

Department of Economics

A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions

George Kapetanios and Massimiliano Marcellino

Working Paper No. 489

April 2003

ISSN 1473-0278



Queen Mary
University of London

A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions

George Kapetanios
Queen Mary, University of London

Massimiliano Marcellino*
IEP-Bocconi University, IGIER and CEPR

March 2003

Abstract

The estimation of dynamic factor models for large sets of variables has attracted considerable attention recently, due to the increased availability of large datasets. In this paper we propose a new methodology for estimating factors from large datasets based on state space models, discuss its theoretical properties and compare its performance with that of two alternative estimation approaches based, respectively, on static and dynamic principal components. The new method appears to perform best in recovering the factors in a set of simulation experiments, with static principal components a close second best. Dynamic principal components appear to yield the best fit, but sometimes there are leakages across the common and idiosyncratic components of the series. A similar pattern emerges in an empirical application with a large dataset of US macroeconomic time series.

J.E.L. Classification: C32, C51, E52

Keywords: Factor models, Principal components, Subspace algorithms

*Corresponding author: Massimiliano Marcellino, IGIER - Università Bocconi, Via Salasco 5, 20136, Milano, Italy. Phone: +39-02-5836-3327. Fax: +39-02-5836-3302.

1 Introduction

Recent work in the macroeconometric literature considers the problem of summarising efficiently a large set of variables and using this summary for a variety of purposes including forecasting. Work in this field has been carried out in a series of recent papers by Stock and Watson (2001) (SW) and Forni et al (1998,2000,2001). Factor analysis has been the main tool used in summarising the large datasets.

The main factor model used in the past to extract dynamic factors from economic time series has been a state space model estimated using maximum likelihood. This model was used in conjunction with the Kalman filter in a number of papers carrying out factor analysis (see, among others, Stock and Watson (1989) and Camba-Mendez et al (2001)). However, maximum likelihood estimation of a state space model is not practical when the dimension of the model becomes too large due to the computational cost. For the case considered by SW where the number of time series is greater than the number of observations, maximum likelihood estimation is not practically feasible, even when a sophisticated EM algorithm is used for optimization, as in Quah and Sargent (1993). For this reason, SW have suggested principal component based estimation. This method can accommodate a very large number of time series and there is no need for the number of observations to exceed the number of variables. SW have shown that it can estimate consistently the factor space asymptotically. In small samples and for a finite number of series, the dynamic element of the principal component analysis is not easy to interpret. Forni et al (1998) suggested an alternative procedure based on dynamic principal components (see Chapter 9 of Brillinger (1981)) that incorporates an explicitly dynamic element in the construction of the factors.

In this paper we suggest a third approach for factor estimation that retains the attractive framework of a 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. We analyze the asymptotic properties of the new estimators, first for a fixed number of series, N , and then allowing N to diverge. We show that as long as N grows less than $T^{1/3}$, where T is the number of observations, the subspace algorithm still yields consistent estimators for the space spanned by the factors. Moreover, we suggest a modified subspace algorithm that permits to analyze datasets with N larger

than T , i.e., more series than observations, and evaluate its performance using Monte Carlo simulations. Finally, we develop an information criterion that leads to consistent selection of the number of factors to be included in the model, along the lines of Bai and Ng (2002) for the static principal component approach.

Our second contribution is an extensive simulation study of the relative performance of the three competing estimation methods. We evaluate the dynamic relationship between the true factors and their estimated counterparts both in the time and frequency domain, and we further examine the properties of the resulting idiosyncratic component of the data. We find that the state space based method performs better in a variety of experiments compared to the principal component based methods, also when $N > T$, with the static principle component estimates ranked second. Though these findings may depend on the experimental designs, they appear to be rather robust.

Our final contribution is the analysis of a large dataset of 146 US macroeconomic time series, extracted from SW. As in the simulation experiments, it turns out that the performance of static principal components and state space methods is overall comparable. Moreover, when the state space based factors are included in small scale monetary VARs, more reasonable responses of output gap and inflation to interest rate shocks are obtained.

The paper is organised as follows. Section 2 reviews the static and dynamic principle component methodologies. Section 3 presents the state space model approach and derives the properties of the estimators for the fixed N case. Section 4 deals with the diverging N case and with correlation of the idiosyncratic components. Section 5 develops a modified algorithm to analyze datasets with $N > T$. Section 6 compares the competing methods using Monte Carlo simulations. Section 7 discusses the empirical analysis and section 8 concludes.

2 The dynamic factor model and the available estimators

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

adopt is

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

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 $\text{cov}(\xi_{jt}, u_{st-k}) = 0$ 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 χ_{nt} , ξ_{nt} , and therefore x_{nt} , to have rational spectral density matrices, Σ_n^x , Σ_n^ξ , and Σ_n^x , 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, $\lambda_{n1}^x, \dots, \lambda_{nq}^x$, 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_t to affect infinitely many units.

The static version of this model was analyzed, among others, by Chamberlain and Rothschild (1983), Connor and Korajczyk (1986, 1993). When the idiosyncratic components are uncorrelated across units the model is usually referred to as an exact static model, otherwise it is approximate. Geweke (1977) and Sargent and Sims (1977) studied a dynamic factor model for a limited number of units. Further developments were due to Sargent (1989), Stock and Watson (1991) and Quah and Sargent (1993), but all these methods are not suited when n , the number of variables, is larger than 50-60, while the procedure by FHLR can handle hundreds of variables. We will refer to the procedure by FHLR as dynamic principle component analysis (DPCA)

A competing procedure for the large n case was developed by SW. The model by SW, in its time invariant formulation, can be written as

$$x_{nt} = \Lambda f_t + \xi_{nt}, \quad (2)$$

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 (2) 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 (2), 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 .

Let us now briefly describe the estimation methods suggested by FHLR and SW, more details can be found in the original papers. Five elements are primarily of interest in a factor model: the number of factors, the factors themselves, their loadings, the common component, and the idiosyncratic component.

Let us assume for the moment that the number of common factors is known. Then, FHLR suggest to estimate the common component of χ_{it} with the following step-wise procedure. (i) Estimate the spectral density matrix of x_{nt} as

$$\Sigma_n^T(\theta_h) = \sum_{k=-M}^M \Gamma_{nk}^T \omega_k e^{-ik\theta_h}, \quad \theta_h = 2\pi h/(2M + 1), \quad h = 0, \dots, 2M,$$

where Γ_{nk}^T is the sample covariance matrix of x_{nt} and x_{nt-k} , ω_k is the Bartlett lag window of size M ($\omega_k = 1 - k/(M + 1)$), and M diverges but M/T tends to zero. (ii) Calculate the first q eigenvectors of $\Sigma_n^T(\theta_h)$, $p_{nj}^T(\theta_h)$, $j = 1, \dots, q$, for $h = 0, \dots, 2M$. (iii) Define $p_{nj}^T(L)$ as

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

$p_{nj}^T(L)x_{nt}$, $j = 1, \dots, q$, are the first q estimated dynamic principal components of x_{nt} . (iv) Run an OLS regression of x_{it} on present, past, and future dynamic principal components. The fitted value is the estimated common component of x_{it} , $\hat{\chi}_{it}$. FHLR prove that, under mild conditions, $\hat{\chi}_{it}$ is a consistent estimator of χ_{it} . In practice, M and the number of leads (s) and lags (g) of $p_{nj}^T(L)x_{nt}$ to be included as regressors can be either chosen a priori or determined by minimizing the information criterion

$$\frac{T}{n} \sum_{i=1}^n \log \hat{\sigma}_i + 2q(g + s + 1), \quad (3)$$

where $\hat{\sigma}_i$ is the estimated variance of the residuals from the i^{th} regression. Since the properties of the latter method are not known, in what follows 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, $\hat{\xi}_{it} = x_{it} - \hat{\chi}_{it}$.

The starting point in SW's approach is instead 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 $n^{-1} \sum_{i=1}^n \underline{x}_i \underline{x}_i'$, where $\underline{x}_i = (x_{i1}, \dots, x_{iT})$. Moreover, the optimal estimators of Λ are the OLS estimators of the coefficients in a regression of x_{it} 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.

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 i^{th} 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 $\hat{\xi}_{it} = x_{it} - \hat{\chi}_{it}$. We will refer to this estimation method as PCA.

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 (3), could be minimized also with respect to q .

SW, in a related context, also 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.

3 The state space factor estimator

In this section we present the basic state space approach and then discuss the asymptotic properties of the estimators when T diverges and N is fixed. In the following sections we extend the framework to deal with the N going to infinity case, with the analysis of datasets with a larger cross-section than time-series dimension, and with cross-sectionally or serially correlated idiosyncratic errors.

3.1 The basic state space model

We consider the following state space model.

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

x_{nt} is an n -dimensional vector of strictly stationary zero-mean variables observed at time t . f_t is an k -dimensional vector of unobserved states (factors) at time t , and u_t and v_t are multivariate, mutually uncorrelated, standard orthogonal white noise sequences of dimension, respectively, n and k . D^* is

assumed to be nonsingular. The aim of the analysis is to obtain estimates of the states f_t , for $t = 1, \dots, T$.

This model is quite general. Its aim is to use the states as a summary of the information available from the past on the future evolution of the system. A large literature exists on the identification issues related with the state space representation given in (6). An extensive discussion may be found in Deistler and Hannan (1988). In particular, they show in Chapter 1 that (6) is equivalent to the prediction error representation of the state space model given by

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

This form will be used for the derivation of our estimation algorithm. Note that as at this stage the number of series, N , is large but fixed we need to impose no conditions on the structure of C . Conditions on this matrix will be discussed later when we consider the case of N tending to infinity.

As we have mentioned in the introduction, maximum likelihood techniques, possibly using the Kalman filter, may be used to estimate the parameters of the model under some identification scheme. 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. 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.

There are many subspace algorithms, and vary in many respects, but a unifying characteristic is their view of the state as the interface between the past and the future in the sense that the best linear prediction of the future of the observed series is a linear function of the state. A very good 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).

The starting point of most subspace algorithms is the following representation of the system which follows from the state space representation in (7) and the assumed nonsingularity of D .

$$X_t^f = \mathcal{O}KX_t^p + \mathcal{E}E_t^f \tag{8}$$

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

$$\mathcal{E} = \begin{pmatrix} D & 0 & \dots & 0 \\ CB & D & \ddots & \vdots \\ CAB & \ddots & \ddots & 0 \\ \vdots & & CB & D \end{pmatrix}$$

The derivation of this representation is easy once we note that (i) $X_t^f = \mathcal{O}f_t + \mathcal{E}E_t^f$ and (ii) $f_t = \mathcal{K}X_t^p$. The best linear predictor of the future of the series at time t is given by $\mathcal{O}\mathcal{K}X_t^p$. The state is given in this context by $\mathcal{K}X_t^p$ at time t . The task is therefore to provide an estimate for \mathcal{K} . Obviously, the above representation 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 $\mathcal{F} = \mathcal{O}\mathcal{K}$ may be obtained by regressing $X_{s,t}^f$ on $X_{p,t}^p$. Following that, the most popular subspace algorithms use a singular value decomposition (SVD) of an appropriately weighted version of the least squares estimate of \mathcal{F} , denoted by $\hat{\mathcal{F}}$. In particular the algorithm we will use, due to Larimore (1983), applies an SVD to $\hat{\Gamma}^f \hat{\mathcal{F}} \hat{\Gamma}^p$, where $\hat{\Gamma}^f$, 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 $\hat{\mathcal{F}}$. Then, the estimate of \mathcal{K} is given by

$$\hat{\mathcal{K}} = \hat{S}_k^{1/2} \hat{V}_k' \hat{\Gamma}^p^{-1/2}$$

where $\hat{U} \hat{S} \hat{V}'$ represents the SVD of $\hat{\Gamma}^{f-1/2} \hat{\mathcal{F}} \hat{\Gamma}^{p1/2}$, \hat{V}_k denotes the matrix containing the first k columns of \hat{V} and \hat{S}_k denotes the heading $k \times k$ submatrix of \hat{S} . \hat{S} contains the singular values of $\hat{\Gamma}^{f-1/2} \hat{\mathcal{F}} \hat{\Gamma}^{p1/2}$ in decreasing order. Then, the factor estimates are given by $\hat{\mathcal{K}}X_t^p$. We refer to this method as SSS.

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. So an alternative suggestion is to simply use identity matrices instead of the covariance matrices. It is this suggestion we follow in the Monte Carlo study.

A second point to note is that consistent estimation of the factor space requires that p increases 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 simplified condition for the lower bound would be to set it to $T^{1/r}$ for any $r > 3$. We have found that a range of α between 1.05 and 1.5 for this parameter provides a satisfactory performance. We have used the value $\alpha = 1.25$ for the simulation experiments.

For consistency s is also required to be set so as to satisfy $sN > k$. As N is usually going to be very large for the applications we have in mind, this restriction is not binding. We suggest two options. Either we set $s = 1$ or equal to the number of assumed factors k . In the experiments we have carried out below we set $s = k$ but we also evaluate the performance of the method with $s = 1$. This is particularly relevant to avoid any use of future information, as e.g. in a real time forecasting context.

Once estimates of the factors have been obtained, and if estimates of the parameters (including the factor loadings) are subsequently required, least squares methods may be used to obtain such estimates. These estimates have been proved to be \sqrt{T} -consistent and asymptotically normal in Bauer (1998). We note that the identification scheme used above is implicit and depends on the normalisation used in the computation of the SVD. In particular the SVD used in the simulations normalises the left and right singular value vectors by restricting them to have an identity second moment matrix.

