Department of Economics

Factor Analysis Using Subspace Factor Models: Some Theoretical Results and an Application to UK Inflation Forecasting

George Kapetanios

Working Paper No. 466 November 2002 ISSN 1473-0278

Queen Mary
University of London

Factor analysis using subspace factor models: Some theoretical results and an application to UK inflation forecasting

George Kapetanios*
Queen Mary University of London

November 2002

Abstract

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. This paper provides an alternative method for estimating factors derived from a factor state space model. This model has a clear dynamic interpretation. Further, the method does not require iterative estimation techniques and due to a modification introduced, can accommodate cases where the number of variables exceeds the number of observations. The computational cost and robustness of the method is comparable to that of principal component analysis because matrix algebraic methods are used.

JEL Codes:C13, C32
Keywords: Factor Models, Subspace Methods, State Space Models

*Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS. email: G.Kapetanios@qmul.ac.uk

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 (1998), Forni and Reichlin (2000), Forni, Hallin, Lippi, and Reichlin (2000) and Forni, Hallin, Lippi, and Reichlin (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, Kapetanios, Smith, and Weale (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 Stock and Watson (1998) where the number of time series is greater than the number of observations, maximum likelihood estimation is not practically feasible. For this reason, Stock and Watson (1998) have suggested an approximate dynamic factor model based on principal component analysis. This model can accommodate a very large number of time series and there is no need for the number of obsevations to exceed the number of variables. Nevertheless, the principal component model is not, strictly speaking, a dynamic model. Stock and Watson (1998) have shown that it can estimate consistently the factor space asymptotically (but the number of time series has to tend to infinity). In small samples and for a finite number of series, the dynamic element of the principal component analysis is not easy to interpret. Forni and Reichlin (2000) suggested an alternative procedure based on dynamic principal components (see Brillinger (1981, ch. 9)). This method incorporates an explicitly dynamic element in the construction of the factors.

This paper provides an alternative method for estimating factors derived from a factor state space model. This model has a clear dynamic interpretation. Further, the method does not require iterative estimation techniques and due to a modification introduced, can accommodate cases where the number of variables exceeds the number of observations. The computational cost and robustness of the method is comparable to that of principal component analysis because matrix algebraic methods are used. The method forms parts of a large set of algorithms used in the engineering literature for estimating state space models caled subspace algorithms. Another advantage of the method is that the asymptotic distribution and therefore the standard errors of the factor estimates are available. Further, as the factor analysis is carried out within a general model, forecasting is easier to carry out than in the currently available procedures where a forecasting model needs to be specified.

The structure of the paper is as follows: Section 2 describes the elements of the suggested factor extraction method. Section 3 discusses the asymptotic properties of the new method. Section 4 discusses possible improved estimation of the factor estimates. Section 5 presents an application of the method to the forecasting of UK inflation in the recent past. The results are compared to the forecasts produced by the principal component factor extraction analysis of Stock and Watson. Section 6 concludes.

2 Theoretical considerations

2.1 The method

We consider the following state space model.

$$x_t = Cf_t + Du_t, \quad t = 1, \ldots, T$$

$$f_t = Af_{t-1} + Bu_{t-1} \tag{1}$$

x_t is an n-dimensional vector of strictly stationary zero-mean variables observed at time t. f_t is an m-dimensional vector of unobserved states (factors) at time t and u_t is a multivariate standard white noise sequence of dimension n. The aim of the analysis is to obtain estimates of the states f_t, for $t = 1, \ldots, T$.

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 (1). An extensive discussion may be found in Hannan and Deistler (1988). As we have mentioned in the introduction, maximum likelihood techniques either using the Kalman filter or otherwise may be used to estimate the parameters of the model under some identification scheme. For large datasets this is likely to be computationally intensive. Subspace algorithms 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 repects 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 and the assumed nonsingularity of D.

$$X_t^f = \mathcal{O}\mathcal{K}X_t^p + \mathcal{E}E_t^f \tag{2}$$

where $X_t^f = (x_t', x_{t+1}', x_{t+2}', \ldots)'$, $X_t^p = (x_{t-1}', x_{t-2}', \ldots)'$, $E_t^f = (u_t', u_{t+1}', \ldots)'$, $\mathcal{O} = [C', A'C', (A^2)'C', \ldots]'$, $\mathcal{K} = [\bar{B}, (A - \bar{B}C)\bar{B}, (A - \bar{B}C)^2\bar{B}, \ldots]$, $\bar{B} = BD^{-1}$ and

$$\mathcal{E} = \begin{pmatrix} D & 0 & \ldots & 0 \\ CB & D & \ddots & \vdots \\ CAB & \ddots & \ddots & 0 \\ \vdots & & CB & D \end{pmatrix} \tag{3}$$

