ESTIMATION OF APPROXIMATE FACTOR MODELS: IS IT IMPORTANT TO HAVE A LARGE NUMBER OF VARIABLES?

CHRIS HEATON[†] AND VICTOR SOLO[‡]

[†]*Department of Economics, Macquarie University and School of Economics, University of New South Wales, E-mail: cheaton@efs.mq.edu.au*
[‡]*Department of Electrical Engineering and Computer Science, Department of Statistics, University of Michigan, Email: vsolo@umich.edu*

DRAFT: 15th February 2006

ABSTRACT: The use of principal component techniques to estimate approximate factor models with large cross-sectional dimension is now well established. However, recent work by Inklaar, R. J., J. Jacobs and W. Romp (2003) and Boivin, J. and S. Ng (2005) has cast some doubt on the importance of a large cross-sectional dimension for the precision of the estimates. This paper presents some new theory for approximate factor model estimation. Consistency is proved and rates of convergence are derived under conditions that allow for a greater degree of cross-correlation in the model disturbances than previously published results. The rates of convergence depend on the rate at which the cross-sectional correlation of the model disturbances grows as the cross-sectional dimension grows. The consequences for applied economic analysis are discussed.

1. Introduction

In this paper we consider the estimation of factor models of the form

(1) $X = FB' + n$
(2) $y = F\beta + v$

where X is a T×N matrix containing T observations on N variables denoted x_t, y is a T×1 vector of predicted variables denoted y_t, F is a T×k matrix containing T realisations of k unobservable factors denoted f_t, n is a T×N matrix of disturbances denoted n_t, v is a T×1 vector of disturbances denoted v_t, and B is a N×k matrix of factor loadings. It will be assumed throughout that $E(f_t f_t') = I_k$, $E(f_t n_t') = 0$, $E(v_t^2) = \sigma_v^2 < \infty$, $E(f_t v_t) = 0$ and $E(n_t v_t) = 0$. In the case that $\Psi = \frac{1}{T} E(n'n)$ is a diagonal matrix, Equation 1 is referred to as a strict factor model (Chamberlain, G. and M. Rothschild (1983)). If F and n are also serially uncorrelated then Equation 1 is the classical factor analysis model (Lawley, D. N. and A. E. Maxwell (1971), Jöreskog, K. G. (1967)). This model is well understood and has found wide application in a range of empirical disciplines.

In cases in which N is large, classical factor analysis is unsatisfactory for two main reasons. Firstly, the standard asymptotic theory for factor analysis models (Anderson, T. W. and H. Rubin (1956), Gill, R. D. (1977)) assumes that N is fixed and T approaches infinity – in the case that N and T are both large, an asymptotic theory in which N and T jointly approach infinity would be more appropriate. Secondly, the larger N becomes, the more difficult it is likely to be to justify the assumption that $\Psi = \frac{1}{T} E(n'n)$ is diagonal. In cases where N is large, an approximate factor model (Chamberlain, G. and M. Rothschild (1983)), which allows for some degree of correlation between the elements of the disturbance vector (n_t) is likely to be more appropriate.

Estimation theory for the approximate factor model has been provided by Stock, J. H. and M. W. Watson (2002). They show that the principal components of the covariance matrix of X converge in probability to the true factors (up to a sign change), that the OLS estimates of the principal components regressed on y converge in probability to β (up to a sign change) and that forecasts of y based on OLS regressions of principal components of X on y converge in probability to the vector-product of the true factors and β. Bai, J. and S. Ng (2002) derive criteria for the estimation of the number of factors, and prove their consistency. Bai, J. (2003) develops central limit theorems for principal component estimates of factors. Bai, J. and S. Ng (2005) prove central limit theorems for the least squares estimates of principal component regression coefficients and the least squares principal component forecasts. Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000), Forni, M. and M. Lippi (2001) and Forni, M., M. Hallin, M. Lippi and L. Reichlin (2004) consider a dynamic version of Equation 1 in which the common component, FB' is a polynomial filter of a white noise factor. They prove that sample dynamic principal components are consistent estimators of the factors and they derive rates of convergence.

There now exist many applications of this approach to factor estimation and forecasting. The Federal Reserve Bank of Chicago produces an index of national activity constructed from 85 macroeconomic series which is updated monthly and is available from their website (http://www.chicagofed.org/economic_research_and_data/cfnai.cfm). Altissimo, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Lippi, L. Reichlin and G. Veronese (2001) construct a coincident indicator for the Euro area using 246 series. Their index, named EuroCOIN, is published monthly by the CEPR. Forni, M., M. Hallin, M. Lippi and L. Reichlin (2003) use 447 series from the main countries of the Euro area to forecast industrial production and inflation. Bernanke, B. S. and J. Boivin (2003) use data sets containing 78 series and use the unbalanced panel of 215 series used by Stock, J. H. and M. W. Watson (2002) to forecast industrial production and inflation in an analysis of US monetary policy. Gillitzer, C., J. Kearns and A. Richards (2005) construct monthly and quarterly coincident indicators of the Australian business cycle using balanced panels of 29 and 25 series respectively. They compare these to the results obtained from a balanced panels of up to 111 series.

Recently some doubts have been raised about the importance of having a large cross-sectional dimension when using principal components to estimate approximate factor models. Boivin, J. and S. Ng (2005) conduct Monte Carlo simulations for variables with different degrees of disturbance cross-sectional correlation. They find that increasing the cross-sectional dimension does not necessarily improve the precision of the estimator if it comes at the cost of increased cross-sectional correlation in the disturbances. They also compute correlation coefficients for residuals from a factor model estimated using 147 US macroeconomic variables and find that many of them are quite high. They find that as few as 40 carefully chosen variables can produce forecasts of comparable quality to those computed using the full data set of 147 variables. Inklaar, R. J., J. Jacobs and W. Romp (2003) find that they can construct a coincident index for Europe just as well from 38 carefully chosen variables as they can from a data set of 246 variables. These results are not easily understood from the perspective of the published theory on approximate factor model estimation.

The objective of this paper is to develop a better understanding of the circumstances in which principal component techniques are likely to provide estimates of factor quantities of sufficient precision to be of practical interest. To this end we develop theory linking population principal components to population factors. We also prove the convergence of sample principal components to population principal components for dual sequences in N and T. Combining these approaches we develop new consistency theorems for sample principal component quantities as estimates of factor quantities as N and T jointly approach infinity. We discuss the assumptions that have been made in the existing literature about the degree of cross-sectional correlation between the model disturbance terms, and show that our theorem allows for a much higher degree of correlation – albeit at the cost of possibly slow convergence. In the final section of the paper we discuss the implications of our findings for applied work.

2. Theory

The theory of principal component estimation of approximate factor model quantities will be developed in four parts. In the first part, the relationship between population factors and population principal components will be explored. It will be shown that the 'closeness' of principal components to factors depends on the magnitude of a certain noise to signal ratio. In the second part, the relationship between factor structure, the number of variables and the noise to signal ratio is explored. In the third part, theorems are developed which show that sample principal component quantities are consistent estimates of their population counterparts in a large-N, large-T setting. In the fourth part, the theoretical results are combined to provide consistency results for sample principal component quantities as estimates of their population factor counterparts. Finally, we define a concept of strong cross-sectional correlation for the disturbance terms and demonstrate that our results allow for much stronger growth in the number of strongly correlated disturbances than previously published results.

2.1 Population Principal Components and Population Factors

The idea that principal components might be closely related to factors is not new. It has often been noted in the psychology literature that principal component techniques and factor analysis sometimes give similar results, especially when the number of variables in the analysis is large[1]. However, it is only relatively recently that the nature of the relationship between principal components and factors has been formally investigated. Chamberlain, G. and M. Rothschild (1983) define an approximate factor model as one for which the disturbance eigenvalues are uniformly bounded and show that the eigenvectors of the population covariance matrix of x_t are asymptotically equivalent to the factor loadings as N gets large. Bentler, P. M. and Y. Kano (1990) consider a single factor model and show that as $N \to \infty$ the correlation between the population principal component and the population factor converges to one and the principal component loading vector converges to the factor loading vector. In a significant paper Schneeweiss, H. and H. Mathes (1995) develop measures of distance of population principal components and factors and show that these measures converge to zero as the smallest eigenvalue of $B'B$ gets large relative to the largest eigenvalue of Ψ. Schneeweiss, H. (1997) produces similar results and also shows that population principal components converge to population factors when the ratio of the largest eigenvalue of Ψ to the smallest eigenvalue of Ψ approaches one. The remainder of this section presents some new theory linking population principal component quantities to their analogous population factor quantities. Unlike the results mentioned above, the new results are not based on limit theorems.

We will assume throughout that the predicted variable y_t and the predictor variables x_t have finite, time-invariant second moments. We consider the factor representation given

---

[1] See SNOOK, S. C., and R. L. GORSUCH (1989): "Component Analysis Versus Common Factor Analysis: A Monte Carlo Study," *Psychological Bulletin*, 106, 148-154. for a survey.

by equations 1 and 2 and the moment conditions $E(f_t f_t') = I_k$, $E(f_t n_t') = 0$, $E(v_t^2) = \sigma_v^2 < \infty$, $E(f_t v_t) = 0$, $E(n_t v_t) = 0$ and $E(n_t n_t') = \Psi$ a positive definite finite matrix. We denote the covariance of x_t as $E(x_t x_t') = \Omega = BB' + \Psi$. Let $D = \mathrm{diag}(d_1,...,d_k)$ be the k×k diagonal matrix of ordered eigenvalues of BB' and U be the corresponding N×k matrix of eigenvectors. Let $\Lambda_1 = \mathrm{diag}(\lambda_1 ... \lambda_k)$ be the k×k diagonal matrix of the first k ordered eigenvalues of Ω and Q_1 be the corresponding N×k matrix of eigenvectors. Λ_2 and Q_2 contain the remaining eigenvalues and eigenvectors. We therefore have the spectral decompositions

$$BB' = UDU' \text{ and } \Omega = Q_1 \Lambda_1 Q_1' + Q_2 \Lambda_2 Q_2'.$$

We define σ^2 to be the maximum eigenvalue of Ψ. We then define the noise to signal ratio to be

$$\rho = \frac{\sigma^2}{\lambda_k}$$

In what follows L will be a diagonal matrix of ± 1's. That is, $\mathrm{abs}(L) = I$. We will consider the estimate of the factor vector given by the population principal component vector

$$s_t = \Lambda_1^{-\frac{1}{2}} Q_1' x_t,$$

We will also consider the estimate of the mean square efficient factor forecast $\beta' f_{T+h}$ given by the population principal component forecast

$$\hat{y}_{T+h} = \tilde{\beta}_S' s_{T+h}$$

where

$$\tilde{\beta}_S = \frac{1}{T} \sum_{t=1}^{T} s_t y_t$$

We define $r^2 = \dfrac{\|\beta\|^2}{\|\beta\|^2 + \sigma_v^2}$. Note that r^2 is the proportion of the variance of y_t that is explained by the factors. Thus, it may be interpreted as the population analogue of the R^2 statistic from regression analysis. We also denote

$$\gamma = \sum_{j=1}^{\infty}\left|\frac{E(y_t y_{t-j})}{E(y_t^2)}\right| + \sup_i \sum_{j=1}^{\infty}\left|\frac{E(y_t s_{i,t-j})}{\sqrt{E(y_t^2)E(s_{it}^2)}}\right|$$

and

$$c = \max_{i,j:i\neq j} \frac{\lambda_j}{|\lambda_i - \lambda_j|} \text{ for } k > 1 \text{ and } c = 0 \text{ for } k = 1,$$

and we define the forecast deviation as $e_{T+h} = \hat{y}_{T+h} - \beta' f_{T+h}$.

Theorem 1

(a) For the factor model described above $1-\rho \le \frac{d_i}{\lambda_i} \le 1$ for i=1,..,k.