It is worth pointing out that the estimated parameters can be used with the Kalman filter on the state space model to obtain both filtered and smoothed estimates of the factors. We note that the SSS method produces factor estimates at time t conditional on data available at time $t - 1$. Therefore, it may be reasonable to suggest that smoothed estimates from the Kalman filter may be superior to those obtained by the SSS method. However, the parameter estimates are conditional on the factor estimates obtained in the first step by the SSS method. Limited experimentation using the Monte Carlo setup reported in the next sections suggests that the loss in performance of the smoothed Kalman filter factor estimate because of the use of estimated parameters from the SSS method, is roughly similar to the benefit of using all the data. In general, factors estimated using the SSS method outperform filtered Kalman filter factor estimates.

Finally, we must note that the SSS method is also applicable in the case of unbalanced panels. In analogy to the work of SW, use of the EM algorithm, described there, can be made to provide estimates both of the factors and of the missing elements in the dataset.

3.2 Asymptotic Properties

We now discuss the asymptotic properties of the factor estimators, derive their standard errors, and suggest a consistent criterion for the selection of the number of factors.

Let us denote the true number of factors by k^0 and investigate in more detail OLS estimation of the multivariate regression model

$$X_{s,t}^f = \mathcal{F}X_{p,t}^p + \mathcal{E}E_t^f \quad (9)$$

for fixed $s \geq k^0$. Estimation of the above is equivalent to estimation of each equation separately. We make the following assumptions

Assumption 1 u_t is an i.i.d. $(0, \Sigma_u)$ sequence with finite fourth moment.

Assumption 2 $p_1 \leq p \leq p_2$ where $p_1 = O(T^{1/r})$, $r > 3$ and $p_2 = o(T^{1/3})$

Theorem 1 (Consistency). If we define $\hat{f}_t = \hat{\mathcal{K}}X_t^p$, then \hat{f}_t converges, in probability, to the space spanned by the true factors

Proof. By (7) and (8) we can see that $\mathcal{K}X_t^p$ spans the space of the true factors. So we need to concentrate on the properties of $\hat{\mathcal{K}}$ as an

estimator of \mathcal{K} . By Theorem 4 of Berk (1974), who provides a variety of results for parameter estimates in infinite autoregressions, we have that $\hat{\mathcal{F}}$ is consistent for \mathcal{F} and that $\sqrt{T - Np}(\hat{\mathcal{F}} - \mathcal{F})$ has an asymptotic normal distribution with the standard OLS covariance matrix. This result follows straightforwardly from equation (2.17) of Berk (1974) once we note that the sum of the absolute values of the coefficients in each regression multiplied by $p^{1/2}$ tends to zero. This follows by the fact that the absolute value of the maximum eigenvalue of \mathcal{F} , denoted $|\lambda_{\max}(\mathcal{F})|$, is less than one implying exponentially declining coefficients with respect to p . This implies consistent estimation of the factors since $\hat{\mathcal{K}}$ is a continuous function of $\hat{\mathcal{F}}$ for large enough T . Since both T and p grow, by assumption 2 the rate of convergence of the factor estimates lies between $(T - Np)^{1/2-1/2r}$ and $(T - Np)^{1/3}$. We will denote the square of the rate of convergence by T^* .

■

A remark on the nature of the consistency of the factor estimates is in order. As we have seen the analysis of consistency and asymptotic normality is carried out in terms of the matrix $\hat{\mathcal{K}}$. If this matrix is consistent

and asymptotically normal then the estimated factors have the same properties. This clearly assumes some identification scheme. In fact there are two identification schemes involved. The first is the identification scheme that identifies the matrices A, B, C, D from the unique transfer function of the data x_{nt} . Deistler and Hannan (1988) discuss in detail conditions needed for this identification. The second identification scheme is implicit through the normalisation carried out to obtain the SVD. These identification schemes imply the existence of a nonsingular matrix such that \hat{f} consistently estimates $H^k f$. Then, we have the following theorem on the asymptotic distribution of the factor estimates.

Theorem 2 (Asymptotic distribution). If $s = 1$, the asymptotic distribution of $\sqrt{T^*}(\text{vec}(\hat{f}) - \text{vec}(H^k f))$ with $f = (f_1, \dots, f_T)'$ is $N(0, V_f)$, with

$$V_f = (I_k \otimes X^p) \frac{\partial g}{\partial (A_1 \mathcal{F} A_2)} (A_2' \otimes A_1) (\Gamma^p \otimes \Sigma) (A_2 \otimes A_1') \frac{\partial g'}{\partial (A_1 \mathcal{F} A_2)} (I_k \otimes X^{p'})$$

where g is defined in the proof of the Theorem and for some nonsingular matrix H^k

Proof. Asymptotic normality of the estimators follows from asymptotic normality of $\hat{\mathcal{K}}$ which follows from the asymptotic normality of $\sqrt{T - Np}(\hat{\mathcal{F}} - \mathcal{F})$ proved in Theorem 4 of Berk (1974). The normality of $\hat{\mathcal{K}}$ follows by using a simple Taylor expansion of the function implicitly defined by the SVD of $\hat{\mathcal{F}}$. Denote this function by g . The existence of the Taylor expansion follows from continuity and differentiability of g which follows from Theorems 5.6 and 5.8 of Chatelin (1983).

The variance calculations will be carried out conditional on X_t^p , as when obtaining variances of regression coefficients conditional on the regressors. We assume a normalisation such that $f = X^p \hat{\mathcal{K}}'$. Then, simple manipulations indicate that

$$V \left(\sqrt{T^*}(\text{vec}(\hat{f}) - \text{vec}(f)) \right) = (I_k \otimes X^p) V \left(\sqrt{T^*} \left(\text{vec}(\hat{\mathcal{K}}') - \text{vec}(\mathcal{K}') \right) \right) (I_k \otimes X^{p'})$$

We need to derive the asymptotic variance of $V \left(\sqrt{T^*} \left(\text{vec}(\hat{\mathcal{K}}') - \text{vec}(\mathcal{K}') \right) \right)$. In general, $\hat{\mathcal{K}}'$ is a function of the SVD of $A_1 \hat{\mathcal{F}} A_2$, where A_1 and A_2 are weighting matrices discussed before. Note the importance of $sn \geq k$ for the calculation of the SVD. Moreover, the proof is valid for $s = 1$ only as there is serial correlation in the error terms in (8) for $s > 1$. This case is discussed in more detail in Section 9.0.2 below.

We define formally the function $g(\cdot)$ such that $vec(\hat{\mathcal{K}}') = g\left(vec(A_1\hat{\mathcal{F}}A_2)\right)$. By a first order Taylor expansion of $g(vec(A_1\hat{\mathcal{F}}A_2))$ and $g(vec(A_1\mathcal{F}A_2))$ around $A_1\mathcal{F}^*A_2$, possible since $g(\cdot) \in C^\infty$ and where each element of \mathcal{F}^* lies between the respective elements of \mathcal{F} and $\hat{\mathcal{F}}$, we have that

$$V\left(\sqrt{T^*}\left(vec(\hat{\mathcal{K}}') - vec(\mathcal{K}')\right)\right) = \frac{\partial g}{\partial(A_1\mathcal{F}A_2)}$$

$$V\left(\sqrt{T^*}\left(vec(A_1\hat{\mathcal{F}}A_2) - vec(A_1\mathcal{F}A_2)\right)\right) \frac{\partial g'}{\partial(A_1\mathcal{F}A_2)}$$

Consistency and a $\sqrt{T^*}$ rate of convergence of the parameter estimates $\hat{\mathcal{F}}$ to their true values implies that the remainder of the Taylor approximation is $o_p(1)$. So we need to derive the variance of $\sqrt{T}\left(vec(A_1\hat{\mathcal{F}}A_2) - vec(A_1\mathcal{F}A_2)\right)$. Again simple manipulations imply that

$$V\left(\sqrt{T^*}\left(vec(A_1\hat{\mathcal{F}}A_2) - vec(A_1\mathcal{F}A_2)\right)\right) = (A_2' \otimes A_1) V\left(\sqrt{T^*}\left(vec(\hat{\mathcal{F}}) - vec(\mathcal{F})\right)\right) (A_2 \otimes A_1')$$

From multivariate regression analysis we know that

$$V\left(\sqrt{T^*}\left(vec(\hat{\mathcal{F}}) - vec(\mathcal{F})\right)\right) = (\Gamma^p \otimes \Sigma)$$

where Σ is the variance covariance matrix of the regression error, which yields the result. ■

We now develop an information criterion that yields consistent selection of the number of factors. Let us define the criterion as

$$IC(k) = \ln|\hat{\Sigma}_u^k| + c_T(k), \quad (10)$$

where we assume that the penalty function $c_T(k)$ satisfies

Assumption 3 $c_T(k) = o(1)$.

We have

Theorem 3 (Consistency of IC(m)). When T diverges, $arg\,k \in_{k \in \{1, \dots, k^{max}\}} IC(k) = k^0, \forall k^{max}$.

Proof. We wish to prove (i) $\text{plim}_{T^* \rightarrow \infty} \ln|\hat{\Sigma}_u^k| / \ln|\hat{\Sigma}_u^{k^0}| > 1$ for $k < k^0$ and (ii) $\text{plim}_{T^* \rightarrow \infty} \ln|\hat{\Sigma}_u^k| / \ln|\hat{\Sigma}_u^{k^0}| = 1$ for $k > k^0$. For (ii) we note that by

the fact that the singular values of $\hat{\mathcal{F}}$ tend to their true values at a rate of $\sqrt{T - Np}$ and the fact that any $(k^0 + i)$ -th largest singular value of \mathcal{F} is equal to zero, any observations of the $(k^0 + i)$ -th factor series will tend to zero at rate $\sqrt{T - Np}$. As a result asymptotically any regression that involves more than k^0 factors will have a singular regressor matrix. We assume that any such regression will not be further considered and therefore the probability of picking $k > k^0$ tends to zero asymptotically.

We now have to prove (i). We note that even for $k < k^0$ the first k factors are consistently estimated as they are obtained from the unrestricted OLS estimates of \mathcal{F} . In fact, for a given sample the factor estimates for the first k factors in a model which assumes $k_1 > k$ factors are identical to those in a model with $k_2 > k$ factors. This is easy to see once it is noted that irrespective of the assumed number of factors the same SVD is carried out. We examine the probability of the event $\ln|\hat{\Sigma}_u^k|/\ln|\hat{\Sigma}_u^{k^0}| > 1$ We have that

$$Pr(\ln|\hat{\Sigma}_u^k|/\ln|\hat{\Sigma}_u^{k^0}| < 1 + \epsilon) = Pr\left(\frac{\ln|1/(T - Np)X'\hat{M}^kX|}{\ln|1/(T - Np)X'\hat{M}^{k^0}X|} < 1 + \epsilon\right) \quad (11)$$

where $X = (x_1, \dots, x_T)'$, $\hat{M}^k = I - \hat{f}^k(\hat{f}^{k'}\hat{f}^k)^{-1}\hat{f}^{k'}$, $\hat{M}^{k^0} = I - \hat{f}^{k^0}(\hat{f}^{k^0'}\hat{f}^{k^0})^{-1}\hat{f}^{k^0'}$, $\hat{f}^k = (\hat{f}_1^k, \dots, \hat{f}_T^k)$ and $\hat{f}_t^k = (\hat{f}_{1,t}, \dots, \hat{f}_{k,t})'$. If we show that the probability in (11) is equal to

$$Pr\left(\frac{\ln|1/(T - Np)X'M^kX|}{\ln|1/(T - Np)X'M^{k^0}X|} < 1 + \epsilon\right)$$

where M^k and M^{k^0} are defined in the obvious way then the fact that the above probability is less than ϵ for all $\epsilon > 0$ follows from standard regression results on uniform convergence and asymptotic normality of regression parameters and the analysis of, e.g. Sin and White (1996). To show that, we need to show that

$$\|1/(T - Np)X'\hat{M}^kX - 1/(T - Np)X'M^kX\| = o_p(1)$$

for all k . This follows if we show that $1/(T - Np)\|\hat{M}^k - M^k\|$ for all k . But this follows from the fact that $1/(T - Np)\|f - \hat{f}\| = o_p(1)$ ■

4 The case: $N \rightarrow \infty$

In this section we firstly discuss correlation of the idiosyncratic errors. Then we investigate whether the SSS method can yield consistent factor estimators

when N diverges or, more generally, what are the conditions for consistency when N diverges. Finally, we derive a proper modification of the IC criterion for the selection of the number of factors for the diverging N case.

4.1 Correlation in the idiosyncratic errors

In this subsection we discuss the case of cross-sectional and/or serial correlation of the idiosyncratic errors. This extension can be rather simply handled within the state space method. Basically, the idiosyncratic errors can be treated as additional pseudo-factors that enter only a few of the variables via restrictions on the matrix of loadings C . These pseudo-factors can be serially correlated processes or not depending on A .

The problem becomes one of distinguishing common factors and pseudo-factors, i.e., cross-sectionally correlated idiosyncratic errors. This is virtually impossible for finite N , while when N diverges a common factor is one which enters an infinite number of series, i.e, the column of the, now infinite dimensional, matrix C associated with a common factor will have an infinity of non-zero entries, and likewise a pseudo-factor will only have a finite number of non-zero entries in the respective column of C . Let k_1 denote the number of common factors thus defined and k_2 the number of pseudo-factors. Note that k_2 may tend to infinity but not faster than N . Then, following Forni et al. (2000), we make the following assumption.

