DETERMINING THE NUMBER OF FACTORS IN APPROXIMATE FACTOR MODELS

Jushan Bai   Serena Ng

Department of Economics
Boston College
Chestnut Hill
MA 02467

January 31, 2000

Abstract

In this paper we develop some statistical theory for factor models of large dimensions. The focus is the determination of the number of factors, which is an unresolved issue in the rapidly growing literature on multifactor models. We propose a panel C_p criterion and show that the number of factors can be consistently estimated using the criterion. The theory is developed under the framework of large cross-sections (N) and large time dimensions (T). No restriction is imposed on the relation between N and T. Simulations show that the proposed criterion yields almost precise estimates of the number of factors for configurations of the panel data encountered in practice.

JEL Classification: C13, C33, C43

1 Introduction

The idea that variations in a large number of economic variables can be modelled by a small number of reference variables is appealing and is used in many economic analysis. In the finance literature, the arbitrage pricing theory (APT) of Ross (1976) assumes that a small number of factors can be used to explain a large number of asset returns.[1] The observed comovements in a large number of macroeconomic time series have likewise generated enormous interests in developing ways to account for business cycle dynamics by a small number of reference variables. In demand analysis, Lewbel (1991) showed that linear engel curves can be expressed in terms of a finite number of factors. Factor analysis also provide a convenient way to model the aggregate implications of microeconomic behavior. For example, Forni and Lippi (1997) showed that explicit consideration of the cross section units can explain excess sensitivity in aggregate consumption. Central to both the theoretical and the empirical validity of factor models is the correct specification of the number of factors. To date, this crucial parameter is often assumed rather than determined by the data.[2] This paper develops a formal statistical procedure which can consistently estimate the number of factors from observed data. The theory is developed under the assumption that $N \to \infty$ and $T \to \infty$. We demonstrate that the penalty for overfitting must be a function of both N and T in order to consistently estimate the number of factors. Consequently the usual AIC and BIC which are functions of T alone do not work for factor models.

There are additional motivations to studying the dimension of factor models. Stock and Watson (1999) showed that the forecast mean squared error of many macroeconomic variables can be reduced by including diffusion indexes, or factors, in structural as well as non-structural forecasting models. Knowledge of the number of factors can also be used to test the validity of economic assumptions and models. For example, if a demand system has one common factor, budget shares should be independent of the level of income. Therefore if more than one factor is found in the data, homothetic preferences can be rejected. From an econometric perspective, cross-country and sectoral datasets are becoming increasingly available. Data on asset returns are also available over an increasingly long span. In many cases, the time dimension of such datasets, although small relative to the cross section

[1] Cochrane (1999) stressed that financial economists now recognize that there are multiple sources of risk, or factors, that give rise to high returns. Backus, Forsei, Mozumdar and Wu (1997) made similar conclusions in the context of the market for foreign assets.

[2] Lehmann and Modest (1988), for example, tested the APT for 5, 10 and 15 factors. Stock and Watson (1989) assumed there is one factor underlying the coincident index. Ghysels and Ng (1998) tested the affine term structure model assuming two factors.

dimension, is too large to justify the assumption of a fixed T.

A small number of papers in the literature have also considered the problem of determining the number of factors, but the present analysis differs from these works in important ways. Connor and Korajcyk (1993) developed a test for the number of factors in asset returns, but their test is valid for large N and fixed T. Furthermore, because their test is based on the comparison of variances over different time periods, covariance stationarity and homokedasticity are not only technical assumptions, but are crucial for the validity of their test. Also under the assumption that $N \to \infty$ for fixed T, Forni and Reichlin (1998) suggested a graphic approach to identify the number of factors, but no theory is available. Assuming $N, T \to \infty$ with $\sqrt{N}/T \to \infty$, Stock and Watson (1998) showed that a modification to the BIC can be used to select the number of factors optimal for forecasting a single series. Their criterion is restrictive not only because it requires $N >> T$, but also because there can be factors that are pervasive for a set of data and yet have no predictive ability for an individual data series. Thus, their rule is inadequate outside of forecasting framework. Forni, Hallin, Lippi and Reichlin (1999) suggested a multivariate variant of the AIC but the theoretical properties of the criterion are not known. Lewbel (1991) and Donald (1997) used the rank of a matrix to test for the number of factors, but these theories assume either N or T is fixed. Cragg and Donald (1997) also developed procedures to select the number of factors based on test statistics and information criterion with a fixed dimension. In addition, their tests is based on the rank of the data matrix after being projected onto a set of explanatory variables. As their simulations showed, standard information criteria such as the AIC and the BIC do not have good properties, especially when the model dimension is large. The theory we develop below estimate the number of factors in the observed data directly and performs well for many configurations of the data.

We set up the determination of factors as a model selection problem. In consequence, the proposed criteria depend on the usual trade-off between good fit and parsimony. However, the problem is non-standard not only because account needs to be taken of the sample size in both the cross section and the time series dimensions, but also because the factors are not observed. Section 2 sets up the preliminaries and introduces notation and assumptions. Estimation of the factors is considered in Section 3 and the estimation of the number of factors is studied in Section 4. A number of specific criteria are also proposed in Section 4. Simulations are used to illustrate the finite sample properties of the proposed criteria and results are reported in Section 5. Concluding remarks are provided in Section 6. All the proofs are given in the Appendix.

2 Preliminaries

Let X_{it} be the observed data for i^{th} cross section unit at time t, for $i = 1, \ldots N$, and $t = 1, \ldots T$. Granger (1987) defined a sequence of random variables to be a dominant common factor of X_{it} if the variance of the sum of the first N terms in the sequence increases at rate N^2. Consider the following model

$$X_{it} = \lambda_i' F_t + e_{it}. \tag{1}$$

Since $\sum_{i=1}^{N} X_{it}$ has variance dominated by $var(NF_t)$, F_t are common factors of X_{it} in the sense of Granger (1987). Then $\lambda_i' F_t$ is the common component of X_{it}, λ_i are the factor loadings associated with F_t, and e_{it} is the idiosyncratic component of X_{it}. Equation (1) is then the factor representation of the data. Note that the factors, their loadings, as well as the idiosyncratic errors are not observable.

Let F_t^0, λ_i^0 and r denote the true common factors, factor loadings, and true number of factors, respectively. Note that F_t^0 is r dimensional. At a given t, we have

$$\underset{(N \times 1)}{X_t} = \underset{(N \times r)}{\Lambda^0} \underset{(r \times 1)}{F_t^0} + \underset{(N \times 1)}{e_t}. \tag{2}$$

where $X_t = (X_{1t}, X_{2t}, ..., X_{Nt})'$, $\Lambda^0 = (\lambda_1^0, \lambda_2^0, ..., \lambda_N^0)'$, and $e_t = (e_{1t}, e_{2t}, ..., e_{Nt})'$. The objective is to determine the true number of factors, r. The classical factor analysis (e.g., **?**) assumes a fixed N, the independence of factors and the errors e_t, as well as diagonality of the covariance of e_t. Normalizing the covariance matrix of F_t to be an identity matrix, we have $\Sigma = \Lambda^0 \Lambda^{0\prime} + \Omega$, where Σ and Ω are the covariance matrices of X_t and e_t, respectively. The classical factor analysis starts with a root-T consistent and asymptotically normal estimator of Σ, say the sample covariance matrix $\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})(X_t - \bar{X})'$ and makes inference on r based on $\widehat{\Sigma}$. However, when $N \to \infty$, consistent estimation of Σ (a $N \times N$ matrix) is not a well defined problem. For example, when $N > T$, the rank of $\widehat{\Sigma}$ is no more than T, whereas the rank of Σ can always be N. Fortunately, when T is fixed, this $N \times N$ problem can be turned into a $T \times T$ problem, as noted by Connor and Korajcyk (1993) and others. The essentials of the classical factor analysis carry over to the case of large N but fixed T. In either case, the theory developed for a fixed dimension delivers poor performance for moderately large N and T, as documented by the Monte Carlo simulations in Cragg and Donald (1997) as well as by our simulations.

In this paper, we develop the theory of factor analysis for large dimensional factor models. Classical factor analysis does not apply to this situation. In addition, we allow for cross-sectional and serial dependence in e_t and cross-sectional and serial heteroskedasticity. Some weak dependence between the factors and the errors are also permitted. Simulation shows that the theory of large dimensional factor analysis works well even if one of the dimensions is small (as small as 10, not reported in the simulations).

Before proceeding further, we consider several concrete examples where model (1) is pertinent.

1. *Arbitrage pricing theory.* In this case, X_{it} represents the return of asset i at time t, F_t represents the vector of factor returns and e_{it} is the idiosyncratic returns. The number of factors can be determined when N and T are large. The factor returns F_t can also be consistently estimated (up to a invertible transformation).