The derivation of this representation is easy to see 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 $X_{s,t}^f = (x_t', x_{t+1}', x_{t+2}', \ldots, x_{t+s-1}')'$ and $X_{q,t}^p = (x_{t-1}', x_{t-2}', \ldots, x_{t-q}')'$. Then an estimate of $\mathcal{F} = \mathcal{O}\mathcal{K}$ may be obtained by regressing $X_{s,t}^f$ on $X_{q,t}^p$. Following that, the most popular subspace algorithms use a singular value decomposition 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 a singular value decomposition 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 $X_{q,t}^p$ 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

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

where $\hat{U}\hat{S}\hat{V}$ represents the singular value decomposition of $\hat{\Gamma}^f\hat{\mathcal{F}}\hat{\Gamma}^p$, $\hat{S}_m$ denotes the matrix containing the first m columns of $\hat{S}$ and $\hat{V}_m$ denotes the heading $m \times m$ submatrix of $\hat{V}$. $\hat{S}$ contains the singular values of $\hat{\Gamma}^f\hat{\mathcal{F}}\hat{\Gamma}^p$ in decreasing order. Then, the factor estimates are given by $\hat{\mathcal{K}}X_t^p$. For what follows it is important to note that the choice of the weighting matrices are important but not crucial for the asymptotic properties of the estimation method. They are only required to be nonsingular. A second thing to note is that consistent estimation of the factor space requires that q tends to infinity at a certain rate as T tends to infinity as pointed out by Bauer (1998, pp. 54). Once estimates of the factors have been obtained and if estimates of the parameters (including the factor loadings) are subsequently required, it is easy to see that 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, ch.4). We note that the identification scheme used above is implicit and depends on the normalisation used in the computation of the singular value decomposition. Finally, we must note that the method is also applicable in the case of unbalanced panels. In analogy to the work of Stock and Watson (1998) 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.

2.2 Dealing with large datasets

Up to now we have outlined an existing method for estimating factors which requires that the number of observations be larger than the number of elements in X_t^p. Given the work of Stock and Watson (1998) this is rather restrictive. We therefore suggest a modification of the existing methodology to allow the number of series in X_t^p be larger than the number of observations. The problem arises in this method because the least squares estimate of $\mathcal{F}$ does not exists due to rank deficiency of $X^{p\prime}X^p$ where $X^p = (X_1^p, \ldots, X_T^p)'$. As we mentioned in the previous section we do not neccesarily 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 singular value decomposition 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 $X^p\mathcal{F}'$ is given by

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

where $X^f = (X_1^f, \ldots, X_T^f)'$ and A^+ denotes the unique Moore-Penrose inverse of matrix A. Once this step is modified then the estimate of the factors may be straightforwardly obtained by applying a singular value decomposition to $\widehat{X^p\mathcal{F}'}$. We choose to set both weighting matrices to the identity matrix in this case.

2.3 Number of factors

A very important question relates to the determination of the number of factors, i.e. the dimension of the state vector. This issue has only recently received attention in the econometric literature. Stock and Watson (1998) suggest using information criteria for determining this dimension. Bai and Ng (2002) provide modified information criteria and justification for their use in the case where the number of variables goes to infinity as well as the number of observations. We suggest a simple information theoretic method for determining the number of factors in our model. Its simplicity comes from the fact that both the number of series and factors are assumed to be finite.

The search simply involves (i) fixing a maximum number of factors f^{max} to search over, (ii) estimating the factors for each assumed number of factors $m = 1, \ldots, m^{max}$ and (iii) minimising the negative penalised loglikelhood of the regression

$$x_t = C\hat{f}_t + u_t,$$

i.e. minimising $ln|\hat{\Sigma}_u^m| + c_T(m)$ where $\hat{\Sigma}_u^m$ is the estimated covariance matrix of u_t and $c_T(m)$ is a penalty term depending on the choice of the information criterion used. We prove consistency of the method under some conditions on $c_T(m)$ in the next section.

We briefly discuss an alternative class of testing procedures for determining the number of factors prevalent in the state space model literature. The testing procedures are based on the well known fact that the rank of certain block matrices referred to as Hankel matrices is equal to the dimension of the state vector. The most familiar Hankel matrix is the covariance Hankel matrix. The autocovariance Hankel matrix is a block matrix made up of the autocovariances of the observed process x_t. It is given by

$$\begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \ldots \\ \Gamma_2 & \Gamma_3 & \ldots \\ \Gamma_3 & \ldots & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

where Γ_i denotes the i-th autocvariance of x_t. Its finite truncation may be estimated by $1/TX^{f'}X^p$. Tests of rank may be used to estimate the rank of the covariance Hankel matrix from its estimate. A thorough investigation of the properties of the information criteria and the testing procedures in determining the rank of the Hankel matrix may be found in Camba-Mendez and Kapetanios (2001b). Further issues are discussed in Camba-Mendez and Kapetanios (2001a). A related discussion of the tests of rank used may also be found in Camba-Mendez, Kapetanios, Smith, and Weale (2000).