Assumption 4 *The matrix \mathcal{OK} in (8) has k_1 singular values tending to infinity as N tends to infinity and k_2 non-zero finite singular values.*

For example, the condition in the assumption is satisfied if k_1 common factors enter a non zero fraction, bN , $0 < b < 1$, of the series x_{nt} , in the state space model given by (6), while $k_2(N)$ pseudo-factors enter a vanishing proportion of the series x_{nt} , i.e. each such factors enter $c(N)N$ of the series x_{nt} where $\lim_{N \rightarrow \infty} c(N)N = 0$ and $k_2(N)$ is at most $O(N)$.

4.2 Consistency of the SSS estimator

To prove consistency of the SSS estimator, we need to add an assumption to those in the previous section. In particular, we require

Assumption 5 $Np = o(T^{1/3})$; $p = O(T^{1/r})$, $r > 3$;

Then we have

Theorem 4 (Consistency when $N \rightarrow \infty$). If N is $o(T^{1/3-1/r})$, then when N and T diverge $\hat{f}_t = \hat{\mathcal{K}}X_t^p$ converges in probability to the space spanned by the true factors.

Proof. Consistent estimation of the coefficients of the model in (9) by OLS, and therefore of the factors, holds if the number of regressors in each of the N s equations tends to infinity at a rate lower than $T^{1/3}$ but the number of lags, p , grows at a minimum rate of $T^{1/r}$ where $r > 0$. Since the number of regressors is Np we see that N can grow at rates of at most $T^{1/3-1/r}$. Under these conditions the estimates of the factors will be consistent at rate $(T/Np)^{1/2}$ as the results by Berk (1974) applied to every equation separately hold. ■

Thus, divergence of N requires to be accompanied by a faster divergence of T for the SSS factor estimators to remain consistent.

4.3 Choice of the number of factors

With reference to the determination of the number of factors, in the previous section we have suggested a simple information theoretic method for the fixed N case. The method requires (i) to fix a maximum number of factors f^{max} to search over, (ii) to estimate the factors for each assumed number of factors $k = 1, \dots, k^{max}$, and (iii) to minimise the negative penalised loglikelihood of the regression

$$x_t = C\hat{f}_t + u_t, \quad (12)$$

i.e. minimising $\ln|\hat{\Sigma}_u^k| + c_T(k)$ where $\hat{\Sigma}_u^k$ is the estimated covariance matrix of u_t and $c_T(k)$ is a penalty term depending on the choice of the information criterion used.

We now propose a more general criterion that works also for the N diverging case. For this problem the analysis of Bai and Ng (2002) is relevant. We will show that an information criterion of the form

$$IC(k_1) = V(k_1, \hat{f}^{k_1}) + k_1 g(N, T) \quad (13)$$

where

$$V(k_1, \hat{f}^{k_1}) = (NT)^{-1} \sum_{t=1}^T tr[(x_{nt} - \hat{C}\hat{f}_t^{k_1})(x_{nt} - \hat{C}\hat{f}_t^{k_1})'] \quad (14)$$

$\hat{f}_t^{k_1}$ are the factor estimates for the k_1 first common factors (according to the singular values), \hat{C} is the OLS estimate of C based on $\hat{f}_t^{k_1}$ and $g(N, T)$ is a penalty term, provides a consistent estimate of the number of factors.

Before examining the properties of this criterion, note that any set of up to k_1^0 factor estimates are consistent estimates of the respective set of true factors up to a nonsingular transformation determined by the normalisation used in the SVD carried out during the estimation and the identification of the state space model, see SW for a similar point . Thus, denoting the matrix of the k_1 first true factors by f^{0, k_1} , we have that

$$(T/Np)^{1/2} \|f_t^{k_1} - H^{k_1'} f_t^{0, k_1}\| = O_p(1)$$

for some nonsingular matrix H^{k_1} . This follows from Theorem 4. Then, the following theorem holds

Theorem 5. Let the factors be estimated by the SSS method. Let $\hat{k}_1 = \operatorname{argmin}_{0 \leq k \leq k_{max}} IC(k_1)$. Then, $\lim_{T \rightarrow \infty} Pr(\hat{k}_1 = r) = 1$ if i) $g(N, T) \rightarrow 0$ and ii) $Ng(N, T) \rightarrow \infty$ as $N, T \rightarrow \infty$.

Proof. Denote the true number of common factors k_1^0 . We must prove that $\lim_{N(T), T \rightarrow \infty} Pr(IC(k_1) < IC(k_1^0)) = 0$ for all $k_1 \neq k_1^0$, $k_1 < k^{max}$. First, denoting the matrix of the first k_2 true idiosyncratic factors by $f^{0, 2, k_2}$, we examine

$$V(k_1, (f^{0, k_1}, f^{0, 2, k_2})) - V(k_1, (f^{0, k_1}))$$

for any finite k_2 . By the uncorrelatedness of f^{0, k_1} and $f^{0, 2, k_2}$ we know that for all elements of x_{nt} in which $f^{0, 2, k_2}$ does not enter

$$1/T \sum_{t=1}^T (x_{i, nt} - \hat{C}'_{i, 1, 2} (f_t^{0, k_1'}, f_t^{0, 2, k_2'})')^2 - 1/T \sum_{t=1}^T (x_{i, nt} - \hat{C}'_{i, 1} f_t^{0, k_1})^2 = O_p(T^{-1})$$

For a finite number of elements of x_{nt}

$$1/T \sum_{t=1}^T (x_{i, nt} - \hat{C}'_{i, 1, 2} (f_t^{0, k_1'}, f_t^{0, 2, k_2'})')^2 - 1/T \sum_{t=1}^T (x_{i, nt} - \hat{C}'_{i, 1} f_t^{0, k_1})^2 = O_p(1)$$

So overall

$$V(k_1, (f^{0, k_1}, f^{0, 2, k_2})) - V(k_1, (f^{0, k_1})) = O_p(N^{-1}) \quad (15)$$

First consider $k_1 < k_1^0$. Then

$$IC(k_1) - IC(k_1^0) = V(k_1, \hat{f}^{k_1}) - V(k_1^0, \hat{f}^{k_1^0}) - (k_1^0 - k_1)g(N, T)$$

and the required condition for the result is

$$Pr[V(k_1, \hat{f}^{k_1}) - V(k_1^0, \hat{f}^{k_1^0}) < (k_1^0 - k_1)g(N, T)] = 0$$

as $N(T), T \rightarrow \infty$. Now

$$\begin{aligned} V(k_1, \hat{f}^{k_1}) - V(k_1^0, \hat{f}^{k_1^0}) &= [V(k_1, \hat{f}^{k_1}) - V(k_1, f^{k_1} H^{k_1})] + [V(k_1, f^{k_1} H^{k_1}) - V(k_1^0, f^{k_1^0} H^{k_1^0})] + \\ &\quad [V(k_1^0, f^{k_1^0} H^{k_1^0}) - V(k_1^0, \hat{f}^{k_1^0})] \end{aligned}$$

By the rate of convergence of the factor estimates and Lemma 2 of [?] we have

$$V(k_1, \hat{f}^{k_1}) - V(k_1, f^{k_1} H^{k_1}) = O_p((T/Np)^{-1})$$

and

$$V(k_1^0, \hat{f}^{k_1^0}) - V(k_1^0, f^{k_1^0} H^{k_1^0}) = O_p((T/Np)^{-1})$$

Then $V(k_1, f^{k_1} H^{k_1}) - V(k_1^0, f^{k_1^0} H^{k_1^0})$ can be written as $V(k_1, f^{k_1} H^{k_1}) - V(k_1^0, f^{k_1^0} H^{k_1^0})$ which has positive limit by Lemma 3 of Bai and Ng (2002). Thus as long as $g(N, T) \rightarrow 0$ $Pr(IC(k_1) < IC(k_1^0)) = 0$ for all $k_1 < k_1^0$. Then to prove $Pr(IC(k_1) < IC(k_1^0)) = 0$ for all $k_1 > k_1^0$ we have to prove that

$$Pr[V(k_1^0, \hat{f}^{k_1^0}) - V(k_1, \hat{f}^{k_1}) < (k_1 - k_1^0)g(N, T)] = 0$$

By (15) we know that asymptotically the analysis of the state space model will be equivalent to the case of a model where there are no idiosyncratic factors up to an order of probability of N^{-1} . Then

$$|V(k_1^0, \hat{f}^{k_1^0}) - V(k_1, \hat{f}^{k_1})| \leq 2max_{k_1^0 < k_1 \leq kmax} |V(k_1, \hat{f}^{k_1}) - V(k_1, f^{0, k_1^0})|$$

By following the analysis of Lemma 4 of Bai and Ng (2002) we know that

$$max_{k_1^0 < k_1 \leq kmax} |V(k_1, \hat{f}^{k_1}) - V(k_1, f^{0, k_1^0})| = O_p((T/Np)^{-1})$$

Combining this with (15) and the fact that Np grows slower than $T^{1/3}$, gives the required result since then $(T/Np)^{-1} < N^{-1}$.

■

5 Dealing with Large Datasets

Up to now we have outlined a method for estimating factors which requires the number of observations to be larger than the number of elements in X_t^p . Given the work of SW and FHLR this is rather restrictive. We therefore suggest a modification of the methodology to allow the number of series be larger than the number of observations.

The problem arises because the least squares estimate of \mathcal{F} does not exist due to rank deficiency of $X^{p'}X^p$ where $X^p = (X_1^p, \dots, X_T^p)'$. As we mentioned in the previous section we do not necessarily want an estimate of \mathcal{F} but an estimate of the states $X^p\mathcal{K}'$. That could be obtained if we had an estimate of $X^p\mathcal{F}'$ and used a SVD of that. But it is well known (see e.g. Magnus and Neudecker (1988)) that although $\hat{\mathcal{F}}$ may not be estimable, $X^p\mathcal{F}'$ always is using least squares methods. In particular, the least squares estimate of $\widehat{X^p\mathcal{F}'}$ is given by

$$\widehat{X^p\mathcal{F}'} = X^p(X^{p'}X^p)^+X^{p'}X^f \quad (16)$$

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 $\widehat{X^p\mathcal{F}'} = X^f$. 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 principle component method described in an earlier section. We will refer to this method as SSS0. This method is static, abstracting from the fact that s may be larger than 1, thereby leading to a decomposition involving leads of x_{nt} .

Alternative solutions exist to this problem. In particular, note that we are after a subspace decomposition of the estimate of the fitted value $X^p\mathcal{F}'$. Essentially, we are after a reduced rank approximation of $X^p\mathcal{F}'$ and several possibilities exist. The main requirement is that as the assumed rank (number of factors) tends to the full rank of the estimate of the fitted value, the approximation should tend to the estimated fitted value $\widehat{X^p\mathcal{F}'} = X^p(X^{p'}X^p)^+X^{p'}X^f = X^f$. The alternative decomposition we suggest is a SVD on $X^{f''}X^p(X^{p'}X^p)^+ = \hat{U}\hat{S}\hat{V}'$. Then the estimated factors are given by $\hat{\mathcal{K}}X_t^p$ where $\hat{\mathcal{K}}$ is obtained as before but using the SVD of $X^{f''}X^p(X^{p'}X^p)^+$. We choose to set both weighting matrices to the identity matrix in this case. We also refer to this decomposition as SSS.

Clearly, the decomposition we suggest is simply a generalisation of the

method in the previous section and if $Np < T$ it reduces to that method. As k tends to $\min(Ns, Np)$ the set of factor estimates tends to the OLS estimated fitted value X^f . Its performance needs to be judged in terms of its small sample properties in approximating (linear combinations of) the true factors, which can be determined via simulations. More generally, we see that the essence of the method is to provide a reduced rank approximation of the expectation of the future of the process conditional on its past. We will concentrate on SSS given that one of the focuses of this paper is on explicitly dynamic factor extraction methods.

We now briefly discuss the relationship between the SSS method and ML estimation. Although in standard regression \mathcal{F} is not estimable because there are more parameters to determine than observations to use, this does not extend to reduced rank regression. In particular note that since \mathcal{F} has reduced rank structure there are fewer free parameters to estimate. If the number of free parameters is less than NT then it is clear that \mathcal{F} is estimable by maximum likelihood. ML estimation, here, cannot be proven to be equivalent to the standard reduced rank ML estimator since that requires nonsingularity of the second moment matrix of the regressors, but is easily carried out using iterative techniques. In our context, estimability is likely since the number of factors, which is equal to the rank of \mathcal{F} , is much smaller than the number of series, N . Therefore, in principle \mathcal{F} is estimable. However, as we noted earlier, due to the dynamic nature of the model unbiasedness of this estimate does not hold in finite samples and therefore no formal argument on the properties of ML estimation of \mathcal{F} is possible. However, assume that a similar situation arises in i.i.d. data where we consider a reduced rank regression of Y on X . Then, reduced rank ML estimation is unbiased. Denote this ML estimator of \mathcal{F} by $\hat{\mathcal{F}}^{ML}$. Denote $\hat{\mathcal{F}}^{MP} = (X'X)^+X'Y$. We know that in this case, $E(X \hat{\mathcal{F}}^{ML'}) = E(X \hat{\mathcal{F}}^{MP'})$. Since the structure is of a reduced rank form it follows that $E(X \hat{\mathcal{F}}^{ML'}) = E(X \hat{\mathcal{F}}_r^{MP'})$ where $\hat{\mathcal{F}}_r^{MP}$ is the reduced rank approximation to $\hat{\mathcal{F}}^{MP}$ obtained via SVD, and where r is the known true rank of \mathcal{F} . Conditioning on X gives $X E(\hat{\mathcal{F}}^{ML'} - \hat{\mathcal{F}}_r^{MP'}) = 0$. This holds if and only if either $E(\hat{\mathcal{F}}^{ML'} - \hat{\mathcal{F}}_r^{MP'}) = 0$ or $E(\hat{\mathcal{F}}^{ML'} - \hat{\mathcal{F}}_r^{MP'})$ belongs to the null space of X . It is, thus, seen that there is a close relationship between the estimation method we suggest and ML estimation.