(b) For the factor model described above, if $k = 1$ or $c \le \frac{1}{2\rho}\sqrt{\frac{1-\rho}{k-1}}$, then there exists a sign matrix L such that $\|Q_1 - UL\|^2 \le k\left(\rho + 4c^2\rho^2(k-1)\right)$

(c) For the factor model described above, if $k = 1$ or $c \le \frac{1-\rho}{2\rho\sqrt{(k-1)(1-\rho)}}$,
then there exists a sign matrix L such that
$E\|s_t - Lf_t\|^2 \le k\left(2\rho + \rho^2\left(4c^2(k-1)(1-\rho) - 1\right)\right)$

(d) For the factor model described above, if f_t, v_t and n_t are Gaussian and $\gamma < \infty$, then $\frac{E|e_{T+h}|}{\sqrt{\beta'\beta}} \le \sqrt{k\left(\rho^2 + \frac{2k\gamma}{r^2 T}\right)} + \sqrt{k\rho\left(1 + \frac{2\gamma}{r^2 T}\right)}$

Proofs: See Appendix

The following asymptotic results follow trivially from these theorems:

Corollaries to Theorem 1

a) $\frac{d_i}{\lambda_i} \to 1$ as $\rho \to 0$;
b) There exists a sign matrix L such that $Q_1 \to UL$ as $\rho \to 0$;
c) There exists a sign matrix L such that $s_t \xrightarrow{p} Lf_t$ as $\rho \to 0$;

d) $e_{T+h} = \hat{y}_{T+h} - \beta' f_{T+h} \xrightarrow{p} 0$ as $\left(\frac{1}{T}, \rho\right) \to (0,0)$.

It follows from a) and b) that

e) There exists a sign matrix L such that $Q_1 \Lambda_1^{\frac{1}{2}} \to BL$ as $\rho \to 0$.

Corollaries a), b) and c) were previously proved by Schneeweiss, H. (1997). Results d) and e) are new. Importantly, Theorem 1 is new and provides rates of convergence which are necessary for subsequent theorems.

Note that Theorems 1(b) and 1(c) require the eigenvalues of Ω to be distinct so that c is bounded. The distance between the relevant quantities in these theorems depends on the closeness of adjacent eigenvalues and on the noise to signal ratio. Theorem 1(d) assumes Gaussianity. It is quite likely that this assumption could be replaced by the assumption of an upper bound on sums of fourth moments, however Gaussianity produces a result which is more easily interpretable. In any case, in the asymptotic arguments that follow, the assumption of Gaussianity will not be needed. In order for the principal component forecast and the theoretically optimal forecast to be close, we need the noise to signal ratio to be fairly small and the sample size to be reasonably large. Precisely how large the sample size needs to be will depend on the magnitude of the autocovariances of the data and the proportion of the variance of the forecast variable that is determined by the factors.

2.2 N and the Noise to Signal Ratio

Consider the factor model given by equation 1. The covariance matrix is $E(x_t x_t') = \Omega = BB' + \Psi$. As above, we have the spectral decomposition $BB' = UDU'$, and we denote the maximum eigenvalue of Ψ by σ^2. The following simple result links the noise to signal ratio to the number of variables in the model.

Theorem 2

For the factor model described above, if

1. $\exists d_L \in \mathbb{R}^+ \ni N d_L < d_j$, j=1,..,k;
2. $\sigma^2 \sim O\left(N^{1-\alpha}\right)$ where $0 < \alpha \le 1$;

then $\rho \sim O\left(N^{-\alpha}\right)$. Furthermore $\tilde{\rho} = \frac{\sigma^2}{d_k} \sim O\left(N^{-\alpha}\right)$.

Proof: see Appendix.

Note that it is a consequence of the assumptions that $\operatorname{tr}(\Omega) \sim O(N)$. This is consistent with the variables being bounded in probability. Furthermore $\operatorname{tr}(BB') \sim O(N)$ and $\operatorname{tr}(\Psi) \sim O(N)$ so that neither the common component nor the disturbances dominate asymptotically.

Given this result, in cases where factor structure is believed to exist, it seems reasonable to believe that the noise to signal ratio is decreasing in N. However, the rate at which the noise to signal ratio shrinks to zero could be very slow. Importantly, it should be noted that there is no particular value of N which guarantees a small noise to signal ratio. Consequently there is no number of variables that can generally be considered to be "large" when estimating approximate factor models. In the case of a strict factor model the disturbance terms are uncorrelated so the maximum eigenvalue of the disturbance covariance is simply the largest disturbance variance. As such, it is reasonable to assume that it is bounded and that the noise to signal ratio will decline quite rapidly as N grows. In the case of the approximate factor model however, the maximum eigenvalue may be much larger than the largest disturbance variance and may well grow as N increases. In such cases, the shrinkage of the noise to signal ratio as N grows could be quite slow.

Boivin, J. and S. Ng (2005) and Inklaar, R. J., J. Jacobs and W. Romp (2003) found that in practice raising N above a certain level produced little improvement in the quality of the estimator. The simulation study of Boivin, J. and S. Ng (2005) suggested that this phenomenon was related to increasing cross-correlation as N gets large. They describe a situation in which an analyst increases N by adding to the model relatively noisy variables and variables which are closely related to variables already in the model. They state that the consequence of this is that the relative size of the common component may be reduced, reducing the precision of the estimator, as N increases. The above theory characterizes this phenomenon more precisely. What matters when using principal component quantities to estimate their factor analysis counterparts is not the number of variables per se, but rather the noise to signal ratio. While the existence of factor structure is a good reason to expect the noise to signal ratio to decline as the number of variables increases, there is no value for N which guarantees a small noise to signal ratio.

2.3 Sample Principal Components and Population Principal Components

It is well known that when N is fixed, sample eigenvalues are consistent estimates of population eigenvalues and sample eigenvectors and principal components are consistent estimates of orthogonal transformations of their population counterparts. However, it is not known whether these results apply in a setting in which N is of the same order of magnitude as T. Below, a new asymptotic theory of principal components in a large-N, large-T framework is developed. It is shown that the scaled sample eigenvalues converge in probability to the scaled population eigenvalues. Results about the behaviour of sample eigenvectors and principal components then follow.

Let X be a T×N matrix containing T observations on N variables. The rows of X are denoted x_t'. Assume that

1. $E(X)=0$;
2. $\frac{1}{T}E(X'X)=\Omega<\infty$ a full-rank matrix;
3. $\sum_{j=0}^{N} ABS\left[\operatorname{cov}\left\{\operatorname{vec}(x_t x_t'), \operatorname{vec}(x_{t-j}x_{t-j}')\right\}\right]<\infty$

Let $\Omega=Q\Lambda Q'$ where Λ is the diagonal matrix of eigenvalues in descending order and Q the corresponding matrix of eigenvectors. The sample eigenvalues and eigenvectors are similarly defined as $\frac{1}{T}X'X=\hat{Q}\hat{\Lambda}\hat{Q}'$. The population principal components of X are defined as $S=\Lambda^{-\frac{1}{2}}Q'X$ which consists of rows f_t'. The sample principal components of X are defined as $\hat{S}=\hat{\Lambda}^{-\frac{1}{2}}\hat{Q}'X$ which consists of rows $\hat{f}_t'$.

<u>Theorem 3</u>

(a) Under assumptions 1 to 3, $\frac{1}{N}\hat{\lambda}_j=\frac{1}{N}\lambda_j+O_p\left(T^{-\frac{1}{2}}\right)$ where $\hat{\lambda}_j$ is the j^{th} eigenvalue of $\frac{1}{T}X'X$.

(b) Under assumptions 1 to 3, $\hat{s}_{jt}=\hat{\lambda}^{-\frac{1}{2}}\hat{q}'x_t=L_{jj}s_{jt}+O_p\left(T^{-\frac{1}{2}}\right)$, j=1,..,k where k is a fixed value and $L_{jj}=\pm 1$

<u>Proofs:</u> See Appendix

Note that the assumptions of Theorem 2 are reasonably general. The zero-mean assumption may be satisfied by simply subtracting sample means from the variables. The existence of finite second moments is required simply so that the principal components concept makes sense. The assumption of finite sums of fourth moments implies that the sample second moments converge in probability to their population counterparts. Importantly, there is no requirement that the noise to signal ratio be small or that there be growing gaps between eigenvalues. Note that convergence is at a rate independent of N. What the theorem shows is that $\sqrt{T}$-consistency holds in the presence of growing N. It does not suggest that increasing N improves the estimate.

2.4 Sample Principal Components and Population Factors

The previous three subsections provide us with a framework for constructing a theory linking sample principal components to population factors. We have shown that population principal components are close to population factors when the noise to signal ratio is small, that under certain conditions the noise to signal ratio will be small when N is large, and that sample principal components are close to population principal components when T is large. Combining these ideas, the following results are proved. For the factor model

$$\begin{aligned} X &= FB' + n \\ y &= F\beta + v \end{aligned}$$

we denote $D = \text{diag}(d_1, ..., d_k)$ to be the eigenvalues of BB', $E(n_t n_t') = \Psi$ and $\sigma^2 = \max \text{eig}\left(\Psi\right)$. We assume the following:

Assumptions

1. $\exists d_U, d_L \in \mathbb{R}^+ \ni Nd_L < d_j < Nd_U$ for j=1,..,k. Furthermore, $d_1 > d_2 > ... > d_k$;
2. $\frac{1}{N}\text{tr}\left(\Psi\right) < \infty$;
3. $\sigma^2 \sim O\left(N^{1-\alpha}\right), 0 < \alpha \leq 1$;
4. $E\left(F\right) = 0$, $E\left(n\right) = 0$, $E\left(v\right) = 0$;
5. $\frac{1}{T}E\left(F'F\right) = I_k$, $\frac{1}{T}E\left(v'v\right) = \sigma_v^2 < \infty$, $\frac{1}{T}E\left(F'n\right) = 0$, $\frac{1}{T}E\left(F'v\right) = 0$;
6. $\sum_{j=0}^{N} \text{ABS}\left[\text{cov}\left\{\text{vec}\left(z_t z_t'\right), \text{vec}\left(z_{t-j} z_{t-j}'\right)\right\}\right] < \infty$ where $z_t = \left(n_t' \quad f_t' \quad v_t\right)'$;
7. $\beta \sim O\left(1\right)$.

Theorem 4

Under the above assumptions

(a) $\frac{1}{N}\hat{\lambda}_j = \frac{1}{N}d_j + O_p\left[\max\left(T^{-\frac{1}{2}}, N^{-\alpha}\right)\right]$ for j=1,..,k;

(b) $\frac{1}{T}X'\hat{S}_1 = BL + O_p\left[\max\left(T^{-\frac{1}{2}}, N^{\frac{-\alpha}{2}}\right)\right]$ where $\hat{S}_1$ is a T×k matrix containing the first k columns of $\hat{S}$.

(c) $\hat{s}_t = Lf_t + O_p\left[\max\left(T^{-\frac{1}{2}}, N^{\frac{-\alpha}{2}}\right)\right]$;

(d) $\hat{\beta}_S = L\beta + O_p\left[\max\left(T^{-\frac{1}{2}}, N^{\frac{-\alpha}{2}}\right)\right]$ where $\hat{\beta}_S = \frac{1}{T}\hat{S}_1' y$;

(e) $\hat{\beta}_S' \hat{s}_{T+h} = \beta' f_{T+h} + O_p\left[\max\left(T^{-\frac{1}{2}}, N^{\frac{-\alpha}{2}}\right)\right]$.

