $\chi_{it} = x_{it} - \xi_{it}$ is the common component of x_{it} and ξ_{it} is its idiosyncratic component. More precisely, u_t is an orthonormal white noise process, so that $\text{var}(u_{jt}) = 1$, $\text{cov}(u_t, u_{t-k}) = 0$ and $\text{cov}(u_{jt}, u_{st-k}) = 0$ for any $j \neq s, t$ and k . $\xi_N = \{\xi_{1t}, \dots, \xi_{Nt}\}'$ is a wide sense stationary process for any N , and $\text{cov}(\xi_{it}, u_{st-k}) = 0$ for any j, s, t and k . $b_i(L)$ is a $q \times 1$ vector of square summable, bilateral filters, for any i . Hence, $x_{Nt} = \{x_{1t}, \dots, x_{Nt}\}'$ is also a stationary vector process.

FHLR also require x_{Nt} , ξ_{Nt} and χ_{Nt} to have rational spectral density matrices $\sum_N^x$, $\sum_N^\xi$ and $\sum_N^\chi$ respectively. To achieve (asymptotic) identification, they assume that the first (largest) idiosyncratic dynamic eigenvalue, λ_{N1}^ξ is uniformly bounded and that the first (largest) q common dynamic eigenvalues, $\lambda_{N1}^\chi, \dots, \lambda_{Nq}^\chi$, diverge, where dynamic eigenvalues are the eigenvalues of the spectral density matrix, see e.g. Brillinger (1981,

Chap. 9). In words, the former condition limits the effects of ξ_{it} on other cross-sectional units. The latter, instead, requires u_i to affect infinitely many units.

Let us assume for the moment that the number of common factors is known. Then, FHLR suggest to estimate the common component of x_{it} with the following step-wise procedure.

(i) Estimate the spectral density matrix of x_{Nt} by $\sum_N^T(\theta_h) = \sum_{k=-M}^M \Gamma_{Nk}^T w_k e^{-ik\theta_h}$, $\theta_h = 2\pi h/(2M+1)$, $h=0, \dots, 2M$, where Γ_{Nk}^T is the sample covariance matrix of x_{Nt} , and x_{Nt-k} , w^k is the Bartlett lag window of size $M(w_k = 1 - k/(M+1))$, and M diverges but M/T tends to zero.

(ii) Calculate the first q eigenvectors of $\sum_N^T(\theta_h)$, $p_{Nj}^T(\theta_h)$, $j=1, \dots, q$, for $h=0, \dots, 2M$.

(iii) Define $p_{Nj}^T(L)$ as $p_{Nj}^T(L) = \sum_{k=-M}^M p_{Nj,k}^T L^k$, $p_{Nj,k}^T = \frac{1}{2M+1} \sum_{h=0}^{2M} p_{Nj}^T(\theta_h) e^{ik\theta_h}$, $k=-M, \dots, M$.

$p_{Nj}^T(L)x_{Nt}$, $j=1, \dots, q$, are the first q estimated dynamic principal components of x_{Nt} .

(iv) The estimated common component of x_{it} , χ_{it} , is the projection of x_{it} on the first q estimated past, present and future dynamic principal components.

FHLR prove that, under mild conditions, the estimator of the common component is consistent when N and T diverge. In practice, M and the number of leads (s) and lags

(g) of $p_{Nj}^T(L)x_{Nt}$ to be included in the projection can be either chosen a priori or determined by minimizing the information criterion

$$\frac{T}{N} \sum_{i=1}^N \log \hat{\sigma}_i + 2g(q+s+1) \quad (2)$$

where $\hat{\sigma}_i$ is the estimated variance of the residuals from the i^{th} equation. Since the properties of the latter method are not known, in the simulation experiments we fix $M=3$, $s=3$ and $g=3$, but a sensitivity analysis indicated that the results are rather robust to this choice. Notice that for a static factor model M can be set to zero.

Once the common component is estimated, the idiosyncratic one is obtained simply as a residual, namely, $\hat{\xi}_{it} = x_{it} - \hat{\chi}_{it}$.