6 Monte Carlo experiments

6.1 Setup

In this section we investigate the small sample properties of the factor estimates obtained through the three factor extraction methods. We concentrate on two aspects of the estimated factors. The first aspect concerns the relationship between the true and estimated common components obtained from the three methods. The common component estimate for the FHLR method is a direct product of their method. For the other two methods which provide estimates of the factors, we estimate the common components by a regression of the series on the estimated factors. The fitted values from this regression are the estimates of the common components. As we carry out a Monte Carlo exercise we know the true common components conditional on the data generation process we use, which will be discussed below. We use two measures to quantify the relationship between true and estimated common components. The first is a simple correlation between them. The second is the spectral coherency between them for selected frequencies.

The second aspect of interest is the set of properties of the idiosyncratic component of each series. We examine whether this component is white noise as in the DGP by using an LM(4) test and presenting the probability values of the test. The statistics we present are averages over all series in each Monte Carlo replication and over all Monte Carlo replications.

A very important question concerns the choice of the data generation process. As the FHLR and SW methods are nonparametric they provide little guidance as to the model that should generate the data. A very general but, of course, parametric model underlies the state space factor extraction method. Therefore, we adopt a state space model to generate the data. We assume a VARMA model for the transition equation. We therefore have the following model:

$$\begin{aligned}x_t &= Cf_t + \epsilon_t, \quad t = 1, \dots, T \\ A(L)f_t &= B(L)u_t\end{aligned}\tag{17}$$

where $A(L) = I - A_1(L) - \dots - A_p(L)$, $B(L) = I + B_1(L) + \dots + B_q(L)$. Of course, this can be rewritten in the standard state space format where the (redefined) state vector f_t follows a VAR(1) process. Note that in this case the number of factors in the VAR(1) representation will be equal to $p + q$.

We consider four groups of experiments and in general we present results for the following (N, T) pairs: $(50, 50)$, $(50, 100)$, $(50, 300)$, $(100, 50)$ and $(100, 100)$. Using the criterion $p = \log T^{1.25}$ yields $Np > T$ so that in general we adopt the modified SSS procedure of Section 5. To consider the standard SSS method of Section 3 we also consider an experiment where $(N, T) = (50, 300)$ but the lag order p is set to 5, and one with the lead order s set to 1 rather than to k (the number of factors), to evaluate the role of future information.

In the first group of experiments we assume that we have a single VARMA factor with 6 different specifications (these are labelled experiments 1-6):

- $a_1 = 0.2, b_1 = 0.4$
- $a_1 = 0.7, b_1 = 0.2$
- $a_1 = 0.3, a_2 = 0.1, b_1 = 0.15, b_2 = 0.15$
- $a_1 = 0.2, b_1 = -0.4$
- $a_1 = 0.7, b_1 = -0.2$
- $a_1 = 0.3, a_2 = 0.1, b_1 = -0.15, b_2 = -0.15$

In the second group we investigate the case of serially correlated idiosyncratic errors. The DGP for that 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 7-12. 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. Again the specification of Experiments 1-6 is used for the factors. For this group we restrict attention to the following (N, T) pairs: $(50, 50)$ and $(50, 100)$.

In the fourth group of experiments 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.

For this group we restrict attention to the following (N, T) pairs: (50,50) and (100,100).

For each group of experiments, the C matrix is generated using standard normal variates as elements. The error terms are generated as standard normal pseudo-random variables for experiments 1-6. The same holds for the errors terms in the AR processes of the idiosyncratic errors in experiments 7-12. 500 replications for each experiment are undertaken.

Finally, we need to specify the number of factors assumed in the estimation. Since the small sample properties of the information criteria for factor determination are not yet well known, using a data dependent method for determining the number of factors is likely to dilute the results of the Monte Carlo exercises because it will not be clear whether the factor selection criterion or the factor estimation method is the cause of the observed performance. Thus, we fix a priori the number of factors to $p+q$, which is larger than the true number of factors in the FHLR setup and should provide a reasonable approximation for SW too.

6.2 Results

The simulation results for the first two groups of experiments (using the modified SSS method) are summarized in Tables 1-5.

Starting with the smaller datasets ($N=50$, Tables 1-3), the SSS method clearly outperforms the other two in terms of correlation between true and estimated common components. The maximum improvement is obtained for experiment 6 (ARMA(2,2) factor) and $T=300$ where the SSS correlation is 12% larger than the PCA correlation. No method outperforms the SSS method for $N=50$ with the exception of experiment 2 for some sample sizes where the PCA correlation exceeds the SSS correlation by up to 9%. There is little evidence that the idiosyncratic component is serially correlated for any of the methods for experiments 1-6. The DPCA method provides slightly more evidence for serial correlation. In the case of serially correlated errors (Experiments 7-12) the same ranking of methods arises indicating that at least for low serial correlation the methods perform satisfactorily. Note that rejection of the LM test in the case of serially correlated idiosyncratic errors is clearly something to be expected.

Moving on to the $N=100$ experiments (Tables 4-5) we see that the SSS method retains superiority albeit the gap with PCA shrinks. The maximum improvement for correlation is around 6% for experiment 6 and $T=100$.

Again there is little evidence of serial correlation in the estimated idiosyncratic component with the exception of the DPCA method in general, for experiments 1-6 and the SSS method for experiment 2, where some weak evidence of serial correlation exists. Similar results arise for experiments 7-12.

We briefly comment on the estimated coherencies between true and estimated common components for the three methods, (figures are available upon request). Overall, all methods seem to have higher coherencies at lower frequencies indicating that the lower frequency components of the common components are easier to capture. This is indicated by the negative slope of the estimated coherencies. The only exception to this is experiment 5 (ARMA(1,1) factor with negative MA coefficient). Another interesting feature is that even for experiment 2 where the PCA method performs better than the SSS method, the SSS method has a better than or equal performance to the PCA method for lower frequencies. Finally, the DPCA method has in many cases a non-monotonic coherency especially for large datasets ($N=100$).

The DPCA method shows consistently lower correlation between true and estimated common components. It shows, in general, more evidence of serial correlation, although not to any significant extent. 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.

We can now compare the performance of the modified and standard SSS methods by comparing Tables 3 and 6 ($N=50$, $T=300$, but $p=5$ in Table 6). The differences are always minor and not statistically significant, though in general larger than those for PCA and DPCA which are only due to Monte Carlo variability. The SSS method still systematically outperforms PCA and DPCA in Table 6. In Table 7 we then report the results for $s = 1$ rather than

$s = k$. In this case the performance of SSS slightly deteriorates, it remains better than DPCA but is often beaten by PCA, though the differences are small.

Cross correlation in the errors (experiments for group 3, results in Tables 8 and 9) yields again the ranking SSS-PCA-DPCA. It is worth mentioning that for all methods there is no deterioration of performance with respect to the idiosyncratic errors case, compare the figures in Tables 1 and 2 with those in Tables 8 and 9.

To conclude, we briefly discuss the results from the factor-VAR experiments (group 4) in Table 10. In this case the three methods have very similar properties in terms of the estimated factors. This may indicate that for simple autoregressive processes with short lags there is little to choose between them.

7 An empirical example

In this section we use a dynamic factor model estimated with the three methods to analyze a large balanced dataset of 146 US macroeconomic variables, over the period 1959:1-1998:12. The dataset is extracted by the one in SW, dropping series with missing observations and other data irregularities. A list of all the variables, together with the stationarity transformation applied, is presented in the Appendix, see SW for additional details.

To start with, we estimate the common component of each variable according to the three methods, and then compute the resulting (adjusted) R^2 and the correlation among the three common components. SW showed that the first two SW factors are the most relevant for forecasting several variables in the dataset, while Favero, Marcellino and Neglia (2002) found that 3 or 4 FHLR factors are sufficient according to the dynamic eigenvalues criterion described in Section 2. Since it is better to overestimate the number of factors rather than underestimate it, we have chosen to use six factors. We have also repeated the analysis with three factors, but similar results are obtained. For the SSS method the number of lags (p) was set to 3, that allows to use the standard procedure of Section 3, the number of leads (s) to 1, to avoid any use of future information, and an AR specification is assumed for each of the factors. For FHLR we set $M = 3$, $g = s = 2$, as in Favero et al. (2002).

Table 11 reports the results. Focusing on the R^2 first, the performance

of SSS and PCA is comparable, the latter is slightly better than the former on average (0.44 versus 0.39), while DPCA is ranked first, with an average R^2 over all variables of about 0.52. A similar pattern emerges from a more disaggregate analysis, DPCA yields a higher R^2 for most variables. The better fit of DPCA could be explained by the longer sample available, which improves substantially the multivariate spectrum estimation, and by the use of future information in the computation of the spectrum. On the other hand, as the Monte Carlo results show, the better fit may be an artefact of the tendency of the DPCA method to soak up part of the idiosyncratic component in the data.

The correlation among the estimated common components is highest for SSS-PCA, with an average value of 0.93, slightly lower but still considerable for PCA-DPCA, 0.76, and SSS-DPCA, 0.73. Overall, these values are in line with the Monte Carlo simulations, which showed a higher similarity of PCA and SSS.

The second exercise we consider is the inclusion of the estimated factors in a monetary VAR to evaluate the response of inflation and the output gap to unexpected monetary shocks, see e.g. Stock and Watson (2002) for a recent review of the methodology. Favero and Marcellino (2001) showed that for several European countries responses more in line with economic theory are obtained when the SW factors are included in the VAR, and the standard error around the responses decreases. Favero et al. (2002) obtained a similar result in the case of the US, though the gains are smaller, and DPCA seems to perform slightly better than PCA. The rationale for the better results with the factors in the VAR is that they summarize the information contained in large datasets, thus reducing the common problem of omitted variable bias in small scale VARs. Here we are interested in the relative performance of the SSS method.

The VARs we consider are for the output gap (USGAP), inflation (US-INFL), a commodity price index, the effective exchange rate, the federal fund rate (USPR) and six factors treated as exogenous regressors. Four lags are included for each endogenous variable and the VAR is estimated over the sample 1980:1-1998:12 to cover a relatively homogenous period from the monetary policy point of view but long enough to obtain reliable estimates of the parameters. Impulse response functions are obtained with a Choleski decomposition with the variables ordered as listed above, see Favero et al (2002) for additional details.

The responses of USGAP, USINFL and USPR to a one standard deviation

shock in USPR are graphed in Figure 1 for the cases where the factors are excluded from the VAR (base), and when they are included as exogenous regressors and estimated according to each of the three methods. To use a comparable information set, the DPCA are lagged three periods, since two future quarters are used to compute the spectrum, while the PCA and SSS only once.

The base case shows a positive (though not significant) response of US-INFL for about 3 years, what is commonly named price puzzle since inflation should instead decrease. The positive reaction of USGAP is also not in line with standard economic theory. The inclusion of the dynamic principal components does not change sensibly the pattern of response; with static principal components the USGAP decreases (after deletion of the first factor); and with the SSS factors also the price puzzle is eliminated. To obtain such a result with PCA or DPCA a larger number of factors has to be included in the VAR, up to 12.

In summary, this section supports the usefulness of factor models for the analysis of large datasets of variables but also to address important economic questions. In the particular applications we consider, the performance of the three factor estimation methods is rather similar, with a slight advantage for our SSS.

8 Conclusions

In this paper we have developed an estimation method for dynamic factor models of large dimension based on a subspace algorithm applied to the state space representation of the model (SSS). We have derived the asymptotic properties of the estimators, formulae for their standards errors, and information criteria for a consistent selection of the number of factors.

Then we have undertaken an extensive comparative analysis of the performance of alternative factor estimation methods. We have focused on the relationship between true and estimated common components as measured by the correlation and coherency between them. Additionally, we have examined the serial correlation properties of the estimated idiosyncratic components. Our main conclusion is that the SSS method, which takes explicit account of the dynamic nature of the data generating process, performs better than alternative approaches for a number of experimental setups. Static principal components seem to perform satisfactorily overall, while dynamic principal

components appear slightly less able to distinguish between common and idiosyncratic factors, in the particular setup we have considered which is, nevertheless, quite general.

Finally, we have provided an empirical application with a large dataset for the US, that further confirms the good empirical performance of the SSS method and, more generally, the usefulness of the dynamic factor model as a modelling tool for datasets of large dimension.

There are many other interesting directions for research in this field, and we are currently working on two of them. First, structural identification of the factors or of the underlying shocks, along the lines of the structural VAR literature. This is simpler in our model based framework than in other non-parametric approaches to large scale factor modelling. Second, the common assumption of weakly stationary variables can be relaxed, since state space modelling of nonstationary data and estimation of linear dynamic models with non-stationary regressors, the two main components of our SSS approach, are well developed.

9 Appendix - Some additional results

9.0.1 SSS and SSS0 in small samples

We undertake a small Monte Carlo experiment to evaluate the relative properties of the SSS0 and SSS decompositions. We generate data according to the model

$$\begin{aligned} Y &= F\delta' + \epsilon, \\ F &= X\gamma, \end{aligned}$$

where Y, X, ϵ are $T \times N$ matrices, and X and ϵ are matrices of independent standard normal random variables and independent normal variables of variance a respectively. The parameter a controls the signal to noise ratio and in our experiment takes the values 0.1, 0.5, 1, 5, 15, 30. We evaluate the approximation error of SSS0 and SSS by the average eigenvalue of $G'G$ where G is the difference between $X\gamma\delta'$ and the estimated fitted value assuming one factor. In other words we approximate the fitted value by its rank-1 approximation. We set $T = 20$ and vary N as $N = 10, 11, \dots, 50$. Figure 1 presents the approximation error. Some interesting results arise. For high signal to noise ratios (low a) SSS0 performs better than SSS but only in the vicinity of $N = 20$. Otherwise the two methods work equally well. For low signal to noise ratios the SSS works better than SSS0 as expected given that the presence of a large ϵ leads to the deterioration of the performance of SSS0. The finding for high signal to noise ratios generalises to other values of T . It seems that SSS works well as long as N and T are different for this particularly simple setup. Whether the inverse of $X'X$ exists or not does not appear to be relevant in these experiments.