2.5 Strong Disturbance cross-correlation

It is of particular interest to compare the assumptions that we make about cross-correlation of the disturbances (n) to those made in the previously published literature. Chamberlain, G. and M. Rothschild (1983) define an approximate factor model as one for which the eigenvalues of the disturbance covariance matrix (Ψ) are bounded. Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000) make an analogous assumption in a dynamic setting. Since we allow for the maximum eigenvalue to grow at a rate of $N^{1-\alpha}$ where $0 < \alpha \leq 1$, our theorem clearly allows for a greater degree of disturbance cross-correlation. Stock, J. H. and M. W. Watson (2002), Bai, J. and S. Ng (2002), and Bai, J. (2003) assume that $\frac{1}{N}\sum_{i=1}^{N} R_{iN}$ is bounded where $R_{iN} = \sum_{j=1}^{N}\left|\Psi_{ij}\right|$ is the i^{th} absolute row sum of Ψ. Since $\max \operatorname{eig}(\Psi) \leq \max_i R_{iN}$ our conditions are satisfied if $\max_i R_{iN} \sim O\left(N^{1-\alpha}\right)$. Thus, while Stock, J. H. and M. W. Watson (2002), Bai, J. and S. Ng (2002), and Bai, J. (2003) require that the average absolute row sum is bounded, our theorem allows it to grow in N, provided that the maximum row sum grows at a rate strictly less than N.

It is easier to appreciate the differences in the assumptions made about disturbance cross-correlation if we consider the disturbance vector in the following way. We assume that $\exists \bar{\psi} < \infty \ni \left|\Psi_{ij}\right| < \bar{\psi}\ i = 1,..,N,\ j = 1,..N$ and divide the elements of the disturbance vector n into two mutually exclusive and exhaustive sets: a set of strongly correlated disturbances and a set of weakly correlated disturbances. Let $\Phi = \left\{j : n_j \text{ is weakly correlated}\right\}$. We then define a disturbance n_i to be weakly correlated if $\sum_{j\in\Phi}\left|\Psi_{ij}\right| < c < \infty$ where c is a bound that applies uniformly to all weakly correlated disturbances. It follows that if n_i is strongly correlated then $R_{iN} \leq M_N\bar{\Psi} + c$ where M_N is the number of strongly correlated disturbances in the model. If n_i is weakly correlated then $R_{iN} \leq c$. We now consider restrictions on M_N which will satisfy the different assumptions about Ψ that have been made in the literature. Since $\max \operatorname{eig}(\Psi) \leq \max_i R_{iN}$ the approximate factor model of Chamberlain, G. and M. Rothschild (1983) and Connor, G. and R. A. Korajczyk (1986) can be produced by assuming that the number of strongly correlated disturbances in the model (M_N) is bounded as N grows. Stock, J. H. and M. W. Watson (2002), Bai, J. and S. Ng (2002), and Bai, J. (2003) require that $\frac{1}{N}\sum_{i=1}^{N} R_{iN}$ is bounded. From the above discussion $\frac{1}{N}\sum_{i=1}^{N} R_{iN} \le \frac{M_N}{N}\left(M_N\overline{\psi}+c\right)+\frac{c}{N}\left(N-M_N\right)=\frac{M_N^2}{N}\overline{\psi}+c$. Therefore, the assumption required by these theorems can be achieved by allowing the number of strongly correlated variables (M_N) to grow at a rate as high as $\sqrt{N}$. This is interesting since it allows for a much higher degree of disturbance cross-correlation than the original approximate factor model proposed by Chamberlain, G. and M. Rothschild (1983), a point which does not appear to have been made by the above authors.

Assumption 3 of Theorem 4 in the present paper allows the maximum eigenvalue of the disturbance covariance to grow at a rate of $N^{1-\alpha}$ where $0 < \alpha \le 1$. Since $\max \mathrm{eig}\left(\Psi\right) \le \max_i R_{iN}$, this can be achieved by setting $M_N \sim O\left(N^{1-\alpha}\right)$. Thus, the theory presented in this paper allows the number of strongly correlated disturbances to grow at a rate strictly less than N. This is clearly much greater than the rate of $\sqrt{N}$ imposed to satisfy the assumptions of Stock, J. H. and M. W. Watson (2002), Bai, J. and S. Ng (2002), and Bai, J. (2003), and is far more general than the original approximate factor model of Chamberlain, G. and M. Rothschild (1983). It should be noted however, that while the above theorem applies to models with far more disturbance cross-correlation than is allowed in the previously published literature, the higher is the growth rate of the number of strongly correlated variables, the slower is the rate of convergence of the estimate. This is an important consideration for applied work and provides a possible theoretical explanation of the empirical results of Inklaar, R. J., J. Jacobs and W. Romp (2003) and Boivin, J. and S. Ng (2005).

3. Discussion and Conclusions

The title of this paper poses the question "Is it important to have a large number of variables?" Not surprisingly, the answer to this question is not a simple yes or no. Rather, it depends very much on the application at hand.

Consideration should be given to the importance of the interpretation of sample principal components as estimates of factors. As was shown in the previous section, the (T,N)-asymptotics in which sample principal components converge to population factors can be considered as a combination of two mechanisms which operate more or less independently of each other. Firstly, under certain restrictions on the growth rate of the maximum eigenvalue of the disturbance covariance, as N gets large population principal components converge to population factors. Since this is a property of population quantities, its operation is independent of the number of observations that an analyst might have available. Secondly, as T gets large sample principal components converge to population principal components. This property holds for all sequences in (T,N), but it does not require N to be necessarily large or even growing; nor does it require factor structure. Accordingly, even in cases in which the disturbances are highly cross-correlated and doubts are held as to whether N is sufficiently large for population principal components to be close to population factors, provided T is sufficiently large, the sample principal components can still be interpreted as estimates of the population principal components. In applications for which such an interpretation is meaningful, one need not worry about whether N is large enough.

As a possible example, consider the Chicago Fed National Activity Index (CFNAI). This is constructed as the first principal component of 85 monthly indicators of national economic activity[2], and is interpreted by the Chicago Federal Reserve as an estimate of a latent factor. However, instead of specifying and estimating a factor model of the economy, a more traditional (and perhaps more frequently used) approach to constructing a composite indicator is simply to take a weighted mean of a collection of indicator variables. In practice, the weights are often chosen in an ad-hoc manner, however one particular choice of weights which might be considered would be those that minimise the distance of the weighted composite indicator from the indicator variables, i.e. to choose the composite variable ι_t and the $N\times 1$ vector of weights w to minimize $E\left\|x_t - \iota_t w\right\|^2$ where x_t is the $N\times 1$ vector of indicator variables. This choice of composite indicator is, of course, the first principal component of the vector of indicator variables. If such a choice of weights is regarded as acceptable, then the CFNAI might be regarded as a useful composite indicator of economic activity, even if N is not sufficiently large to overcome the disturbance cross-correlation and provide a reasonable estimate of a single factor model.

For situations in which the interpretation of the sample principal components as estimates of factors is critical, the question of whether N is sufficiently large is more important. In particular, in forecasting applications, unless the population principal components are close to the population factors, there is no good reason to believe that the correlation between the predictor variables and the variable being forecast is explained by the first few principal components. However, as shown in the previous section, the magnitude of N is not particularly informative unless it is accompanied by knowledge of the rate at which the maximum eigenvalue of the disturbance covariance matrix grows with N. Accordingly, there is no particular value of N which can be considered "large." In some applications – particularly those in which the disturbance covariance matrix is diagonal – a couple of dozen variables might be sufficient. In other applications, for which disturbances are highly cross-correlated, thousands of variables might be insufficient for principal components to be good estimates of factors. More to the point, as shown empirically by Inklaar, R. J., J. Jacobs and W. Romp (2003) and Boivin, J. and S. Ng (2005), in some applications, increasing the number of variables in the analysis may not be wise if the new variables introduce increased cross-correlation in the disturbances.

Unfortunately, apart from sounding a note of caution, this provides the applied economist with little in the way of concrete guidance. While it is important to have a large number of variables if principal components techniques are to be used to estimate approximate factor models, it is not generally possible to say how many variables might be considered enough. Perhaps the best advice that can be given to practitioners is to consider the bound

---

[2] see http://www.chicagofed.org/economic_research_and_data/files/cfnai_technical_report.pdf for a list of variables.

provided in Theorem 1(c). Provided that the first k eigenvalues of the covariance of x_t are not too close to each other, $E\|s_t - Lf_t\|^2$ will have an upper bound close to $2k\rho$ where k is the number of factors and ρ is the ratio of the largest eigenvalue of the disturbance covariance to the k^{th} eigenvalue of the covariance of the observable variables, x_t. Direct estimation of ρ is only possible in special cases, for example in the case where the disturbance covariance is diagonal. However, it is possible to consistently estimate a lower bound on ρ. It is shown in the proof to Theorem 1(a) that $d_j + \sigma^2 \geq \lambda_j \ \forall j = 1,..,N$. Since $d_{k+1} = 0$, $\lambda_{k+1} \leq \sigma^2$. We therefore have that $\frac{\lambda_{k+1}}{\lambda_k} \leq \rho$ where λ_j is the j^{th} eigenvalue of the covariance matrix of x_t. This expression makes it clear that in order for the noise to signal ratio to be small, implying that principal components are close to factors, there must exist a large gap between the k^{th} and $(k+1)^{th}$ eigenvalues of the covariance matrix of the observable variables. Rather than asking whether N is large enough in any particular application, it might make more sense for analysts to examine ratios of sample eigenvalues and to ensure that the k^{th} ratio is sufficiently small.

Appendix – Proofs of Theorems

We define $R_1 = U'Q_1$ and $R_2 = Q'\hat{Q}$ and denote the largest eigenvalue of a matrix M by $\lambda_{\max}(M)$. We also define the following notation:

S = the T×N matrix of population principal components;
$\hat{S}$ = the T×N matrix of sample principal components;
$\hat{\lambda}_j$ = the j^{th} diagonal element of $\hat{\Lambda}$;
$\hat{q}_j$ = the j^{th} column of $\hat{Q}$;
q_j = the j^{th} column of Q;
s_t' = the t^{th} row of S
s_j = the j^{th} column of S;
s_{jt} = the j,t^{th} column of S;
$\tilde{\Lambda}_j$ = Λ with the j^{th} row and j^{th} column removed;
$\tilde{Q}_j$ = Q with the j^{th} column removed;
$\tilde{S}_j$ = S with the j^{th} column removed;
$\tilde{s}_{jt}'$ = the t^{th} row of $\tilde{S}_j$;
$\tilde{\rho} = \frac{\sigma^2}{d_k}$;

We may then write $X = s_j \lambda_j^{\frac{1}{2}} q_j' + \tilde{S}_j \tilde{\Lambda}_j^{\frac{1}{2}} \tilde{Q}_j'$ and $x_t = q_j \lambda_j^{\frac{1}{2}} s_{jt} + \tilde{Q}_j \tilde{\Lambda}_j^{\frac{1}{2}} \tilde{s}_{jt}$.

The proofs of the theorems will make use of the following lemmas.

Lemma 1: If $x \sim N(0,\Gamma)$ and α and β are vectors of conformable dimension then $E(\alpha' x)^2(\beta' x)^2 = \alpha'\Gamma\alpha\beta'\Gamma\beta + 2(\alpha'\Gamma\beta)^2$. This is a standard property of Gaussian distributions. See, e.g. Johnson, N. L. and S. Kotz (1972).

Corollary of Lemma 1: $\text{var}(\alpha' x\beta' x) = \alpha'\Gamma\alpha\beta'\Gamma\beta + (\alpha'\Gamma\beta)^2$. The proof is elementary.

Lemma 2: If $z_t = \begin{pmatrix} w_t \\ u_t \end{pmatrix}$ is Gaussian and $E(z_t z_{t-j}') = \Gamma^{(j)} = \begin{pmatrix} \Gamma_w^{(j)} & \Gamma_{wu}^{(j)} \\ \Gamma_{uw}^{(j)} & \Gamma_u^{(j)} \end{pmatrix}$, then using Lemma 1, and the Cauchy-Schwarz inequality,

$$\text{var}(a' S_{wu} b) = \text{var}\left(\frac{1}{T}\sum_{t=1}^{T} a' w_t b' u_t\right) \le \frac{2}{T}\left(a'\Gamma_w^{(0)} a b'\Gamma_u^{(0)} b + \sum_{j=1}^{T-1} a'\Gamma_w^{(j)} a b'\Gamma_u^{(j)} b + \sum_{j=1}^{T-1} a'\Gamma_{wu}^{(j)} a b'\Gamma_{wu}^{(-j)} b \right)$$

where a and b are vectors of conformable dimension.