3 Asymptotic Properties

In this section we discuss the asymptotic properties of the factor estimates including estimation of standard errors and consistent estimation of the number of factors. 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 $p_1 = O\left(ln(T)^{\alpha}\right)$, $\alpha > 1$ and $p_2 = o(T^{1/3})$*

We denote the true number of factors by m^0. Then, we investigate OLS estimation of the multivariate regression model given below

$$X^f_{s,t} = \mathcal{F}X^p_{q,t} + \mathcal{E}E^f_t \tag{4}$$

for fixed $s \geq m^0$. Estimation of the above is equivalent to estimation of each equation separately. Then, by Theorem 4 of Berk (1974), who provides a variety of results for parameter estimates in infinite autoregressions, we have that $\sqrt{T-p}(\hat{\mathcal{F}} - \mathcal{F})$ has an asymptotic normal distribution with the standard OLS covariance matrix. Define $T^* = T - p$ as the effective number of observations. Below we derive the asymptotic distribution of the factor estimates. We distinguish between the case where the number of effective observations is larger than the number of series multiplied by p and the case where the number of effective observations is smaller than the number of series multiplied by p. Note, however, that, unlike other work in the literature, we assume that the number of factors and series is finite. We view this as an advantage of our method. Other methodologies do not choose to have the number of series tend to infinity. This assumption is needed for consistency of the factor estimates. Therefore, the case where the number of series multiplied by p is larger than T is included for completeness and to provide estimates of the factor estimate standard errors in small samples. In the first case, the factor estimates, denoted by $\hat{f}_t$, are given by $\hat{\mathcal{K}}X^p_t$. Note that the variance calculations will be carried out conditional on X^p_t. This is implicitly reflecting the standard treatment of obtaining variances of regression coefficients conditional on the regressors. In particular we wish to derive the asymptotic distribution of $\sqrt{T^*}(vec(\hat{f}) - vec(f))$, where $f = (f_1, \ldots, f_T)'$. In what folows we concentrate on the asymptotic variance. Asymptotic normality of the estimates follows by the asymptotic normality of $\sqrt{T-p}(\hat{\mathcal{F}} - \mathcal{F})$. We have that $f = X^p\hat{\mathcal{K}}'$. Simple manipulations indicate that

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

We need to derive the asymptotic variance of $V\left(\sqrt{T^*}\left(vec(\hat{\mathcal{K}}') - vec(\mathcal{K}')\right)\right)$. In general $\hat{\mathcal{K}}'$ is a function of the singular value decomposition of $A_1\hat{\mathcal{F}}A_2$, where A_1 and A_2 are weighting matrices discussed before. Note the importance of $sn \geq m$ for the calculation of the singular value decomposition. Define a function $g(.)$ such that $vec(\hat{\mathcal{K}}') = g\left(vec(A_1\hat{\mathcal{F}}A_2)\right)$. By a first order Taylor expansion[1] of $g(vec(A_1\hat{\mathcal{F}}A_2))$ and $g(vec(A_1\mathcal{F}A_2))$ around $A_1\mathcal{F}^*A_2$, 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. Thus,

$$V\left(\sqrt{T^*}(vec(\hat{f}) - vec(f))\right) = (I_m \otimes X^p)\frac{\partial g}{\partial(A_1\mathcal{F}A_2)}(A_2' \otimes A_1)(\Gamma^p \otimes \Sigma)$$

$$(A_2 \otimes A_1')\frac{\partial g'}{\partial(A_1\mathcal{F}A_2)}(I_m \otimes X^{p'})$$

[1]Possible since $g(.) \in C^\infty$.

We now move on to the case of the number of regressors exceeding the number of observations. From the above we know that $\hat{f}$ is estimated from a SVD of $\widehat{X^p\mathcal{F}'}$. We define a function $h(.)$ as $vec(\hat{f}) = h(\widehat{X^p\mathcal{F}'})$. Then, by a first order Taylor expansion as above, we have that

$$V\left(\sqrt{T}\left(vec(\hat{f}) - vec(f)\right)\right) = \frac{\partial h}{\partial(A_1\mathcal{F}A_2)} V\left(\sqrt{T}\left(vec(\widehat{X^p\mathcal{F}'}) - vec(X^p\mathcal{F}')\right)\right)\frac{\partial h'}{\partial(A_1\mathcal{F}A_2)}$$

Then, by simple manipulations of the variance derivation of Magnus and Neudecker (1988, pp. 262) we get that

$$V\left(\sqrt{T}\left(vec(\hat{f}) - vec(f)\right)\right) = \frac{\partial h}{\partial(A_1\mathcal{F}A_2)}(\Sigma \otimes X^p(X^{p'}X^p)^+X^{p'})\frac{\partial h'}{\partial(A_1\mathcal{F}A_2)}$$