9.0.2 Improved Factor Estimation

There exists potential for improving upon the standard method of estimating the factors. This is related to the structure of the covariance matrix of the error term of (8). If the lead truncation index, s , is greater than 1, or D is not diagonal, then (8) should be estimated by generalised least squares as there is serial correlation in these error terms. We conjecture that, following results in standard regression analysis, consistency, the rate of convergence and asymptotic normality of the factor estimates are not affected by the presence of serial correlation. Of course the asymptotic variance of the factor

estimates will be different reflecting the use of GLS. Further, the asymptotic variance of feasible GLS is likely to be different to that of infeasible GLS.

We will address only the case of large datasets (i.e. $Np > T$) since the standard case follows easily from that if we note that the Moore-Penrose inverse of a matrix is equal to the inverse of a matrix if that exists. Define $\mathcal{X}^p = I \otimes X^p$, $\mathcal{X}^f = \text{vec}(X^f)$, $E^f = (E_1^f, E_2^f, \dots, E_T^f)'$ and $\mathcal{E}^f = \text{vec}(E^f)$. We first note that the covariance of \mathcal{E}^f follows from the definition of \mathcal{E} in (8). It is a complicated function of A, B, C, D . The parameters involved can be estimated from the estimation of the state space model following OLS estimation of the factors. Once we have an estimate of the covariance matrix of \mathcal{E} , denoted $\hat{\mathcal{V}}$ we can use this to obtain the best affine estimate of $\text{vec}(X^p \mathcal{F})$ as derived by Magnus and Neudecker (1988) and given by

$$\text{vec}(\widehat{X^p \mathcal{F}}) = \mathcal{X}^p (\mathcal{X}^{p'} \hat{\mathcal{V}}^+ \mathcal{X}^p)^+ \mathcal{X}^{p'} \hat{\mathcal{V}}^+ \mathcal{X}^f \quad (18)$$

Then, we can apply the SVD to $(\mathcal{X}^{p'} \hat{\mathcal{V}}^+ \mathcal{X}^p)^+ \mathcal{X}^{p'} \hat{\mathcal{V}}^+ \mathcal{X}^f$ as before. Even if no improved estimation is undertaken $\hat{\mathcal{V}}$ maybe used to obtain the correct standard errors for the factors under OLS estimation for $s > 1$.

References

- [1] Bai, J. and S. Ng (2002), “Determining the number of factors in approximate factor models”, *Econometrica*, 70, 191-223.
- [2] Bauer, D. (1998), “Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms”, Ph.d. Thesis.
- [3] Berk K. N. (1974), “Consistent Autoregressive Spectral Estimates” , *Annals of Statistics*, 2(3) , 489–502.
- [4] Brillinger D.R. (1981), Time Series: Data Analysis and Theory, *McGraw-Hill*.
- [5] Camba-Mendez G., G. Kapetanios, R. J. Smith and M. R. Weale (2001), ”An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries”, *Econometrics Journal*, 4(1), S56-90.
- [6] Chatelin (1983), Spectral Approximation of Linear Operators, *Academic Press*
- [7] Chamberlain, G. and M. Rothschild (1983), “Arbitrage factor structure, and mean variance analysis of large asset markets”, *Econometrica*, 51, 1281-1304.
- [8] Connor, G. and R.A. Korajczyk (1986), “Performance measurement with the arbitrage pricing theory”, *Journal of Financial Economics*, 15, 373-394.
- [9] 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.
- [10] Deistler M. and E.J. Hannan (1988), “The Statistical Analysis of Linear Systems”, John Wiley.
- [11] Favero, C.A. and Marcellino, M. (2001), “Large datasets, small models and monetary policy in Europe”, CEPR WP 3098.

- [12] Favero, C.A., Marcellino, M. and Neglia, F.H. (2002), “Principal components at work: the empirical analysis of monetary policy with large datasets”, IGIER WP 223.
- [13] Forni, M. and L. Reichlin (1996), “Dynamic common factors in large cross-sections”, *Empirical Economics*, 21, 27-42.
- [14] Forni, M. and L. Reichlin (1997), “National policies and local economies: Europe and the United States”, CEPR WP 1632.
- [15] 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.
- [16] Forni, M., Hallin, M., Lippi, M. and L. Reichlin (1998), “The generalized factor model: identification and estimation”, *The Review of Economic and Statistics*, forthcoming.
- [17] Forni, M., Hallin, M., Lippi, M. and L. Reichlin (1999), “Reference cycles: The NBER methodology revisited”, manuscript.
- [18] Geweke, J. (1977), “The dynamic factor analysis of economic time series”, ch. 19 in Aigner, D.J. and A.S. Goldberger (eds.), *Latent variables in socio-economic models*, Amsterdam: North Holland.
- [19] W.E. Larimore (1983), “System Identification, Reduced Order Filters and Modelling via Canonical Variate Analysis”, Proc. 1983 Amer. Control Conference, pp. 445-451.
- [20] J. Magnus and H. Neudecker (1988), “Matrix Differential Calculus with Applications to Statistics and Econometrics”, John Wiley.
- [21] P. Van Overschee and B. De Moor (1996), “Subspace Identification for Linear Systems”, Kluwer Academic Publishers.
- [22] Quah, D. and T.J. Sargent (1993), “A dynamic index model for large cross-sections”, ch. 7 in Stock, J.H. and M.W. Watson (eds.), *Business cycles, indicators and forecasting*, University of Chicago Press for the NBER.

- [23] Sargent, T.J. and C.A. Sims (1977), “Business cycle modelling without pretending to have too much a-priori economic theory”, in Sims, C.A. (ed.), *New methods in business cycle research*, Minneapolis: Federal Reserve Bank of Minneapolis.
- [24] Sin, C. Y. and H. White (1996), “Information criteria for selecting possibly misspecified parametric models”, *Journal of Econometrics*, 71, 207-225.
- [25] Stock, J.H. and M.W. Watson (1989), “New indexes of coincident and leading economic indicators”, *NBER Macroeconomic Annual*, 351-393.
- [26] Stock, J.H. and M.W. Watson (1991), “A probability model of the coincident economic indicators”, ch.4 in Lahiri, K. and G.H. Moore (eds.), *Leading economic indicators: New approaches and forecasting records*, New York: Cambridge University Press.
- [27] Stock, J.H. and M.W. Watson (2001), “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business and Economic Statistics*, 20, 147-162.

Table 1: Results for case: N=50, T=50

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.821 _(0.055)	0.852 _(0.058)	0.727 _(0.055)	0.068 _(0.035)	0.069 _(0.037)	0.093 _(0.039)
Exp 2	0.861 _(0.049)	0.857 _(0.067)	0.782 _(0.055)	0.073 _(0.037)	0.097 _(0.090)	0.102 _(0.044)
Exp 3	0.738 _(0.049)	0.783 _(0.055)	0.630 _(0.052)	0.069 _(0.037)	0.087 _(0.046)	0.137 _(0.052)
Exp 4	0.804 _(0.053)	0.842 _(0.055)	0.702 _(0.051)	0.067 _(0.035)	0.066 _(0.036)	0.096 _(0.042)
Exp 5	0.826 _(0.053)	0.851 _(0.055)	0.733 _(0.053)	0.068 _(0.035)	0.071 _(0.038)	0.102 _(0.043)
Exp 6	0.708 _(0.052)	0.753 _(0.067)	0.597 _(0.051)	0.066 _(0.034)	0.076 _(0.039)	0.132 _(0.050)
Exp 7	0.814 _(0.055)	0.839 _(0.059)	0.722 _(0.053)	0.104 _(0.040)	0.114 _(0.045)	0.133 _(0.052)
Exp 8	0.850 _(0.052)	0.841 _(0.067)	0.774 _(0.056)	0.104 _(0.042)	0.140 _(0.095)	0.133 _(0.048)
Exp 9	0.729 _(0.057)	0.759 _(0.067)	0.627 _(0.060)	0.083 _(0.039)	0.100 _(0.050)	0.161 _(0.055)
Exp 10	0.801 _(0.051)	0.844 _(0.055)	0.700 _(0.048)	0.110 _(0.045)	0.133 _(0.051)	0.149 _(0.053)
Exp 11	0.817 _(0.053)	0.833 _(0.059)	0.728 _(0.054)	0.104 _(0.044)	0.121 _(0.053)	0.142 _(0.049)
Exp 12	0.697 _(0.055)	0.734 _(0.075)	0.589 _(0.054)	0.083 _(0.040)	0.116 _(0.049)	0.158 _(0.054)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 2: Results for case: N=50, T=100

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.857 _(0.043)	0.895 _(0.042)	0.782 _(0.045)	0.058 _(0.032)	0.061 _(0.034)	0.070 _(0.036)
Exp 2	0.892 _(0.038)	0.873 _(0.057)	0.832 _(0.042)	0.060 _(0.033)	0.140 _(0.139)	0.073 _(0.037)
Exp 3	0.777 _(0.046)	0.843 _(0.044)	0.690 _(0.048)	0.055 _(0.031)	0.066 _(0.036)	0.086 _(0.039)
Exp 4	0.842 _(0.043)	0.888 _(0.044)	0.757 _(0.044)	0.055 _(0.032)	0.056 _(0.034)	0.066 _(0.035)
Exp 5	0.862 _(0.041)	0.884 _(0.048)	0.788 _(0.044)	0.060 _(0.034)	0.077 _(0.068)	0.073 _(0.037)
Exp 6	0.751 _(0.043)	0.838 _(0.043)	0.657 _(0.043)	0.057 _(0.031)	0.060 _(0.034)	0.083 _(0.038)
Exp 7	0.852 _(0.041)	0.887 _(0.042)	0.778 _(0.042)	0.213 _(0.055)	0.245 _(0.063)	0.237 _(0.060)
Exp 8	0.881 _(0.041)	0.851 _(0.070)	0.823 _(0.043)	0.205 _(0.056)	0.348 _(0.161)	0.228 _(0.061)
Exp 9	0.768 _(0.046)	0.823 _(0.048)	0.684 _(0.047)	0.161 _(0.051)	0.207 _(0.072)	0.228 _(0.061)
Exp 10	0.834 _(0.044)	0.885 _(0.046)	0.752 _(0.044)	0.228 _(0.059)	0.257 _(0.060)	0.259 _(0.059)
Exp 11	0.852 _(0.044)	0.871 _(0.048)	0.779 _(0.044)	0.210 _(0.059)	0.255 _(0.078)	0.238 _(0.063)
Exp 12	0.746 _(0.043)	0.830 _(0.045)	0.653 _(0.044)	0.165 _(0.054)	0.238 _(0.064)	0.226 _(0.058)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 3: Results for case: N=50, T=300

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.888 _(0.033)	0.923 _(0.031)	0.838 _(0.035)	0.051 _(0.032)	0.074 _(0.053)	0.056 _(0.033)
Exp 2	0.913 _(0.028)	0.827 _(0.100)	0.875 _(0.031)	0.053 _(0.032)	0.383 _(0.232)	0.058 _(0.032)
Exp 3	0.816 _(0.037)	0.889 _(0.031)	0.756 _(0.040)	0.050 _(0.032)	0.096 _(0.085)	0.061 _(0.034)
Exp 4	0.872 _(0.032)	0.928 _(0.031)	0.813 _(0.035)	0.051 _(0.033)	0.052 _(0.032)	0.058 _(0.033)
Exp 5	0.893 _(0.033)	0.900 _(0.040)	0.841 _(0.036)	0.054 _(0.032)	0.141 _(0.129)	0.058 _(0.034)
Exp 6	0.791 _(0.036)	0.903 _(0.031)	0.725 _(0.038)	0.052 _(0.033)	0.054 _(0.034)	0.063 _(0.035)
Exp 7	0.885 _(0.034)	0.921 _(0.032)	0.833 _(0.036)	0.729 _(0.066)	0.782 _(0.065)	0.743 _(0.066)
Exp 8	0.910 _(0.031)	0.833 _(0.080)	0.873 _(0.032)	0.722 _(0.064)	0.886 _(0.074)	0.735 _(0.064)
Exp 9	0.808 _(0.036)	0.883 _(0.033)	0.750 _(0.038)	0.668 _(0.068)	0.781 _(0.076)	0.718 _(0.067)
Exp 10	0.870 _(0.034)	0.929 _(0.032)	0.811 _(0.037)	0.739 _(0.063)	0.757 _(0.060)	0.757 _(0.061)
Exp 11	0.888 _(0.034)	0.892 _(0.043)	0.839 _(0.036)	0.723 _(0.069)	0.805 _(0.081)	0.741 _(0.066)
Exp 12	0.782 _(0.039)	0.897 _(0.035)	0.719 _(0.040)	0.677 _(0.069)	0.753 _(0.062)	0.725 _(0.060)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 4: Results for case: N=100, T=50