Corollary of Lemma 2: $E(\alpha' u)^2(\beta' v)^2 = \alpha'\Gamma_u\alpha\beta'\Gamma_v\beta + 2(\alpha'\Gamma_{uv}\beta)^2 \le 3\alpha'\Gamma_u\alpha\beta'\Gamma_v\beta$

Lemma 3: If Z is a random vector and e_i is a $k\times 1$ vector of zeros but with a 1 in position i, and M is a $k\times k$ constant, then

$$E(Z'M'MZ) = E(Z'M'\sum_{i=1}^{k} e_i e_i' MZ) = \sum_{i=1}^{k} E(Z'M'e_i)(e_i' MZ) = \sum_{i=1}^{k} E(e_i' MZ)^2$$

Lemma 4: $\Lambda = Q_1'(UDU' + \Psi)Q_1$

Proof of Lemma 4: $\Omega Q_1 = Q_1\Lambda_1 \Rightarrow (UDU' + \Psi)Q_1 = Q_1\Lambda_1$. Premultiplying by Q_1' gives the result. #

Lemma 5: If $M = D^{\frac{1}{2}} R_1 \Lambda_1^{-1} R' D^{\frac{1}{2}}$, then the eigenvalues of $I - M$ are equal to the eigenvalues of $\Lambda_1^{-\frac{1}{2}} Q_1' \Psi Q_1 \Lambda_1^{-\frac{1}{2}}$.

Proof of Lemma 5: The eigenvalues of $I - M$ are the solutions of

$$0 = |\lambda I - (I - M)| = |(\lambda - 1)I + M| = \left|(\lambda - 1)I + D^{\frac{1}{2}} R_1 \Lambda_1^{-1} R_1' D^{\frac{1}{2}}\right|$$

$$= |(\lambda - 1)I + R_1\Lambda_1^{-1} R_1' D| = |(\lambda - 1)I + I - Q_1'\Psi Q_1 \Lambda_1^{-1}| \text{ from Lemma 4}$$

$$= \left|\lambda I - Q_1'\Psi Q_1 \Lambda_1^{-1}\right| = \left|\lambda I - \Lambda_1^{-\frac{1}{2}} Q_1'\Psi Q_1 \Lambda_1^{-\frac{1}{2}}\right|. \quad \#$$

Lemma 6: The eigenvalues of $D^{\frac{1}{2}} R_1 \Lambda_1^{-2} R_1' D^{\frac{1}{2}}$ are equal to the eigenvalues of $\Lambda^{-1} R_1' D R_1 \Lambda^{-1}$.

Proof of Lemma 6: The eigenvalues of $D^{\frac{1}{2}} R_1 \Lambda^{-2} R_1' D^{\frac{1}{2}}$ are the solutions for λ of

$$0 = \left|\lambda I - D^{\frac{1}{2}} R_1 \Lambda_1^{-2} R_1' D^{\frac{1}{2}}\right| = \left|\lambda I - R_1 \Lambda_1^{-2} R_1' D\right| = \left|\lambda I - \Lambda_1^{-1} R_1' D R_1 \Lambda_1^{-1}\right| \quad \#$$

Lemma 7: $\left|q_i{}' u_j\right| \le 2c\rho$ for $i \ne j$ where q_i is the i^{th} column of Q_1 and u_j is the j^{th} column of U, and $c = \max_{i,j:i\ne j} \frac{\lambda_j}{\left|\lambda_i - \lambda_j\right|}$.

Proof of Lemma 7: $Q_1\Lambda_1 Q_1' + Q_2 \Lambda_2 Q_2' = UDU' + \Psi$. Premultiplying by Q_1', postmultiplying by $U\Lambda_1^{-1}$, and subtracting $Q_1'U$ yields

$$\Lambda_1 Q_1' U \Lambda_1^{-1} - Q_1' U = Q_1'\Psi U \Lambda_1^{-1} - Q_1' U (I - D\Lambda_1^{-1}) \tag{A1}$$

We now consider each of the right hand side terms. Let e_i be a vector of zeros with a 1 in the i^{th} element only. We have

$$\left(e_i{}' Q_1 \Psi U \Lambda_1^{-1} e_j\right)^2 \le \mathrm{tr}\left(U'\Psi Q_1 e_i e_i{}' Q_1' \Psi U \Lambda_1^{-2}\right) \le \frac{1}{\lambda_k^2} \mathrm{tr}\left(e_i{}' Q_1' \Psi U U' \Psi Q_1 e_i\right)$$
$$\le \frac{1}{\lambda_k^2} e_i{}' Q_1' \Psi^2 Q_1 e_i \le \frac{\sigma^4}{\lambda_k^2}$$

$$\therefore \quad \left|e_i{}' Q_1 \Psi U \Lambda_1^{-1} e_j\right| \le \rho \tag{A2}$$

For the other right hand side term we have
$\left|e_i' Q_1' U (I - D\Lambda_1^{-1}) e_j\right| = \left|q_i' u_j \left(1 - \frac{d_j}{\lambda_j}\right)\right| = \left|q_i' u_j\right| \left(1 - \frac{d_j}{\lambda_j}\right)$ since $\lambda_j \ge d_j$. But $1 - \frac{d_j}{\lambda_j} \le \rho$ from Theorem 1 so

$\left|q_i' u_j\right| \left(1 - \frac{d_j}{\lambda_j}\right) \le \rho \left|q_i' u_j\right|$ and $\left|q_i' u_j\right| \le 1$ by the Cauchy-Schwarz inequality, so

$$\left|e_i'Q_1'U(I-D\Lambda_1^{-1})e_j\right| \le \rho \qquad \text{(A3)}$$

Combining equations A1, A2, and A3,

$$\left|e_i'\left(\Lambda_1 Q_1'U\Lambda_1^{-1} - Q_1'U\right)e_j\right| = \left|e_i'Q_1'\Psi U\Lambda_1^{-1}e_j - e_i'Q_1'U(I-D\Lambda_1^{-1})e_j\right|$$
$$\le \left|e_i'Q_1'\Psi U\Lambda_1^{-1}e_j\right| + \left|e_i'Q_1'U(I-D\Lambda_1^{-1})e_j\right| \le 2\rho$$

i.e. $\left|\left(\frac{\lambda_i}{\lambda_j}-1\right)q_i'u_j\right| \le 2\rho \quad \Rightarrow \quad \left|q_i'u_j\right| \le 2c\rho$ for $i \ne j$. #

Lemma 8: $1-\rho \le \sum_{j=1}^{k}\left(q_i'u_j\right)^2$

Proof of Lemma 8: $Q_1\Lambda_1Q_1' + Q_2\Lambda_2Q_2' = UDU' + \Psi$ **.** Premultiply by $Q_1'UU'$, postmultiply by Q_1, and substitute $R_1 = U'Q_1$ to get $R_1'R_1\Lambda_1 = R_1'DR_1 + R_1'U'\Psi Q_1$
(A4)

Also $\Lambda = Q_1'\Omega Q_1 \Rightarrow \Lambda = Q_1'(UDU'+\Psi)Q_1 \Rightarrow \Lambda = R_1'DR_1 + Q_1'\Psi Q_1 \qquad$ (A5)

Subtract A4 from A5 and postmultiply by Λ_1^{-1} to get

$$R_1'R_1 - I = R_1'U'\Psi Q_1\Lambda^{-1} - Q_1'\Psi Q_1\Lambda_1^{-1} = Q_1'UU'\Psi Q_1\Lambda_1^{-1} - Q_1'\Psi Q_1\Lambda_1^{-1} = Q_1'(UU'-I)\Psi Q_1\Lambda_1^{-1}$$
$$= -Q_1'U_\perp U_\perp'\Psi Q_1\Lambda_1^{-1}$$

so $e_i'\left(I - R_1'R_1\right)e_i = e_i'\left(Q_1'U_\perp U_\perp'\Psi Q_1\Lambda_1^{-1}\right)e_i \le \frac{\sigma^2}{\lambda_k} = \rho$.

i.e. $1-\rho \le \sum_{j=1}^{k}\left(q_i'u_j\right)^2$ #

Lemma 9: Let $z_j \sim O_p\left(T^k\right)$, $\alpha_i \ge 0$ for $i = 1,\ldots,N$, and $\sum_{i=1}^{N}\alpha_i \le \bar{\alpha} < \infty$. Then $\sum_{i=1}^{N}\alpha_i z_i \sim O_p\left(T^k\right)$.

Proof of Lemma 9:

From the assumptions, $\forall\varepsilon>0\ \exists\delta_i\in\mathbb{R}^+ \ni P\left(|z_i|\geq T^k\delta_i\right)\leq\frac{\varepsilon}{N}$. Therefore

$\forall\varepsilon>0\ \exists\delta_i\in\mathbb{R}\ni P\left(|\alpha_i z_i|\geq T^k\delta_i\alpha_i\right)\leq\frac{\varepsilon}{N}$ for $i=1,\dots,N$.

Let $E_i=\left\{z_i:|\alpha_i z_i|\leq\alpha_i T^k\delta_i\right\}$ for $i=1,\dots,N$, $E=\left\{J:|J|\leq T^k\sum_{i=1}^{N}\alpha_i\delta_i\right\}$, and

$\Xi=\left\{J:|J|\leq T^k\bar{\alpha}\delta_{max}\right\}$ where $\delta_{max}=\sup_i\delta_i$. Since $\bigcap_{i=1}^{N}E_i\subset E\subset\Xi$, it follows from the implication rule (Lukacs, E. (1975)) that

$P(\Xi)\geq P\left(\bigcap_{i=1}^{N}E_i\right)=1-P\left(\bigcup_{i=1}^{N}\bar{E}_i\right)\geq 1-\sum_{i=1}^{N}P\left(|\alpha_i z_i|\geq\alpha_i T^k\delta_i\right)\geq 1-\varepsilon$ where $\bar{E}_i$ is the complement of E_i. Let $m=\delta_{max}\bar{\alpha}$. Then $\forall\varepsilon>0\ \exists m\in\mathbb{R}^+\ni P\left(|J|\geq T^k m\right)\leq\varepsilon$, i.e. $J\sim O_p\left(T^k\right)$. #

Lemma 10: Under the assumptions $\frac{1}{N}\Lambda^{\frac{1}{2}}\frac{1}{T}S'S\Lambda^{\frac{1}{2}}=\frac{1}{N}\Lambda+Z_1$ where $\|Z_1\|_F\sim O_p\left(T^{-\frac{1}{2}}\right)$.

Proof of Lemma 10: By standard arguments under the assumptions

$\frac{1}{TN}X'X=\frac{1}{TN}Q\Lambda^{\frac{1}{2}}S'S\Lambda^{\frac{1}{2}}Q'=\frac{1}{N}Q\Lambda Q'+\frac{1}{N}O_p\left(T^{-\frac{1}{2}}\right)$ so $\frac{1}{TN}\Lambda^{\frac{1}{2}}S'S\Lambda^{\frac{1}{2}}=\frac{1}{N}\Lambda+Z_1$ where

$Z_1\sim\frac{1}{N}Q'Z_2Q$ and $Z_2\sim O_p\left(T^{-\frac{1}{2}}\right)\Rightarrow\frac{1}{N}tr(Z_2)\sim O_p\left(T^{-\frac{1}{2}}\right)$.