2. *The rank of demand system.* Consumer demand theory postulates that the demand system may be described by $X_{it} = a_{i1}G_1(z_t) + \cdots a_{ir}G_r(z_t) + e_{it}$, where X_{it} is the consumption good i's buget share for the t^th consumer, $G_j(z)$ is a nonparametric function of observable variable z, which includes income and relative prices. The number r is called the rank of the demand system. Let $F_t = (G_1(z_t), ..., G_r(z_t))'$. The system is of the form of (1). When the number of goods (N) is large, the theory of this paper promises a consistent estimation of the rank of the demand system without the need of estimating the nonparametric functions $G_j(\cdot)$. Nevertheless, as a by product, the nonparametric functions evaluated at z_t, i.e., F_t is also consistently estimable. Furthermore, the nonparametric functions $G_j(\cdot)$ may be recovered (up to a matrix transformation) from $\widehat{F}_t(t = 1, .., T)$ via nonparametric estimation, especially when z_t is of small dimension. For example, z_t is the income variable when the data are taken at the same time period for all consumers because of the relative prices are the same for all consumers.

3. *Forecasting with diffusion indices.* Stock and Watson (1998) consider forecasting inflation with diffusion indices ("factors") constructed from large number of macroeconomic series. The underlying premise is that the movement of a large number of macroeconomic series may be driven by a small number of unobservable factors. The factors can be extracted from these series and then used as an input in the forecasting equation. To be specific, consider the scalar series

$$y_{t+1} = \alpha' F_t + \beta' W_t + \epsilon_t,$$

which is the forecasting equation. The variable W_t is observable. Let

$$X_t = \Lambda F_t + e_t \quad (t = 1, 2, ..., T) \tag{3}$$

where X_t is $N \times 1$ for some large N. Extract F_t from system (3) and denote it by $\widehat{F}_t$. Then regress y_t on $\widehat{F}_{t-1}$ and W_{t-1} to obtain the coefficients $\widehat{\alpha}$ and $\widehat{\beta}$. The forecast is formed by

$$\widehat{y}_{T+1|T} = \widehat{\alpha}'\widehat{F}_T + \widehat{\beta}W_T$$

This approach of forecasting outperforms many competing forecasting methods, see Stock and Watson (1998, 1999) and thus is quite promising. Note the vector series X_t may have a structural representation determined by F_t and X_t itself. We interpret (3) as the reduced form representation of X_t in terms of the unobservable factors. In this paper, we show that the factors F_t can be consistently estimated (up to a matrix transformation). This result can be used to show that the forecast $\widehat{y}_{T+1|T}$ is a consistent estimation of the conditional mean of y_{T+1} conditional on the information up to time T.

We now consider other representations of model (1) as well as assumptions imposed on the model. Let $\underline{X}_i$ be a $T \times 1$ vector of time series observations for the i^{th} cross section unit. For a given i, we have

$$\underset{(T\times 1)}{\underline{X}_i} = \underset{(T\times r)}{F^0} \; \underset{(r\times 1)}{\lambda_i^0} + \underset{(T\times 1)}{\underline{e}_i}, \tag{4}$$

where $\underline{X}_i = (X_{i1}, X_{i2}, ..., X_{iT})'$, $F^0 = (F_1^0, F_2^0, ..., F_T^0)'$ and $\underline{e}_i = (e_{i1}, e_{i2}, ..., e_{iT})'$. For the panel of data $X = (\underline{X}_1, \ldots, \underline{X}_N)$, we have

$$\underset{(T\times N)}{X} = \underset{(T\times r)}{F^0} \; \underset{(r\times N)}{\Lambda^{0\prime}} + \underset{(T\times N)}{e}, \tag{5}$$

with $e = (\underline{e}_1, ..., \underline{e}_N)$.

Let $tr(A)$ denote the trace of A. The norm of the matrix A is then $||A|| = tr(A'A)^{1/2}$. The following assumptions are made:

Assumption A: Factors

$E\|F_t^0\|^4 < \infty$ and $T^{-1}\sum_{t=1}^T F_t^0 F_t^{0\prime} \to \Sigma_F$ for some positive definite matrix Σ_F.

Assumption B: Factor Loadings

$\|\lambda_i\| \le \bar{\lambda} < \infty$, and $||\Lambda^{0\prime}\Lambda^0/N - D|| \to 0$ for some $r \times r$ positive definite matrix D.

Assumption C: Time and Cross-Section Dependence and heteroskedasticity

There exists a positive constant $M < \infty$, not necessarily the same throughout, such that for all N and T,

1. $E(e_{it}) = 0$, $E|e_{it}|^8 \le M$;

2. $E(e_s' e_t/N) = E(N^{-1}\sum_{i=1}^N e_{is}e_{it}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \le M$ for all s, and

$$T^{-1}\sum_{s=1}^T\sum_{t=1}^T |\gamma_N(s,t)| \le M;$$

3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \le |\tau_{ij}|$ for some τ_{ij} and for all t. In addition,

$$N^{-1}\sum_{i=1}^N\sum_{j=1}^N |\tau_{ij}| \le M;$$

4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T |\tau_{ij,ts}| \le M$;

5. For every (t,s), $E|N^{-1/2}\sum_{i=1}^N \Big[e_{is}e_{it} - E(e_{is}e_{it})\Big]|^4 \le M$.

Assumption D: Weak dependence between factors and idiosyncratic errors

$$E\Big(\frac{1}{N}\sum_{i=1}^N \|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^0 e_{it}\|^2\Big) \le M.$$

Assumption A is standard for factor models. Assumption B ensures that each factor has a non-trivial contribution to the variance of X_t. We only consider non-random factor loadings for simplicity. Our results still hold when λ_i is random, provided they are independent of the factors and idiosyncratic errors, and $E||\lambda_i||^4 \le M$. Assumption C allows for limited time series and cross section dependence in the idiosyncratic component. Heteroskedasticities in both the time and cross section dimensions are also allowed. Under stationarity in the time dimension, $\gamma_N(s,t) = \gamma_N(s-t)$, though the condition is not necessary. Given Assumption C1, the remaining assumptions in C are easily satisfied if the e_{it} are independent for all i and t. The allowance for some correlation in the idiosyncratic components sets up a model to have an *approximate factor structure* as defined in Chamberlain and Rothschild (1983). It is more general than a *strict factor model* which assumes e_{it} is uncorrelated across i, the framework in which the APT theory of Ross (1976) was based. Thus, the results to be developed will also apply to strict factor models. When the factors and idiosyncratic errors are independent (a standard assumption for conventional factor models), Assumption D is implied by Assumptions A and C. Independence is not required for D to be true. For example, suppose that $e_{it} = \epsilon_{it}\|F_t\|$ with ϵ_{it} being independent of F_t and ϵ_{it} satisfies Assumption C, then Assumption D holds. Finally, we note that the model being analyzed is static, in the sense that X_{it} has a contemporaneous relationship with the factors. The analysis of dynamic models is beyond the scope of this paper.

3 Estimating the common factors

To determine the number of factors, it is necessary to examine the consistency property of the estimated common factors. This differs from the classical factor analysis, in which a root-T consistent and asymptotically normal estimator for covariance matrix of X_t is available. So thattest statistics can be constructed based on the eigenvalues of the matrix. Because of the normality assumption, chi-square limiting distribution is obtained. Such a luxury is not at our possession for large models, and a different strategy is called for. It turns out that we need to explore the consistency property of the estimated common factors. In any case, the behavior of the estimated common factors is of important interest on its own right.

When N is small, common factors are often expressed in its state space form, normality is assumed, and the parameters are estimated by maximum likelihood. For example, Stock and Watson (1989) used $N = 4$ variables to estimate one factor, the coincident index. The drawback of the approach is that, because the number of parameters increases with N,[3] computational difficulties make it necessary to abandon information on many series even though they are available. But common factors can also be estimated (non-parametrically) by the method of asymptotic principal components for large N, both when T is small and when it is large.[4]

Since the true number r is unknown, we start with an arbitrary number k ($k < min\{N, T\}$). The superscript in λ_i^k and F_t^k signifies the allowance of k factors in the estimation. Estimates of λ^k and F^k are obtained by solving the optimization problem

$$V(k) = \min_{\Lambda, F^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^k F_t^k)^2$$

subject to the normalization of either $\Lambda^{k\prime}\Lambda^k/N = I_k$ or $F^{k\prime}F^k/T = I_k$. If we concentrating out Λ^k and use the normalization that $F^{k\prime}F^k/T = I_k$, the optimization problem is identical to maximizing $tr(F^{k\prime}(N^{-1}XX')F^k)$. The estimated factor matrix, denoted by $\widetilde{F}^k$, is $\sqrt{T}$ times eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix XX'. The estimated factors are thus the first k principal components. Given $\widetilde{F}^k$, $\widetilde{\Lambda}^{k\prime} = (\widetilde{F}^{k\prime}\widetilde{F}^k)^{-1}\widetilde{F}^{k\prime}X = \widetilde{F}^{k\prime}X/T$ are the corresponding factor loadings.

[3] Gregory, Head and Raynauld (1997) backed out a world factor and seven country specific factors from output, consumption, and investment for each of the G7 countries. The exercise involves estimation of 92 parameters and has perhaps stretched the state-space model to its limit.

[4] The method of asymptotic principal components of Chamberlain and Rothschild (1983) was used in Connor and Korajcyk (1986) and Connor and Korajcyk (1988) for fixed T. Forni et al. (1999) and Stock and Watson (1998) considered the method for large T.

A mathematically equivalent estimator, denoted by $\bar{\Lambda}^k$, can be constructed as $\sqrt{N}$ times the eigenvectors corresponding to the k largest eigenvalues of the $N \times N$ matrix $X'X$. Using the normalization that $\Lambda^{k\prime}\Lambda^k/N = I_k$, we have $\bar{F}^k = X\bar{\Lambda}^k/N$. The second set of calculations is computationally less costly when $T > N$, while the first is less intensive when $T < N$.[5]

Define

$$\widehat{F}^k = \bar{F}^k(\bar{F}^{k\prime}\bar{F}^k/T)^{1/2},$$

a rescaled estimator of the factors. The following Theorem summarizes the asymptotic properties of the estimated factors.

Theorem 1 *For any fixed $k \geq 1$, there exists a $(r \times k)$ matrix H^k with $rank(H^k) = \min(k, r)$, and $C_{NT} = \min(\sqrt{N}, \sqrt{T})$, such that*

$$C_{NT}^2 \left(\frac{1}{T} \sum_{t=1}^{T} ||\widehat{F}_t^k - H^{k\prime}F_t^0||^2 \right) = O_p(1).$$

Because the true factors (F^0) can only be identified up to scale, what is being considered is a rotation of F^0. The theorem establishes that the time average of the squared deviations between estimated factors and those that lie in the true factor space vanish as $N, T \to \infty$. The rate of convergence is determined by the smaller of N or T, and thus depends on the panel structure.

Using similar argument as in Theorem 1 and under an additional assumption that $\sum_{s=1}^{T} \gamma_N(s,t)^2 \leq M$ for all t and T, it can be shown that[6]

$$C_{NT}^2 \|\widehat{F}_t - H^{k\prime}F_t^0\|^2 = O_p(1) \qquad \text{for each } t \tag{6}$$

However, there is no guarantee that such a convergence is uniform over all t. As a consequence, Theorem 1 and (6) are not equivalent. Uniform convergence is considered by Stock and Watson (1998), which has a much slower convergence rate and the result requires $\sqrt{N} >> T$.

An important insight of this paper is that, to consistently estimate the number of factor, neither (6) nor uniform convergence is required. It is the average convergence rate of Theorem 1 that is essential. Theorem 1 has important ramifications. Using Theorem 1, it is possible to obtain the limiting distribution of the estimated common factors and common components (i.e., $\widehat{\lambda}_i'\widehat{F}_t$). This result is under further investigation.

[5]A more detailed account of computation issues, including how to deal with unbalanced panels, is given in Stock and Watson (1998).

[6]The proof is actually simpler than that of Theorem 1 and is thus omitted to avoid repetition.

4 Estimating the number of factors

To compare with the standard model selection problem, suppose for a moment that we observe all potentially informative factors but not the factor loadings. Then the problem is simply to choose k factors that best capture the variations in X and estimate the corresponding factor loadings. Since the model is linear and the factors are observed, λ_i can be estimated by applying ordinary least squares to each equation. This is then a classical model selection problem. A model with $k+1$ factors can fit no worse than a model with k factors, but efficiency is lost as more factor loadings have to be estimated. Let F^k be a matrix of k factors, and

$$V(k, F^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k\prime} F_t^k)^2$$

be the sum of squared residuals (divided by NT) from the cross-section regressions of $\underline{X}_i$ on k factors. Then a loss function $V(k, F^k) + kg(N,T)$, where $g(N,T)$ is the penalty for overfitting can be used to determine k. Because the estimation of λ_i is classical, it can be shown that the BIC with $g(N,T) = ln(T)/T$ can consistently estimate r. On the other hand, the AIC with $g(N,T) = 2$ may choose $k > r$ even in large samples. The result is the same as in Geweke and Meese (1981) derived for $N = 1$. The penalty factor does not need to take into account of the sample size in the cross-section dimension. Our main result is to show that this will no longer be true when the factors have to be estimated, and even the BIC will not always consistently estimate r.

Since $\widetilde{F}^k$, $\bar{F}^k$ and $\widehat{F}^k$ span the same column space, without loss of generality, we let

$$V(k, \widehat{F}^k) = \min_{\Lambda} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k\prime} \widehat{F}_t^k)^2 \tag{7}$$

denote the sum of squared residuals (divided by NT) when k factors are entertained. It should be clear that $V(k, \widetilde{F}^k) = V(k, \bar{F}^k) = V(k, \widehat{F}^k)$. We want to find penalty functions, $g(N,T)$, such that criteria of the form

$$IC(k) = V(k, \widehat{F}^k) + kg(N,T)$$

can consistently estimate r. Let $kmax$ be an bounded integer such that $r < kmax$.

Theorem 2 *Suppose that Assumptions A–D hold and that the k factors are estimated by principal components. Let* $\widehat{k} = \text{argmin}_{0 \le k \le kmax} IC(k)$. *Then* $\lim_{N,T\to\infty} Prob[\widehat{k} = r] = 1$ *if (i)* $g(N,T) \to 0$ *and (ii)* $C_{NT}^2 \cdot g(N,T) \to \infty$ *as* $N, T \to \infty$, *where* $C_{NT} = \min(\sqrt{N}, \sqrt{T})$.

A formal proof is provided in the Appendix. The crucial element in consistent estimation of r is a penalty factor that vanishes on the one hand, but still dominates the difference in the sum of squared residuals between the true and the overparameterized model. Let $\widehat{\sigma}^2$ be a consistent estimate of $(TN)^{-1}\sum_{i=1}^N\sum_{t=1}^T E(e_{it}^0)^2$. Consider the following criteria:

$$\begin{aligned}
PC_{p1}(k) &= V(k,\widehat{F}^k) + k\,\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right);\\
PC_{p2}(k) &= V(k,\widehat{F}^k) + k\,\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\cdot\ln C_{NT}^2;\\
PC_{p3}(k) &= V(k,\widehat{F}^k) + k\,\widehat{\sigma}^2\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right).
\end{aligned}$$

Since $V(k,\widehat{F}^k) = N^{-1}\sum_{i=1}^N \widehat{\sigma}_i^2$, where $\widehat{\sigma}_i^2 = \underline{\widehat{e}}_i'\underline{\widehat{e}}_i/T$, the criteria generalize the C_p criterion of Mallows (1973) developed for selection of models in strict time series or cross section contexts to a panel data setting. For this reason, we refer to these statistics as Panel C_p (PC_p) criteria. Like the C_p criterion, $\widehat{\sigma}^2$ provides the proper scaling to the penalty term. In applications, it can be replaced by $V(kmax,\widehat{F}^{kmax})$.

The proposed penalty functions are based on the sample size in the smaller of the two dimensions. All three criteria satisfy conditions (i) and (ii) of Theorem 2 since $C_{NT}^{-2}\approx \frac{N+T}{NT}\to 0$ as $N,T\to\infty$. However, in finite samples, $C_{NT}^{-2}\le\frac{N+T}{NT}$. Hence, the three criteria, although asymptotically equivalent, will have different properties in finite samples.

To understand the conditions imposed by Theorem 2, we also consider:

$$\begin{aligned}
Test_1(k) &= V(k,\widehat{F}^k) + k\,\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln(N+T);\\
Test_2(k) &= V(k,\widehat{F}^k) + k\,\widehat{\sigma}^2\left(\frac{2}{T}\right);\\
Test_3(k) &= V(k,\widehat{F}^k) + k\,\widehat{\sigma}^2\left(\frac{\ln T}{T}\right).
\end{aligned}$$

Consider first $Test_1$ and suppose $N>T$. Then $\frac{N+T}{NT}\approx\frac{1}{T}$ for large N. The condition that $g(N,T)\to 0$ would fail if $ln(N+T)/T\not\to 0$. For example, if $N=\exp(T)$, r will not be consistently estimated in theory, even though such a data configuration is unusual in practice. But for $N = T^\alpha$, with α an arbitrary positive constant, then $Test_1$ meets the conditions of Theorem 2, we therefore expect $Test_1$ to perform well for a wide range of N and T. $Test_2$ and $Test_3$ resemble the AIC and the BIC respectively. Although $g(N,T) \to 0$ as $T \to \infty$, $Test_2$ fails the second condition for all N and T. When $N << T$ such that $N \log(T)/T \not\to \infty$, $Test_3$ also fails condition (ii) of Theorem 2. Thus we expect $Test_2$ will not work for all N and T and $Test_3$ will not work for small N relative to T.

5 Simulations

We simulate data from the following model:

$$\begin{aligned} X_{it} &= \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it} \\ &= c_{it} + e_{it}, \end{aligned}$$

where the factors are $T \times r$ matrices of $N(0,1)$ variables, and the factor loadings are $N(0,1)$ variates. Hence, the common component of X_{it}, denoted by c_{it} has variance r. Results with λ_{ij} uniformly distributed are similar and will not be reported. Our base case assumes that the idiosyncratic component has the same variance as the common component (i.e. $\theta = r$). We consider 15 configurations of the data. The first five simulates plausible asset pricing applications with five years of data ($T = 60$) on 100 to 2000 asset returns. We then increase T to 100. The next two configurations with N=60, T=100 and 200 are plausible size of datasets for sectors, states, regions, and countries. The large T configurations are reported for completeness. All computations were performed using Matlab Version 5.3.

Reported in Tables 1 to 3 are the average $\widehat{k}$ over 1000 replications, for $r = 1$, 3, and 5 respectively, assuming that e_{it} is homoskedastic $N(0,1)$. Of the three criteria that satisfy Theorem 2, PC_{p3} is less robust than PC_{p1} and PC_{p2} when N or T is small. The term $\frac{NT}{N+T}$ provides a small sample correction to the asymptotic convergence rate of C_{NT}^2 and has the effect of adjusting the penalty upwards. The simulations show this adjustment to be desirable. For all cases, the maximum of number factors, $kmax$, is set to 8.

Of the three criteria that do not satisfy all aspects of Theorem 1, $Test_1$ evidently performs very well. The reason, as explained earlier, is that $Test_1$ fails if N is at least of order $exp(T)$, but such a data combination is not considered in the simulations. Although this is also an unlikely case in practice, it is preferable to consider PC_{p1} and PC_{p2} since they have the same empirical properties and are preferred on theoretical grounds. $Test_2$ fails miserably because $C_{NT}^2 \cdot g(N,T)$ does not diverge. $Test_3$ fairs better when $T < N$, but fails when $N < T$, as theory predicts.

Table 4 relaxes the assumption of homokedasticity. Instead, we let $e_{it} = e_{it}^1$ for t odd, and $e_{it} = e_{it}^1 + e_{it}^2$ for t even, where e_{it}^1 and e_{it}^2 are independent $N(0,1)$. Thus, in even periods, the variance in the even periods is twice as large as the odd periods. Without loss of generality, we only report results for $r = 3$. PC_{p1} and PC_{p2} continue to select the true number of factors very accurately, and dominates the remaining criteria considered.

We also consider $\theta = 2r$ and $\theta = r/2$ to assess the robustness of the results to θ, the variance of the idiosyncratic errors. When $\theta > r$, the variance of the idiosyncratic component is larger than the common component. It is conceivable that the common factors are estimated with less precision. Nonetheless, PC_{p1} and PC_{p2} still gives the correct estimate of r. When $\theta < r$, the common component has larger variance. In such a case, all three proposed criteria give a precise estimate of r.

The preferred criteria, from both theoretical and empirical grounds, are thus PC_{p1} and PC_{p2}. It should be emphasized that the results reported in Tables 1-6 are the average of $\widehat{k}$ over 1000 simulations. This average can equal r only if $\widehat{k} = r$ in every replication, and our preferred criteria accomplishes this precision.

6 Concluding Remarks

A characteristic of a panel of data that has a r factor representation is that the first r largest population eigenvalues of the $N \times N$ covariance of X_t diverge as N increases to infinity, but the $(r+1)^{th}$ eigenvalue is bounded, see Chamberlain and Rothschild (1983). This would seem to suggest that a test based on the sample eigenvalues of the matrix $\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})(X_t - \bar{X})'$ can be used to test for the number of factors. But it can be shown that all non-zero sample eigenvalues (not just the first r) of the matrix $\widehat{\Sigma}$ increase with N, and a test based on the sample eigenvalues is thus not feasible. For this reason, our test is based directly on the factor model. The main appeal of our results is that they are developed under the assumption that $N, T \to \infty$ and thus appropriate for many datasets typically used in macroeconomic analysis. The test should be useful in applications in which the number of factors has traditionally been assumed rather than determined by the data.

The foregoing analysis has assumed a static relationship between the observed data and the factors. Sargent and Sims (1977) and Geweke (1977) extended the static strict factor model to allow for dynamics. Stock and Watson (1998) suggest how dynamics can be introduced into factor models, although their empirical applications assume a static factor structure. Forni et al. (1999) allowed X_{it} to also depend on the leads of the factors. When the method developed in this paper applied to a dynamic model, the estimated number of factors gives an upper bound of the true number of factors. Consider the data generating process $X_{it} = af_t + bf_{t-1} + e_{it}$. From the dynamic point of view, there is only one factor. The static approach treats the model as having two factors. While this may not have much practical consequence, it illustrates the theoretical restriction of the static approach. A more intriguing example is $X_{it} = aX_{it-1} + bf_t + e_{it}$ ($|a| \neq 1$, otherwise use ΔX_{it}) so that X_{it} is an infinite moving average of a single factor. Depending on how fast the moving average coefficients decay to zero, the model may be approximated by a finite number of factors. Still, the limitation of the static approach is apparent. That is, the static approach applied to true dynamic model will only give an upper bound on the true number of factors. Nevertheless, this paper takes an important step toward a solution to dynamic models. Developing a factor selection rule in a dynamic setting is a non-trivial task and will continue to be the subject of investigation.

In summary, this paper has made some contribution to the analysis of factor models of large dimensions. We show that the common factors as well as the number of factors can be consistently estimated. Although serial correlations are allowed, the model is static in nature. The study of the dynamic model is a subject of ongoing research. THIS SUMMARY NEED TO BE REWRITTEN.

Appendix

To prove the main results we need the following lemma.

Lemma 1 *Under assumptions A-C, we have for some $M < \infty$, and for all N and T,*

$$\begin{aligned}
&(i) && T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 \le M,\\
&(ii) && E\Big(T^{-1}\sum_{t=1}^{T}\|N^{-1/2}e_t'\Lambda^0\|\Big) = E\Big(T^{-1}\sum_{t=1}^{T}\|N^{-1/2}\sum_{i=1}^{N}e_{it}\lambda_i^0\|\Big) \le M\\
&(iii) && E\Big(T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(N^{-1}\sum_{i=1}^{N}X_{it}X_{is})^2\Big) \le M,\\
&(iv) && E\Big\|(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i^0\Big\| \le M.
\end{aligned}$$

Proof: Consider (i). Let $\rho(s,t) = \gamma_N(s,t)/[\gamma_N(s,s)\gamma_N(t,t)]^{1/2}$. Then $|\rho(s,t)| \le 1$. From $\gamma_N(s,s) \le M$,

$$\begin{aligned}
T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 &= T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,s)\gamma_N(t,t)\rho(s,t)^2\\
&\le MT^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,s)\gamma_N(t,t)|^{1/2}|\rho(s,t)|\\
&= MT^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\gamma_N(s,t)| \le M^2
\end{aligned}$$

by Assumption C2. Consider (ii).

$$E\|N^{-1/2}\sum_{i=1}^{N}e_{it}\lambda_i^0\|^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}E(e_{it}e_{jt})\lambda_i^0\lambda_j^0 \le \bar\lambda^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}| \le \bar\lambda^2 M$$

by Assumptions B and C3. For (iii), it is sufficient to prove $E|X_{it}|^4 \le M$ for all (i,t). Now $E|X_{it}|^4 \le 8E(\lambda_i^{0\prime}F_t^0)^4 + 8E|e_{it}|^4 \le 8\bar\lambda^4E\|F_t^0\|^4 + 8E|e_{it}|^4 \le M$ for some M by Assumptions A, B and C1. Finally for (iv),

$$\begin{aligned}
E\|(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}\lambda_i^0\|^2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E(e_{it}e_{js})\lambda_i^{0\prime}\lambda_j^0\\
&\le \bar\lambda^2\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \le \bar\lambda^2 M
\end{aligned}$$

by Assumption C4.

Proof of Theorem 1 We use the mathematical identity $\widehat{F}^k = N^{-1}X\widetilde{\Lambda}^k$, and $\widetilde{\Lambda}^k = T^{-1}X'\widetilde{F}^k$. From the normalization $\widetilde{F}^{k\prime}\widetilde{F}^k/T = I_k$, we also have $T^{-1}\sum_{t=1}^T ||\widetilde{F}_t^k||^2 = O_p(1)$. For $H^k = (\widetilde{F}^{k\prime}F^0/T)(\Lambda^{0\prime}\Lambda^0/N)$, we have:

$$\begin{aligned}
\widehat{F}_t^k - H^{k\prime}F_t^0 &= T^{-1}\sum_{s=1}^T \widetilde{F}_s^k\gamma_N(s,t) + T^{-1}\sum_{s=1}^T \widetilde{F}_s^k\zeta_{st} + T^{-1}\sum_{s=1}^T \widetilde{F}_s\eta_{st} + T^{-1}\sum_{s=1}^T \widetilde{F}_s\xi_{st} \\
\text{where } \zeta_{st} &= \frac{e_s'e_t}{N} - \gamma_N(s,t), \\
\eta_{st} &= F_s^{0\prime}\Lambda'e_t/N, \\
\xi_{st} &= F_t^{0\prime}\Lambda'e_s/N
\end{aligned}$$

Note that H^k depends on N and T. Throughout, we will suppress this dependence to simplify the notation. We also note that $\|H^k\|$ is bounded. Because $(x+y+z+u)^2 \le 4(x^2+y^2+z^2+u^2)$, $||\widehat{F}_t^k - H^{k\prime}F_t^0||^2 \le 4(a_t + b_t + c_t + d_t)$, where

$$\begin{aligned}
a_t &= T^{-2}||\sum_{s=1}^T \widetilde{F}_s^k\gamma_N(s,t)||^2, \\
b_t &= T^{-2}||\sum_{s=1}^T \widetilde{F}_s^k\zeta_{st}||^2, \\
c_t &= T^{-2}||\sum_{s=1}^T \widetilde{F}_s^k\eta_{st}||^2, \\
d_t &= T^{-2}||\sum_{s=1}^T \widetilde{F}_s^k\xi_{st}||^2.
\end{aligned}$$

Now $||\sum_{s=1}^T \widetilde{F}_s^k\gamma_N(s,t)||^2 \le (\sum_{s=1}^T ||\widetilde{F}_s^k||^2)\cdot(\sum_{s=1}^T \gamma_N^2(s,t))$. Thus,

$$\begin{aligned}
T^{-1}\sum_{t=1}^T a_t &\le T^{-1}\Big(T^{-1}\sum_{s=1}^T ||\widetilde{F}_s^k||^2\Big)\cdot T^{-1}\Big(\sum_{t=1}^T\sum_{s=1}^T \gamma_N(s,t)^2\Big) \\
&= O_p(T^{-1})
\end{aligned}$$

by Lemma 1(i).

For b_t, we have that

$$\begin{aligned}
\sum_{t=1}^T b_t &= T^{-2}\sum_{t=1}^T ||\sum_{s=1}^T \widetilde{F}_s^k\zeta_{st}||^2 \\
&= T^{-2}\sum_{t=1}^T\sum_{s=1}^T\sum_{u=1}^T \widetilde{F}_s^{k\prime}\widetilde{F}_u^k\zeta_{st}\zeta_{ut} \\
&\le \left(T^{-2}\sum_{s=1}^T\sum_{u=1}^T(\widetilde{F}_s^{k\prime}\widetilde{F}_u^k)^2\right)^{1/2}\left[T^{-2}\sum_{s=1}^T\sum_{u=1}^T(\sum_{t=1}^T\zeta_{st}\zeta_{ut})^2\right]^{1/2}
\end{aligned}$$

$$\leq \quad (T^{-1}\sum_{s=1}^{T}||\widetilde{F}_s^k||^2)\cdot\left[T^{-2}\sum_{s=1}^{T}\sum_{u=1}^{T}(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut})^2\right]^{1/2}.$$

Since $E(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut})^2 = E(\sum_{t=1}^{T}\sum_{v=1}^{T}\zeta_{st}\zeta_{ut}\zeta_{sv}\zeta_{uv}) \leq T^2\max_{s,t}E|\zeta_{st}|^4$. But

$$E|\zeta_{st}|^4 = \frac{1}{N^2}E|N^{-1/2}\sum_{i=1}^{N}(e_{it}e_{is} - E(e_{it}e_{is})|^4 \leq N^{-2}M$$

by Assumption C5. We have

$$\sum_{t=1}^{T}b_t \leq O_p(1)\cdot\sqrt{\frac{T^2}{N^2}} = O_p(\frac{T}{N}),$$

$T^{-1}\sum_{t=1}^{T}b_t = O_p(N^{-1})$. For c_t, we have

$$\begin{aligned} c_t &= T^{-2}||\sum_{s=1}^{T}\widetilde{F}_s^k\eta_{st}||^2 = T^{-2}||\sum_{s=1}^{T}\widetilde{F}_s^kF_s^{0\prime}\Lambda^{0\prime}e_t/N||^2 \\ &\leq N^{-2}\|e_t'\Lambda^0\|(T^{-1}\sum_{s=1}^{T}\|\widetilde{F}_s^k\|^2)(T^{-1}\sum_{s=1}^{T}\|F_s^0\|^2) \\ &= N^{-2}||e_t'\Lambda^0||^2O_P(1). \end{aligned}$$

It follows that

$$T^{-1}\sum_{t=1}^{T}c_t = O_p(1)N^{-1}T^{-1}\sum_{t=1}^{T}||\frac{e_t'\Lambda^0}{\sqrt{N}}||^2 = O_p(N^{-1}).$$

by Lemma 1 (ii). The term $d_t = O_p(N^{-1})$ can be proved similarly. Combining these results, we have $T^{-1}\sum_{t=1}^{T}(a_t + b_t + c_t + d_t) = O_p(N^{-1}) + O_p(T^{-1})$.

To prove Theorem 2, we need additional results.

Lemma 2 *For any k, $1 \leq k \leq r$, and H^k be the matrix defined in Theorem 1,*

$$V(k,\widehat{F}^k) - V(k, F^0H^k) = O_p(C_{NT}^{-1}).$$

Proof For the true factor matrix with r factors and H^k defined in Theorem 1, let M_{FH}^0 denote the projection matrix spanned by null space of F^0H^k. Correspondingly, let $M_{\widehat{F}}^k = I_T - \widehat{F}^k(\widehat{F}^{k\prime}\widehat{F}^k)^{-1}\widehat{F}^k$. Then

$$\begin{aligned} V(k,\widehat{F}^k) &= N^{-1}T^{-1}\sum_{i=1}^{N}\underline{X}_i{\prime}M_{\widehat{F}}^k\underline{X}_i, \\ V(k,F^0H^k) &= N^{-1}T^{-1}\sum_{t=1}^{N}\underline{X}_i'M_{FH}^0\underline{X}_i, \\ V(k,\widehat{F}^k) - V(k,F^0H^k) &= N^{-1}T^{-1}\sum_{i=1}^{N}\underline{X}_i'(P_{FH}^0 - P_{\widehat{F}}^k)\underline{X}_i. \end{aligned}$$

Let $D_k = \widehat{F}^{k\prime}\widehat{F}^k/T$ and $D_0 = H^k F^{0\prime} F^0 H^{k\prime}/T$. Then

$$\begin{aligned}
P^k_{\widehat{F}} - P^0_{FH} &= T^{-1}\widehat{F}^k \left(\frac{\widehat{F}^{k\prime}\widehat{F}^k}{T}\right)^{-1} \widehat{F}^k - T^{-1}F^0 H^k \left(\frac{H^k F^{0\prime} F^0 H^{k\prime}}{T}\right)^{-1} H^{k\prime} F^{0\prime} \\
&\equiv T^{-1}\left[\widehat{F}^k D_k^{-1}\widehat{F}^k - F^0 H^k D_0^{-1} H' F^{0\prime}\right], \\
&= T^{-1}\left[\widehat{F}^k - F^0 H^k + F^0 H^k) D_k^{-1}(\widehat{F}^k - F^0 H^k + F^0 H^k) - F^0 H^k D_0 H^{k\prime} F^{0\prime}\right], \\
&= T^{-1}\Big[(\widehat{F}^k - F^0 H^k) D_k^{-1}(\widehat{F}^k - F^0 H^k)' + (\widehat{F}^k - F^0 H^k) D_k^{-1} H^k F^{0\prime} \\
&\quad + F^0 H^k D_k^{-1}(\widehat{F}^k - F^0 H^k)' + F^0 H^k (D_k^{-1} - D_0^{-1}) H^{k\prime} F^{0\prime}\Big].
\end{aligned}$$

Thus, $N^{-1}T^{-1}\sum_{i=1}^N \underline{X}_i'(P^k_{\widehat{F}} - P^0_{FH})\underline{X}_i = I + II + II + IV$. We consider each term in turn.

$$\begin{aligned}
I &= N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T(\widehat{F}^k_t - H^{k\prime}F^0_t)' D_k^{-1}(\widehat{F}^k_s - H^{k\prime}F^0_s) X_{it}X_{it} \\
&\le \left(T^{-2}\sum_{t=1}^T\sum_{s=1}^T\left[(\widehat{F}^k_t - H^{k\prime}F^0_t)' D_k^{-1}(\widehat{F}^k_s - H^{k\prime}F^0_s)\right]^2\right)^{1/2} \cdot \left[T^{-2}\sum_{t=1}^T\sum_{s=1}^T(N^{-1}\sum_{i=1}^N X_{it}X_{is})^2\right]^{1/2} \\
&\le \left(T^{-1}\sum_{t=1}^T ||F^k_t - H^{k\prime}F^0_t||^2\right) \cdot ||D_k^{-1}|| \cdot O_p(1) = O_p(C_{NT}^{-2}).
\end{aligned}$$

by Theorem 1 and Lemma 1(iii).

$$\begin{aligned}
II &= N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T(\widehat{F}^k_t - H^{k\prime}F^0_t)' D_k^{-1} H^{k\prime} F^0_s X_{it}X_{is} \\
&\le \left(T^{-2}\sum_{t=1}^T\sum_{s=1}^T ||\widehat{F}^k_t - H^{k\prime}F^0_t||^2 \cdot ||H^{k\prime}F^0_s||^2 \cdot ||D_k^{-1}||^2\right)^{1/2} \cdot \left[T^{-2}\sum_{t=1}^T\sum_{s=1}^T(N^{-1}\sum_{i=1}^N X_{it}X_{is})^2\right]^{1/2} \\
&\le \left(T^{-1}\sum_{t=1}^T ||\widehat{F}^k_t - H^{k\prime}F^0_t||^2\right)^{1/2} \cdot ||D_k^{-1}|| \cdot \left(T^{-1}\sum_{t=1}^T ||H^{k\prime}F^0_s||^2\right)^{1/2} \cdot O_p(1) \\
&= \left(T^{-1}\sum_{t=1}^T ||\widehat{F}^k_t - H^{k\prime}F^0_t||^2\right)^{1/2} \cdot O_p(1) = O_p(C_{NT}^{-1}).
\end{aligned}$$

It can be verified that III is also $O_p(C_{NT}^{-1})$.

$$\begin{aligned}
IV &= N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T F^{0\prime}_t H^k (D_k^{-1} - D_0^{-1}) H^{k\prime} F^0_s X_{it}X_{is} \\
&\le ||D_k^{-1} - D_0^{-1}|| N^{-1}\sum_{i=1}^N\left(T^{-1}\sum_{t=1}^T ||H^{k\prime}F^0_t|| \cdot |X_{it}|\right)^2 \\
&= ||D_k^{-1} - D_0^{-1}|| \cdot O_p(1).
\end{aligned}$$

Next, we prove that $\|D_k^1 - D_0\| = O_p(C_{NT})$.

$$\begin{aligned}
D_k - D_0 &= \frac{\widehat{F}^{k\prime}\widehat{F}^k}{T} - \frac{H^{k\prime}F^{0\prime}F^0H^k}{T} \\
&= T^{-1}\sum_{t=1}^{T}\left[\widehat{F}_t^k\widehat{F}_t^{k\prime} - H^{k\prime}F_t^0F_t^0H^k\right] \\
&= T^{-1}\sum_{t=1}^{T}(\widehat{F}_t^k - H^{k\prime}F_t^0)(\widehat{F}_t^k - H^{k\prime}F_t^0)' \\
&\quad + T^{-1}\sum_{t=1}^{T}(\widehat{F}_t^k - H^{k\prime}F_t^0)F_t^0H^k + T^{-1}\sum_{t=1}^{T}H^kF_t^0(\widehat{F}_t^k - H^{k\prime}F_t^0)', \\
||D_k - D_0|| &\leq T^{-1}\sum_{t=1}^{T}||\widehat{F}_t^k - H^{k\prime}F_t^0||^2 + \\
&\quad 2\left(T^{-1}\sum_{t=1}^{T}||\widehat{F}_t^k - H^{k\prime}F_t^0||^2\right)^{1/2} \cdot \left(T^{-1}\sum_{t=1}^{T}||H^{k\prime}F_t^0||^2\right)^{1/2} \\
&= O_p(C_{NT}^{-2}) + O_p(C_{NT}^{-1}) = O_p(C_{NT}^{-1}).
\end{aligned}$$

Because ${F^0}'F^0/T$ converges to a positive definite matrix, and because $rank(H^k) = k \leq r$, D_0 converges to a positive definite matrix. From $D_k - D_0 = O_p(C_{NT}^{-1})$, D_k also converges to a positive definite matrix. This in turn implies that $D_K^{-1} - D_0^{-1} = O_p(C_{NT}^{-1})$.

Lemma 3 *For the matrix H^k defined in Theorem 1, and for each k with $k < r$, there exists a $\tau_k > 0$ such that*

$$\text{plim}\inf_{N,T\to\infty} V(k, F^0H^k) - V(r, F^0) = \tau_k.$$

Proof

$$\begin{aligned}
V(k, F^0H^k) - V(r, F^0) &= N^{-1}T^{-1}\sum_{i}^{N}\underline{X}_i'(P_F^0 - P_{FH}^0)\underline{X}_i \\
&= N^{-1}T^{-1}\sum_{i-1}^{N}(F^0\lambda_i^0 + \underline{e}_i)'(P_F^0 - P_{FH}^0)(F^0\lambda_i^0 + \underline{e}_i) \\
&= N^{-1}T^{-1}\sum_{i=1}^{N}\lambda_i^0F^{0\prime}(P_F^0 - P_{FH}^0)F^0\lambda_i^0 \\
&\quad + 2N^{-1}T^{-1}\sum_{i=1}^{N}\underline{e}_i'(P_F^0 - P_{FH}^0)F^0\lambda_i^0 \\
&\quad + N^{-1}T^{-1}\sum_{i=1}^{N}\underline{e}_i'(P_F^0 - P_{FH}^0)\underline{e}_i \\
&= I + II + III.
\end{aligned}$$

First, note that $P_F^0 - P_{FH}^0 \geq 0$. Hence, $III \geq 0$. For the first two terms,

$$\begin{aligned}
I &= tr\left[T^{-1}F^{0\prime}(P_F^0 - P_{FH}^0)F^0 N^{-1}\sum_{i=1}^N \lambda_i^0\lambda_i^{0\prime}\right] \\
&= tr\left(\left[\frac{F^{0\prime}F^0}{T} - \frac{F^{0\prime}F^0H^k}{T}\left(\frac{H^{k\prime}F^{0\prime}F^0H^k}{T}\right)^{-1}\frac{H^{k\prime}F^{0\prime}F^0}{T}\right]\cdot N^{-1}\sum_{i=1}^N \lambda_i^0\lambda_i^{0\prime}\right) \\
&\to tr\left(\left[\Sigma_F - \Sigma_F H_0^k(H_0^{k\prime}\Sigma_F H_0^k)^{-1}H_0^{k\prime}\Sigma_F\right]\cdot D\right) \\
&= tr(A\cdot D),
\end{aligned}$$

where $A = \Sigma_F - \Sigma_F H_0^k(H_0^{k\prime}\Sigma_F H_0^k)^{-1}H_0^{k\prime}\Sigma_F$ and H_0^k is the limit of H^k with $\text{rank}(H_0^k) = k < r$ [see, Stock and Watson (1998)]. Now $A \neq 0$ because $\text{rank}(\Sigma_F) = r$ (Assumption A). Also, A is semipositive and $D > 0$ (Assumption B). This implies that $tr(A\cdot D) > 0$.

Now $II = 2N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' P_F^0 F^0\lambda_i^0 - 2N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' P_{FH}^0 F^0\lambda_i^0$. Consider the first term.

$$\begin{aligned}
|N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' P_F^0 F^0\lambda_i^0| &= |N^{-1}T^{-1}\sum_{i=1}^N\sum_{t=1}^T e_{it}F_t^{0\prime}\lambda_i^0| \\
&\leq (T^{-1}\sum_{t=1}^T ||F_t^0||^2)^{1/2}\cdot\frac{1}{\sqrt{N}}(T^{-1}\sum_{t=1}^T ||\frac{1}{\sqrt{N}}\sum_{i=1}^N e_{it}\lambda_i^0||^2)^{1/2} \\
&= O_p(\frac{1}{\sqrt{N}}).
\end{aligned}$$

The last equality follows from Lemma 1 (ii). The second term is also $O_p(\frac{1}{\sqrt{N}})$, and hence $II = O_p(\frac{1}{\sqrt{N}}) \to 0$.

Lemma 4 *For any fixed k with $k \geq r$, $V(k,\widehat{F}^k) - V(r,\widehat{F}^r) = O_p(\frac{1}{N}+\frac{1}{T})$.*

Proof:

$$\begin{aligned}
|V(k,\widehat{F}^k) - V(r,\widehat{F}^r)| &\leq |V(k,\widehat{F}^k) - V(r,F^0)| + |V(r,F^0) - V(r,\widehat{F}^r)| \\
&\leq 2\max_{r\leq k\leq kmax}|V(k,\widehat{F}^k) - V(r,F^0)|.
\end{aligned}$$

Thus, it is sufficient to prove for each k with $k \geq r$,

$$V(k,\widehat{F}^k) - V(r,F^0) = O_p(\frac{1}{N}+\frac{1}{T}).$$

Let H^k be as defined in Theorem 1, now with rank r because $k \geq r$. Let H^{k+} be the generalized inverse of H^k such that $H^kH^{k+} = I_r$. From $\underline{X}_i = F^0\lambda_i^0 + \underline{e}_i$, we have $\underline{X}_i = F^0H^kH^{k+}\lambda_i^0 + \underline{e}_i$. This implies

$$\begin{aligned}\underline{X}_i &= \widehat{F}^k H^{k+}\lambda_i^0 + \underline{e}_i - (\widehat{F}^k - F^0 H^k)H^{k+}\lambda_i^0 \\ &= \widehat{F}^k H^{k+}\lambda_i^0 + \underline{u}_i,\end{aligned}$$

where $\underline{u}_i = \underline{e}_i - (\widehat{F}^k - F^0 H^k)H^{k+}\lambda_i^0$.

Note that

$$V(k,\widehat{F}^k) = N^{-1}T^{-1}\sum_{i=1}^N \underline{u}_i' M_{\widehat{F}}^k \underline{u}_i,$$

$$V(r,\widehat{F}^0) = N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' M_{F}^0 \underline{e}_i.$$

$$\begin{aligned}V(k,\widehat{F}^k) &= N^{-1}T^{-1}\sum_{i=1}^N \Big(\underline{e}_i - (\widehat{F}^k - F^0H^k)H^{k+}\lambda_i^0\Big)' M_{\widehat{F}}^k \Big(\underline{e}_i - (\widehat{F}^k - F^0H^k)H^{k+}\lambda_i^0\Big) \\ &= N^{-1}T^{-1}\sum_{i=1}^N \underline{e}_i' M_{\widehat{F}}^k \underline{e}_i - 2N^{-1}T^{-1}\sum_{i=1}^N \lambda_i^{0\prime} H^{k+\prime}(\widehat{F}^k - F^0H^k)' M_{\widehat{F}}^k \underline{e}_i \\ &\quad + N^{-1}T^{-1}\sum_{i=1}^N \lambda_i^{0\prime} H^{k+\prime}(\widehat{F}^k - F^0H^k)' M_{\widehat{F}}^k (\widehat{F}^k - F^0H^k)H^{k+}\lambda_i^0 \\ &\equiv a+b+c.\end{aligned}$$

Because $M_{\widehat{F}}^k$ is a projection matrix, $x'M_{\widehat{F}}^k x \le x'x$. Thus,

$$\begin{aligned}c &\le N^{-1}T^{-1}\sum_{i=1}^N \lambda_i^{0\prime} H^{k+\prime}(\widehat{F}^k - F^0H^k)'(\widehat{F}^k - F^0H^k)H^{k+}\lambda_i^0 \\ &\le T^{-1}\sum_{t=1}^T ||\widehat{F}_t^k - H^{k\prime}F_t^0||^2 \cdot \Big(N^{-1}\sum_{i=1}^N ||\lambda_i^0||^2 ||H^{k+}||^2\Big) \\ &= O_p(C_{NT}^{-2}) \cdot O_p(1)\end{aligned}$$

by Theorem 1. For the term b, we use the fact that $|\text{tr(A)}| \le \text{r}\|\text{A}\|$ for any $r \times r$ matrix A. Thus

$$\begin{aligned}b &= 2T^{-1}\text{tr}\Big(H^{k+}(\widehat{F}^k - F^0H^k)' M_{\widehat{F}}^k (N^{-1}\sum_{i=1}^N \underline{e}_i\lambda_i^0)\Big) \\ &\le 2r||H^{k+}|| \cdot ||\frac{\widehat{F}^k - F^0H^k}{\sqrt{T}}|| \cdot ||\frac{1}{\sqrt{T}N}\sum_{i=1}^N \underline{e}_i\lambda_i^0|| \\ &\le 2r||H^{k+}|| \cdot \Big(T^{-1}\sum_{t=1}^T ||\widehat{F}_t^k - H^{k\prime}F_t^0||^2\Big)^{1/2} \cdot \frac{1}{\sqrt{N}} ||\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T e_{it}\lambda_i^0|| \\ &= O_p(C_{NT}^{-1}) \cdot \frac{1}{\sqrt{N}} = O_p(C_{NT}^{-2})\end{aligned}$$

by Theorem 1 and Lemma 1(iv). Therefore,

$$V(k, \widehat{F}^k) \;=\; N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i' M_{\widehat{F}}^k \underline{e}_i + O_p(C_{NT}^{-2})$$

Using the fact that $V(k, \widehat{F}^k) - V(r, F^0) \leq 0$ for $k \geq r$,

$$\begin{aligned} 0 \;\geq\; & V(k, \widehat{F}^k) - V(r, F^0) = N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i'(P_{\widehat{F}}^k - P_F^0)\underline{e}_i + O_p(C_{NT}^{-2}) \\ =\; & N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i' P_{\widehat{F}}^k \underline{e}_i - N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i' P_F^0 \underline{e}_i + O_p(C_{NT}^{-2}). \end{aligned}$$

Note that

$$\begin{aligned} N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i' P_F^0 \underline{e}_i \;\leq\; & \|(F^{0\prime}F^0/T)^{-1}\| \cdot N^{-1}T^{-2}\sum_{i=1}^{N} \underline{e}_i' F^0 F^{0\prime} \underline{e}_i \\ =\; & O_p(1)T^{-1}N^{-1}\sum_{i=1}^{N} \|T^{-1/2}\sum_{t=1}^{T} F_t^0 e_{it}\|^2 = O_p(T^{-1}) \leq O_p(C_{NT}^{-2}) \end{aligned}$$

by Assumption D. Thus

$$0 \geq N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i' P_{\widehat{F}}^k \underline{e}_i + O_p(C_{NT}^{-2}).$$

This implies that $N^{-1}T^{-1}\sum_{i=1}^{N} \underline{e}_i' P_{\widehat{F}}^k \underline{e}_i = O_p(C_{NT}^{-2})$. In summary

$$V(k, \widehat{F}^k) - V(r, F^0) = O_p(C_{NT}^{-2}) = O_p(\frac{1}{N} + \frac{1}{T}) = O_p(\frac{N+T}{NT}).$$

Proof of Theorem 2 We shall prove that $\lim_{N,T\to\infty} P(IC(k) < IC(r)) = 0$ for all $k \neq r$ and $k \leq kmax$. Consider $k < r$. Since

$$IC(k) - IC(r) = V(k, \widehat{F}^k) - V(r, \widehat{F}^r) - (r-k)g(N,T),$$

the required condition when $k < r$ is $Prob[V(k, \widehat{F}^k) - V(r, \widehat{F}^r) < (r-k)g(N,T)] = 0$ as $N, T \to \infty$. Now

$$\begin{aligned} V(k, \widehat{F}^k) - V(r, \widehat{F}^r) \;=\; & [V(k, \widehat{F}^k) - V(k, F^0H^k)] + \\ & [V(k, F^0H^k) - V(r, F^0H^r)] + [V(r, F^0H^r) - V(r, \widehat{F}^r)]. \end{aligned}$$

Lemma 2 mplies that the first and the third terms are both $O_p(C_{NT}^{-1})$. Next, consider the second term. Because F^0H^r and F^0 span the same column space, $V(r, F^0H^r) = V(r, F^0)$.

Thus the second term can be rewritten as $V(k, F^0H^k) - V(r, F^0)$, which has a positive limit by Lemma 3. Hence, $Prob[IC(k) < IC(r)] \to 0$ if $g(N,T) \to 0$ as $N, T \to \infty$. Next, for $k \geq r$,

$$Prob[IC(k) - IC(r) < 0] = Prob[V(r, \widehat{F}^r) - V(k, \widehat{F}^k) > (k-r)g(N,T)].$$

By Lemma 4, $V(r, \widehat{F}^r) - V(k, \widehat{F}^k) = O_p(C_{NT}^{-2})$. For $k > r$, $(k-r)g(N,T) \geq g(N,T)$, which converges to zero at a slower rate that C_{NT}^{-2}. Thus for $k > r$, $Prob[IC(k) < IC(r)] \to 0$ as $N, T \to \infty$.

Table 1: DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}$
$r = 1; \theta = 1.$

N	T	PC_{p1}	PC_{p2}	PC_{p3}	$Test_1$	$Test_2$ (AIC)	$Test_3$ (BIC)
100	60	1.000	1.000	2.407	1.000	8.000	2.407
200	60	1.000	1.000	1.000	1.000	8.000	1.000
500	60	1.000	1.000	1.000	1.000	5.213	1.000
1000	60	1.000	1.000	1.000	1.000	1.004	1.000
2000	60	1.000	1.000	1.000	1.000	1.000	1.000
100	100	1.000	1.000	3.209	1.000	8.000	3.209
200	100	1.000	1.000	1.000	1.000	8.000	1.000
500	100	1.000	1.000	1.000	1.000	8.000	1.000
1000	100	1.000	1.000	1.000	1.000	1.079	1.000
2000	100	1.000	1.000	1.000	1.000	1.000	1.000
60	100	1.000	1.000	2.284	1.000	8.000	8.000
60	200	1.000	1.000	1.000	1.000	8.000	8.000
60	500	1.000	1.000	1.000	1.000	8.000	8.000
60	1000	1.000	1.000	1.000	1.000	8.000	8.000
60	2000	1.000	1.000	1.000	1.000	8.000	8.000

Table 2: DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}$
$r = 3; \theta = 3.$

N	T	PC_{p1}	PC_{p2}	PC_{p3}	$Test_1$	$Test_2$ (AIC)	$Test_3$ (BIC)
100	60	3.000	3.000	3.543	3.000	8.000	3.543
200	60	3.000	3.000	3.000	3.000	8.000	3.000
500	60	3.000	3.000	3.000	3.000	5.961	3.000
1000	60	3.000	3.000	3.000	3.000	3.000	3.000
2000	60	3.000	3.000	3.000	3.000	3.000	3.000
100	100	3.000	3.000	4.217	3.000	8.000	4.217
200	100	3.000	3.000	3.000	3.000	8.000	3.000
500	100	3.000	3.000	3.000	3.000	8.000	3.000
1000	100	3.000	3.000	3.000	3.000	3.014	3.000
2000	100	3.000	3.000	3.000	3.000	3.000	3.000
60	100	3.000	3.000	3.501	3.000	8.000	8.000
60	200	3.000	3.000	3.000	3.000	8.000	8.000
60	500	3.000	3.000	3.000	3.000	8.000	8.000
60	1000	3.000	3.000	3.000	3.000	8.000	8.000
60	2000	3.000	3.000	3.000	3.000	8.000	8.000

Table 1–Table 6 report the estimated number of factors averaged over 1000 simulations. The true number of factors is r and $kmax = 8$.

Table 3: DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}$
$r = 5; \theta = 5.$

N	T	PC_{p1}	PC_{p2}	PC_{p3}	$Test_1$	$Test_2$ (AIC)	$Test_3$ (BIC)
100	60	5.000	4.998	5.065	4.942	8.000	5.065
200	60	5.000	5.000	5.000	4.987	8.000	5.000
500	60	5.000	5.000	5.000	4.988	6.883	5.000
1000	60	5.000	5.000	5.000	4.975	5.000	5.000
2000	60	5.000	5.000	5.000	4.940	5.000	5.000
100	100	5.000	5.000	5.443	4.999	8.000	5.443
200	100	5.000	5.000	5.000	5.000	8.000	5.000
500	100	5.000	5.000	5.000	5.000	8.000	5.000
1000	100	5.000	5.000	5.000	5.000	5.000	5.000
2000	100	5.000	5.000	5.000	5.000	5.000	5.000
60	100	5.000	4.999	5.055	4.940	8.000	8.000
60	200	5.000	5.000	5.000	4.984	8.000	8.000
60	500	5.000	5.000	5.000	4.995	8.000	8.000
60	1000	5.000	5.000	5.000	4.969	8.000	8.000
60	2000	5.000	5.000	5.000	4.934	8.000	8.000

Table 4: DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}$
$e_{it} = e_{it}^1 + \delta_t e_{it}^2 \quad (\delta_t = 1 \text{ for } t \text{ even}, \delta_t = 0 \text{ for } t \text{ odd})$
$r = 3; \theta = 3.$

N	T	PC_{p1}	PC_{p2}	PC_{p3}	$Test_1$	$Test_2$ (AIC)	$Test_3$ (BIC)
100	60	3.000	3.000	4.931	3.000	8.000	4.931
200	60	3.000	3.000	3.000	3.000	8.000	3.000
500	60	3.000	3.000	3.000	3.000	8.000	3.000
1000	60	3.000	3.000	3.000	3.000	7.929	3.000
2000	60	3.000	3.000	3.000	2.998	4.686	3.000
100	100	3.000	3.000	5.772	3.000	8.000	5.772
200	100	3.000	3.000	3.000	3.000	8.000	3.000
500	100	3.000	3.000	3.000	3.000	8.000	3.000
1000	100	3.000	3.000	3.000	3.000	8.000	3.000
2000	100	3.000	3.000	3.000	3.000	4.110	3.000
60	100	3.000	3.000	4.305	2.999	8.000	8.000
60	200	3.000	3.000	3.000	3.000	8.000	8.000
60	500	3.000	3.000	3.000	3.000	8.000	8.000
60	1000	3.000	3.000	3.000	2.998	8.000	8.000
60	2000	3.000	3.000	3.000	2.999	8.000	8.000

Table 5: DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}$
$r = 5; \theta = r \times 2.$

N	T	PC_{p1}	PC_{p2}	PC_{p3}	$Test_1$	$Test_2$ (AIC)	$Test_3$ (BIC)
100	60	4.781	4.408	5.060	3.500	8.000	5.060
200	60	4.899	4.787	4.996	3.614	8.000	4.996
500	60	4.963	4.947	4.991	3.386	6.874	4.991
1000	60	4.979	4.975	4.991	3.133	5.000	4.991
2000	60	4.985	4.981	4.989	2.651	5.000	4.989
100	100	4.957	4.677	5.430	4.128	8.000	5.430
200	100	5.000	4.994	5.000	4.681	8.000	5.000
500	100	5.000	5.000	5.000	4.900	8.000	5.000
1000	100	5.000	5.000	5.000	4.904	5.000	5.000
2000	100	5.000	5.000	5.000	4.789	5.000	5.000
60	100	4.749	4.394	5.052	3.456	8.000	8.000
60	200	4.908	4.809	4.997	3.575	8.000	8.000
60	500	4.973	4.957	4.999	3.419	8.000	8.000
60	1000	4.969	4.960	4.987	3.073	8.000	8.000
60	2000	4.975	4.974	4.983	2.637	8.000	8.000

Table 6: DGP: $X_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta} e_{it}$
$r = 5; \theta = r/2.$

N	T	PC_{p1}	PC_{p2}	PC_{p3}	$Test_1$	$Test_2$ (AIC)	$Test_3$ (BIC)
100	60	5.000	5.000	5.066	5.000	8.000	5.066
200	60	5.000	5.000	5.000	5.000	8.000	5.000
500	60	5.000	5.000	5.000	5.000	6.877	5.000
1000	60	5.000	5.000	5.000	5.000	5.000	5.000
2000	60	5.000	5.000	5.000	5.000	5.000	5.000
100	100	5.000	5.000	5.444	5.000	8.000	5.444
200	100	5.000	5.000	5.000	5.000	8.000	5.000
500	100	5.000	5.000	5.000	5.000	8.000	5.000
1000	100	5.000	5.000	5.000	5.000	5.000	5.000
2000	100	5.000	5.000	5.000	5.000	5.000	5.000
60	100	5.000	5.000	5.058	5.000	8.000	8.000
60	200	5.000	5.000	5.000	5.000	8.000	8.000
60	500	5.000	5.000	5.000	5.000	8.000	8.000
60	1000	5.000	5.000	5.000	5.000	8.000	8.000
60	2000	5.000	5.000	5.000	5.000	8.000	8.000

References

Backus, D., Forsei, S., Mozumdar, A. and Wu, L. (1997), Predictable Changes in Yields and Forward Rates, mimeo, Stern School of Business.

Chamberlain, G. and Rothschild, M. (1983), Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets, *Econometrica* **51**, 1305–1324.

Cochrane, J. (1999), New Facts in Finance, and Portfolio Advice for a Multifactor World, NBER Working Paper 7170.

Connor, G. and Korajcyk, R. (1986), Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis, *Journal of Financial Economics* **15**, 373–394.

Connor, G. and Korajcyk, R. (1988), Risk and Return in an Equilibrium APT: Application to a New Test Methodology, *Journal of Financial Economics* **21**, 255–289.

Connor, G. and Korajcyk, R. (1993), A Test for the Number of Factors in an Approximate Factor Model, *Journal of Finance* **XLVIII:4**, 1263–1291.

Cragg, J. and Donald, S. (1997), Inferring the Rank of a Matrix, *Journal of Econometrics* **76**, 223–250.

Donald, S. (1997), Inference Concerning the Number of Factors in a Multivariate Nonparameteric Relationship, *Econometrica* **65:1**, 103–132.

Forni, M. and Lippi, M. (1997), *Aggregation and the Microfoundations of Dynamic Macroeconomics*, Oxford University Press, Oxford, U.K.

Forni, M. and Reichlin, L. (1998), Let's Get Real: a Factor-Analytic Approach to Disaggregated Business Cycle Dynamics, *Review of Economic Studies* **65**, 453–473.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (1999), The Generalized Dynamic Factor Model: Identification and Estimation, mimeo, Université Libre de Bruxelles.

Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time Series, *in* D. J. Aigner and A. S. Goldberger (eds), *Latent Variables in Socio Economic Models*, North Holland, Amsterdam.

Geweke, J. and Meese, R. (1981), Eestimating Regression Models of Finite but Unknown Order, *International Economic Review* **23:1**, 55–70.

Ghysels, E. and Ng, S. (1998), A Semi-parameteric Factor Model for Interest Rates and Spreads, *Review of Economics and Statistics* **80:4**, 489–502.

Granger, C. W. (1987), Implications of Aggregation with Common Factors, *Econometric Theory* **3**, 208–222.

Gregory, A., Head, A. and Raynauld, J. (1997), Measuring World Business Cycles, *International Economic Review* **38:3**, 677–701.

Lehmann, B. N. and Modest, D. (1988), The Empirical Foundations of the Arbitrage Pricing Theory, *Journal of Financial Economics* **21**, 213–254.

Lewbel, A. (1991), The Rank of Demand Systems: Theory and Nonparameteric Estimation, *Econometrica* **59**, 711–730.

Mallows, C. L. (1973), Come Comments on C_p, *Technometrics* **15**, 661–675.

Ross, S. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Finance* **13**, 341–360.

Sargent, T. and Sims, C. (1977), Business Cycle Modelling without Pretending to have too much a Priori Economic Theory, *in* C. Sims (ed.), *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

Stock, J. H. and Watson, M. (1989), New Indexes of Coincident and Leading Economic Indications, *in* O. J. Blanchard and S. Fischer (eds), *NBER Macroeconomics Annual 1989*, M. I. T. Press, Cambridge.

Stock, J. H. and Watson, M. W. (1998), Diffusion Indexes, NBER Working Paper 6702.

Stock, J. H. and Watson, M. W. (1999), Forecasting Inflation, *Journal of Monetary Economics, forthcoming.*