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.843 _(0.037)	0.860 _(0.038)	0.744 _(0.040)	0.067 _(0.025)	0.069 _(0.027)	0.102 _(0.031)
Exp 2	0.870 _(0.038)	0.836 _(0.071)	0.788 _(0.043)	0.072 _(0.025)	0.108 _(0.100)	0.107 _(0.034)
Exp 3	0.762 _(0.046)	0.778 _(0.055)	0.643 _(0.050)	0.070 _(0.025)	0.088 _(0.034)	0.159 _(0.043)
Exp 4	0.822 _(0.040)	0.844 _(0.040)	0.715 _(0.039)	0.064 _(0.024)	0.064 _(0.025)	0.098 _(0.032)
Exp 5	0.840 _(0.041)	0.841 _(0.048)	0.739 _(0.044)	0.070 _(0.025)	0.073 _(0.039)	0.106 _(0.033)
Exp 6	0.735 _(0.042)	0.743 _(0.078)	0.610 _(0.042)	0.067 _(0.027)	0.079 _(0.030)	0.145 _(0.040)
Exp 7	0.830 _(0.040)	0.840 _(0.046)	0.735 _(0.042)	0.103 _(0.030)	0.112 _(0.034)	0.138 _(0.036)
Exp 8	0.861 _(0.039)	0.829 _(0.057)	0.786 _(0.043)	0.102 _(0.031)	0.148 _(0.093)	0.137 _(0.037)
Exp 9	0.749 _(0.043)	0.754 _(0.052)	0.637 _(0.046)	0.083 _(0.029)	0.098 _(0.035)	0.172 _(0.042)
Exp 10	0.818 _(0.039)	0.846 _(0.041)	0.714 _(0.038)	0.112 _(0.032)	0.128 _(0.033)	0.155 _(0.039)
Exp 11	0.832 _(0.038)	0.816 _(0.061)	0.738 _(0.042)	0.101 _(0.031)	0.114 _(0.049)	0.144 _(0.038)
Exp 12	0.725 _(0.042)	0.723 _(0.083)	0.606 _(0.043)	0.086 _(0.029)	0.110 _(0.038)	0.170 _(0.046)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 5: Results for case: N=100, T=100

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.877 _(0.029)	0.892 _(0.028)	0.798 _(0.033)	0.058 _(0.023)	0.063 _(0.028)	0.073 _(0.026)
Exp 2	0.904 _(0.028)	0.829 _(0.085)	0.844 _(0.031)	0.060 _(0.023)	0.192 _(0.186)	0.075 _(0.027)
Exp 3	0.805 _(0.031)	0.835 _(0.032)	0.708 _(0.036)	0.059 _(0.024)	0.069 _(0.038)	0.091 _(0.028)
Exp 4	0.861 _(0.031)	0.891 _(0.031)	0.773 _(0.033)	0.057 _(0.023)	0.057 _(0.023)	0.068 _(0.027)
Exp 5	0.879 _(0.031)	0.869 _(0.052)	0.802 _(0.034)	0.058 _(0.024)	0.087 _(0.086)	0.072 _(0.027)
Exp 6	0.784 _(0.032)	0.843 _(0.032)	0.678 _(0.034)	0.056 _(0.022)	0.060 _(0.024)	0.089 _(0.028)
Exp 7	0.869 _(0.031)	0.883 _(0.032)	0.793 _(0.032)	0.211 _(0.042)	0.250 _(0.057)	0.233 _(0.044)
Exp 8	0.897 _(0.030)	0.815 _(0.095)	0.839 _(0.033)	0.209 _(0.042)	0.389 _(0.179)	0.228 _(0.042)
Exp 9	0.798 _(0.034)	0.817 _(0.036)	0.706 _(0.036)	0.163 _(0.037)	0.199 _(0.067)	0.227 _(0.042)
Exp 10	0.854 _(0.031)	0.888 _(0.031)	0.768 _(0.031)	0.229 _(0.044)	0.252 _(0.045)	0.262 _(0.045)
Exp 11	0.867 _(0.033)	0.852 _(0.047)	0.794 _(0.033)	0.209 _(0.040)	0.270 _(0.096)	0.240 _(0.043)
Exp 12	0.775 _(0.033)	0.831 _(0.036)	0.675 _(0.034)	0.168 _(0.038)	0.232 _(0.052)	0.234 _(0.042)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 6: Results for case: N=50, T=300, p=5

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.887 _(0.033)	0.885 _(0.034)	0.836 _(0.036)	0.053 _(0.032)	0.041 _(0.028)	0.058 _(0.032)
Exp 2	0.919 _(0.030)	0.929 _(0.027)	0.878 _(0.033)	0.054 _(0.032)	0.039 _(0.027)	0.059 _(0.034)
Exp 3	0.814 _(0.035)	0.863 _(0.032)	0.755 _(0.038)	0.053 _(0.032)	0.040 _(0.029)	0.064 _(0.034)
Exp 4	0.793 _(0.038)	0.835 _(0.035)	0.727 _(0.041)	0.052 _(0.031)	0.043 _(0.029)	0.065 _(0.037)
Exp 5	0.892 _(0.032)	0.894 _(0.032)	0.842 _(0.033)	0.054 _(0.033)	0.041 _(0.027)	0.057 _(0.033)
Exp 6	0.872 _(0.035)	0.856 _(0.036)	0.813 _(0.036)	0.052 _(0.032)	0.042 _(0.028)	0.057 _(0.035)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 7: Results for case: N= 50 T= 300, s=1, p=5

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.889 _(0.032)	0.847 _(0.034)	0.837 _(0.035)	0.051 _(0.031)	0.039 _(0.027)	0.056 _(0.033)
Exp 2	0.917 _(0.028)	0.898 _(0.030)	0.877 _(0.032)	0.053 _(0.032)	0.040 _(0.028)	0.055 _(0.032)
Exp 3	0.817 _(0.035)	0.787 _(0.037)	0.758 _(0.038)	0.052 _(0.030)	0.039 _(0.028)	0.062 _(0.034)
Exp 4	0.788 _(0.039)	0.744 _(0.038)	0.723 _(0.040)	0.051 _(0.030)	0.039 _(0.027)	0.061 _(0.033)
Exp 5	0.893 _(0.033)	0.853 _(0.035)	0.842 _(0.035)	0.051 _(0.032)	0.038 _(0.027)	0.055 _(0.032)
Exp 6	0.870 _(0.033)	0.810 _(0.036)	0.811 _(0.036)	0.050 _(0.031)	0.037 _(0.027)	0.055 _(0.033)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 8: Results for cross-correlated idiosyncratic errors (Case: N=50
T=50)

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	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 2	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 3	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 4	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 5	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 6	0.705 _(0.053)	0.761 _(0.065)	0.598 _(0.052)	0.067 _(0.036)	0.075 _(0.041)	0.128 _(0.046)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 9: Results for cross-correlated idiosyncratic errors (Case: N=50
T=100)

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
Exp 1	0.849 _(0.043)	0.891 _(0.045)	0.775 _(0.044)	0.058 _(0.034)	0.059 _(0.034)	0.070 _(0.036)
Exp 2	0.884 _(0.036)	0.870 _(0.054)	0.825 _(0.038)	0.061 _(0.034)	0.141 _(0.134)	0.073 _(0.036)
Exp 3	0.767 _(0.044)	0.844 _(0.044)	0.685 _(0.046)	0.058 _(0.031)	0.066 _(0.041)	0.087 _(0.040)
Exp 4	0.834 _(0.043)	0.889 _(0.042)	0.752 _(0.042)	0.054 _(0.033)	0.057 _(0.034)	0.067 _(0.036)
Exp 5	0.854 _(0.045)	0.884 _(0.043)	0.782 _(0.047)	0.057 _(0.031)	0.066 _(0.050)	0.069 _(0.036)
Exp 6	0.738 _(0.045)	0.835 _(0.045)	0.649 _(0.044)	0.059 _(0.033)	0.064 _(0.035)	0.081 _(0.039)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

Table 10: Results for trivariate VAR(1) case

Exp.	Corr. with True ^a			Serial Correlation ^b		
	PCA	SSS	DPCA	PCA	SSS	DPCA
N=50,T=50	0.976 _(0.008)	0.966 _(0.010)	0.949 _(0.013)	0.075 _(0.039)	0.072 _(0.039)	0.122 _(0.050)
N=100,T=100	0.988 _(0.003)	0.956 _(0.012)	0.974 _(0.005)	0.062 _(0.025)	0.094 _(0.060)	0.081 _(0.028)

^aMean Correlation between true and estimated common component

^bMean probability value of LM serial correlation test of idiosyncratic component

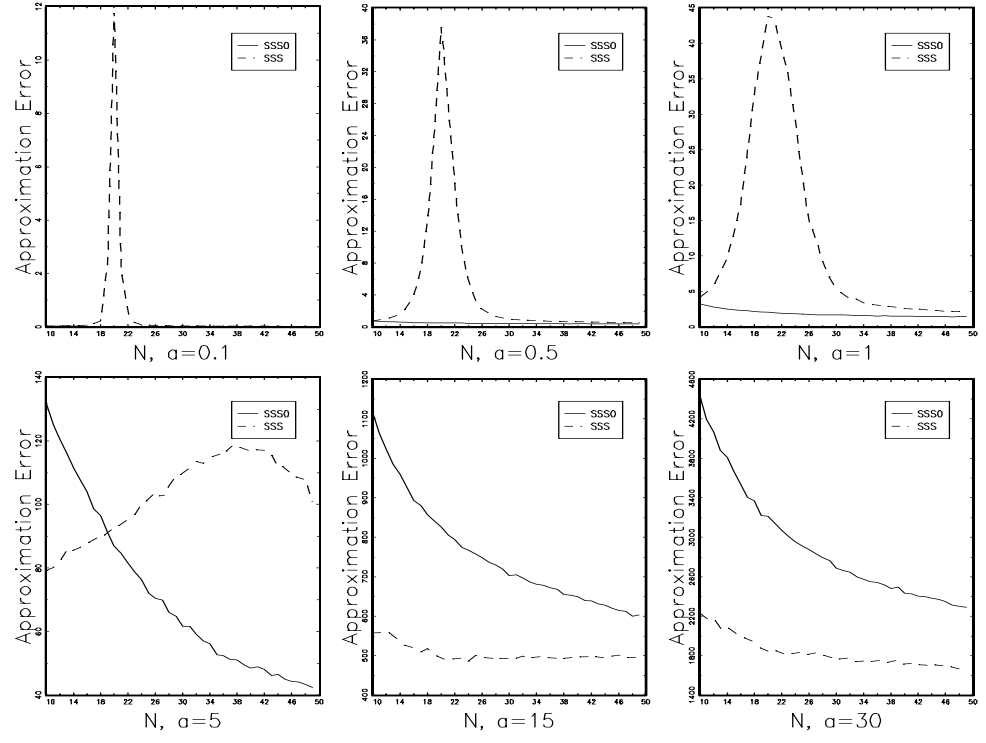


Figure 2: Comparison of SSS and SSS0

Table 11: US dataset – Fit of factor model and correlations of permanent components