Note that by virtue of the fact the p tends to infinity the asymptotic variances for the $\sqrt{T^*}$ normalised factor estimates tend to infinity at the same rate as T thereby implying that the factor estimates are effectively $O_P(pT^{*-1/2})$-consistent. By assumption 2 the rate of convergence of the factor estimates lies between $T^{*1/2}\left(ln(T^*)\right)^{\alpha}$ and $T^{*1/6}$. The standard error derivations given above are valid only for the case $s = 1$ as there is serial correlation in the error terms in (2) for $s > 1$. This case is discussed in more detail in Section 4 below.

We now discuss the asymptotic properties of the determination of the number of factors using information criteria. We assume that

Assumption 3 $\lim_{T^*\to\infty} T^* c_T(m) = \infty$ *and* $c_T(m) = o(1)$.

We want to show that $argmin_{m\in\{1,\ldots,m^{max}\}} IC(m) = m^0, \forall m^{max}$, where $IC(m) = ln|\hat{\Sigma}_u^m| + c_T(m)$. We wish to prove (i) $\text{plim}_{T^*\to\infty} ln|\hat{\Sigma}_u^m|/ln|\hat{\Sigma}_u^{m^0}| > 1$ for $m < m^0$ and (ii) $\text{plim}_{T^*\to\infty} ln|\hat{\Sigma}_u^m|/ln|\hat{\Sigma}_u^{m^0}| = 1$ for $m > m^0$ 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 $\sqrt{T^*}$ and the fact that any $(m^0 + i)$-th largest singular value of $\mathcal{F}$ is equal to zero, any observations of the $(m^0 + i)$-th factor series will tend to zero at rate $\sqrt{T^*}$. As a result asymptotically any regression that involves more than m^0 factors will have a singular regressor matrix. We assume that any such regression will be rejected and therefore the probability of picking $m > m^0$ tends to zero asymptotically. We now wish to prove (i). We note that even for $m < m^0$ the first m 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 m factors in a model which assumes $m_1 > m$ factors are identical to those in a model with $m_2 > m$ factors. We examine the probability of the event $ln|\hat{\Sigma}_u^m|/ln|\hat{\Sigma}_u^{m^0}| > 1$ We have that

$$Pr(ln|\hat{\Sigma}_u^m|/ln|\hat{\Sigma}_u^{m^0}| < 1+\epsilon) = Pr\left(\frac{ln|1/TX'\hat{M}^m X|}{ln|1/TX'\hat{M}^{m^0} X|} < 1+\epsilon\right) \quad (5)$$

where $X = (x_1, \ldots, x_T)'$, $\hat{M}^m = I - \hat{f}^m\left(\hat{f}^{m'}\hat{f}^m\right)^{-1}\hat{f}^{m'}$, $\hat{M}^{m^0} = I - \hat{f}^{m^0}(\hat{f}^{m^{0'}}\hat{f}^{m^0})^{-1}\hat{f}^{m^{0'}}$, $\hat{f}^m = (\hat{f}_1^m, \ldots, \hat{f}_T^m)$ and $\hat{f}_t^m = (\hat{f}_{1,t}, \ldots, \hat{f}_{m,t})'$. If we show that the probability in (5) is equal to

$$Pr\left(\frac{ln|1/TX'M^m X|}{ln|1/TX'M^{m^0} X|} < 1+\epsilon\right)$$

where M^m and M^{m^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

$$||1/TX'\hat{M}^m X - 1/TX'M^m X|| = o_p(1)$$

for all m. But this easily follows from the fact that $1/T||f - \hat{f}|| = o_p(1)$.

4 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 (2) given by (3). If the lead truncation index, s, is greater than 1, or D is not diagonal, then (2) should be estimated by generalised least squares as there is serial correlation in these error terms. Of course, consistency, the rate of convergence and asymptotic normality of the factor estimates are not affected by the presence of serial correlation. We will address only the case of large datasets (i.e. $Nq > T$) 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 $\mathcal{X}^p = I \otimes X^p$, $\mathcal{X}^f = vec(X^f)$, $E^f = (E_1^f, E_2^f, \ldots, E_T^f)'$ and $\mathcal{E}^f = vec(E^f)$. We first note that the covariance of $\mathcal{E}^f$ follows from the definition of $\mathcal{E}$ in (3). 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 $vec(X^p\mathcal{F})$ as derived by Magnus and Neudecker (1988, Ch.13,Th.13) and given by

$$vec(\widehat{X^p\mathcal{F}}) = \mathcal{X}^p(\mathcal{X}^{p\prime}\hat{\mathcal{V}}^+\mathcal{X}^p)^+\mathcal{X}^{p\prime}\hat{\mathcal{V}}^+\mathcal{X}^f \tag{6}$$

Then, we apply the singular value decomposition 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 $s > 1$.

5 An Application: Forecasting Inflation

In this section we provide an application of the proposed dynamic factor methodology to the forecasting of UK inflation defined for our purposes as RPIX inflation (RPI inflation minus mortgage interest payments). We construct a large dataset of variables. These are grouped into six different categories (number of variables in parentheses): Exchange rates (plus share prices) (4), interest rates (4), monetary variables (7), prices (14), real activity (34), surveys (17). 80 variables are used in total. The sample period is 1988Q1-1997Q4. The forecast evaluation period is 1998Q1-2000Q4. For some minimal robustness analysis we repeat the whole process for a sample period of 1988Q1-199Q2 and evaluation period 1997Q3-2001Q2. Crucially, the data for this exercise have been obtained following large revisions carried out as part of the major data revision exercise by the UK Office of National Statistics in September 2001. The data for the first forecasting exercise were obtained prior to the September 2001 revision exercise.

The variables are differenced as many times as needed to reduce them to stationarity according to the ADF unit root test. Any variable with an outlier, defined as being an observation whose distance from the median of the sample is as large as or larger than six times the interquartile range, is rejected. No such variable is found in the set. All variables are normalised to have mean zero and variance one. Using the subspace algorithm and setting $p = f = 1$ we extract factors that summarise information in each of these categories. The number of factors for each category is: exchange rates (2), interest rates (2), monetary variables (3), prices (4), real activity (4), surveys (3), all variables (6). There are 24 factor variables.

The factors extracted from each category together with the factors extracted from all variables viewed as a whole, are then used as the set of variables on which the following procedure is applied. We construct all possible combinations up to a maximum size of 4 variables from the set of 24 factor variables. Each combination together with inflation is then viewed as a VAR model[2]. Then, for each model we use the component factor variables to forecast inflation. We produce forecasts one step ahead. We have 12950 different models. We repeat the whole exercise using factors extracted using principal components as suggested by Stock and Watson (1998).

We compare the forecasting performance of the two sets of models. The relative performance is used to indicate the relative ability of the two factor approaches in forecasting UK inflation. We use two measures of forecasting

[2]Note the equivalence of a forecast produced by this model and a simple dynamic regression model for one-step ahead forecasts.

peformance. The first is the root mean square forecast error and the second is the Diebold and Mariano (1995) test of predictive ability[3]. The Diebold-Mariano test compares the forecasting performance of two models by testing the null hypothesis that they have equal predictive ability. The test is based on a series of differences between the losses arising out of each model for some loss function. For the current application we will use squared forecast errors as the preferred loss function. As we have a very large number of models we need to have a rule for pairing models so as to carry out the comparisons. We do this as follows: We rank models from each set according to their root mean square performance. Since we have the same number of models we then pair off the ranked models and carry out pairwise Diebold-Mariano tests. We use two different algorithms for the subspace factor method. In the first, the weighting matrix is made up from the covariances of the data as discussed in the previous section. This set is denoted by Γ. In the second case we use the identity matrix as the weighting matrix. This set is denoted by I.

5.1 Results

We concentrate our analysis on the top 5% of the models as selected by their root mean square forecasting performance. These are the top 625 models. Out of these, 42.8 % of the Γ and 100% of the I models outperform their pair model from the principal component factor model set. The respective percentage for the Γ set goes up to 99% for the top 100 models. In fact, all of the top 25% I models outperform their pair principal component models.

Going to the Diebold-Mariano tests we see that for top 5% of the models we have no rejection of the null hypothesis of equal predictive ability for the Γ models. On the other hand, for the I models, 1.2% of the top 650, 2.7%

[3]This test is known to have bad small sample size properties. We have addressed this issue by considering the corrected Diebold-Marino test suggested by Harvey, Leybourne, and Newbold (1997). There was no qualitative difference in the results.

of the top 300, 5% of top 100 and 8% of the top 50 models reject the null hypothesis of equal predictive ability at the 10% significance level in favour of the subspace models. Note that the power of the Diebold-Mariano test is likely to be low given the very small sample of 12 observations in the forecast evaluation period. It is clear that although both weighting matrices for the subspace method provide clear advantages over the principal component methodology, the better choice is the identity weighting matrix. We therefore need to point out that both matrices are asymptotically acceptable choices and that the optimality of the covariance weighting matrix relates simply to the asymptotic relative efficiency of the parameter estimates of the state space model. Such a property is of little guidance for finite sample forecasting exercises.

We repeat the exercise and use data that have undergone significant revisions, as part of the regular revision efforts caried out by the Office of National Statistics, in September 2001. The results suggest a superior forecasting ability for the dynamic factor method. The respective percentages for better Γ and I models in the top 625 are 44.7 % and 98.6 % and in the top 100 models they are 99% and 91%.

6 Conclusion

In this paper we have suggested and briefly evaluated a new factor based method for forecasting time series. Our work follows closely in spirit the work of Stock and Watson (1998), Stock and Watson (1999) and subsequent, as yet unpublished papers by these authors and their co-authors on the one hand and the work by Forni and Reichlin (2000), Forni, Hallin, Lippi, and Reichlin (2000) and Forni, Hallin, Lippi, and Reichlin (2001) on the other hand. The innovation lies in providing an alternative method for obtaining factor estimates.

One strand of the literature on factor extraction relies on explicitly dynamic state space models to estimate factors via computationally expensive and, in small samples, non-robust maximum likelihood estimation. The other strand of the literature based on the work of Stock and Watson (1998) uses principal components to extract the factors. This methodology is robust, computationally feasible with very large datasets and asymptotically valid for dynamic settings. Unfortunately, these methods are approximately dynamic in that the dynamic structure of the factors is not explicitly modelled in finite samples but captured only asymptotically where both the number of observations and the number of series used, grows to infinity. We propose a new methodology which while sharing all the advantages of the principal component extraction method is explicitly dynamic. This method is based on linear algebraic techniques for estimating the state and, if need be, the parameters of a general linear state space model.

We evaluate the new methodology on a very large number of forecasting models for UK inflation. We find that the methodology compares favorably to the principal component methodology of Stock and Watson (1998) for a great majority of the forecasting models considered.

References

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1).

BAUER, D. (1998): "Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms," Ph.d. Thesis.

BERK, K. N. (1974): "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2(3), 489–502.

Brillinger, D. (1981): *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.

Camba-Mendez, G., and G. Kapetanios (2001a): "Statistical Testing Procedures and Information Criteria in Rank Determination for System Identification," Mimeo.

——— (2001b): "Testing the Rank of the Hankel Covariance matrix: A Statistical Approach," *Institute of Electrical and Electronic Engineers Transactions on Automatic Control*, 46(2), 331–336.

Camba-Mendez, G., G. Kapetanios, R. J. Smith, and M. R. Weale (2000): "Tests of Rank in Reduced Rank Regression Models," Forthcoming in Journal of Business and Economic Statistics.

——— (2001): "An Automatic Leading Indicator of Economic Activity: Forecasting GDP growth for European Countries," *Econometrics Journal*, 4, 56–90.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalized Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82(4), 540–554.

——— (2001): "The Generalised Dynamic Factor Model: One Sided-Estimation and Forecasting," Mimeo, Université Libre de Bruxelles.

Forni, M., and L. Reichlin (2000): "Let's Get Real: A Factor Analytic Approach to Business Cycle Dynamics," *Review of Economic Studies*, 65(3), 453–473.

Hannan, E. J., and M. Deistler (1988): *The Statistical Theory of Linear Systems*. John Wiley.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1997): "Testing the Equality of Prediction Mean Square Errors," *International Journal of Forecasting*, 13, 273–281.

Larimore, W. E. (1983): "System Identification, Reduced Order Filters and Modelling via Canonical Variate Analysis," *Proc. 1983 Amer. Control Conference*, 2, 445–451.

Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications to Statistics and Econometrics*. John Wiley.

Sin, C. Y., and H. White (1996): "Information Criteria for Selecting Possibly Misspecified Parametric Models," *Journal of Econometrics*, 71, 207–225.

Stock, J. H., and M. Watson (1989): "New Indexes (Sic) of Coincident and Leading Indicators," *NBER Macroeconomics annual*, 4, 351–394.

Stock, J. H., and M. W. Watson (1998): "Diffusion Indices," Forthcoming in Journal of Business and Economic Statistics.

——— (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44(2), 293–335.

Van Overschee, P., and B. De Moor (1996): *Subspace Identification for Linear Systems*. Kluwer Academic Publishers.

Data Appendix

Exchange rates and share prices

"266305K " "United Kingdom" "SHARE PRICES FT-SE-A NON-FINANC Index publication base /Share prices Industrials Total Total United Kingdom /1995Y"

"267001K " "United Kingdom" "REAL EFFECTIVE EXCHANGE RATES Index publication base /Currency Conversions Real Effective Exchange..United Kingdom /1995Y"

”267003D ” ”United Kingdom” ”US DOLLAR EXCHANGE RATE MTH AVG Quantum (non-additive or stock figures) /Currency Conversions US$ e.United Kingdom /pd/$”

”267009D ” ”United Kingdom” ”US DOLLAR EXCHANGE RATE:FORWARD Quantum (non-additive or stock figures) /Currency Conversions US$ e.United Kingdom /$/pd”

subsection*Interest rates

”266213D ” ”United Kingdom” ”OVERNIGHT INTERBANK RATE Quantum (non-additive or stock figures) /Interest Rates Immediate rates (¡.United Kingdom /

”266215D ” ”United Kingdom” ”LONDON CLEARING BANKS’ RATE Quantum (non-additive or stock figures) /Interest Rates Immediate rates.United Kingdom /

”266225D ” ”United Kingdom” ”3-MONTH INTERBANK LOANS Quantum (non-additive or stock figures) /Interest Rates 3-mth or 90-day rat.United Kingdom /

”266261D ” ”United Kingdom” ”10-YEAR GOVT BONDS Quantum (non-additive or stock figures) /Interest Rates Long-term (1 yr or more).United Kingdom /

Monetary variables

(All variables are seasonally adjusted and in levels)

Notes and Coins

M0

M3

M4

Bank Lending

M4 Lending

Bank Retail Deposits

Prices

”265011K ” ”United Kingdom” ”PPI MFG OUTPUT FOOD Index publication base /Producer Prices Food products, beverages & tobacco, te.United Kingdom /1995Y”

”265018K ” ”United Kingdom” ”PPI MFG INPUT FUEL Index publication base Producer Prices Input to production Fuel Total United Kingdom /1995Y”

”265023K ” ”United Kingdom” ”PPI MFG INPUT RAW MATERIALS Index publication base /Producer Prices Input to production Raw materi.United Kingdom /1995Y”

"265025K " "United Kingdom" "PPI MFG INPUT TOTAL EXCL FOOD Index publication base /Producer Prices Input to production Input to.United Kingdom /1995Y"

"265045K " "United Kingdom" "PPI MFG OUTPUT ALL PRODUCTS Index publication base /Producer Prices Industry aggregates Manufactur.United Kingdom /1995Y"

"265046K " "United Kingdom" "PPI MFG OUTPUT TOTAL EXCL FOOD Index publication base /Producer Prices Industry aggregates Manufac.United Kingdom /1995Y"

"265061K " "United Kingdom" "PPI MFG OUTPUT CHEMICALS Index publication base /Producer Prices Other transportable gds, excl. me.United Kingdom /1995Y"

"265201K " "United Kingdom" "CPI BEVERAGE & TOBACCO Index publication base /Consumer Price Index Goods Alcoholic beverages & Tob.United Kingdom /1995Y"

"265209K " "United Kingdom" "CPI FUEL & ELECTRICITY Index publication base /Consumer Price Index Goods Utilities (Electricity, G.United Kingdom /1995Y"

"265213K " "United Kingdom" "CPI FOOD (OECD) Index publication base /Consumer Price Index Goods Food Food (incl. restaurants) United Kingdom /1995Y"

"265221K " "United Kingdom" "CPI HOUSING Index publication base /Consumer Price Index Services Housing - rental services Total i.United Kingdom /1995Y"

"265239K " "United Kingdom" "CPI ALL ITEMS LESS SEAS FOOD Index publication base /Consumer Price Index All items All items exclu.United Kingdom /1995Y"

"265241K " "United Kingdom" "CPI ALL ITEMS Index publication base Consumer Price Index All items Total Total United Kingdom /1995Y"

"265247K " "United Kingdom" "CPI EXCLUDING MORTAGE INTEREST Index publication base /Consumer Price Index All items All items exc.United Kingdom /1995Y"

Real activity

"261023NSA" "United Kingdom" "GDP CONSUMERS' EXPENDITURE SA National currency annual Level SA /National Accounts GDP by Expendit.United Kingdom /MN 1990 pd"

"261025NSA" "United Kingdom" "GDP GOVERNMENT EXPENDITURE SA National currency annual Level SA /National Accounts GDP by Expendit.United Kingdom /MN 1990 pd"

"261027NSA" "United Kingdom" "GDP CONSTRUCTION SA National currency annual Level SA /National Accounts GDP by Expenditure (const.United

Kingdom /MN 1990 pd”

”261029NSA” ”United Kingdom” ”GDP EXPORTS GOODS & SERVICES SA National currency annual Level SA /National Accounts GDP by Expendi.United Kingdom /MN 1990 pd”

”261031NSA” ”United Kingdom” ”GDP IMPORTS GOODS & SERVICES SA National currency annual Level SA /National Accounts GDP by Expendi.United Kingdom /MN 1990 pd”

”261035NSA” ”United Kingdom” ”GDP GROSS FIXED INVESTMENT SA National currency annual Level SA /National Accounts GDP by Expendit.United Kingdom /MN 1990 pd”

”261037NSA” ”United Kingdom” ”GDP 1995 PRICES sa National currency annual Level SA /National Accounts GDP by Expenditure (constan.United Kingdom /MN 1990 pd”

”262007K ” ”United Kingdom” ”IIP MANUFACTURING Index publication base /Production Production by economic activities (IIP or Real.United Kingdom /1995Y”

”262017KSA” ”United Kingdom” ”REAL GDP TRANSPORT + COMM Index publication base SA /Production Production by economic activities (.United Kingdom /1995Y”

”262019KSA” ”United Kingdom” ”REAL GDP TOTAL SERVICES Index publication base SA /Production Production by economic activities (II.United Kingdom /1995Y”

”262021KSA” ”United Kingdom” ”REAL GDP DISTRIBUTION Index publication base SA /Production Production by economic activities (IIP..United Kingdom /1995Y”

”262023KSA” ”United Kingdom” ”REAL GDP FINANCIAL + BUSINESS Index publication base SA /Production Production by economic activiti.United Kingdom /1995Y”

”262027KSA” ”United Kingdom” ”IIP TOTAL sa Index publication base SA /Production Production by economic activities (IIP or Real G.United Kingdom /1995Y”

”262111KSA” ”United Kingdom” ”IIP DURABLE CONSUMER GOODS SA Index publication base SA /Production Production by type of good (II.United Kingdom /1995Y”

”262117KSA” ”United Kingdom” ”IIP NON DURABLE CONS GOODS SA Index publication base SA /Production Production by type of good (II.United Kingdom /1995Y”

”262119KSA” ”United Kingdom” ”IIP INTERMEDIATE GOODS SA Index publication base SA /Production Production by type of good (IIP or.United Kingdom /1995Y”

”262121KSA” ”United Kingdom” ”IIP INVESTMENT GOODS SA Index publication base SA /Production Production by type of good (IIP or R.United Kingdom /1995Y”

"262201B " "United Kingdom" "PRODUCTION PASSENGER CARS Monthly Level Production Commodity Output Passenger cars Total United Kingdom /number"

"262205B " "United Kingdom" "PRODUCTION COMMERCIAL VEHICLES Monthly Level /Production Commodity Output Vehicles Commercial vehic.United Kingdom /number"

"263209KSA" "United Kingdom" "SALES ENGINEERING DOMEST VOL SA Index publication base SA /Manufacturing Sales Volume Domestic United Kingdom /1995Y"

"263211KSA" "United Kingdom" "SALES ENGINEERING EXPORT VOL SA Index publication base SA /Manufacturing Sales Volume Export United Kingdom /1995Y"

"263213KSA" "United Kingdom" "SALES ENGINEERING TOTAL VOL SA Index publication base SA /Manufacturing Sales Volume Total United Kingdom /1995Y"

"263341OSA" "United Kingdom" "STK CHANGE MFG FINISHED GDS SA National currency sum over component sub-periods SA /Manufacturing .United Kingdom /MN 1995 pd mln"

"263343OSA" "United Kingdom" "STK CHANGE MFG WORK IN PROG SA National currency sum over component sub-periods SA /Manufacturing .United Kingdom /MN 1995 pd mln"

"263427KSA" "United Kingdom" "NEW ORD ENGINEERING DOM VOL SA Index publication base SA /Manufacturing New orders Volume Domestic United Kingdom /1995Y"

"263429KSA" "United Kingdom" "NEW ORD ENGINEERING EXP VOL SA Index publication base SA /Manufacturing New orders Volume Export United Kingdom /1995Y"

"263431KSA" "United Kingdom" "NEW ORD ENGINEERING TOT VOL SA Index publication base SA /Manufacturing New orders Volume Total United Kingdom /1995Y"

"263547KSA" "United Kingdom" "CON NEW ORDERS TOTAL VOL(GB) SA Index publication base SA /Construction New orders Total constructi.United Kingdom /1995Y"

"263601KSA" "United Kingdom" "RETAIL SALES TOTAL (GB) SA Index publication base SA /Domestic Demand Retail trade Value Total United Kingdom /1995Y"

"263641B " "United Kingdom" "NEW PASSENGER CAR REG (GB) Monthly Level /Domestic Demand Registrations Passenger cars Total United Kingdom /number"

"264011D " "United Kingdom" "EMPLOYMT SERVICE INDUSTRIES GB Quantum (non-additive or stock figures) /Employment Employees (depe.United Kingdom /'000 persons"

"264103DSA" "United Kingdom" "REGISTERED UNEMPLOYMENT SA

Quantum (non-additive or stock figures) SA /Unemployment Level of Unempl.United Kingdom /'000 persons"

"264111DSA" "United Kingdom" "UNEMP

"264235DSA" "United Kingdom" "VACANCIES AT JOB CENTRES"

Surveys

Survey data are taken from the CBI (Confederation of British Industry) Quarterly Industrial Trends. The balances from the following questions have been included in the dataset: Q1, Q2, Q3, Q5, Q7, Q7a, Q7b, Q8, Q9a, Q9b. Additionally, the balance of the GfK consumer confidence survey is included.

Queen Mary
University of London

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

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