$\|Z_1\|_F=\sqrt{\sum_{j=1}^{N}eig_j(Z_1)^2}\leq\sum_{j=1}^{N}eig_j(Z_1)=tr(Z_1)=\frac{1}{N}tr(Z_2)\sim O_p\left(T^{-\frac{1}{2}}\right)$. #

Proof of Theorem 1(a): Let $\Phi=\sigma^2 I_p-\Psi$. Then $\Phi+\Omega=BB'+\sigma^2 I_p$. Note that $eig_j(\Phi)=\sigma^2-\sigma_j^2$, where $eig_j(.)$ denotes the j^{th} ordered eigenvalue of its matrix argument, so $eig_j(\Phi)\geq 0\ \forall j=1,..,p$. Thus, Φ is positive semi-definite. It follows from Magnus and Neudecker (1991, p.208, Theorem 9) that $eig_j(BB'+\sigma^2 I_p)\geq eig_j(\Omega)$, i.e. $d_j+\sigma^2\geq\lambda_j\ \forall j=1,..,p$. It also follows from Magnus and Neudecker (1991, p.208, Theorem 9) that $\lambda_i\geq d_i\ \forall i=1,\dots,k$. The result follows.

Proof of Theorem 1(b): $\|Q_1-UL\|^2=tr\left[\left(Q_1'-LU'\right)\left(Q_1-UL\right)\right]=2tr\left(I-LQ_1'U\right)$ (A6)

From Lemma 8 $1-\rho-\sum_{j\neq i}^{k}\left(q_i'u_j\right)^2 \le \left(q_i'u_i\right)^2$. If k = 1 then c = 0 and the result holds from A6 with sign(L_{ii})=sign($q_i'u_j$). If k > 1 then using Lemma 7 $1-\rho-4c^2\rho^2(k-1) \le \left(q_i'u_i\right)^2$. With $c \le \frac{1}{2\rho}\sqrt{\frac{1-\rho}{k-1}}$ the left hand side is non-negative and $\sqrt{1-\rho-4c^2\rho^2(k-1)} \le \left|q_i'u_i\right|$. (A7)

If we choose L so that $\text{sign}(L_{ii}) = \text{sign}(q_i'u_j)$ then from A6 and A7 we get $\left\|Q-UL\right\|^2 \le k-k\sqrt{(1-\rho)-4c^2\rho^2(k-1)}$. Multiplying this by $\frac{1+\sqrt{(1-\rho)-4c^2\rho^2(k-1)}}{1+\sqrt{(1-\rho)-4c^2\rho^2(k-1)}}$ yields the result. #

Proof of Theorem 1(c):

$$E\left\|s_t-Lf_t\right\|^2 = \text{tr}\left[\left(\Lambda_1^{-\frac{1}{2}}Q_1'UD^{\frac{1}{2}}-L\right)\left(D^{\frac{1}{2}}U'Q_1\Lambda_1^{-\frac{1}{2}}-L\right)+\Lambda_1^{-\frac{1}{2}}Q_1'\Psi Q_1\Lambda_1^{-\frac{1}{2}}\right]$$

$$= 2\text{tr}\left(I-L\Lambda_1^{-\frac{1}{2}}Q_1'UD^{\frac{1}{2}}\right)$$

As in the proof to Theorem 2, $1-\rho-4c^2\rho^2(k-1) \le \left(q_i'u_i\right)^2$. From Theorem 1 $1-\rho \le \frac{d_i}{\lambda_i}$, so $\left(1-\rho\right)^2-4c^2\rho^2(k-1) \le \frac{d_i}{\lambda_i}\left(q_i'u_i\right)^2$. If $c \le \frac{1-\rho}{2\rho\sqrt{(k-1)(1-\rho)}}$ then the left hand side is non-negative and $\sqrt{\left(1-\rho\right)^2-4c^2\rho^2(k-1)} \le \sqrt{\frac{d_i}{\lambda_i}}\left|q_i'u_i\right|$

If we choose S so that $\text{sign}(L_{ii}) = \text{sign}(q_i'u_j)$, we get $E\left\|s_t-Lf_t\right\|^2 \le k-k\sqrt{(1-\rho)^2-4c^2\rho^2(k-1)(1-\rho)}$. Multiplying this by $\frac{1+\sqrt{(1-\rho)^2-4c^2\rho^2(k-1)(1-\rho)}}{1+\sqrt{(1-\rho)^2-4c^2\rho^2(k-1)(1-\rho)}}$ yields the result. #

Proof of Theorem 1(d): Defining $S_{xy} = \frac{1}{T}\sum_{t=1}^{T}x_t y_t$, the forecast deviation is

$$e_{T+h} = \tilde{\beta}'s_{T+h}-\beta'f_{T+h} = \left(\Lambda_1^{-\frac{1}{2}}Q_1'S_{xy}\right)'\Lambda_1^{-\frac{1}{2}}Q_1'x_{T+h}-\beta'f_{T+h} = S_{xy}'Q_1\Lambda_1^{-1}Q_1(Bf_{T+h}+n_{T+h})-\beta'f_{T+h}$$

$= e_a + e_b$ where

$e_a = \left(S'_{xy}Q_1\Lambda_1^{-1}Q_1'B - \beta'\right)f_{T+h}$ and $e_b = S'_{xy}Q_1\Lambda_1^{-1}Q_1'n_{T+h}$.

First consider e_b. From the Cauchy-Schwarz inequality we have $\left|e_b\right| \le \left\|S'_{xy}Q_1\Lambda_1^{-1}\right\|\left\|Q_1'n_{T+h}\right\|$ and $\left(E\left|e_b\right|\right)^2 \le E\left\|S'_{xy}Q_1\Lambda_1^{-1}\right\|^2 E\left\|Q_1'n_{T+h}\right\|^2$ (A8)

We have that $E\left\|Q_1'n_{T+h}\right\|^2 = tr(Q_1'\Psi Q_1) \le \sigma^2 k$. (A9)

Also

$$E\left\|S'_{xy}Q_1\Lambda_1^{-1}\right\|^2 = E\left(S'_{xy}Q_1\Lambda_1^{-2}Q_1'S_{xy}\right) = \sum_{i=1}^{k} E\left[\left(e_i'\Lambda_1^{-1}Q_1'S_{xy}\right)^2\right] \text{ from Lemma 3}$$

$$= \sum_{i=1}^{k}\left[var\left(e_i'\Lambda_1^{-1}Q_1'S_{xy}\right) + E\left(e_i'\Lambda_1^{-1}Q_1'S_{xy}\right)^2\right] = \sum_{i=1}^{k}\left[var\left(e_i'\Lambda_1^{-1}Q_1'S_{xy}\right) + \left(e_i'\Lambda_1^{-1}Q_1'B\beta\right)^2\right] \quad \text{(A10)}$$

Now $\sum_{i=1}^{k}\left(e_i'\Lambda_1^{-1}Q_1'B\beta\right)^2 = \beta'BQ_1\Lambda_1^{-2}Q_1'B\beta = \beta'D^{\frac{1}{2}}U'Q_1\Lambda_1^{-2}Q_1'UD^{\frac{1}{2}}\beta$

$$= \beta D^{\frac{1}{2}}R\Lambda_1^{-2}R'D^{\frac{1}{2}}\beta \le \beta'\beta\lambda_{max}\left(\Lambda_1^{-1}R'DR\Lambda_1^{-1}\right) \text{ by Lemma 6}$$

$$\le \beta'\beta\lambda_{max}\left(\Lambda_1^{-1}\Lambda_1\Lambda_1^{-1}\right) \text{ by Lemma 4}$$

$$\le \beta'\beta\lambda_k^{-1} \quad \text{(A11)}$$

Also, from Lemma 2

$$\sum_{i=1}^{k} var\left(e_i'\Lambda_1^{-1}Q_1'S_{xy}\right) \le \Upsilon_1 + \Upsilon_2 + \Upsilon_3 \text{ where}$$

$$\Upsilon_1 = \frac{2}{T}\sum_{i=1}^{k} e_i'\Lambda_1^{-1}Q_1'\Omega Q_1\Lambda_1^{-1}e_i\sigma_y^{(0)2}$$

$$\Upsilon_2 = \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\left|e_i'\Lambda_1^{-1}Q_1'E(x_t x'_{t-j})Q_1\Lambda_1^{-1}e_i\sigma_y^{(j)2}\right|$$

$$\Upsilon_3 = \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\left|e_i'\Lambda_1^{-1}Q_1'E(x_t y_{t-j})E(y_t x'_{t-j})Q_1\Lambda_1^{-1}e_i\right|$$

where $\sigma_y^{(j)2} = E(y_t y_{t-j})$. We have that

$$\Upsilon_1 = \frac{2}{T}\sigma_y^{(0)2} tr\left(\Lambda_1^{-1}\Lambda_1\Lambda_1^{-1}\right) \le \frac{2}{T}\sigma_y^{(0)2}\lambda_k^{-1}$$

$$\Upsilon_2 = \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\left| e_i'\Lambda_1^{-\frac{1}{2}}E(s_t s_{t-j}')\Lambda_1^{-\frac{1}{2}}e_i\sigma_y^{(j)2}\right|$$ where $s_t = \Lambda_1^{-\frac{1}{2}}Q_1'x_t$ is the principal component vector of x_t. So

$$\Upsilon_2 = \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\left|\frac{E(s_t s_{t-j}')}{\lambda_i}\sigma_y^{(j)2}\right| \le \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\frac{1}{\lambda_i}\left|E(s_t s_{t-j}')\right|\left|\sigma_y^{(j)2}\right| \le \frac{2}{T}\lambda_k^{-1}\sum_{j=1}^{T-1}\left|\sigma_y^{(j)2}\right|$$

by Cauchy-Schwarz, and

$$\Upsilon_3 = \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\frac{1}{\lambda_i}\left|E(s_{it}y_{t-j})E(y_t s_{i,t-j}')\right| \le \frac{2}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\frac{1}{\lambda_i}\left|E(s_{it}y_{t-j})\right|\left|E(y_t s_{i,t-j}')\right| \le \frac{2\sigma_y^{(0)2}}{T}\sum_{i=1}^{k}\sum_{j=1}^{T-1}\frac{1}{\lambda_i}\left|E(y_t s_{i,t-j}')\right|$$

$$\le \frac{2\sigma_y^{(0)2}}{T}\lambda_k^{-1}\sup_i\sum_{j=1}^{T-1}\left|E(y_t s_{i,t-j}')\right|$$

so $$\sum_{i=1}^{k}\operatorname{var}\left(e_i'\Lambda_1^{-1}Q_1'S_{xy}\right) \le \frac{2}{T\lambda_k}\left(\sum_{j=0}^{T-1}\left|\sigma_y^{(j)2}\right| + \sigma_y^{(0)}\sup_i\sum_{j=1}^{T-1}\left|E(y_t s_{i,t-j}')\right|\right) = \frac{2}{T\lambda_k}\sigma_y^{(0)2}\gamma \qquad \text{(A12)}$$

where $$\gamma = \sum_{j=1}^{T-1}\left|\frac{E(y_t y_{t-j})}{E(y_t^2)}\right| + \sup_i\sum_{j=1}^{T-1}\left|\frac{E(y_t s_{i,t-j})}{\sqrt{E(y_t^2)E(s_{it}^2)}}\right|$$

Equations A10, A11, and A12 yield

$$E\left\|S_{xy}'Q_1\Lambda_1^{-1}\right\|^2 \le \lambda_k^{-1}\left(\frac{2\sigma_y^{(0)2}\gamma}{T} + \beta'\beta\right)$$

which when combined with A8 and A9 yield

$$\frac{\left(E\left|e_b\right|\right)^2}{\|\beta\|^2} \le \frac{\sigma^2 k}{\lambda_k}\left(\frac{2}{T}\frac{\sigma_y^{(0)2}\gamma}{\|\beta\|^2} + 1\right).$$

We now consider e_a. By the Cauchy-Schwarz inequality we have

$$\left(E\left|e_a\right|\right)^2 \le E\left\|S_{xy}'Q_1\Lambda_1^{-1}Q_1B - \beta'\right\|^2 E\left\|f_h\right\|^2 = kE\left\|S_{xy}'Q_1\Lambda_1^{-1}Q_1B - \beta'\right\|^2 \qquad \text{(A13)}$$

Now $E\left\|S_{xy}'Q_1\Lambda_1^{-1}Q_1B - \beta'\right\|^2 = \sum_{i=1}^{k}E\left[e_i'\left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy} - \beta\right)^2\right]$ from Lemma 3

$$= \sum_{i=1}^{k} \mathrm{var}\left[e_i' \left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy} - \beta \right) \right] + \sum_{i=1}^{k} E\left[e_i' \left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy} - \beta \right) \right]^2 \qquad \text{(A14)}$$

but from Lemma 2

$$= \sum_{i=1}^{k} \mathrm{var}\left[e_i' \left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy} - \beta \right) \right] \le \Delta_1 + \Delta_2 + \Delta_3 \quad \text{where}$$

$$\Delta_1 = \sum_{i=1}^{k} \frac{2}{T} e_i'B'Q_1\Lambda_1^{-1}Q_1'\Omega Q_1\Lambda_1^{-1}Q_1'Be_i\sigma_y^{(0)2}$$

$$\Delta_2 = \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \left| e_i'B'Q_1\Lambda_1^{-1}Q_1'E(x_t x_{t-j})Q_1\Lambda_1^{-1}Q_1'Be_i\sigma_y^{(j)2} \right|$$

$$\Delta_3 = \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \left| e_i'B'Q_1\Lambda_1^{-1}Q_1'E(x_t y_{t-j})e_i'B'Q_1\Lambda_1^{-1}Q_1'E(x_t y_{t-j}) \right|$$

We have that

$$\Delta_1 = \frac{2\sigma_y^{(0)2}}{T} \mathrm{tr}\left(B'Q_1\Lambda_1^{-1}Q_1'B \right) \le \frac{2\sigma_y^{(0)2}}{T} \mathrm{tr}\left(B'\Omega^{-1}B \right) \le \frac{2k\sigma_y^{(0)2}}{T}$$

$$\Delta_2 = \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \left| e_i'B'Q_1\Lambda_1^{-\frac{1}{2}}E(s_t s_{t-j})\Lambda_1^{-\frac{1}{2}}Q_1'Be_i\sigma_y^{(j)2} \right|$$

let $v_i = e_i'B'Q_1\Lambda_1^{-\frac{1}{2}} = \begin{pmatrix} 0 & \ldots & 0 & \nu_i & 0 & \ldots & 0 \end{pmatrix}$ then

$$\Delta_2 = \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \left| v_i E(s_t s_{t-j}) v_i' \sigma_y^{(j)2} \right| = \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \left| \nu_i^2 E(s_t s_{t-j})_i \sigma_y^{(j)2} \right| \le \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \nu_i^2 \left| \sigma_y^{(j)2} \right|$$
$$\le \frac{2}{T} \sum_{i=1}^{k} e_i'BQ_1\Lambda_1^{-1}Q_1'Be_i \sum_{j=1}^{T-1} \left| \sigma_y^{(j)2} \right| \le \frac{2}{T} \sum_{i=1}^{k} \mathrm{tr}(B'\Omega^{-1}B) \sum_{j=1}^{T-1} \left| \sigma_y^{(j)2} \right| \le \frac{2k}{T} \sum_{j=1}^{T-1} \left| \sigma_y^{(j)2} \right|$$

$$\Delta_3 = \frac{2}{T} \sum_{i=1}^{k} \sum_{j=1}^{T-1} \left| e_i'B'Q_1\Lambda_1^{-\frac{1}{2}}E(s_t y_{t-j})E(y_t s_{t-j}')\Lambda_1^{-\frac{1}{2}}Q_1'Be_i \right| = \frac{2}{T} \sum_{i=1}^{k} e_i'B'Q_1\Lambda_1^{-1}Q_1'Be_i \sum_{j=1}^{T-1} \left| E(s_t y_{t-j})E(y_t s_{t-j}') \right|$$
$$\le \frac{2}{T} \mathrm{tr}(B'\Omega^{-1}B)\sigma_y^{(0)} \sup_i \sum_{j=1}^{T-1} \left| E(y_t s_{t-j}') \right|$$

So

$$\sum_{i=1}^{k} \mathrm{var}\left[e_i'\left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy}-\beta\right)\right] \le \frac{2k}{T}\sum_{j=1}^{T-1}\left|\sigma_y^{(j)2}\right| + \sigma_y^{(0)} \sup_i \sum_{j=1}^{T-1}\left|E(y_t s_{t-j}')\right| = \frac{2k}{T}\sigma_y^{(0)}\gamma \quad \text{(A15)}$$

Also $\sum_{i=1}^{k} E\left[e_i'\left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy}-\beta\right)\right] = e_i'(M-I)\beta$ where $M = D^{\frac{1}{2}}R\Lambda_1^{-1}R'D^{\frac{1}{2}}$

So $\sum_{i=1}^{k} E\left[e_i'\left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy}-\beta\right)\right]^2 = \beta'(M-I)^2\beta$.

But $M^2 \le B'Q_1\Lambda_1^{-1}Q_1'\Omega Q_1\Lambda_1^{-1}Q_1'B = M$ so $(M-I)^2 \le (I-M) \Rightarrow M \le I_k$

Therefore $\sum_{i=1}^{k} E\left[e_i'\left(B'Q_1\Lambda_1^{-1}Q_1'S_{xy}-\beta\right)\right]^2 \le \beta'(M-I)^2\beta \le \beta'\beta\left[\lambda_{max}(I-M)\right]^2$

$$= \beta'\beta\left[\lambda_{max}\left(\Lambda_1^{-\frac{1}{2}}Q_1'\Psi Q_1\Lambda_1^{-\frac{1}{2}}\right)\right]^2 \text{ from Lemma 5}$$

$$\le \beta'\beta\left(\frac{\sigma^2}{\lambda_k}\right)^2 \quad \text{(A16)}$$

Combining (A13)-(A16) yields

$$\frac{\left(E\left|e_a\right|\right)^2}{\|\beta\|^2} \le k\left(\left(\frac{\sigma^2}{\lambda_k}\right)^2 + \frac{2k}{T}\frac{\sigma_y^{(0)2}\gamma}{\|\beta\|^2}\right)$$

Noting that $\rho = \frac{\sigma^2}{\lambda_k}$ and $\frac{1}{r^2} = \frac{\sigma_y^2}{\|\beta\|^2}$, combining the above results yields the result of the theorem. #

Proof of Theorem 2

From Magnus and Neudecker (1991, p.208, Theorem 9) $\lambda_k \ge d_k \ge Nd_L$ so

$$\rho = \frac{\sigma^2}{\lambda_k} \le \frac{\sigma^2}{d_k} \le \frac{\sigma^2}{Nd_U} \sim O\left(N^{\tau-1}\right).$$

Proof of Theorem 3(a)

Let $\Psi_j = \left\{r_j : r'r_j = 1 \text{ if } j = i, 0 \text{ otherwise}\right\}$. Then

$$\frac{1}{N}\hat{\lambda}_j = \sup_{r_j\in\Psi_j} r_j'\frac{1}{TN}XX'r_j = \sup_{r_j\in\Psi_j} r_j'\frac{1}{TN}Q\Lambda^{\frac{1}{2}}S'S\Lambda^{\frac{1}{2}}Q'r_j = \sup_{r_j\in\Psi_j} r_j'\Lambda^{\frac{1}{2}}\frac{1}{TN}S'S\Lambda^{\frac{1}{2}}r_j$$

$$\le \sup_{r_j\in\Psi_j} r_j'\frac{1}{N}\Lambda r_j + \sup_{r_j\in\Psi_j} r_j'Z_1r_j = \frac{1}{N}\lambda_j + O_p\left(T^{-\frac{1}{2}}\right) \text{ from Lemma 10.}$$

Also $\frac{1}{N}\hat{\Lambda}_j = \frac{1}{TN}\hat{Q}_j'X'X\hat{Q}_j = \frac{1}{N}\hat{Q}_j'\Omega\hat{Q}_j + \frac{1}{N}\hat{Q}_j'Z_3\hat{Q}_j$ where $Z_3 \sim O_p\left(T^{-\frac{1}{2}}\right)$, so

$\frac{1}{N}tr\left(\hat{\Lambda}_j\right) = \frac{1}{N}tr\left(\hat{Q}_j'\Omega\hat{Q}_j\right) + \frac{1}{N}tr\left(\hat{Q}_j'Z_3\hat{Q}_j\right) \le \frac{1}{N}tr\left(\Lambda_j\right) + O_p\left(T^{-\frac{1}{2}}\right)$, i.e

$$\frac{1}{N}\sum_{j\ne i}^{N}\hat{\lambda}_j \le \frac{1}{N}\sum_{j\ne i}^{N}\lambda_j + O_p\left(T^{-\frac{1}{2}}\right) \text{ for i=1,..,N} \qquad \text{(A17)}$$

and we have already shown that $\frac{1}{N}\hat{\lambda}_j \le \frac{1}{N}\lambda_j + O_p\left(T^{-\frac{1}{2}}\right)$. (A18)

But we also have that $\frac{1}{N}tr\left(\hat{\Lambda}\right) = \frac{1}{N}tr\left(\hat{Q}'\Omega\hat{Q}\right) + \frac{1}{N}tr\left(\hat{Q}'Z_3\hat{Q}\right)$ so

$\frac{1}{N}\sum_{j=i}^{N}\hat{\lambda}_j \le \frac{1}{N}\sum_{j=i}^{N}\lambda_j + O_p\left(T^{-\frac{1}{2}}\right)$ for i=1,..,N. Thus, strict equality must hold for equations A17 and A18, proving the theorem. #

Lemma 11

Under assumptions 1 to 3, $R_2 = Q'\hat{Q} = L + O_p\left(T^{-\frac{1}{2}}\right)$ where L is a N×N sign matrix.

Proof of Lemma 11

$\frac{1}{T}X'X\hat{Q} = \hat{Q}\hat{\Lambda}$ so $\Lambda^{\frac{1}{2}}\frac{1}{T}S'S\Lambda^{\frac{1}{2}}R_2 = R_2\hat{\Lambda}$ where $R_2 = Q'\hat{Q}$ and

$\frac{1}{N}\Lambda R_2 + Z_1R_2 = \frac{1}{N}R_2\Lambda + \frac{1}{N}R_2\left(\hat{\Lambda} - \Lambda\right)$ from Lemma 10 so

$\frac{1}{N}\Lambda R_2 + Z_1R_2 = \frac{1}{N}R_2\Lambda + O_p\left(T^{-\frac{1}{2}}\right)$ from Lemma 10 and Theorem 3(a). We write this as $\Lambda r_j = \lambda_j r_j + O_p\left(T^{-1}\right)$ for $j = 1,...,N$ where r_j is the j^{th} column of R. The solution set for r_j is $r_j = \left(I_N - \left(\lambda_j I_N - \Lambda\right)^{+}\left(\lambda_j I_N - \Lambda\right)\right)z_j + O_p\left(T^{-1}\right)$, $j = 1,...,N$ where z_j is an arbitrary N×1 vector and $^{+}$ denotes the Moore-Penrose generalised inverse. This implies that $R_2 = \bar{R} + O_p\left(T^{-1}\right)$ where $\bar{R}$ is a diagonal matrix.

We also have $\frac{1}{N}\hat{\Lambda}=\hat{Q}'\frac{1}{TN}X'X\hat{Q}=\hat{Q}'Q\Lambda^{\frac{1}{2}}\frac{1}{TN}S'S\Lambda^{\frac{1}{2}}Q'\hat{Q}=R_2'\frac{1}{N}\Lambda R_2+R_2'Z_1R_2$ from Lemma 10, so $\frac{1}{N}\Lambda=R_2'\frac{1}{N}\Lambda R_2+O_p\left(T^{-\frac{1}{2}}\right)$. We can write $R_2=\bar{R}+M$ where $M\sim O_p\left(T^{-1}\right)$ so

$$\frac{1}{N}\Lambda=\bar{R}'\frac{1}{N}\Lambda\bar{R}+\bar{R}'\frac{1}{N}\Lambda M+M'\frac{1}{N}\Lambda\bar{R}+M'\frac{1}{N}\Lambda M+O_p\left(T^{-\frac{1}{2}}\right)=\bar{R}'\frac{1}{N}\Lambda\bar{R}+O_p\left(T^{-\frac{1}{2}}\right)$$

since $M'\frac{1}{N}\Lambda M\sim O_p\left(T^{-1}\right)$ and $\bar{R}'\frac{1}{N}\Lambda M\sim O_p\left(T^{-\frac{1}{2}}\right)$. We therefore have $\frac{1}{N}\lambda_j=\bar{r}_j^2\frac{1}{N}\lambda_j+O_p\left(T^{-\frac{1}{2}}\right)$. It follows that $R_2=Q'\hat{Q}=L+O_p\left(T^{-\frac{1}{2}}\right)$ where L is a N×N sign matrix. #

Lemma 12: Under assumptions 1 to 3 $\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'\hat{q}_j\hat{\lambda}_j^{-\frac{1}{2}}\sim O_p\left(T^{-\frac{1}{2}}\right)$ for j=1,..,k.

Proof of Lemma 12: $\frac{1}{T}X'X\hat{q}_j=\hat{q}_j\hat{\lambda}_j$ so $\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'\frac{1}{T}X'X\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}=\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'\hat{q}_j\hat{\lambda}_j^{-\frac{1}{2}}$ but

$$\frac{1}{T}X'X=\lambda_jq_j\frac{1}{T}s_j's_jq_j'+q_j\lambda_j^{\frac{1}{2}}\frac{1}{T}s_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'+\tilde{Q}\tilde{\Lambda}_j^{\frac{1}{2}}\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}q_j'+\tilde{Q}\tilde{\Lambda}_j^{\frac{1}{2}}\frac{1}{T}\tilde{S}_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'$$ so

$$\tilde{Q}_j\frac{1}{T}X'X=\tilde{\Lambda}_j^{\frac{1}{2}}\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}q_j'+\tilde{\Lambda}_j^{\frac{1}{2}}\frac{1}{T}\tilde{S}_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'$$ and

$$\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}_j'\hat{q}_j\hat{\lambda}_j^{-\frac{1}{2}}=\tilde{s}_{jt}'\tilde{\Lambda}_j\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}q_j'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}+\tilde{s}_{jt}'\tilde{\Lambda}_j\frac{1}{T}\tilde{S}_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}$$. We have

$\tilde{s}_{jt}'\tilde{\Lambda}_j\frac{1}{T}\tilde{S}_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}=\alpha_1+\alpha_2$ where $\alpha_1=\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{3}{2}}\tilde{Q}'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}$ and $\alpha_2=\tilde{s}_{jt}'\tilde{\Lambda}_jZ_1\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}$ from Lemma 10. Now $\alpha_2^2\le\left\|N^{-1}\tilde{s}_{jt}'\tilde{\Lambda}_j\right\|_2^2\left\|Z_1\right\|_2^2\left\|N^{-\frac{1}{2}}\tilde{\Lambda}_j^{\frac{1}{2}}\right\|_2^2\left\|\tilde{Q}'\hat{q}_j\right\|_2^2N^{\frac{3}{2}}\hat{\lambda}_j^{-\frac{3}{2}}$ and we have $E\left\|N^{-1}\tilde{s}_{jt}'\tilde{\Lambda}_j\right\|_2^2=\frac{1}{N^2}E\left(\tilde{s}_{jt}'\tilde{\Lambda}_j^2\tilde{s}_{jt}\right)=\frac{1}{N^2}\sum_{i\neq j}^{N}\lambda_i^2\sim O(1)\Rightarrow\left\|N^{-1}\tilde{s}_{jt}'\tilde{\Lambda}_j\right\|_2^2\sim O_p(1)$. Also $\left\|N^{-\frac{1}{2}}\tilde{\Lambda}_j^{\frac{1}{2}}\right\|_2^2=\max\text{eig}\left(\frac{1}{N}\tilde{\Lambda}_j\right)\sim O(1)$, $\left\|\tilde{Q}'\hat{q}_j\right\|_2^2=\hat{q}'\tilde{Q}_j\tilde{Q}'\hat{q}_j\le1$, $N^{\frac{3}{2}}\hat{\lambda}_j^{-\frac{3}{2}}\sim O_p(1)$ and $\left\|Z_1\right\|_2^2\sim O_p\left(T^{-1}\right)$ from Lemma 10. It follows that $\alpha_2\sim O_p\left(T^{-\frac{1}{2}}\right)$.

From Lemma 11 $\tilde{Q}'\hat{q}_j \sim O_p\left(T^{-\frac{1}{2}}\right)$ so, since $\tilde{s}_{jt} \sim O_p(1)$, $\text{diag}\left(\tilde{s}_{jt}\right)\tilde{Q}'\hat{q}_j \sim O_p\left(T^{-\frac{1}{2}}\right)$ where diag(v) is a diagonal matrix constructed from the vector v. Also $N^{\frac{3}{2}}\hat{\lambda}_j^{-\frac{3}{2}} \sim O_p(1)$ and $\frac{1}{N^{\frac{3}{2}}}\sum_{j=1}^{N}\lambda_i^{\frac{3}{2}} \le \frac{\lambda_1^{\frac{1}{2}}}{N^{\frac{1}{2}}}\sum_{j=1}^{N}\frac{\lambda_i}{N} < \infty$ so from Lemma 11

$\alpha_1 = \text{DIAG}\left(\frac{1}{N^{\frac{3}{2}}}\tilde{\Lambda}_j^{\frac{3}{2}}\right)\text{diag}\left(\tilde{s}_{jt}\right)\tilde{Q}'\hat{q}_j N^{\frac{3}{2}}\hat{\lambda}_j^{-\frac{3}{2}} \sim O_p\left(T^{-\frac{1}{2}}\right)$ where DIAG(M) is the vector formed from the diagonal elements of the square matrix M. Combining results yields $\tilde{s}_{jt}'\tilde{\Lambda}_j\frac{1}{T}\tilde{S}_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{Q}'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}} \sim O_p\left(T^{-\frac{1}{2}}\right)$.

Also $\left\|\tilde{s}_{jt}'\tilde{\Lambda}_j\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}q_j'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}}\right\|_2^2 \le \left\|\frac{1}{N^{\frac{1}{2}}}\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\right\|_2^2\left\|\frac{1}{N}\tilde{\Lambda}_j^{\frac{1}{2}}\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}\right\|_2^2\left\|q_j'\hat{q}_j\right\|_2^2\hat{\lambda}_j^{-3}N^3$. We have $E\left\|N^{-\frac{1}{2}}\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\right\|_2^2 = \frac{1}{N}E\left(\tilde{s}_{jt}'\tilde{\Lambda}_j\tilde{s}_{jt}\right) = \frac{1}{N}\sum_{i\ne j}^{N}\lambda_i \sim O(1) \Rightarrow \left\|N^{-\frac{1}{2}}\tilde{s}_{jt}'\tilde{\Lambda}_j^{\frac{1}{2}}\right\|_2^2 \sim O_p(1)$, $\left\|q_j'\hat{q}_j\right\|_2^2 \le 1$ and $\hat{\lambda}_j^{-3}N^3 \sim O_p(1)$. $\frac{1}{T}s_j'\tilde{S}_j$ is the jth row of $\frac{1}{T}S'S$ with the jth column excluded. Therefore $\lambda_j\frac{1}{T}s_j'\tilde{S}_j\tilde{\Lambda}_j^{\frac{1}{2}}$ is the jth row of Z_1 with the jth column excluded. Thus $\left\|\frac{1}{N}\tilde{\Lambda}_j^{\frac{1}{2}}\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}\right\|_2^2 = \frac{1}{TN}\lambda_j^{\frac{1}{2}}s_j'\tilde{S}_j\tilde{\Lambda}_j\frac{1}{TN}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}} \le \left[Z_1'Z_1\right]_{jj} \le \text{tr}\left(Z_1'Z_1\right) \sim O_p(1)$ from Lemma 10. Therefore $\tilde{s}_{jt}'\tilde{\Lambda}_j\frac{1}{T}\tilde{S}_j's_j\lambda_j^{\frac{1}{2}}q_j'\hat{q}_j\hat{\lambda}_j^{-\frac{3}{2}} \sim O_p\left(T^{-\frac{1}{2}}\right)$ and the Lemma is proved. #

Proof of Theorem 3(b)

$\hat{s}_{jt} = \hat{\lambda}_j^{-\frac{1}{2}}\hat{q}_j'x_t = \hat{\lambda}_j^{-\frac{1}{2}}\hat{q}_j'q_j\lambda_j^{-\frac{1}{2}}s_{jt} + \hat{\lambda}_j^{-\frac{1}{2}}\hat{q}_j'\tilde{Q}_j\tilde{\Lambda}_j^{\frac{1}{2}}\tilde{s}_{jt}$. The second term is $O_p\left(T^{-\frac{1}{2}}\right)$ by Lemma 12. For j=1,..,k $N^{\frac{1}{2}}\hat{\lambda}_j^{-\frac{1}{2}} = N^{\frac{1}{2}}\lambda_j^{-\frac{1}{2}} + z_j$ where $z_j \sim O_p\left(T^{-\frac{1}{2}}\right)$ so

$$\hat{\lambda}_j^{-\frac{1}{2}}\hat{q}_j'q_j\lambda_j^{-\frac{1}{2}}s_{jt} = L_{jj}s_{jt} + m_{jj}s_{jt} + z_jL_{jj}\frac{\lambda_j^{\frac{1}{2}}}{N^{\frac{1}{2}}}s_{jt} + z_jm_{jj}\frac{\lambda_j^{\frac{1}{2}}}{N^{\frac{1}{2}}}s_{jt} \text{ where}$$

$m_{jj} = \hat{q}_j'q_j - L_{jj} \sim O_p\left(T^{-\frac{1}{2}}\right)$ from Lemma 11; so $\hat{s}_{jt} = Ls_{jt} + O_p\left(T^{-\frac{1}{2}}\right)$. #

Lemma 13: $1-\tilde{\rho} \le \frac{d_i}{\lambda_i} \le 1$ for i=1,..,k.

Proof of Lemma 13: As shown in the proof of Theorem 1(a) $\lambda_i \ge d_i \ \forall i = 1,...,k$. Therefore $\tilde{\rho} = \frac{\sigma^2}{d_k} \ge \frac{\sigma^2}{\lambda_k} = \rho \Rightarrow 1-\tilde{\rho} \le 1-\rho$. The result then follows from Theorem 1(a). #

Lemma 14: $\left|q_i'u_j\right| \le 2\tilde{c}\tilde{\rho}$ for $i \ne j$ where q_i is the i[th] column of Q_1 and u_j is the j[th] column of U, and $\tilde{c} = \max_{i,j:i\ne j}\frac{d_j}{\left|d_i - d_j\right|}$.

Proof of Lemma 14: $Q_1\Lambda_1Q_1' + Q_2\Lambda_2Q_2' = UDU' + \Psi$. Premultiplying by $D^{-1}Q_1'$, postmultiplying by UD, and subtracting $Q_1'U$ yields

$$D^{-1}Q_1'UD - Q_1'U = \left(D^{-1}\Lambda_1 - I_k\right)Q_1'U - D^{-1}Q_1'\Psi U \qquad \text{(A19)}$$

We now consider each of the right hand side terms. Let e_i be a vector of zeros with a 1 in the i[th] element only. We have

$$\left(e_i'D^{-1}Q_1'\Psi Ue_j\right)^2 \le tr\left(U'\Psi D^{-1}e_ie_i'D^{-1}Q_1'\Psi U\right) \le \frac{1}{d_k^2}e_i'D^{-1}Q_1'\Psi UU'\Psi Q_1D^{-1}e_i$$
$$\le \frac{1}{d_k^2}e_i'Q_1'\Psi^2Q_1e_i \le \frac{\sigma^4}{d_k^2}$$

$$\therefore \quad \left|e_i'D^{-1}Q_1'\Psi Ue_j\right| \le \tilde{\rho} \qquad \text{(A20)}$$

For the other right hand side term we have

$\left|e_i'(D^{-1}\Lambda_1 - I_k)Q_1'Ue_j\right| = \left|\left(\frac{\lambda_j}{d_j} - 1\right)q_i'u_j\right| = \left|q_i'u_j\right|\left|\frac{\lambda_j}{d_j} - 1\right|$. From the proof to Theorem 1(a) we have $\lambda_j \le d_j + \sigma^2$ for j=1,..,k. Dividing by d_j yields $\frac{\lambda_j}{d_j} \le 1 + \frac{\sigma^2}{d_j} \le \frac{\sigma^2}{d_k} \le \tilde{\rho}$ and $\left|q_i'u_j\right| \le 1$ by the Cauchy-Schwarz inequality, so

$$\left|e_i'(D^{-1}\Lambda_1 - I_k)Q_1'Ue_j\right| \le \tilde{\rho} \qquad \text{(A21)}$$

Combining equations A19, A20, and A21,

$$\begin{aligned}\left|e_i'\left(D^{-1}Q_1'UD - Q_1'U\right)e_j\right| &= \left|e_i'\left(D^{-1}\Lambda_1 - I_k\right)Q_1'Ue_j - e_i'D^{-1}Q_1'\Psi Ue_j\right| \\ &\le \left|e_i'\left(D^{-1}\Lambda_1 - I_k\right)Q_1'Ue_j\right| + \left|e_i'D^{-1}Q_1'\Psi Ue_j\right| \le 2\tilde{\rho}\end{aligned}$$

i.e. $\left|\left(\frac{d_i}{d_j} - 1\right)q_i'u_j\right| \le 2\tilde{\rho} \quad \Rightarrow \quad \left|q_i'u_j\right| \le 2\tilde{c}\tilde{\rho}$ for $i \ne j$. #

Lemma 15: $1 - \tilde{\rho} \le \sum_{j=1}^{k}\left(q_i'u_j\right)^2$

Proof of Lemma 15: As shown in the proof to Lemma 13 $1 - \tilde{\rho} \le 1 - \rho$ so the result follows from Lemma 8. #

Lemma 16: If k = 1 or $\tilde{c} \le \frac{1 - \tilde{\rho}}{2\tilde{\rho}\sqrt{(k-1)(1-\tilde{\rho})}}$, then there exists a sign matrix L such that $E\left\|s_t - Lf_t\right\|^2 \le k\left(2\tilde{\rho} + \tilde{\rho}^2\left(4\tilde{c}^2(k-1)(1-\tilde{\rho}) - 1\right)\right)$

Proof of Lemma 16:

$$\begin{aligned}E\left\|s_t - Sf_t\right\|^2 &= \mathrm{tr}\left[\left(\Lambda_1^{-\frac{1}{2}}Q_1'UD^{\frac{1}{2}} - L\right)\left(D^{\frac{1}{2}}U'Q_1\Lambda_1^{-\frac{1}{2}} - L\right) + \Lambda_1^{-\frac{1}{2}}Q_1'\Psi Q_1\Lambda_1^{-\frac{1}{2}}\right] \\ &= 2\mathrm{tr}\left(I - L\Lambda_1^{-\frac{1}{2}}Q_1'UD^{\frac{1}{2}}\right)\end{aligned}$$

From Lemma 15 $1-\tilde{\rho}-\sum_{j\neq i}^{k}\left(q_i'u_j\right)^2 \leq \left(q_i'u_i\right)^2$. If k = 1 then $\tilde{c}$ = 0 and the result holds with sign(L_{jj})=sign(q′u). If k > 1 then using Lemma 14 $1-\tilde{\rho}-4\tilde{c}^2\tilde{\rho}^2(k-1)\leq\left(q_i'u_i\right)^2$. From Theorem 13

$1-\tilde{\rho}\leq\frac{d_i}{\lambda_i}$, so $\left(1-\tilde{\rho}\right)^2-4\tilde{c}^2\tilde{\rho}^2(k-1)\leq\frac{d_i}{\lambda_i}\left(q_i'u_i\right)^2$. If $\tilde{c}\leq\frac{1-\tilde{\rho}}{2\tilde{\rho}\sqrt{(k-1)(1-\tilde{\rho})}}$ then the left hand side is non-negative and $\sqrt{\left(1-\tilde{\rho}\right)^2-4\tilde{c}^2\tilde{\rho}^2(k-1)}\leq\sqrt{\frac{d_i}{\lambda_i}}\left|q_i'u_i\right|$

If we choose L so that $\text{sign}(L_{ii})=\text{sign}(q_i'u_j)$, we get

$E\left\|s_t-Lf_t\right\|^2\leq k-k\sqrt{(1-\tilde{\rho})^2-4\tilde{c}^2\tilde{\rho}^2(k-1)(1-\tilde{\rho})}$. Multiplying this by $\frac{1+\sqrt{(1-\tilde{\rho})^2-4\tilde{c}^2\tilde{\rho}^2(k-1)(1-\tilde{\rho})}}{1+\sqrt{(1-\tilde{\rho})^2-4\tilde{c}^2\tilde{\rho}^2(k-1)(1-\tilde{\rho})}}$ yields the result. #

Proof of Theorem 4: Under assumptions 2,4,5 and 6 $\frac{1}{T}X'X=\Omega+O_p\left(T^{-\frac{1}{2}}\right)$. From assumptions 1 and 2 and using the proof of Theorem 1(a) we have

$\frac{1}{N}\sum_{j=1}^{N}\lambda_j\leq\frac{1}{N}\sum_{j=1}^{k}d_j+\frac{1}{N}\text{tr}\left(\Psi\right)<\infty$. Under these conditions Lemma 10, Lemma 11, Theorem 3(a) and Theorem 3(b) hold. From Lemma 13 and assumption 1 we have $\frac{\lambda_i}{N}=\frac{d_i}{N}+O\left(\tilde{\rho}\right)$ for i=1,..,k. Combining this with the results of Theorem 3(a) and Theorem 2 gives $\frac{\hat{\lambda}_i}{N}=\frac{d_i}{N}+O\left(N^{\tau-1}\right)+O_p\left(T^{-\frac{1}{2}}\right)$ proving Theorem 4(a).

Combining the result of Theorem 3(b) with Lemma 16 and Theorem 2 yields

$\hat{s}_{jt}=L_{jj}f_{jt}+O_p\left(N^{\frac{\tau-1}{2}}\right)+O_p\left(T^{-\frac{1}{2}}\right)$ proving Theorem 4(c).

From Theorem 4(c) $\frac{1}{T}X'\hat{S}_1=\frac{1}{T}X'FL+\frac{1}{T}X'Z_3$ where $Z_3\sim O_p\left[\max\left(T^{-\frac{1}{2}},N^{\tau-1}\right)\right]$. Under assumptions 4, 5 and 6 $\frac{1}{T}X'FL=B+O_p\left(T^{-\frac{1}{2}}\right)$ and $\frac{1}{T}X'Z_3\sim O_p\left[\max\left(T^{-\frac{1}{2}},N^{\tau-1}\right)\right]$ proving Theorem 4(b).

Also $\hat{\beta}_S = \frac{1}{T}\hat{S}_1'y = \frac{1}{T}LF'y + \frac{1}{T}Z_3 y$ where $Z_3 \sim O_p\left[\max\left(T^{-\frac{1}{2}}, N^{\tau-1}\right)\right]$. Under assumptions 4, 5 and 6 $\frac{1}{T}F'y = \beta + O_p\left(T^{-\frac{1}{2}}\right)$ and $\frac{1}{T}Z_3 y \sim O_p\left[\max\left(T^{-\frac{1}{2}}, N^{\tau-1}\right)\right]$ proving Theorem 4(d).

From Theorems 4(c) and (d) $\hat{\beta}_S'\hat{s}_{T+h} = \beta' f_{T+h} + \beta' L z_2 + z_1' L f_{T+h} + z_1' z_2$ where $z_1 = \hat{\beta}_S - L\beta$ and $z_2 = \hat{s}_{T+h} - Lf_{T+h}$. Under assumptions 4, 5, 6 and 7 and using Theorems 4(c) and (d), Theorem 4(e) is proved. #

References

ALTISSIMO, F., A. BASSANETTI, R. CRISTADORO, M. FORNI, M. LIPPI, L. REICHLIN, and G. VERONESE (2001): "Eurocoin: A Real Time Coincident Indicator of the Euro Area Business Cycle," CEPR.

ANDERSON, T. W., and H. RUBIN (1956): "Statistical Inference in Factor Analysis," *Third Berkeley Symposium on Mathematical Statistics and Probability*, 5, 111-150.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-71.

BAI, J., and S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

— (2005): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions."

BAI, Z. D. (1999): "Methodologies in Spectral Analysis of Large Dimensional Random Matrices, a Review," *Statistica Sinica*, 9, 611-677.

BENTLER, P. M., and Y. KANO (1990): "On the Equivalence of Factors and Components," *Multivariate Behavioral Research*, 25, 67-74.

BERNANKE, B. S., and J. BOIVIN (2003): "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics*, 50, 525-546.

BOIVIN, J., and S. NG (2005): "Are More Data Always Better for Factor Analysis," *Journal of Econometrics*, forthcoming.

CHAMBERLAIN, G., and M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281-1304.

CONNOR, G., and R. A. KORAJCZYK (1986): "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373-94.

FORNI, M., M. HALLIN, M. LIPPI, and L. REICHLIN (2000): "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540-54.

— (2003): "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?," *Journal of Monetary Economics*, 50, 1243-1255.

— (2004): "The Generalized Dynamic Factor Model Consistency and Rates," *Journal of Econometrics*, 119, 231-255.

FORNI, M., and M. LIPPI (2001): "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory*, 17, 1113-41.

GILL, R. D. (1977): "Consistency of Maximum Likelihood Estimators of the Factor Analysis Model, When the Observations Are Not Multivariate Normally Distributed," in *Recent Developments in Statistics*, ed. by J. B. Barra, F. Brodeau, G. Romier, and B. Van Cutsem. Amsterdam: North-Holland.

GILLITZER, C., J. KEARNS, and A. RICHARDS (2005): "The Australian Business Cycle: A Coincident Indicator Approach," Reserve Bank of Australia.

INKLAAR, R. J., J. JACOBS, and W. ROMP (2003): "Business Cycle Indexes: Does a Heap of Data Help?," University of Groningen.

JOHNSON, N. L., and S. KOTZ (1972): *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

JÖRESKOG, K. G. (1967): "Some Contributions to Maximum Likelihood Factor Analysis," *Psychometrika*, 32, 443-482.

LAWLEY, D. N., and A. E. MAXWELL (1971): *Factor Analysis as a Statistical Method*. Butterworths.

LUKACS, E. (1975): *Stochastic Convergence*. New York: Academic Press.

SCHNEEWEISS, H. (1997): "Factors and Principal Components in the near Spherical Case," *Multivariate Behavioral Research*, 32, 375-401.

SCHNEEWEISS, H., and H. MATHES (1995): "Factor Analysis and Principal Components," *Journal of Multivariate Analysis*, 55, 105-124.

SNOOK, S. C., and R. L. GORSUCH (1989): "Component Analysis Versus Common Factor Analysis: A Monte Carlo Study," *Psychological Bulletin*, 106, 148-154.

STOCK, J. H., and M. W. WATSON (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167-79.

— (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147-62.