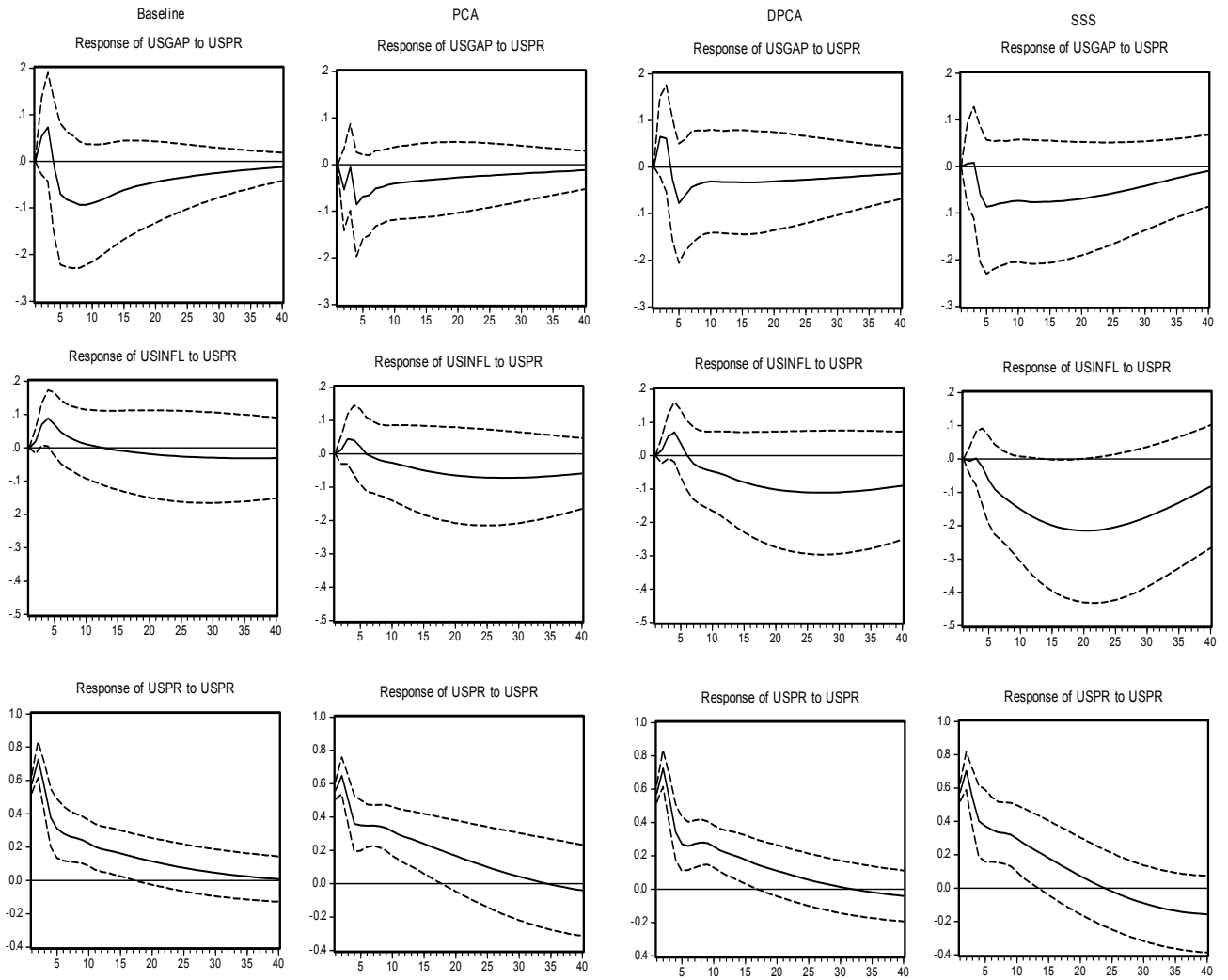
Var.	Adjusted-R ²			Correlations		
	SSS	PCA	DPCA	PCA-DPCA	PCA-SSS	SSS-DPCA
1	0.6814	0.7699	0.829	0.9329	0.9375	0.8917
2	0.6761	0.7129	0.8129	0.9004	0.9454	0.871
3	0.6305	0.6523	0.758	0.8853	0.9541	0.8634
4	0.5283	0.5195	0.7094	0.8214	0.9486	0.8212
5	0.4465	0.4743	0.6009	0.8522	0.9524	0.8309
6	0.2114	0.1758	0.3822	0.6444	0.9379	0.6742
7	0.4345	0.4805	0.4896	0.912	0.9426	0.8652
8	0.3726	0.4408	0.496	0.8939	0.9031	0.8326
9	0.4658	0.5665	0.6118	0.9312	0.9278	0.8785
10	0.3188	0.3673	0.4421	0.8646	0.9158	0.8139
11	0.6825	0.7838	0.8472	0.9326	0.9316	0.8862
12	0.6071	0.7032	0.7573	0.9241	0.9328	0.8742
13	0.4208	0.452	0.5672	0.8653	0.9163	0.8375
14	0.0329	0.0322	0.0646	0.5548	0.9292	0.5701
15	0.0387	0.0631	0.0915	0.4253	0.8912	0.3895
16	0.6165	0.7973	0.8476	0.9427	0.9017	0.8279
17	0.2577	0.3234	0.4394	0.8328	0.8695	0.7011
18	0.634	0.7441	0.8345	0.9348	0.894	0.8326
19	0.239	0.2603	0.3577	0.784	0.9049	0.733
20	0.2608	0.2972	0.3569	0.8385	0.9057	0.7743
21	0.5834	0.8001	0.9311	0.9004	0.8785	0.7624
22	0.669	0.6364	0.8275	0.8727	0.8244	0.8456
23	0.7052	0.8005	0.8323	0.9164	0.8914	0.8877
24	0.7946	0.8686	0.9241	0.9263	0.8816	0.8962
25	0.758	0.8185	0.9228	0.9103	0.8575	0.8709
26	0.7738	0.8629	0.9146	0.9309	0.8761	0.8826
27	0.5916	0.7607	0.7532	0.9388	0.902	0.8422
28	0.6228	0.7719	0.7692	0.9365	0.8983	0.8435
29	0.6263	0.7738	0.7673	0.9495	0.9058	0.8601
30	0.2442	0.2805	0.3475	0.7501	0.8559	0.6864
31	0.5725	0.7312	0.7513	0.9447	0.916	0.8468
32	0.5314	0.673	0.6878	0.9442	0.921	0.8534
33	0.3525	0.4671	0.532	0.8795	0.888	0.7646
34	0.3113	0.4196	0.4553	0.8605	0.9055	0.7614
35	0.3166	0.4155	0.4809	0.867	0.8857	0.7529
36	0.1767	0.2019	0.3267	0.7337	0.8858	0.7101
37	0.1837	0.2109	0.2597	0.8029	0.9155	0.7393
38	0.179	0.1859	0.1657	0.796	0.858	0.7797
39	0.3901	0.2453	0.8319	0.5377	0.7654	0.6717
40	0.3541	0.1466	0.8492	0.4288	0.7255	0.6556
41	0.0134	0.0151	0.1761	0.32	0.8822	0.3223
42	0.0113	0.0103	0.1243	0.3305	0.928	0.3022
43	0.5367	0.6399	0.8634	0.8325	0.8705	0.771
44	0.3702	0.3897	0.6201	0.7503	0.8831	0.7532
45	0.4422	0.4908	0.5396	0.911	0.9119	0.8671
46	0.5508	0.6057	0.7857	0.8327	0.8641	0.7868
47	0.3253	0.4393	0.6302	0.7968	0.8302	0.7102
48	0.5487	0.5976	0.8732	0.8053	0.8541	0.7778
49	0.3917	0.3029	0.564	0.6836	0.8812	0.7745

50	0.7778	0.7765	0.8364	0.9283	0.981	0.9221
51	0.6335	0.6354	0.7501	0.8992	0.9787	0.8885
52	0.5694	0.5818	0.6858	0.9052	0.9766	0.8894
53	0.285	0.2709	0.3764	0.7855	0.9718	0.7875
54	0.297	0.31	0.3867	0.7735	0.9806	0.7679
55	0.3232	0.3262	0.4428	0.8221	0.9649	0.8081
56	0.1686	0.1761	0.175	0.6944	0.9894	0.6878
57	0.3444	0.3476	0.5994	0.6969	0.9797	0.6861
58	0.154	0.1604	0.4729	0.5593	0.972	0.5403
59	0.3281	0.4199	0.4618	0.8007	0.9194	0.7203
60	0.3017	0.408	0.3864	0.8422	0.9052	0.7473
61	0.2657	0.3732	0.3544	0.8636	0.9079	0.7856
62	0.1351	0.1558	0.1316	0.6577	0.9228	0.5539
63	0.1047	0.1085	0.108	0.7662	0.9244	0.7121
64	0.0877	0.1388	0.3009	0.5727	0.9037	0.4793
65	0.7791	0.7885	0.8375	0.9234	0.9798	0.918
66	0.6465	0.6664	0.7411	0.9114	0.9742	0.8984
67	0.293	0.3	0.3778	0.7703	0.9852	0.7703
68	0.2321	0.2513	0.551	0.5616	0.9784	0.5478
69	0.6014	0.7659	0.8804	0.9329	0.9073	0.838
70	0.4828	0.7023	0.8258	0.9167	0.8953	0.7923
71	0.4903	0.7081	0.8179	0.9328	0.8882	0.8041
72	0.5319	0.5553	0.6797	0.8897	0.9108	0.8541
73	0.539	0.6093	0.6724	0.9241	0.9177	0.8666
74	0.6329	0.7839	0.8631	0.9348	0.9211	0.8475
75	0.4916	0.5131	0.6241	0.8864	0.9243	0.8533
76	0.5661	0.5849	0.6218	0.9348	0.9748	0.9136
77	0.5717	0.6329	0.8017	0.8488	0.9704	0.8122
78	0.2186	0.244	0.4698	0.627	0.9437	0.596
79	0.6991	0.7512	0.8295	0.8987	0.9819	0.8716
80	0.5231	0.5734	0.7408	0.8193	0.9743	0.7883
81	0.5658	0.6267	0.7947	0.8458	0.9732	0.8114
82	0.4699	0.5224	0.7255	0.7895	0.9685	0.7563
83	0.4102	0.4062	0.4811	0.8065	0.9888	0.8133
84	0.2102	0.2173	0.3042	0.7549	0.9823	0.7533
85	0.4266	0.4769	0.5993	0.8394	0.9431	0.8024
86	0.4131	0.4635	0.5853	0.8357	0.94	0.7987
87	0.146	0.2514	0.3088	0.7986	0.8595	0.6641
88	0.1631	0.1878	0.4318	0.6002	0.9298	0.5773
89	0.1549	0.1815	0.4328	0.5939	0.9268	0.5687
90	0.0694	0.0812	0.2015	0.5179	0.9669	0.492
91	0.058	0.0626	0.1771	0.4467	0.9338	0.4074
92	0.0335	0.0361	0.1034	0.3441	0.9027	0.3414
93	0.2214	0.2675	0.4387	0.6479	0.9247	0.5903
94	0.0137	0.02	0.1346	0.2883	0.9453	0.2732
95	0.014	0.0157	0.1845	0.2313	0.9432	0.2251
96	0.0517	0.0668	0.1584	0.4858	0.9656	0.4595
97	0.3773	0.4453	0.4041	0.7251	0.9554	0.6749
98	0.384	0.4497	0.401	0.7304	0.9579	0.6828
99	0.3673	0.43	0.3856	0.7272	0.9579	0.6797
100	0.3121	0.3641	0.3231	0.7255	0.9566	0.6743
101	0.313	0.3582	0.286	0.7649	0.9599	0.7398
102	0.4412	0.4389	0.8244	0.6871	0.8703	0.7029
103	0.4357	0.4135	0.7201	0.7238	0.8734	0.7474
104	0.4303	0.473	0.5235	0.7507	0.9726	0.741

105	0.4134	0.4591	0.5108	0.733	0.9712	0.7214
106	0.452	0.5108	0.4867	0.7773	0.9634	0.7531
107	0.4545	0.5214	0.4437	0.8104	0.95	0.7654
108	0.3232	0.3427	0.4046	0.7212	0.9657	0.7204
109	0.2168	0.2478	0.1945	0.6611	0.9165	0.6907
110	0.1852	0.1882	0.1817	0.6266	0.9296	0.6689
111	0.1829	0.1912	0.1947	0.6069	0.9281	0.6443
112	0.1316	0.1785	0.0957	0.6858	0.871	0.6749
113	0.0796	0.0759	0.0786	0.6324	0.9539	0.6083
114	0.5288	0.6087	0.6156	0.9139	0.9311	0.8603
115	0.264	0.2741	0.2966	0.7347	0.9901	0.7302
116	0.2833	0.2949	0.316	0.7504	0.9904	0.7439
117	0.0218	0.0219	0.0615	0.4527	0.9649	0.4569
118	0.6339	0.6569	0.5729	0.8286	0.986	0.8158
119	0.0793	0.0814	0.0569	0.6599	0.9912	0.6518
120	0.2821	0.2967	0.2632	0.7351	0.9852	0.7235
121	-0.0004	0.0007	0.1036	0.2642	0.9734	0.2445
122	0.6449	0.6732	0.5692	0.8349	0.9889	0.8213
123	0.0242	0.0263	0.0588	0.5574	0.9742	0.542
124	0.0364	0.0356	0.2172	0.3279	0.9671	0.3277
125	0.3036	0.3109	0.2654	0.771	0.9848	0.757
126	0.6334	0.6511	0.5024	0.8488	0.9916	0.8431
127	0.5824	0.6004	0.5246	0.8198	0.9857	0.808
128	0.6372	0.6647	0.5359	0.8291	0.9895	0.819
129	0.0163	0.0209	0.0454	0.5307	0.958	0.4847
130	0.6828	0.7039	0.5455	0.8449	0.9907	0.8359
131	0.0852	0.0925	0.1441	0.5238	0.9776	0.5111
132	0.2675	0.3751	0.4279	0.8325	0.9008	0.7214
133	0.3626	0.3897	0.5481	0.7041	0.9841	0.6871
134	0.23	0.247	0.4103	0.6295	0.9864	0.6137
135	0.182	0.191	0.4429	0.5765	0.9797	0.5624
136	0.1009	0.1119	0.1045	0.7097	0.9448	0.7012
137	0.1908	0.2008	0.3419	0.5938	0.9852	0.5813
138	0.6577	0.7324	0.6881	0.9026	0.94	0.8659
139	0.71	0.8267	0.7605	0.9401	0.9255	0.8699
140	0.7093	0.8429	0.8057	0.9315	0.9231	0.8546
141	0.5761	0.7534	0.7613	0.9163	0.906	0.8063
142	0.6671	0.852	0.8994	0.9305	0.8844	0.8361
143	0.6792	0.8552	0.9104	0.9301	0.8775	0.8409
144	0.7032	0.8496	0.9087	0.9325	0.8691	0.8632
145	0.6404	0.8514	0.8978	0.9426	0.8644	0.8439
146	0.6837	0.8601	0.8902	0.941	0.8799	0.8527
<i>mean</i>	0.3875	0.4365	0.5179	0.7644	0.9295	0.7288
<i>sd</i>	0.2207	0.2557	0.2587	0.1682	0.0468	0.1513

Note: The table reports the adjusted R2 in a regression of the variable on the common component and the correlation between each pair of estimated common components
See the Data Appendix for variable definitions.

Figure 1: The role of factors in monetary VARs



Note: Impulse response function to an interest rate shock in the base case (no factors in VAR), with static principal components (PCA), dynamic principal components (DPCA), and state space factors (SSS).

DATA APPENDIX

This appendix lists the variables used in the empirical analysis, with a short description and the transformation applied.

The transformation codes are: 1 = no transformation; 2 = first difference; 3= second difference; 4 = logarithm; 5 = first difference of logarithm; 6 = second difference of logarithm.