Finally, we have to discuss the determination of the number of factors. No formal testing procedures are available at the moment. FHLR suggest: (i) to estimate the spectral density matrix of x_{jt} , $j=1, \dots, N$; (ii) to calculate the dynamic eigenvalues for a grid of frequencies, $\lambda_{j\theta}^x$; (iii) to chose q on the basis of two properties: (a) when j increases the average over frequencies of the first q dynamic eigenvalues diverges, while the average of the $(q+1)^{\text{th}}$ does not; (b) for $j=N$, there should be a big gap between the variance of x_{Nt} explained by the first q dynamic principal components and that explained by the $(q+1)^{\text{th}}$ principal component. An information criterion could be also used. In particular, the criterion that FHLR suggest for selection of g and s , equation (2), could be minimized also with respect to q .

2.2 Static principal components (SW)

A competing procedure for the large N case was developed by Stock and Watson (2002a, 2002b). The model by SW, in its time invariant formulation, can be written as

$$x_{Nt} = \Lambda f_t + \xi_{Nt}, \quad (3)$$

where f_t is an $r \times 1$ vector of common factors. Contrary to the specification by FHLR, the factors are not required to be uncorrelated in time, and they can be also correlated with the idiosyncratic component, only $\text{var}(f_t) = I$ is imposed for identification. Precise moment conditions on f_t and ξ_{Nt} , and requirements on the loadings, are given in SW.

The specification in (3) is related to the one by FHLR in (1). When $b_i(L)$ is unilateral and of finite order b, say $b_i(L) = b_{0i} - b_{1i}L - b_{bi}L^b$, the model in (1) can be written as in (3), where $f_t = (u_t, u_{t-1}, u_{t-b})$ and the i-th row of Λ has elements b_{0i}, b_{1i}, b_{bi} . Hence, $r = q(b+1)$, and the factors f_t are dynamically singular, i.e., the spectral density matrix of f_t has rank q.

The starting point in SW's approach is the estimation of the factors, f_t , and the loadings Λ . They define the estimators $\hat{f}_t$ as the minimizers of the objective function

$$V_{NT}(f, \Lambda) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \Lambda_i f_t)^2. \quad (4)$$

Under the hypothesis of k common factors, it turns out that the optimal estimators of the factors are ($\sqrt{T}$ times) the k eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix $N^{-1} \sum_{i=1}^N x_i x_i'$, where $x_i = (x_{i1}, \dots, x_{iT})$. Moreover, the optimal estimators of Λ are the OLS estimators of the coefficients in a regression of x_i on the k estimated factors $\hat{f}_t, i = 1, \dots, N$.

SW prove that when $k=r$, i.e. the exact number of common factors is assumed, $\hat{f}_t$ converges in probability to f_t , apart from a full rank $r \times r$ transformation matrix, H. When $k > r$, $k-r$ estimated factors are redundant linear combinations of the elements of f_t . The rate of convergence is T^b , where $b < \min(\rho/2 - 1, 1)$, and ρ is the limiting value of $\log(N)/\log(T)$. Hence, quicker convergence is achieved when the number of cross sectional units grows faster than that of temporal observations. Additional inferential results are provided by Bai (2003).

When $k=r$, the estimators of the loading converge to ΛH^{-1} , as can be easily demonstrated when they are considered as OLS estimators in the regression of x_N on $\hat{f}_t$. Thus, a consistent estimator of the ith common component can be obtained as $\hat{\chi}_{it} = \hat{\Lambda}_i \hat{f}_t$. When $k > r$, it can be shown that it is still possible to obtain consistent

estimators of the common component; this follows from Theorem 2 of SW. A natural choice for the estimator of the idiosyncratic component is $\hat{\xi}_{it} = x_{it} - \hat{\chi}_{it}$.

Finally, for the determination of the number of factors, SW suggest to determine r by minimizing an information criterion. Their proposed measure is

$$\log \hat{\sigma}_i + qw \log T / T^b, \quad (5)$$

where w is a positive constant. They prove that such a criterion leads to a consistent choice of r . Yet, from their simulation experiments, more standard criteria like the AIC or BIC perform better. Bai and Ng (2002) further refined the information criterion approach within the SW framework.

2.3 Subspace algorithm (KM)

Following Deistler and Hannan (1988), Kapetanios and Marcellino (2003) consider the following state space model:

$$\begin{aligned} x_{Nt} &= Cf_t + D^* u_t, & t = 1, \dots, T \\ f_t &= Af_{t-1} + B^* v_{t-1}, \end{aligned} \quad (6)$$

where u_t and v_t are multivariate, mutually uncorrelated, standard orthogonal white noise sequences of dimension, respectively, N and r , D^* is assumed to be nonsingular, the eigenvalues of A are all in modulus smaller than one. No assumptions are required on the C matrix as long as N is finite, while basically the same assumptions as in FHLR are necessary for identification of the common factors when N diverges. When N diverges KM also show that the assumptions can be relaxed to allow for temporal and cross-sectional correlation of the idiosyncratic components, u_t .

The first equation of the model in (6) is virtually identical to the one used by SW, compare equation (3). But there are two relevant differences with respect to SW. First, the model is complemented by a parametric specification for the factors. Second, the factors are driven by lagged rather than contemporaneous errors. The latter hypothesis is quite important to obtain consistent estimators of the factors also for finite N . The consequences of its violation are evaluated empirically in the Monte Carlo experiments, but we can anticipate that they are minor because the proposed estimator remains consistent for Af_{t-1} . Notice also that the model in (6) can be easily modified to allow for a dynamic effect of the factors on the variables, as in the case of SW, the only difference being a reduced rank variance covariance matrix for the extended set of factors.

As we have mentioned in the introduction, maximum likelihood techniques, possibly using the Kalman filter, may be used to estimate the parameters of the factor model in (6). Yet, for large datasets this is very computationally intensive. Quah and Sargent (1993) developed an EM algorithm that allows to consider up to 50-60 variables, but it is still so time-consuming that it is not feasible to evaluate its performance in a simulation experiment. KM note that a convenient solution is provided by subspace algorithms that avoid expensive iterative techniques and instead rely on matrix algebraic methods to provide estimates for the factors as well as the parameters of the state space representation. A review of existing subspace algorithms is given by Bauer (1998) in an

econometric context. Another review with an engineering perspective may be found in Van Overschee and De Moor (1996).

To apply KM's proposed subspace algorithm, it is convenient to adopt the prediction error representation of the model in (6), given by (see e.g. Deistler and Hannan (1988, Chapter 1)):

$$\begin{aligned} x_{Nt} &= Cf_t + Du_t, \\ f_t &= Af_{t-1} + Bu_{t-1}. \end{aligned} \quad t = 1, \dots, T. \quad (7)$$

In turn, the model in (7) can be written as:

$$X_t^f = OKX_t^p + \varepsilon E_t^f \quad (8)$$

where $X_t^f = (x'_{Nt}, x'_{Nt+1}, x'_{Nt+2}, \dots)'$, $X_t^p = (x'_{Nt-1}, x'_{Nt-2}, \dots)'$, $E_t^f = (u'_t, u'_{t+1}, \dots)'$,
 $O = [C', A'C', (A^2)C', \dots]$, $K = [\overline{B}, (A - \overline{B}C)\overline{B}, (A - \overline{B}C)^2\overline{B}, \dots]$, $\overline{B} = BD^{-1}$

and $\varepsilon = \begin{pmatrix} D & 0 & \dots & 0 \\ CB & D & \ddots & \vdots \\ CAB & \ddots & \ddots & 0 \\ \vdots & \dots & CB & D \end{pmatrix}$.

The derivation of this representation is simple once we note that (i) $X_t^f = Of_t + \varepsilon E_t^f$ and (ii) $f_t = KX_t^p$. The best linear predictor of the future of the series at time t is given by OKX_t^p . The state is given in this context by KX_t^p at time t. The task is therefore to provide an estimate for K.

The representation in (8) involves infinite dimensional vectors. In practice, truncation is used to end up with finite sample approximations given by $X_{s,t}^f = (x'_{Nt}, x'_{Nt+1}, x'_{Nt+2}, \dots, x'_{Nt+s-1})'$ and $X_{p,t}^p = (x'_{Nt-1}, x'_{Nt-2}, \dots, x'_{Nt-p})'$. Then an estimate of $F=OK$ may be obtained by regressing $X_{s,t}^f$ on $X_{p,t}^p$.

Following Larimore (1983), KM apply a singular value decomposition (SVD) to $\hat{\Gamma}^f \hat{F} \hat{\Gamma}^p$, where $\hat{F}$ is the estimated F matrix, and $\hat{\Gamma}^f$ and $\hat{\Gamma}^p$ are the sample covariances of $X_{s,t}^f$ and $X_{p,t}^p$ respectively. These weights are used to determine the importance of certain directions in F. Then, the matrix K is estimated by $\hat{K} = \hat{S}_k^{1/2} \hat{V}_k' \hat{\Gamma}^{p^{-1/2}}$, where $\hat{U} \hat{S} \hat{V}'$ represents the SVD of $\hat{\Gamma}^{f^{-1/2}} \hat{F} \hat{\Gamma}^{p^{1/2}}$, $\hat{V}_k$ denotes the matrix containing the first k columns of $\hat{V}$ and $\hat{S}_k$ denotes the heading k×k submatrix of $\hat{S}$. $\hat{S}$ contains the singular values of $\hat{\Gamma}^{f^{-1/2}} \hat{F} \hat{\Gamma}^{p^{1/2}}$ in decreasing order.

Finally, the factor estimators are given by

$$\hat{f} = \hat{K} X_t^p. \quad (9)$$

For what follows it is important to note that the choice of the weighting matrices is important but not crucial for the asymptotic properties of the estimation method. They

are only required to be nonsingular. Therefore, in the simulation experiments we use identity matrices instead of the covariance matrices. A second point to note is that consistent estimation of the factor space requires the parameter p to increase at a rate greater than $\ln(T)^\alpha$, for some $\alpha > 1$ that depends on the maximum eigenvalue of A , but at a rate lower than $T^{1/3}$. A range of α between 1.05 and 1.5 for this parameter provides a satisfactory performance, and the value $\alpha = 1.25$ is used for the simulation experiments. For consistency of the estimator the parameter s is also required to be set so as to satisfy $sN \geq k$. As N is very large when this method is applied, this restriction is not binding and we can set $s = 1$. This is particularly relevant for forecasting applications, because with $s = 1$ no future information is used to estimate the factors.

KM also derive the asymptotic distribution of the estimators of the factors, showing that it is standard normal after a proper standardization, and propose an information criterion for consistent determination of the number of factors, along the lines of Bai and Ng (2002).

Once estimates of the factors have been obtained, if estimates of the parameters (including the factor loadings) are subsequently required, least squares methods can be used. The resulting estimators have been proved to be $\sqrt{T}$ -consistent and asymptotically normal in a related context by Bauer (1998).

Finally, we have to discuss a modification of the KM method to deal with datasets with a larger cross-sectional than temporal dimension, i.e. $N > T$. The problem arises because the least squares estimator of $F = OK$ in (8) does not exist due to the rank deficiency of $X^{p'} X^p$, where $X^p = (X_1^p, \dots, X_T^p)'$. Yet, an estimator of the states $x^p K'$ could be obtained if we had an estimator of $x^p F'$ and used a SVD of that quantity. The least squares estimator of $x^p F'$ is given by

$$x^p F' = X^p (X^{p'} X^p)^+ X^{p'} X^f \quad (10)$$

where $X^f = (X_1^f, \dots, X_T^f)$ and A^+ denotes the unique Moore-Penrose inverse of matrix A . However, when the row dimension of X^p is smaller than its column dimension, $X^p (X^{p'} X^p)^+ X^{p'} = I$ implying that $x^p F' = X^f$. A decomposition of X^f is then easily seen to be similar, but not identical, to the eigenvalue decomposition of the covariance matrix of X^f which is the SW principle component method.

As an alternative, KM suggest an SVD of

$$X^{f'} X^p (X^{p'} X^p)^+ = \hat{U} \hat{S} \hat{V}' \quad (11)$$

Then, the estimated factors are given by $\hat{K} X_t^p$ where $\hat{K}$ is obtained as before but using the SVD of $X^{f'} X^p (X^{p'} X^p)^+$. This alternative estimation method is a generalisation of the previous one, and if $Np < T$ it reduces to that method.

No theoretical results are available for the factor estimators based on (11). Therefore, their performance is evaluated in the Monte Carlo experiments in the next Section, and compared to that of FHLR and SW.

3. A comparison of the factor estimation methods

In this section we present the results of a set of simulation experiments to investigate the small sample properties of the three competing factor extraction methods, i.e. static principal components (PC, SW), dynamic principal components (DPC, FHLR), and the state space approach (SSS, KM). Additional details, other experiments, and results for different combinations of N and T can be found in KM.

3.1. Experimental design

The data generating process (DGP) we use is:

$$\begin{aligned} x_{Nt} &= Cf_t + \varepsilon_t, t = 1, \dots, T \\ A(L)f_t &= B(L)u_t \end{aligned} \quad (12)$$

where $A(L) = I - A_1(L) - \dots - A_p(L)$, $B(L) = I + B_1(L) + \dots + B_q(L)$. The C matrix is generated using standard normal variates as elements, the error terms are generated as uncorrelated standard normal pseudo-random variables, and we set N=50, T=50. Each experiment is replicated 500 times.

Notice that the specification in (12) allows for factors that are determined at time t, in line with the nonparametric context of SW and FHLR, but in contrast with the framework of KM where the factors depended on lagged errors. This biases the comparison against KM, but is important for cross-method evaluation.

To provide a comprehensive evaluation of the relative performance of the three factor estimation methods, we consider four groups of experiments.

In the first group we assume that we have a single VARMA factor with 6 specifications that differ for the extent of serial correlation and the AR and MA order (these are labelled experiments 1-6):

$$\begin{aligned} (1) a_1 &= 0.2, b_1 = 0.4; & (4) a_1 &= 0.2, b_1 = -0.4; \\ (2) a_1 &= 0.7, b_1 = 0.2; & (5) a_1 &= 0.7, b_1 = -0.2; \\ (3) a_1 &= 0.3, a_2 = 0.1, b_1 = 0.15, b_2 = 0.15; & (6) a_1 &= 0.3, a_2 = 0.1, b_1 = -0.15, b_2 = -0.15; \end{aligned}$$

Experiment 7 is as experiment 1 but both the ARMA factor and its lag enter the measurement equation, i.e., the C matrix is $C(L) = C_0 + C_1L$ where L is the lag operator. We fix a priori the number of factors to p+q, which is the true number in the state space representation. It is larger than the true number in the FHLR setup, and it should provide a reasonable approximation for SW too. As a robustness check, we consider the case where the factor is generated as in Experiment 1 but only one factor is assumed to exist rather than p+q. We refer to this experiment as Experiment 8. In the case of experiments 7 and 8, qualitatively similar results are obtained when the mentioned modifications are applied to the parameter specifications 2-6.

In the second group of experiments we investigate the case of serially correlated idiosyncratic errors. The DGP is specified as in experiments 1-6 but with each idiosyncratic error being an AR(1) process with coefficient 0.2 rather than an i.i.d. process. These experiments are labelled 9-14. The results are rather robust to higher values of serial correlation but 0.2 is a reasonable value in practice since usually the common component captures most of the persistence of the series.

In the third group of experiments we investigate the case of cross-correlated errors by assuming that the contemporaneous covariance matrix of the idiosyncratic errors is tridiagonal with diagonal elements equal to 1 and off-diagonal elements equal to 0.2, while the factor is generated as in experiments 1-6. These experiments are labelled 15-20.

Table 1: Comparison of factor estimation methods

Exp.	Correlation with true ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0,827 _(0,051)	0,829 _(0,050)	0,733 _(0,049)	0,066 _(0,035)	0,066 _(0,035)	0,096 _(0,039)
Exp 2	0,858 _(0,047)	0,860 _(0,048)	0,779 _(0,052)	0,069 _(0,035)	0,073 _(0,036)	0,103 _(0,046)
Exp 3	0,737 _(0,052)	0,741 _(0,052)	0,631 _(0,054)	0,067 _(0,035)	0,071 _(0,038)	0,147 _(0,051)
Exp 4	0,810 _(0,052)	0,814 _(0,052)	0,708 _(0,050)	0,064 _(0,037)	0,069 _(0,037)	0,094 _(0,041)
Exp 5	0,823 _(0,055)	0,825 _(0,055)	0,728 _(0,056)	0,068 _(0,036)	0,070 _(0,035)	0,099 _(0,041)
Exp 6	0,713 _(0,053)	0,717 _(0,053)	0,602 _(0,050)	0,066 _(0,035)	0,070 _(0,037)	0,134 _(0,048)
Exp 7	0,897 _(0,027)	0,897 _(0,028)	0,822 _(0,037)	0,066 _(0,037)	0,071 _(0,037)	0,123 _(0,050)
Exp 8	0,907 _(0,060)	0,908 _(0,060)	0,853 _(0,049)	0,068 _(0,036)	0,069 _(0,036)	0,078 _(0,037)
Exp 9	0,815 _(0,054)	0,820 _(0,055)	0,724 _(0,053)	0,101 _(0,043)	0,111 _(0,044)	0,129 _(0,047)
Exp 10	0,852 _(0,051)	0,856 _(0,051)	0,777 _(0,055)	0,103 _(0,044)	0,114 _(0,045)	0,136 _(0,047)
Exp 11	0,727 _(0,058)	0,733 _(0,056)	0,625 _(0,059)	0,084 _(0,042)	0,105 _(0,044)	0,170 _(0,058)
Exp 12	0,801 _(0,056)	0,805 _(0,056)	0,701 _(0,053)	0,110 _(0,042)	0,124 _(0,045)	0,149 _(0,052)
Exp 13	0,813 _(0,056)	0,818 _(0,055)	0,726 _(0,055)	0,104 _(0,045)	0,116 _(0,048)	0,143 _(0,052)
Exp 14	0,707 _(0,050)	0,713 _(0,050)	0,598 _(0,048)	0,087 _(0,039)	0,109 _(0,044)	0,168 _(0,059)
Exp 15	0,822 _(0,049)	0,857 _(0,049)	0,729 _(0,048)	0,069 _(0,036)	0,072 _(0,036)	0,099 _(0,044)
Exp 16	0,851 _(0,049)	0,852 _(0,063)	0,773 _(0,056)	0,073 _(0,038)	0,097 _(0,093)	0,105 _(0,044)
Exp 17	0,729 _(0,053)	0,781 _(0,058)	0,627 _(0,056)	0,070 _(0,036)	0,083 _(0,040)	0,137 _(0,052)
Exp 18	0,801 _(0,054)	0,843 _(0,057)	0,702 _(0,053)	0,062 _(0,033)	0,064 _(0,034)	0,091 _(0,040)
Exp 19	0,818 _(0,053)	0,848 _(0,055)	0,725 _(0,054)	0,069 _(0,036)	0,075 _(0,040)	0,098 _(0,041)
Exp 20	0,705 _(0,053)	0,761 _(0,065)	0,598 _(0,052)	0,067 _(0,036)	0,075 _(0,041)	0,128 _(0,046)
Exp 21	0,975 _(0,008)	0,975 _(0,008)	0,948 _(0,013)	0,076 _(0,040)	0,074 _(0,038)	0,125 _(0,048)

^aMean Correlation between true and estimated common component, with MC st.dev. in ().

^bMean rejection rate of LM serial correlation test of idiosyncratic component, with MC st.dev. in ().

PCA: Principal Component Estimation Method; **DPCA:** Dynamic Principal Component Estimation Method; **SSS:** Subspace algorithm on state space form.

Exp. 1-6 : one factor, different ARMA DGP, no correlation among idiosyncratic components; **Exp 7:** as Exp. 1 but dynamic impact on variables; **Exp 8:** as Exp. 1 but one factor imposed in estimation rather than $p+q$; **Exp. 9-14:** one factor, different ARMA DGP, temporal correlation among idiosyncratic components; **Exp. 15-20:** one factor, different ARMA DGP, cross correlation among idiosyncratic components; **Exp. 21:** three AR factors (non correlated), no correlation among idiosyncratic components.

In the final experiment, 21, we use a 3 dimensional VAR(1) as the data generation process for the factors as opposed to an ARMA process. The coefficient matrix is a diagonal matrix with diagonal elements equal to 0.5.

As far as the results are concerned, we concentrate on the relationship between the true and estimated common components $(Cf_t, \hat{Cf}_t)$, measured by their correlation, and on the properties of the estimated idiosyncratic components $(\hat{\varepsilon}_t)$, using an LM(4) test to evaluate whether they are white noise as in the DGP, and presenting the rejection rate of

the test. The results are summarized in Table 1. The figures are averages over all series in each Monte Carlo replication and over the 500 Monte Carlo replications.

3.2. Results

In experiments 1-6, the SSS method systematically outperforms the other two in terms of correlation between true and estimated common components, with PCA ranked a closed second. KM show that the gains with respect to PCA in general are larger, in the range 5-10%, if the “lead” truncation parameter s in the SSS method is set equal to the number of factors rather than to one 1. The losses from using DPCA are in the range 10-15%, but even with this method the correlations are always larger than .60.

There is little evidence that the idiosyncratic component is serially correlated for any of the methods for experiments 1-6, again with slightly larger rejection rates of the null of no serial correlation for DPCA. Additionally, from results we are not presenting here, the DPCA method has the lowest variance for the idiosyncratic component or, in other words, has the highest explanatory power of the series in terms of the common components. These results seem to indicate that i) part of the idiosyncratic component seems to leak into the estimated common component in the DPCA case, thus reducing the correlation between true and estimated common components and the variance of the idiosyncratic component and ii) some (smaller in terms of variance) part of the common component leaks into the estimated idiosyncratic component thus increasing the serial correlation of the idiosyncratic component. The conclusion from these results is that if one cares about isolating common components as summaries of underlying common features of the data, then a high R^2 may not always be the appropriate guide.

Allowing for a dynamic impact of the factor on the variables or restricting the number of estimated factors to one does not affect significantly the ranking of the methods, compare results for experiments 7 and 8. In both cases even higher values for the correlations are obtained, likely due to the higher explanatory power of the factors in the case of experiment 7 and of the lower estimation uncertainty for experiment 8. The relative performance of DPCA also slightly improves.

In the case of serially correlated errors (experiments 9-14) the same ranking of methods arises and there is only a slight decrease in the correlations between the true and the estimated permanent components, indicating that at least for low serial correlation the methods perform satisfactorily. A similar finding emerges for cross correlation in the errors (experiments 15-20). Note that the low rejection rate of the LM test in the case of serially correlated idiosyncratic errors is due to the low value of T and quickly increases with T .

To conclude, in the case of three independent AR factors the three methods have very similar properties (experiment 21), and yield even higher values for the correlation than in the other experiments. This indicates that when factors are generated by simple autoregressive processes with short lags there is little to choose between the alternative estimation approaches.

4. Conclusions

In this paper we have reviewed three alternative approaches to estimate dynamic factor models of large dimensions, based on static principal components (SW), dynamic

principal components (FHLR) and subspace algorithms (KM). Simulation experiments indicate that SW and KM yield very similar results and are quite reliable already in small samples, while the performance of FHLR is slightly worse across a range of different experimental designs, except when the factors are generated by independent AR processes.

References

- Bai, J. (2003) Inferential Theory for Factor Models of Large Dimensions, *Econometrica*, 71(1), 135-171.
- Bai, J. and S. Ng (2002) Determining the number of factors in approximate factor models, *Econometrica*, 70, 191-223.
- Bauer, D. (1998) Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms, *Ph.D. Thesis*.
- Brillinger D.R. (1981) *Time Series: Data Analysis and Theory*, McGraw-Hill.
- Camba-Mendez, G., Kapetanios, G., Smith, R. J. and M. R. Weale (2001) An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries, *Econometrics Journal*, 4(1), 56-90.
- Chamberlain, G. and M. Rothschild (1983) Arbitrage factor structure, and mean variance analysis of large asset markets, *Econometrica*, 51, 1281-1304.
- Connor, G. and R.A. Korajczyk (1986) Performance measurement with the arbitrage pricing theory, *Journal of Financial Economics*, 15, 373-394.
- Connor, G. and R.A. Korajczyk (1993) A test for the number of factors in an approximate factor model, *Journal of Finance*, 48, 1263-1291.
- Deistler, M. and E.J. Hannan (1988) *The Statistical Analysis of Linear Systems*, John Wiley.
- Forni, M. and L. Reichlin (1996) Dynamic common factors in large cross-sections, *Empirical Economics*, 21, 27-42.
- Forni, M. and L. Reichlin (1997) National policies and local economies: Europe and the United States, *CEPR WP 1632*.
- Forni, M. and L. Reichlin (1998) Let's get real: A dynamic factor analytical approach to disaggregated business cycle, *Review of Economic Studies*, 65, 453-474.
- Forni, M., Hallin, M., Lippi, M. and L. Reichlin (1999) Reference cycles: The NBER methodology revisited, *manuscript*.
- Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2000) The generalised factor model: identification and estimation, *The Review of Economic and Statistics*, 82, 540-554.
- Geweke, J. (1977) The dynamic factor analysis of economic time series, in: *Latent variables in socio-economic models*, Aigner, D.J. and A.S. Goldberger (Eds.), North Holland, Amsterdam, Ch. 19.
- Kapetanios, G. and M. Marcellino (2003) A parametric estimation method for dynamic factor models of large dimensions, *mimeo*.
- Larimore, W.E. (1983) System Identification, Reduced Order Filters and Modelling via Canonical Variate Analysis, *Proc. 1983 Amer. Control Conference*, 445-451.
- Van Overschee, P. and B. De Moor (1996) *Subspace Identification for Linear Systems*, Kluwer Academic Publishers.

- Quah, D. and T.J. Sargent (1993) A dynamic index model for large cross-sections, Ch. 7 in: *Business cycles, indicators and forecasting*, Stock, J.H. and M.W. Watson (Eds.), University of Chicago Press for the NBER.
- Sargent, T.J. and C.A. Sims (1977) Business cycle modelling without pretending to have too much a-priori economic theory, in: *New methods in business cycle research*, C.A. Sims (Ed.), Federal Reserve Bank of Minneapolis, Minneapolis.
- Stock, J.H. and M.W. Watson (1989) New indexes of coincident and leading economic indicators, *NBER Macroeconomic Annual*, 351-393.
- Stock, J.H. and M.W. Watson (1991) A Probability Model of the Coincident Economic Indicators, in: *Leading Economic Indicators: New Approaches and Forecasting Records*, K. Lahiri and G.H. Moore (Eds.), Cambridge University Press, New York, Ch. 4., 63-85.
- Stock, J.H. and M.W. Watson (2002a) Forecasting using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association*, 97, 1167-1179.
- Stock, J.H. and M.W. Watson (2002b) Macroeconomic Forecasting using Diffusion Indexes, *Journal of Business and Economic Statistics*, 20, 147-162.