<i>Variable</i>	<i>Transf</i>
1 INDUSTRIAL PRODUCTION: TOTAL INDEX(1992=100,SA)	5
2 INDUSTRIAL PRODUCTION: PRODUCTS,TOTAL(1992=100,SA)	5
3 INDUSTRIAL PRODUCTION: FINAL PRODUCTS(1992=100,SA)	5
4 INDUSTRIAL PRODUCTION: CONSUMER GOODS(1992=100,SA)	5
5 INDUSTRIAL PRODUCTION: DURABLE CONSUMER GOODS(1992=100,SA)	5
6 INDUSTRIAL PRODUCTION: NONDURABLE CONSUMER GOODS(1992=100,SA)	5
7 INDUSTRIAL PRODUCTION: BUSINESS EQUIPMENT(1992=100,SA)	5
8 INDUSTRIAL PRODUCTION: INTERMEDIATE PRODUCTS(1992=100,SA)	5
9 INDUSTRIAL PRODUCTION: MATERIALS(1992=100,SA)	5
10 INDUSTRIAL PRODUCTION: NONDURABLE GOODS MATERIALS(1992=100,SA)	5
11 INDUSTRIAL PRODUCTION: MANUFACTURING(1992=100,SA)	5
12 INDUSTRIAL PRODUCTION: DURABLE MANUFACTURING(1992=100,SA)	5
13 INDUSTRIAL PRODUCTION: NONDURABLE MANUFACTURING(1992=100,SA)	5
14 INDUSTRIAL PRODUCTION: MINING(1992=100,SA)	5
15 INDUSTRIAL PRODUCTION: UTILITIES(1992=100,SA)	5
16 CAPACITY UTIL RATE: MANUFACTURING,TOTAL(%OF CAPACITY,SA)(FRB)	1
17 PURCHASING MANAGERS' INDEX (SA)	1
18 NAPM PRODUCTION INDEX (PERCENT)	1
19 PERSONAL INCOME (CHAINED) (BIL 92\$, SAAR)	5
20 INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS "(1967=100;SA)"	5
21 EMPLOYMENT: "RATIO;" HELP-WANTED ADS:NO.UNEMPLOYED CLF	4
22 CIVILIAN LABOR FORCE:EMPLOYED,TOTAL (THOUS.,SA)	5
23 CIVILIAN LABOR FORCE:EMPLOYED,NONAGRIC.INDUSTRIES(THOUS.,SA)	5
24 UNEMPLOYMENT RATE:ALL WORKERS,16 YEARS & OVER(% ,SA)	1
25 UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS(SA)	1
26 UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5WKS(THOUS.,SA)	1
27 UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS(THOUS.,SA)	1
28 UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS +(THOUS.,SA)	1
29 UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS(THOUS.,SA)	1
30 EMPLOYEES ON NONAG.PAYROLLS:TOTAL(THOUS.,SA)	5
31 EMPLOYEES ON NONAG.PAYROLLS:TOTAL,PRIVATE (THOUS,SA)	5
32 EMPLOYEES ON NONAG.PAYROLLS:GOODS-PRODUCING(THOUS.,SA)	5
33 EMPLOYEES ON NONAG.PAYROLLS:CONTRACT CONSTRUCTION(THOUS.,SA)	5
34 EMPLOYEES ON NONAG.PAYROLLS:MANUFACTURING(THOUS.,SA)	5
35 EMPLOYEES ON NONAG.PAYROLLS:DURABLE GOODS(THOUS.,SA)	5
36 EMPLOYEES ON NONAG.PAYROLLS:NONDURABLE GOODS(THOUS.,SA)	5
37 EMPLOYEES ON NONAG.PAYROLLS:SERVICE-PRODUCING(THOUS.,SA)	5
38 EMPLOYEES ON NONAG.PAYROLLS:WHOLESALE & RETAIL TRADE (THOUS.,SA)	5
39 EMPLOYEES ON NONAG.PAYROLLS:FINANCE,INSUR.&REAL ESTATE (THOUS.,SA)	5
40 EMPLOYEES ON NONAG.PAYROLLS:SERVICES(THOUS.,SA)	5
41 EMPLOYEES ON NONAG.PAYROLLS:GOVERNMENT(THOUS.,SA)	5
42 AVG. WEEKLY HRS. OF PRODUCTION WKRS.: MANUFACTURING (SA)	1
43 AVG. WEEKLY HRS. OF PROD. WKRS.:MFG., OVERTIME HRS. (SA)	1
44 NAPM employment index (percent)	1
45 MANUFACTURING & TRADE: TOTAL(MIL OF CHAINED 1992 DOLLARS)(SA)	5
46 MANUFACTURING & "TRADE: MANUFACTURING;TOTAL	5

47 MANUFACTURING & "TRADE: MFG;" DURABLE GOODS	5
48 MANUFACT.& "TRADE:MFG;NONDURABLE" GOODS	5
49 MERCHANT WHOLESALERS: TOTAL (MIL OF CHAINED 1992 DOLLARS)(SA)	5
50 MERCHANT WHOLESALERS:DURABLE GOODS TOTAL	5
51 MERCHANT WHOLESALERS:NONDURABLE GOODS	5
52 RETAILTRADE: TOTAL (MIL OF CHAINED 1992 DOLLARS)(SA)	5
53 RETAILTRADE: NONDURABLE GOODS (MIL OF 1992 DOLLARS)(SA)	5
54 PERSONAL CONSUMPTION EXPEND (CHAINED)-TOTAL(BIL 92\$,SAAR)	5
55 PERSONAL CONSUMPTION EXPEND (CHAINED)-TOTAL DURABLES(BIL 92\$,SAAR)	5
56 PERSONAL CONSUMPTION EXPEND (CHAINED)-NONDURABLES(BIL 92\$,SAAR)	5
57 PERSONAL CONSUMPTION EXPEND (CHAINED)-SERVICES(BIL 92\$,SAAR)	5
58 PERSONAL CONS EXPEND (CHAINED)-NEW CARS (BIL 92\$,SAAR)	5
59 HOUSING "STARTS: NONFARM(1947-58);TOTAL" FARM&NONFARM(1959-)(THOUS.,SA	4
60 HOUSING STARTS: NORTHEAST (THOUS.U.)S.A.	4
61 HOUSING STARTS: MIDWEST (THOUS.U.)S.A.	4
62 HOUSING STARTS: SOUTH (THOUS.U.)S.A.	4
63 HOUSING STARTS: WEST (THOUS.U.)S.A.	4
64 HOUSING AUTHORIZED:TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)	4
65 MOBILE HOMES: MANUFACTURERS' SHIPMENTS(THOUS.OF UNITS,SAAR)	4
66 MANUFACTURING & TRADE INVENTORIES:TOTAL(MIL OF CHAINED 1992)(SA)	5
67 INVENTORIES,BUSINESS,MFG(MIL OF CHAINED 1992 DOLLARS, SA)	5
68 INVENTORIES,BUSINESS DURABLES(MIL OF CHAINED 1992 DOLLARS, SA)	5
69 INVENTORIES,BUSINESS,NONDURABLES(MIL OF CHAINED 1992 DOLLARS, SA)	5
70 MANUFACTURING & TRADE INV:MERCHANT WHOLESALERS	5
71 MANUFACTURING & TRADE INV:RETAIL TRADE (MIL OF CHAINED 1992 DOLLARS)(SA)	5
72 RATIO FOR MFG & TRADE:INVENTORY/SALES (CHAINED 1992 DOLLARS, SA)	2
73 RATIO FOR MFG & "TRADE:MFG;INVENTORY/SALES"(87\$)(S.A.)	2
74 RATIO FOR MFG & "TRADE:WHOLESALE;INVENTORY/SALES(87\$)(S.A.)"	2
75 RATIO FOR MFG & TRADE:RETAIL"TRADE;INVENTORY/SALES(87\$)(S.A.)"	2
76 NAPM INVENTORIES INDEX (PERCENT)	1
77 NAPM NEW ORDERS INDEX (PERCENT)	1
78 NAPM VENDOR DELIVERIES INDEX (PERCENT)	1
79 NEW ORDERS (NET)-CONSUMER GOODS & MATERIALS, 1992 DOLLARS(BCI)	5
80 NEW ORDERS, DURABLE GOODS INDUSTRIES, 1992 DOLLARS(BCI)	5
81 NEW ORDERS, NONDEFENSE CAPITAL GOODS,IN 1992 DOLLARS(BCI)	5
82 MFG NEW ORDERS:ALL MANUFACTURING INDUSTRIES,TOTAL(MIL\$,SA)	5
83 MFG NEW ORDERS:MFG INDUSTRIES WITH UNFILLED ORDERS(MIL\$,SA)	5
84 MFG NEW ORDERS:DURABLE GOODS INDUSTRIES, TOTAL(MIL\$,SA)	5
85 MFG NEW ORDERS:DURABLE GOODS INDUST WITH UNFILLED ORDERS(MIL\$,SA)	5
86 MFG NEW ORDERS:NONDURABLE GOODS INDUSTRIES, TOTAL (MIL\$,SA)	5
87 MFG NEW ORDERS:NONDURABLE GDS IND.WITH UNFILLED ORDERS(MIL\$,SA)	5
88 MFG UNFILLED ORDERS: ALL MANUFACTURING INDUSTRIES,TOTAL(MIL\$,SA)	5
89 MFG UNFILLED ORDERS: DURABLE GOODS INDUSTRIES,TOTAL(MIL\$,SA)	5
90 MFG UNFILLED ORDERS: NONDURABLE GOODS INDUSTRIES, TOTAL(MIL\$,SA)	5
91 CONTRACTS & ORDERS FOR PLANT & EQUIPMENT (BIL\$,SA)	5
92 CONTRACTS & ORDERS FOR PLANT & EQUIPMENT IN 1992 DOLLARS(BCI)	5
93 NYSE COMMON STOCK PRICE INDEX: COMPOSITE (12/31/65=50)	5
94 S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)	5
95 S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS(1941-43=10)	5
96 S&P'S COMMON STOCK PRICE INDEX: CAPITAL GOODS (1941-43=10)	5
97 S&P'S COMMON STOCK PRICE INDEX: UTILITIES (1941-43=10)	5
98 S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD(% PER ANNUM)	1
99 S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO(% ,NSA)	1
100 UNITED "STATES;EFFECTIVE" EXCHANGE RATE(MERM)(INDEX NO.)	5
101 FOREIGN EXCHANGE RATE: GERMANY(DEUTSCHE MARK PER U.S.\$)	5

102 FOREIGN EXCHANGE RATE: SWITZERLAND(SWISS FRANC PER U.S.\$)	5
103 FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.\$)	5
104 FOREIGN EXCHANGE RATE: CANADA(CANADIAN \$ PER U.S.\$)	5
105 INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)	2
106 INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)	2
107 BOND YIELD: MOODY'S AAA CORPORATE(%PER ANNUM)	2
108 BOND YIELD: MOODY'S BAA CORPORATE(%PER ANNUM)	2
109 SECONDARY MARKET YIELDS ON FHA MORTGAGES(%PER ANNUM)	2
110 Spread FYCP -FYFF	1
111 Spread FYGM3-FYFF	1
112 Spread FYGM6-FYFF	1
113 Spread FYGT1-FYFF	1
114 Spread FYGT5-FYFF	1
115 Spread FYGT10-FYFF	1
116 Spread FYAAAC-FYFF	1
117 Spread FYBAAC - FYFF	1
118 Spread FYFHA-FYFF	1
119 MONEY STOCK:M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL\$,SA)	6
120 MONEY STOCK:M2(M1+O'NITE RPS,EURO\$,G/P&B/D MMMFS&SAV&SM TIME DEP)	6
121 MONEY STOCK:M3(M2+LG TIME DEP,TERM RP'S&INST ONLY MMMFS)(BIL\$,SA)	6
122 MONEY SUPPLY-M2 IN 1992 DOLLARS (BCI)	5
123 MONETARY BASE,ADJ FOR RESERVE REQUIREMENT CHANGES(MIL\$,SA)	6
124 DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL\$,SA)	6
125 DEPOSITORY INST RESERVES:NONBORROW+EXT CR,ADJ RES REQ CGS(MIL\$,SA)	6
126 NAPM COMMODITY PRICES INDEX (PERCENT)	1
127 PRODUCER PRICE INDEX: FINISHED GOODS(82=100,SA)	6
128 PRODUCER PRICE INDEX: FINISHED CONSUMER GOODS(82=100,SA)	6
129 INDEX OF SENSITIVE MATERIALS PRICES (1990=100)(BCI-99A)	6
130 CPI-U: ALL ITEMS(82-84=100,SA)	6
131 CPI-U: APPAREL & UPKEEP(82-84=100,SA)	6
132 CPI-U: TRANSPORTATION(82-84=100,SA)	6
133 CPI-U: MEDICAL CARE(82-84=100,SA)	6
134 CPI-U: COMMODITIES(82-84=100,SA)	6
135 CPI-U: DURABLES(82-84=100,SA)	6
136 CPI-U: SERVICES(82-84=100,SA)	6
137 CPI-U: ALL ITEMS LESS FOOD (82-84=100,SA)	6
138 CPI-U: ALL ITEMS LESS SHELTER (82-84=100,SA)	6
139 CPI-U: ALL ITEMS LESS MIDICAL CARE (82-84=100,SA)	6
140 PCE,IMPL PR DEFL:PCE (1987=100)	6
141 PCE,IMPL PR "DEFL:PCE;" DURABLES (1987=100)	6
142 PCE,IMPL PR "DEFL:PCE;" NONDURABLES (1987=100)	6
143 PCE,IMPL PR "DEFL:PCE;" SERVICES (1987=100)	6
144 AVG HR EARNINGS OF CONSTR WKRS: CONSTRUCTION (\$,SA)	6
145 AVG HR EARNINGS OF PROD WKRS: MANUFACTURING (\$,SA)	6
146 U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)	1

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2003 George Kapetanios
and Massimiliano Marcellino. All rights reserved.**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm**