CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Determining the Number of Primitive Shocks in Factor Models

Jushan Bai and Serena Ng

CEA@Cass Working Paper Series

WP–CEA–13-2006

Determining the Number of Primitive Shocks in Factor Models

Jushan Bai* Serena Ng †

January 29, 2006

Abstract

A widely held but untested assumption underlying macroeconomic analysis is that the number of shocks driving economic fluctuations, q, is small. In this paper, we associate q with the number of dynamic factors in a large panel of data. We propose a methodology to determine q without having to estimate the dynamic factors. We first estimate a VAR in r static factors, where the factors are obtained by applying the method of principal components to a large panel of data. The eigenvalues of the residual covariance or correlation matrix are then computed. We then test if their eigenvalues satisfy an asymptotically shrinking bound that reflects sampling error. The procedure is applied to determine the number of primitive shocks in a large number of macroeconomic time series. An important aspect of the present analysis is to make precise the relationship between the dynamic factors and the static factors, which is a result of independent interest.

Keywords and phrases: dynamic factor models, static factors, number of factors, common shocks, principal components analysis.

*Jushan Bai is Professor of economics, New York University, 269 Mercer St, New York, NY 10003 (email: Jushan.Bai@nyu.edu).

†Serena Ng is Professor of Economics, University of Michigan, Ann Arbor, MI 48109 (email: Serena.Ng@umich.edu).

The authors would like to thank two anonymous referees for many constructive comments. Financial support from the NSF (grants SES-0137084, SES-0136923) is gratefully acknowledged.

1 Introduction

A common working assumption in macroeconomics is that economic fluctuations are driven by a small number of shocks. It would not be too controversial to suggest that the number of shocks is no larger than four. It is in fact not easy to find a business cycle model built from microfoundations that has more than four shocks. Indeed, macroeconomists have been preoccupied with understanding the transmission and quantifying the importance of three shocks:- technology, monetary and fiscal policy. But just exactly what is the number of primitive shocks in the data? In this paper, a simple testing procedure will be proposed to determine this number, which we will denote by q. More precisely, the q that we determine is the rank of the spectral density matrix of the common components in a large panel of data, or equivalently, the number of common factors in a dynamic factor model. However, we do so without having to estimate a dynamic factor model.

Surprisingly, few test exists to formally evaluate what exactly is q. Using a dynamic index model to analyze quarterly data for fourteen series over the sample 1950:1-1970:1, Sargent and Sims (1977) rejected the one and the two index model in favor of a model with more factors, though they noted that the two index model fits the real variables quite well. In two recent papers, Forni et al. (2003) and Giannone et al. (2005) argued that the number of macroeconomic shocks, which they referred to as the stochastic dimension of the economy, is two. They first estimate common factors from quarterly data on 190 series over the sample 1970-1996. They arrived at the conclusion of two shocks using a reasonable albeit informal judgment that two dynamic factors explain about 60% of the variation in twelve macroeconomic aggregates. However, this does not mean that two factors is optimal for the panel of data from which the factors are extracted. Furthermore, changing the cut-off point from 60% to 80% would lead to a stochastic dimension twice as large. Because there does not exist a formal test for the number of dynamic factors, their conclusion that q is two remains very much an assertion.

The analytical framework used in Forni et al. (2003) and Giannone et al. (2005) is the so called dynamic factor model. Like the static factor model favored by Stock and Watson (2002a), the dynamic factor model also summarizes information in a large panel of data using a small number of factors. The important distinction is that rank of the spectrum of q dynamic factors is always q. Because the $r < \infty$ static factors can be dynamically related, the spectrum of $r \geq q$ static factors has reduced rank. We will see that this rank is actually q, the number of dynamic factors. Accordingly, we refer to q as the number of primitive shocks.

In Section 2, we motivate the procedures in the context of a canonical VAR, in which the number of underlying shocks is less than the number of variables. Tests for the number of dynamic factors are formally developed in Sections 3 and 4. Simulations and an empirical application are considered in Section 5. The last section concludes.

2 The Minimal Number of Primitive Shocks in a VAR

Consider a vector of observed stationary time series, F_t $(r \times 1)$, $t = 1, \ldots T$. Assume that F_t is a VAR process of order p such that

$$A(L)F_t = u_t \tag{1}$$

where $A(L) = I - A_1 L - \ldots - A_p L^p$. Throughout, we assume the roots of $A(L) = 0$ all lie outside of the unit circle, and u_t are iid with $E\|u_t\|^{4+\delta} < M < \infty$ for some $\delta > 0$. We consider the case in which u_t is driven by a vector of lower dimensional shocks. Consideration of such a VAR structure is useful in developing our main analysis, which concerns distinguishing the dynamic factors from the static factors.

Definition 1 *We say that the VAR process F_t is driven by a minimal number of q innovations if there exists a $r \times q$ matrix, R, with rank q such that*

$$u_t = R\varepsilon_t \tag{2}$$

where ε_t is a $q \times 1$ vector of innovations that are mutually uncorrelated, i.e. $\Sigma_\varepsilon = E(\varepsilon_t \varepsilon_t')$ is diagonal. If we define $\Sigma_u = E(u_t u_t')$, then under (2), $\Sigma_u = R\Sigma_\varepsilon R'$ has rank $q \leq r$.

In Bernanke (1986), u_t are the residuals of an estimated VAR, and R is assumed to be full rank. The number of primitive shocks thus equals the number of variables in the system. We allow the rank of R to be less than r. It is in this sense that we are looking for the minimal number of primitive shocks. The number of primitive shocks in u_t is simply the q linearly independent shocks that span u_t.

Suppose B is an arbitrary $r \times r$ matrix with rank q and v_t is vector of $r \times 1$ shocks. Let $u_t = Bv_t$. Then u_t can be expressed as $u_t = R\varepsilon_t$, where R is $r \times q$ and ε_t is $q \times 1$. To determine q, we use as starting point that a $r \times r$ semi-positive definite matrix A of rank q has q non-zero eigenvalues. Let $c_1 > c_2 \geq \ldots \geq c_r \geq 0$ be the ordered eigenvalues (at least one non-zero eigenvalue) and define

$$
\begin{aligned}
D_{1,k} &= \left(\frac{c_{k+1}^2}{\sum_{j=1}^r c_j^2}\right)^{1/2} \\
D_{2,k} &= \left(\frac{\sum_{j=k+1}^r c_j^2}{\sum_{j=1}^r c_j^2}\right)^{1/2}.
\end{aligned}
$$

Under the assumption of $rank(A) = q$, $c_k = 0$ for $k > q$. Thus $D_{1,k} = D_{2,k} = 0$ exactly, for $k \geq q$.

A different interpretation of our test can be obtained using the spectral decomposition of A:

$$A = \sum_{j=1}^{r} c_j \beta_j \beta_j'$$

where β_j is the eigenvector corresponding to c_j. Define the k^{th} pseudo matrix of A as

$$A(k) = \sum_{j=1}^{k} c_j \beta_j \beta_j'$$

If A has $r-q$ eigenvalues that are zero, then $A = A(k)$ for $k = q+1, \ldots r$. If $d_k = \text{vec}(A(k))$, with $d_0 = \text{vec}(A)$, then

$$
\begin{aligned}
D_{1,k} &\equiv \|d_{k+1} - d_k\| \,/\, \|d_0\| \\
D_{2,k} &\equiv \|d_k - d_0\| \,/\, \|d_0\|.
\end{aligned}
$$

It follows that our eigenvalue tests are the square-root of the deviations from the null hypothesis, as measured by the matrix norm. This follows from $\text{trace}(\beta_j \beta_j') = \|\beta_j\|^2 = 1$ by construction, and $d_0^2 = \text{vec}(A)'\text{vec}(A) = \text{tr}(A'A) = \|A\|^2$.

In the next two sections, the A matrix whose rank is to be determined is Σ_u, the covariance matrix of a set of innovations. However, Σ_u and F are not observed, but it can be estimated from the data, denoted by $\widehat{\Sigma}_u$. Let $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ be constructed from the eigenvalues of $\widehat{\Sigma}_u$. It will be shown that $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ converge to zero ($k \geq q$) asymptotically at a rate that depends on the convergence rate of $\widehat{\Sigma}_u$ to Σ_u.

Before turning to our main analysis, a remark on existing rank tests is in order. Available tests seek to determine the rank of a $m \times n$ matrix, say, R, when R is consistently estimated from the regression $u_t = R\varepsilon_t + v_t$. First of all, our focus is in problems for which v_t plays no role, but with unobservable u_t. Furthermore, rank tests tend not to be asymptotically normal, see Anderson (1951), Gill and Lewbel (1992), Cragg and Donald (1996), Cragg and Donald (1997), and Robin and Smith (2000). Recently, Kleibergen and Paap (2003) and Ratsimalahelo (2003) suggest orthogonal rotation of the sample eigenvalues around the origin to restore normality. In doing so, these tests necessitate a consistent estimate of $\text{var}(\widehat{R})$. The matrix whose rank we seek to test is Σ_u. This creates two problems. First, the estimation of $\text{var}(\widehat{\Sigma}_u)$ entails evaluation of a matrix of fourth moments, which tend to be quite imprecisely estimated unless the sample size is extremely large. Second, Σ_u is a variance covariance matrix which only has $r(r+1)/2$ unique elements. Thus, $\text{var}(\Sigma_u)$ and its estimate do not have full rank, an assumption maintained by Kleibergen and Paap (2003). The test of Ratsimalahelo (2003) allows for reduced rank in the variance of apparently matrices that are not symmetric. Our attempts to adopt existing tests have not been successful. This motivates the development of a test using bounds guided by the convergence rate of $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$. We begin with the intermediate case when F_t is observed, but Σ_u is not.

Proposition 1 *Let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t'$, where $\widehat{u}_t$ are the residuals from estimation of a VAR in F_t, where F_t is observed. Let $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ be the estimated $D_{1,k}$ and $D_{2,k}$ using the eigenvalues of $\widehat{\Sigma}_u$. For some $0 < m < \infty$ and $0 < \delta < 1/2$, define*

$$\begin{aligned} \mathcal{K}_1 &= \{k : \widehat{D}_{1,k} < m/T^{1/2-\delta}\} && (3) \\ \mathcal{K}_2 &= \{k : \widehat{D}_{2,k} < m/T^{1/2-\delta}\}. && (4) \end{aligned}$$

Let $\widehat{q}_1 = \min\{k \in \mathcal{K}_1\}$ and $\widehat{q}_2 = \min\{k \in \mathcal{K}_2\}$. Then under H_0 that rank(Σ_u)=q, $\widehat{q}_1 \xrightarrow{p} q$ and $\widehat{q}_2 \xrightarrow{p} q$ as $T \to \infty$.

The proposition follows from $\sqrt{T}(\widehat{\Sigma}_u - \Sigma_u) = O_p(1)$. By continuity of eigenvalues $\widehat{D}_{1,k} = D_{1,k} + O_p(T^{-1/2})$, and likewise for $\widehat{D}_{2,k}$. From $\widehat{D}_{1,k} - D_{1,k} = O_p(T^{-1/2})$, we have $\widehat{D}_{1,k} = O_p(T^{-1/2})$ for $k \geq q$ because $D_{1,k} = 0$. Thus, $\widehat{D}_{1,k} < m/T^{1/2-\delta}$ with probability tending to 1 as $T \to \infty$. This means that $q \in \mathcal{K}$ for large T. But $q-1$ does not belong to $\mathcal{K}$ because $\widehat{D}_{1,k} > c > 0$ and thus greater than $m/T^{1/2-\delta}$ for $k < q$. This gives the consistency result. Essentially, the cut-off point $m/T^{\frac{1}{2}-\delta}$ is the tolerated error induced by sampling variability from estimation of Σ_u. An analogous argument holds for $D_{2,k}$. In large samples, the two tests should arrive at the same conclusion.

Thus far, the VAR process F_t is assumed to be observed. In the next two sections, F_t is a vector of unobserved common factors that are shared by a large number of series x_{it}.

3 Dynamic versus Static Factor Models

There are two types of factor models in the econometrics literature. The *static* model is written as $x_{it} = \Lambda_i' F_t + e_{it}$, where $i = 1, \ldots N, \quad t = 1, \ldots T$. In the language of factor analysis, e_{it} is referred to as the idiosyncratic error, Λ_i is a vector of factor loadings for unit i on the r (static) common factors F_t. The meaning of static factor model refers to the static relationship between x_{it} and F_t, but F_t itself can be a dynamic process. The *dynamic* factor model is written as $x_{it} = \lambda_i'(L) f_t + e_{it}$ where $\lambda_i(L)$ is a vector of dynamic factor loadings of order s. As a matter of notation, the model is a 'dynamic factor model' if s is finite, and a 'generalized dynamic factor model' if s can be infinity. In either case, $f_t = C(L)\varepsilon_t$, where ε_t are iid vectors and $x_{it} = \lambda_i(L)C(L)\varepsilon_t + e_{it}$. In this paper, we consider dynamic factor models (finite s). The dimension of f_t, which is the same as the dimension of ε_t, is called the number of dynamic factors; this is denoted by q. Dynamic factor models with s finite can be written as static factor models with r finite, but the dimension of F_t is in general different from the dimension of f_t since F_t includes the leads and lags of f_t with $r \geq q$. In practice, F_t is estimated using an eigenvalue-eigenvector decomposition of the sample covariance matrix of the data, while the dynamic estimates are based on a eigenvalue decomposition of the spectrum smoothed over various frequencies. Recent research showed that the space spanned by the static as well as the dynamic factors can be consistently estimated when N and T are both large. See, e.g., Forni et al. (2000), Ding and Hwang (2001), Stock and Watson (2002a), Forni et al. (2005), Forni and Lippi (2001), Bai and Ng (2002), and Bai (2003).

The ability to consistently estimate the factor space has opened up new horizons for empirical research. Using the factor estimates to summarize information in a data rich environment has been found useful in forecasting exercises and in understanding the conduct of monetary policy. See, for example, Stock and Watson (2002b), and Bernanke and Boivin (2003). While for forecasting purposes, little is to be gained from a clear distinction between the static and the dynamic factors, many economic analysis hinge on the ability to isolate the primitive shocks, or in other words, the number of dynamic factors.

In Bai and Ng (2002), we showed that under certain conditions, information criteria with appropriately chosen penalties will consistently estimate r, where r was assumed finite. We will ultimately propose a way to determine q from r estimated static factors, where r is assumed finite, or in terms of the parameters of the dynamic model, s is finite. But before we can proceed with such an analysis, we need to make precise the relation between the dynamic and the static factors, treating F_t and f_t as though they are observed. In the remainder of this section, it will be shown that a dynamic factor model always has a static factor representation and in which the dynamics of F_t is characterized by a VAR whose order depends on the dynamics of f_t. We will see from the VAR representation that spectrum of the static factors has rank q.

3.1 Putting the Dynamic Model into Static Form

Consider the dynamic factor model

$$\begin{aligned} x_{it} &= \lambda_{i0}' f_t + \lambda_{i1}' f_{t-1} + \cdots + \lambda_{is}' f_{t-s} + e_{it} \\ &= \lambda_i'(L) f_t + e_{it} \end{aligned} \tag{5}$$

where f_t is q dimensional, and

$$\lambda_i(L) = \lambda_{i1} + \lambda_{i2} L + \cdots + \lambda_{is} L^s. \tag{6}$$

It is clear that we can rewrite (5) in the static form

$$x_{it} = \Lambda_i' F_t + e_{it} \tag{7}$$

where

$$\Lambda_i = \begin{bmatrix} \lambda_{i0} \\ \lambda_{i1} \\ \vdots \\ \lambda_{is} \end{bmatrix} \quad \text{and} \quad F_t = \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-s} \end{bmatrix}. \tag{8}$$

The above is a simple mathematical identity and is true whether f_t itself is AR process or MA process. The dimension of F_t is always equal to

$$r = q(s+1),$$

where q is the dimension of f_t. As mentioned earlier, while the relation between x_{it} and F_t is static; F_t itself can be a dynamic process whose precise characterization will depend on the dynamics of f_t. We consider two cases: f_t is a finite order autoregressive process, and f_t has a moving-average structure.

Case I: f_t is AR(h), h finite, i.e

$$(I_q - B_1 L - \ldots - B_h L^h) f_t = \varepsilon_t \tag{9}$$

Note that f_t is q dimensional, and the number of static factors is r, does not depend on h, the order of the dynamic process governing f_t in (9). Unless $s = 0$, the number of static factors is larger than the number of dynamic factors.

We now want to establish that F_t is a VAR whose order depends on h and s. To see the VAR representation of F_t, we put f_t into a state space form. Let $\kappa = \max(h, s)$, and define $B_{h+1} = \cdots = B_\kappa = 0$, then

$$\begin{pmatrix} f_t \\ f_{t-1} \\ . \\ . \\ f_{t-\kappa} \end{pmatrix} = \begin{pmatrix} B_1 & B_2 & . & . & . & B_{\kappa+1} \\ I_q & 0 & 0 & . & . & 0 \\ 0 & I_q & & . & . & 0 \\ . & . & . & . & . & . \\ 0. & . & . & I_q & . & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ . \\ . \\ f_{t-\kappa-1} \end{pmatrix} + \begin{pmatrix} I_q \\ 0 \\ . \\ . \\ 0 \end{pmatrix} \varepsilon_t.$$

Define $F_t^* = [f_t' \; f_{t-1}' \; \ldots f_{t-\kappa}']'$. We have

$$\begin{aligned} F_t^* &= AF_{t-1}^* + u_t \qquad (10) \\ u_t &= R\varepsilon_t \end{aligned}$$

where A is square matrix of dimension $q \cdot (\kappa + 1)$, and R is a $q(\kappa + 1)$ by q matrix. In traditional state-space representation, $\kappa = h$. In our present case, $\kappa = \max(h, s)$. If $s \geq h$ then

$$F_t \equiv F_t^*$$

so F_t also has a VAR(1) representation. When $s < h$, F_t is a sub-vector of F_t^*. In general, any sub-vector of a VAR is a vector ARMA process, not necessarily VAR. However, due to the special structure of F_t^*, the sub-vector F_t itself is a VAR. This point can easily be made clear with an illustration. Suppose $h = 3$, and $s = 1$, and consider

$$f_t = B_1 f_{t-1} + B_2 f_{t-2} + B_3 f_{t-3} + \varepsilon_t$$

so that f_t is VAR(3). Let $F_t = (f_t', f_{t-1}')'$. Clearly,

$$\begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \end{bmatrix} + \begin{bmatrix} 0 & B_3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f_{t-2} \\ f_{t-3} \end{bmatrix} + \begin{bmatrix} I \\ 0 \end{bmatrix} \varepsilon_t.$$

This implies that F_t is VAR(2). In fact, we can show that for the general situation F_t is VAR(p) with $p = \max(1, h - s)$. Therefore, the dynamic factor model defined by (5)-(6) when f_t is a VAR(h) can be written as a static factor model in the following form

$$F_t = A_1 F_{t-1} + A_2 F_{t-2} + \cdots + A_p F_{t-p} + u_t \qquad (11)$$

$$u_t = (I_q, 0, ..., 0)' \varepsilon_t \qquad (12)$$

Case II: f_t is $MA(h)$ Consider

$$f_t = \varepsilon_t + C_1\varepsilon_{t-1} + \cdots + C_h\varepsilon_{t-h} \tag{13}$$

In this case, $F_t = (f_t', f_{t-1}', ..., f_{t-s}')'$, and F_t is $q(s+1) \times 1$. That is, the number of static factor is $r = q(s+1)$, irrespective of h, the order of f_t.

Under invertibility assumption, the moving average process f_t can be expressed as an VAR(∞) process, which can be approximated by a finite order VAR. This implies that F_t can also be approximated by a finite order VAR as in (11) and (12), see Berk (1974) and Kuersteiner (2004) for theoretical results, particularly when the coefficients of the AR process decay quickly to zero. In general, the lag length should be chosen by a data-dependent method, and is an increasing function of T. While this procedure works well in practice, there are theoretical issues left unexplored when the VAR order increases with T. Proposition 2 below is stated only for VARs of fixed orders, and in effect, it does not cover case II. Nevertheless, we remark that it is possible to extend the theory to include VAR(∞) process with rapidly decaying coefficients. In particular, it is possible to extend the theory to F_t being vector ARMA processes. But this extension is not considered for simplicity. Monte Carlo simulations show that the procedure works well for f_t being MA processes.

The point to highlight is that data generated by the dynamic model can always be mapped into a static model of the form $x_{it} = \Lambda_i' F_t + e_{it}$ by suitably defining a F_t that evolves according to a VAR whose order will depend on the dynamics of f_t. The dimension of F_t is always $r = q(s+1)$ irrespective of the order of the VAR. Given that $A(L)F_t = R\varepsilon_t$, the spectrum of F at frequency ω

$$S_F(\omega) = A(e^{-i\omega})^{-1} R S_\varepsilon(\omega) R' A(e^{i\omega})^{-1},$$

has rank q if $S_\varepsilon(\omega)$ has rank q for all $|\omega| \leq \pi$. Accordingly, the spectrum of the static factors $S_F(\omega)$ will also have q non-zero eigenvalues. We therefore refer to the dynamic factors, q, as the number of primitive shocks.

Although the dynamics of the static factors are in the same form as the observable VAR system in (1) in the sense that both are driven by shocks with dimension less than the dimension of the variables, Proposition I cannot be used immediately to determine q. This is because F_t is not observable, and the convergence rate of $\widehat{\Sigma}_u$ is not $\sqrt{T}$. These issues are dealt with in the next section.

4 Determining q

Let $S_x(\omega)$ be the population spectrum of the N cross-section units. The static model implies

$$S_x(\omega) = \Lambda S_F(\omega)\Lambda' + S_e(\omega), \quad -\pi \leq \omega \leq \pi.$$

Since $S_e(\omega)$ has rank N, $S_x(\omega)$ is also rank N. This would seem to suggest that q cannot be determined without working on $S_F(\omega)$. Such a procedure would necessitate the choice of many auxiliary parameters (such as bandwidth and kernel), and even then, we do not have a formal theory for determining q. We now show how q can be estimated in the time domain, and limiting distributions of the eigenvalues are not necessary.

If F_t were observed, and since it has a VAR representation, Proposition 1 then implies that q can be determined from a spectral decomposition of $\widehat{\Sigma}_u$ provided T is large. What prevents such an analysis is that neither F_t nor its dimension (r) is observed. However, the following holds. Let $\widehat{F}_t^r$ be the r factors obtained by the method of principal components. That is, let $\widehat{\Lambda}$ be a $N \times r$ matrix consisting of the r eigenvectors (multiplied by $\sqrt{N}$) associated with the r largest eigenvalues of the matrix $X'X$ in decreasing order. Then $\widehat{F} = X\widehat{\Lambda}/N$. These principal component estimates adopt the normalization that $\widehat{\Lambda}'\widehat{\Lambda}/N = I_r$. Then under the assumption that (i) $\Sigma_F = E(F_tF_t')$ and $\Lambda'\Lambda/N$ are both rank r, (ii) moment restrictions are satisfied, and (iii) that the time and cross-section correlation in the idiosyncratic errors is weak, Bai and Ng (2002) and Bai (2003) showed that if the data are generated by the static factor model, then as $N, T \to \infty$, there exists a matrix H, of rank r, such that as $N, T \to \infty$ (jointly),

$$\min[N, T]\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{F}_t^r - HF_t\right\|^2\right) = O_p(1)$$

and

$$\text{prob}(\widehat{k} = r)\xrightarrow{p}1$$

where

$$\widehat{k} = \underset{k}{\text{argmin}}\ IC(k) = \underset{k}{\text{argmin}} \quad \log(\sigma_k^2) + kC_{NT}$$

with $C_{NT} \to 0$ but $\min[N, T]C_{NT} \to \infty$ as $N, T \to \infty$.

Importantly, the above large sample results assume that the second moment matrix of Λ_i and F_t are rank r. But Σ_F may have rank less than r. For example, if $F_t = AF_{t-1} + R\varepsilon_t$ is such that $A = \rho I_r$, $|\rho| < 1$, then $\text{var}(F_t) = R\Sigma_\varepsilon R'/(1-\rho^2)$, $\text{var}(F_t)$ only has rank $r^* = q$. In general, when the dynamics of F_t is rich enough, $\text{var}(F_t)$ is of rank r even though the rank of Σ_u is only q. But existing results do not cover cases when $\text{var}(F_t) < r$, which can arise as in the example above when F_t has very simple dynamics. We now extend our results on determining the number of factors to also cover these special cases.

Lemma 1 *Let F_t be a $r \times 1$ vector of factors generated by q primitive common shocks ε_t. Let $q \le r^* \le r$. Let $\Sigma_F = E(F_t F_t')$ and $\Sigma_\Lambda = \text{plim}\,\Lambda'\Lambda/N$. Suppose the $r \times r$ matrix $\Sigma_F \cdot \Sigma_\Lambda$ has rank r^* and the remaining assumptions of Bai and Ng (2002) hold. Let $\widehat{F}_t^{r*}$ be the $r^* \times 1$ vector of factor estimates obtained by the method of principle components. There exists a matrix H^* with rank r^* such that (i) $\min[N,T]\left(\frac{1}{T}\sum_{t=1}^T \left\|\widehat{F}_t^{r*} - H^* F_t\right\|^2\right) = O_p(1)$; (ii) $prob(\widehat{k} = r^*) = 1$ if $\widehat{k} = \text{argmin}_k\ IC(k)$.*

Lemma 1 clarifies that when F_t is reduced rank, the method of principal components will estimate the space spanned by the r^* independent factors. The IC will select $r^* \le r$ factors, since r^* is the rank of Σ_F (assuming Σ_Λ is of full rank).

Consider now the determination of q given $\widehat{F}_t$, where $\widehat{F}_t$ are the $\widehat{r}^*$ factors obtained by the static method of principal components. For notation simplicity, the dimension of $\widehat{F}_t$ is suppressed, but is understood to be of dimension $\widehat{r}^*$, where $\widehat{r}^*$ is determined by the IC. Let $\widehat{u}_t$ be the residuals from estimating a VAR(p) in $\widehat{F}_t$, and let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t \widehat{u}_t$. It should be remarked that since we can only estimate the space spanned by the factors, we need the residuals from estimation of a VAR in $\widehat{F}_t$. Performing $\widehat{r}^*$ univariate autoregressions for each component of $\widehat{F}_t$ will not be appropriate.

Lemma 2 *Consider the model $x_{it} = \lambda_i' F_t + e_{it}$ with $A(L)F_t = u_t$, where $A(L)$ is polynomial in the lag operator of order p. Let $\widehat{F}_t$ be the $r^* \times 1$ factors estimated by the method of principal components under the normalization that $\Lambda'\Lambda/N = I_{r^*}$, where $q \le r^* \le r$. Let $\widehat{u}_t$ be the residuals obtained by least squares estimation of a VAR(p) in $\widehat{F}_t$. Let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t' \widehat{u}_t$, and let $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T u_t u_t'$. Then*

$$\min[\sqrt{N}, \sqrt{T}](\widehat{\Sigma}_u - H^* \widetilde{\Sigma}_u H^{*\prime}) = O_p(1).$$

The ultimate interest is the rank of $\Sigma_u = E(u_t u_t')$. But $\sqrt{T}(\widetilde{\Sigma}_u - \Sigma_u) = O_p(1)$. Lemma 2 thus implies that

$$\left\|\widehat{\Sigma}_u - H^* \Sigma_u H^{*\prime}\right\| = O_p(1/\min[\sqrt{N}, \sqrt{T}]).$$

Note that Σ_u and $H^* \Sigma_u H^{*\prime}$ have the same rank since H^* is of full rank. We have $\widehat{D}_{1,k} - D_{1,k} = O_p(1/\min[\sqrt{N}, \sqrt{T}])$ Because $D_{1,k} = 0$ when $k \ge q$, we also have $\widehat{D}_{1,k} = O_p(1/\min[\sqrt{N}, \sqrt{T}])$. A similar result holds for $\widehat{D}_{2,k}$. Thus, we have

Proposition 2 *Let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t \widehat{u}_t'$, where $\widehat{u}_t$ are the residuals from estimation of a VAR in $\widehat{F}_t$, $\widehat{F}_t$ being the principal components estimator for F_t. For some $0 < m < \infty$ and $0 < \delta < 1/2$, let*

$$\mathcal{K}_3 = \{k : \widehat{D}_{1,k} < m/\min[N^{1/2-\delta}, T^{1/2-\delta}]\} \tag{14}$$

$$\mathcal{K}_4 = \{k : \widehat{D}_{2,k} < m/\min[N^{1/2-\delta}, T^{1/2-\delta}]\}. \tag{15}$$

Let $\widehat{q}_3 = \min\{k \in \mathcal{K}_3\}$ and $\widehat{q}_4 = \min\{k \in \mathcal{K}_4\}$. Then under H_0 with rank(Σ_u)=q, we have $\widehat{q}_3 \xrightarrow{p} q$ and $\widehat{q}_4 \xrightarrow{p} q$ as $N, T \to \infty$.

Our main insight is to exploit the relation between the dynamic and the static factors so that estimation of the dynamic factors is not necessary to determine q. In our setup, if r^* factors explain τ percent of the variation in the data, the q primitive factors will explain the same fraction (up to an error that vanishes asymptotically) of variation in the data. Importantly, r^* and τ are determined using well-defined ceritia. This is in contrast to Giannone et al. (2005), in which q is chosen for a subjectively chosen τ.

Our procedure provides a more formal way of determining the rank of $S_F(\omega)$ and is a useful cross-check to the informal method used in Giannone et al. (2005). In independent work that was completed the same time as the first draft of this paper was written, Stock and Watson (2005) also developed a test for q that uses a rather different approach. Instead of the rank of $S_F(\omega)$, they estimate the rank of the restricted residuals of a $N+r$-dimensional VAR. In our notation, Stock and Watson starts with $x_{it} = \Lambda_i' F_t + \rho_i(L) x_{it-1} + e_{it}$ to allow serial correlation in the idiosyncratic errors. The factor dynamics $A(L)F_t = Ru_t$ where $A(L) = I - A^+(L)L$ implies that $x_{it} = \Lambda_i' A^+(L) F_{t-1} + \Lambda_i' R\varepsilon_t + \rho_i(L) x_{it-1} + e_{it}$. The composite residuals of a VAR in X_t and F_t is of the form $\Lambda_i' R\varepsilon_t + e_{it}$, which has q common factors ε_t. Stock and Watson exploit this factor representation to determine q using the criteria developed in Bai and Ng (2002).

As written, $\widehat{q}_3$ and $\widehat{q}_4$ are the estimated rank of $\widehat{\Sigma}_u$, the sample covariance matrix of $\widehat{u}_t$. But the number of non-zero eigenvalues of $\widehat{\Sigma}_u$ is the same as the number of non-zero eigenvalues of $\widehat{S}_u$, the sample correlation matrix of $\widehat{u}_t$. In our experience, using one or the other matters only for the choice of m. We found $m = 1$ works for both $\widehat{q}_3$ and $\widehat{q}_4$ when $\widehat{\Sigma}_u$ is used. As we will see in the simulations, the preferred values of m are different for $\widehat{q}_3$ and $\widehat{q}_4$ when correlation matrix $\widehat{S}_u$ is used. Statistics based on correlation matrix are scale invariant. We report results for both cases in our simulations.

5 Simulations

We consider four data generating processes:

1. $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)' f_t + e_{it}$, where $f_t = C(L)\varepsilon_t$ and f_t is q dimensional;

2. $x_{it} = (\lambda_{i0} + \lambda_{i1}L)' f_t + e_{it}$, where $A(L)f_t = \varepsilon_t$ and f_t is q dimensional;

3. $x_{it} = \lambda_i' F_t + e_{it}$ where $F_t = A_1 F_{t-1} + u_t$, and $A_1 = \rho I_r$; $u_t = R\varepsilon_t$, $rank(R) = q$.

4. $x_{it} = \lambda_i' F_t + e_{it}$ where $F_t = A_1 F_{t-1} + u_t$, where $A_1 = diag(.2, .375, .55, .725, .9)$ and u_t is the same as in 3.

DGPs 1 and 2 are dynamic factor models considered by Forni et al. (2000) with $q = 2$ dynamic factors. DGP 1 assumes f_t is a bivariate moving average process with MA(1) parameters of .2 and .9 for f_{1t} and f_{2t}, respectively. DGP 2 assumes f_t is a bivariate first order autoregressive process with AR(1) parameters of .2 and .9 for f_{1t} and f_{2t}, respectively. DGP 1 has $r = q(s+1) = 6$ static factors and DGP 2 has $r = 4$ static factors. DGPs 3 and 4 are static factor models. The static factors are driven by $q = 3$ dimensional shocks. This DGP is used in Stock and Watson (2002a) and Bai and Ng (2002), among others. In DGP 3, $r = 5$ but the factors have common dynamics with $\rho = .5$. This implies $r^* = q = 3$. There are many ways of generating u_t of the form $u_t = R\varepsilon_t$. The particular method used is as follows. Let S be an $r \times r$ diagonal matrix of rank q with nonzero elements drawn from uniform $U(0.8, 1.2)$ distribution. Let Γ be an arbitrary orthonormal matrix $\Gamma\Gamma' = I_r$, obtained in matlab via "orth(rand(r,r))". Then u_t is generated as $u_t = \Gamma S \Gamma' v_t$, where v_t is an $r \times 1$ vector of iid normal variables. Note that Γ and S do not vary over t and i. The variance of u_t is $\Gamma S^2 \Gamma'$, having rank q. For DGP 4, A_1 is a diagonal matrix with values .2,.375,.55,.725, .9; the dynamics of F_t are thus richer than that of DGP3. In this case, $q = 3$ and $r = 5$. In addition, $r^* = r$ for large T, but r^* can be less than r for finite T. In all four DGPs, we assume λ_{ij}, e_{it}, and ε_t are iid standard normal.

For all four DGPs, the testing proceeds as follows. Given the data x_{it}, $i = 1, \dots N$, $t = 1, \dots T$, the static factors are estimated using the method of principal components with the normalization that $\Lambda'\Lambda/N = I_{r^*}$. The number of factors is estimated by the IC as in Bai and Ng (2002). Specifically,

$$\widehat{r}^* = \underset{k \in [0, 2r]}{\text{argmin}} \ \log(\widehat{\sigma}_k^2) + k \frac{\log(\min[N, T])}{NT/(N+T)},$$

where $\widehat{\sigma}_k^2 = \frac{1}{NT}\sum_i \sum_t (x_{it} - \widehat{\lambda}_i^k \widehat{F}_t^k)^2$, $\widehat{F}_t^k$ is $k \times 1$. Given $\widehat{F}_t$, a $\widehat{r}^*$ dimensional VAR in $\widehat{F}_t$ is estimated to obtain $\widehat{u}_t$. Selecting too few lags will be problematic as $\widehat{u}_t$ will not be innovations. We report results for VAR(2). Results for higher lags are similar. Given $\widehat{u}_t$, its $\widehat{r}^* \times \widehat{r}^*$ covariance matrix is constructed. Then $\widehat{q}_3$ and $\widehat{q}_4$ are obtained with $\delta = .1$ so that $m^* = \frac{m}{\min[N^{2/5}, T^{2/5}]}$. The number of simulations is 1000.

We report the average values for $\widehat{r}$ and $\widehat{q}$ in Table 1a estimated based on the covariance matrix of $\widehat{u}_t$. Table 1b reports the corresponding results based on the correlation matrix of $\widehat{u}_t$. The four panels correspond to four different data generating processes. For all cases, the number of static factors $\widehat{r}$ is determined by minimizing the IC(k) for k between 0 and $2r$. For small values of $\min[N, T]$, the IC tends to select a large number of static factors. However, even when $\widehat{r}^*$ is overestimated, $\widehat{q}$ can be very close to q for suitable choice of m. As noted earlier, when covariance matrix is used, $m = 1$ is suitable for both q_3 and q_4. But when the correlation matrix is used, m needs to be different for q_3 and q_4. In general, when m is too small, $\widehat{q}$ is larger than q. For all four DGPs we find that $m = 1.25$ works well for $\widehat{q}_3$, and $m = 2.25$ works well for $\widehat{q}_4$ (when using correlation matrices). Between $\widehat{q}_3$ and $\widehat{q}_4$, the former tends to have better properties when N or T is small.

5.1 Empirical Analysis: Shocks in the U.S.

To illustrate, we take data used in Stock and Watson (2005), which can be downloaded at `http://www.princeton.edu/~mwatson`. There are 132 monthly time series available from 1960:1 to 2003:12. The data are transformed (by taking logs, first or second difference) as in Stock and Watson. The objective is to determine the number of primitive, or dynamic factors, in this panel of data.

To get a sense of the importance of the factors in the data, we begin by determining $\widehat{q}$ for $r = 2, 3, \ldots 10$. The value of q is estimated using the correlation matrix. Almost identical results are obtained if covariance matrix is used. But when discrepancy exists, the estimated q from the latter method tends to be higher than that from the former. Thus to assert that q is larger than two, we use a method that is less favorable to the assertion. In this exercise, we do not take a stand on what is the optimal number of static factors in the data. We find that for the full sample of 528 observations, $\widehat{q} = r$ when $r = 2, 3$, $\widehat{q} = 3$, $r = 4, 5, 6$, $\widehat{q} = 4$ when r= 7, 8. Some of the static factors are linearly dependent in a dynamic sense. It is well known that the first two static factors in the data being analyzed are real factors. The finding that $\widehat{q} = 2$ given $r = 2$ indicates that the first two static factors are dynamically distinct.

Next, we allow the number of static factors, r_t, to be determined optimally for each t. We use two concepts of optimality. We first estimate $\widehat{r}_t(\tau)$ static factors, where $\widehat{r}_t(\tau)$ explains closest τ percent of the variation in the data up to time t, and then determine $\widehat{q}_t(\tau)$ given $\widehat{r}_t(\tau)$ factors. Note that $\widehat{r}_t(\tau)$ is not optimal from a statistical point of view. However, the result of Giannone et al. (2005) that $q = 2$ is based on the reasoning that two dynamic factors explain 60% of the variation in twelve variables. It is thus useful to consider results for cut-offs other than $\tau = .6$. We also determine the number of static factors using the IC. This is denoted by $\widehat{r}_t^*$, and the corresponding number of primitive factors is q_t^*. All these statistics are computed for t ranging from 133 to 528, corresponding to estimation ending in 1970:12 and 2003:12, respectively. We thus have 396 statistics, one for every t.

Reported in Table 2 are the mean, minimum and maximum of these statistics over the samples with ending dates from 1970:12 to 2003:12 (sample sizes from $T = 133$ to $T = 528$). $R^2_{\widehat{r}}$ is the average explanatory power of $\widehat{r}$ factors, when $\widehat{r}$ is chosen with cut-off of .3,.4,.5 or .6, and also optimally. The column $R^2_{\widehat{q}}$ is the average explanatory power of the $\widehat{q}$ shocks given $\widehat{r}$ innovations.

The results indicate that four static factors explain .348 of the variation in the full sample data, while six factors explain .438. To explain .619 of the variation in the data would require, on average, 15 factors. When determined optimally by the IC, the data suggest that 7 static factors explain, on average, .460 of the variation in the data over the full sample, and that there four dynamic factors spanning the seven static factors. Using an alternative method but the same data, Stock and Watson (2005) found seven dynamic and static factors. In an earlier version of this paper, the tests were applied to a different dataset for the sample 1960:1-1998:12. We found an average of seven dynamic factors in ten static factors. The evidence is thus very compelling that the number of dynamic factors is larger than two.

We stress once again that there is substantial variation over the sample. Figure 1 depicts the time series plot of $\widehat{r}_t^*, \widehat{q}_{3t}$ and $\widehat{q}_{4t}$. As we can see, $\widehat{r}$ jumped from 4 to 6 around 1973, and has remained roughly at 6 till 2000. On the other hand, $\widehat{q}$ jumped from 3 to 4 and stayed at 4 most of the time. If we had ended the estimation in 2000, we would have $\widehat{r} = 6$ and $\widehat{q} = 4$. However, in the past few years, $\widehat{r}$ seemed to have taken another jump from 6 to 7, though $\widehat{q}$ seems to have stayed at 4.

Figure 2 plots R^2_t, the fraction of variance in the data explained by the factors up to time t. The importance of common shocks exhibit an upward trend, increasing from .25 in the early 70s to peak at about .5 in the early 90s. The results seem to suggest that in the early 1970s, economic fluctuations are dominated by a small number of large common shocks.

More recently, the economy is hit with a larger number of smaller common shocks. Notably, for most of the sample, the stochastic dimension of the economy is at least four. While the full sample analysis obscures the fact that the optimal number of static and dynamic factors have changed over time, it remains the case that the average number of dynamic factors over the sample is more than two.

It remains to reconcile our finding that q exceeds two with the result of Giannone et al. (2005). Our analysis gives the optimal number of factors for the panel of data from which the factors are extracted. Given that $N = 132$ series, our finding suggests six dynamic factors is optimal in explaining the average variation in the data. In contrast, Giannone et al. (2005) first estimated the factors from close to 200 series. They then restrict their attention to only twelve series when arriving at the conclusion that q is two. Their conclusion should not be taken to mean that two dynamic factors best explain the variation in the panel of data from which the factors are extracted.

To highlight the difference, we calculate the explanatory power of the common factors in the full sample for a selected number of series: IPS10 (industrial production), A0M059 (retail trade), A0M057 (manufacturing trade), FYFF (Federal funds drate), PUNEW (CPI), and A0M224r (consumption expenditure). Reported are the R^2 from a regression of x_{it} on a constant and $\widehat{r}$ static factors, where x_{it} is log first difference of IPS10, A0M059, A0M057, and A0M224r, the first difference FYFF, and second difference of the logarithm of PUNEW. It is evident from Table 3 that the explanatory power of the factors tend to be higher for the selected series than for the panel as a whole. If we had focused on these series, fewer static factors would have been necessary. It is conceivable that four or five static factors adequately explain the selected series, and which would have implied two or three dynamic factors.

6 Conclusion

This paper proposes a procedure to determine the number of primitive common shocks in a large number of series. By making precise the link between the dynamic and the static factors, we arrive at a pair of tests that can determine the number of dynamic factors without having to estimate these factors themselves. This enables us to bypass the selection of many auxiliary parameters needed for estimation of the spectrum. The tests are easy to compute. Our tests suggest that the number of dynamic factors in the panel of 132 macroeconomic time series considered is four.

References

Anderson, T. W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics* **22**, 327–351.

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica* **71:1**, 135–172.

Bai, J. and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica* **70:1**, 191–221.

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

Bernanke, B. (1986), "Alternative Explanations of the Money-Income Correlation," *Carnegie Rochester Conference on Public Policy* **25**, 49–100.

Bernanke, B. and Boivin, J. (2003), "Monetary Policy in a Data Rich Environment," *Journal of Monetary Economics* **50:3**, 525–546.

Cragg, J. and Donald, S. (1996), "On the Asymptotic Properties of LDU Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association* **91**, 1301–1309.

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

Ding, A. and Hwang, J. (2001), "Prediction Intervals, Factor Analysis Models and High Dimensional Empirical Linear Prediction," *Journal of the American Statistical Association* **94**, 446–455.

Forni, M. and Lippi, M. (2001), "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory* **17:6**, 1113–1141.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2003), "Opening the Black Box: Identifying Shocks and Propogation Mechanisms in VAR and Factor Models," mimeo.

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

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), "The Generalized Dynamic Factor Model, One Sided Estimation and Forecasting," *Journal of the American Statistical Association* **100**, 830–840.

Giannone, D., Reichlin, L. and Sala, L. (2005), "Monetary Policy in Real Time," *Macroeconomic Annual* **19**, 161–200.

Gill, L. and Lewbel, A. (1992), "Testing the Rank and Definiteness of Estimated Matrices with Applications to Factor, State Space, and ARMA Models," *Journal of the American Statistical Association* **87**, 766–776.

Kleibergen, F. and Paap, R. (2003), "Generalized Reduced Rank Tests using the Singular Value Decomposition," *Journal of Econometrics.*

Kuersteiner, G. (2004), “Automatic Inference for Infinite Order Vector Autoregressions,” *Econometric Theory* **21:1**, 85–115.

Ratsimalahelo, Z. (2003), “Strongly Consistent Determination of the Rank of a Matrix,” unpublished manuscript, University of Franche-Comté.

Robin, J. M. and Smith, R. J. (2000), “Tests of Rank,” *Econometric Theory* **16**, 151–175.

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. W. (2002a), “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J. H. and Watson, M. W. (2002b), “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics* **20:2**, 147–162.

Stock, J. H. and Watson, M. W. (2005), “Implcations of Dynamic Factor Models for VAR Analysis,” *NBER WP 11467*.

Table 1a: Estimated number of dynamic factors

DGP 1: $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)' f_t + e_{it}$, $f_t = \varepsilon_t + C_1\varepsilon_{t-1}$, $q = 2, s = 2, r = 6$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		2	2	1.0	1.0	.5	.5
20	100	10.332	1.137	1.307	1.975	2.090	2.355	3.925
20	200	10.443	1.164	1.339	1.981	2.088	2.283	4.169
50	100	6.078	1.963	1.969	2.000	2.000	2.030	2.068
50	200	6.146	1.978	1.982	2.000	2.000	2.026	2.091
100	100	6.000	2.000	2.000	2.000	2.000	2.000	2.000
100	200	6.011	2.000	2.000	2.000	2.000	2.000	2.001

DGP 2: $x_{it} = (\lambda_{i0} + \lambda_{i1}L)' f_t + e_{it}$, $f_t = A_1 f_{t-1} + u_t$, $q = 2, s = 1, r = 4$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		2	2	1.0	1.0	.5	.5
20	100	6.820	1.122	1.327	1.995	2.223	2.769	3.859
20	200	7.231	1.131	1.321	1.978	2.213	2.703	3.996
50	100	4.356	1.937	1.947	2.000	2.014	2.193	2.300
50	200	4.795	1.955	1.966	2.000	2.026	2.228	2.537
100	100	4.003	2.000	2.000	2.000	2.000	2.002	2.002
100	200	4.075	2.000	2.000	2.000	2.000	2.007	2.032

DGP 3: $x_{it} = \lambda_i' F_t + e_{it}$, $F_t = A_1 F_{t-1} + u_t$, $A_1 = \rho I_r$, $q = 3, r = 5\,(r^* = 3)$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		2	2	1.0	1.0	.5	.5
20	100	4.499	1.122	1.686	2.693	2.787	3.043	3.375
20	200	4.903	1.121	1.782	2.762	2.842	3.020	3.490
50	100	3.000	2.338	2.356	2.999	2.999	3.000	3.000
50	200	3.012	2.405	2.415	3.000	3.000	3.000	3.000
100	100	3.000	2.963	2.963	3.000	3.000	3.000	3.000
100	200	3.000	2.987	2.987	3.000	3.000	3.000	3.000

DGP 4: $x_{it} = \lambda_i' F_t + e_{it}$, $F_t = A_1 F_{t-1} + R u_t$, $q = 3, r = 5$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		2	2	1.0	1.0	.5	.5
20	100	4.596	1.098	1.620	2.573	2.753	3.254	3.570
20	200	4.943	1.096	1.636	2.566	2.777	3.272	3.686
50	100	3.097	2.087	2.148	2.990	2.991	3.013	3.023
50	200	3.289	2.126	2.168	2.992	2.994	3.007	3.061
100	100	3.001	2.768	2.768	3.000	3.000	3.000	3.000
100	200	3.007	2.776	2.776	3.000	3.000	3.000	3.000

Note: The table is obtained based on the covariance matrix of VAR residuals. The entries are the average values over 1000 iterations.

Table 1b: Estimated number of dynamic factors

DGP 1: $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)'f_t + e_{it},\ f_t = \varepsilon_t + C_1\varepsilon_{t-1},\ q = 2, s = 2, r = 6$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		1.25	2.25	1.0	2.0	1.5	2.5
20	100	10.332	1.955	1.866	2.448	2.510	1.642	1.529
20	200	10.443	1.944	1.938	2.259	2.649	1.625	1.578
50	100	6.078	2.062	1.919	2.179	1.977	2.002	1.794
50	200	6.146	2.086	1.944	2.253	1.988	1.999	1.827
100	100	6.000	2.037	1.990	2.076	1.996	2.015	1.978
100	200	6.011	2.047	1.994	2.088	1.998	2.022	1.978

DGP 2: $x_{it} = (\lambda_{i0} + \lambda_{i1}L)'f_t + e_{it},\ f_t = A_1 f_{t-1} + u_t,\ q = 2, s = 1, r = 4$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		1.25	2.25	1.0	2.0	1.5	2.5
20	100	6.820	1.999	1.842	3.960	2.480	1.529	1.464
20	200	7.231	1.880	1.934	4.624	2.693	1.440	1.539
50	100	4.356	2.391	1.936	2.416	2.064	2.316	1.784
50	200	4.795	2.828	2.131	2.857	2.318	2.599	1.948
100	100	4.003	2.032	1.994	2.047	2.012	2.021	1.954
100	200	4.075	2.110	2.018	2.123	2.077	2.093	1.992

DGP 3: $x_{it} = \lambda_i' F_t + e_{it},\ F_t = A_1 F_{t-1} + u_t,\ A_1 = \rho I_r,\ q = 3, r = 5\,(r^* = 3)$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		1.25	2.25	1.0	2.0	1.5	2.5
20	100	4.499	3.124	2.468	3.908	2.791	2.500	2.087
20	200	4.903	2.991	2.762	4.279	3.167	2.474	2.436
50	100	3.000	3.000	2.852	3.000	2.987	3.000	2.303
50	200	3.012	3.012	2.984	3.012	3.011	3.012	2.609
100	100	3.000	3.000	3.000	3.000	3.000	3.000	2.996
100	200	3.000	3.000	3.000	3.000	3.000	3.000	3.000

DGP 4: $x_{it} = \lambda_i' F_t + e_{it},\ F_t = A_1 F_{t-1} + Ru_t,\ q = 3, r = 5$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$	$\widehat{q}_3$	$\widehat{q}_4$
	m		1.25	2.25	1.0	2.0	1.5	2.5
20	100	4.596	3.364	2.244	4.067	2.621	1.920	1.662
20	200	4.943	3.739	2.423	4.405	2.902	1.671	1.841
50	100	3.097	3.067	2.291	3.091	2.524	2.970	2.116
50	200	3.289	3.257	2.471	3.285	2.708	3.133	2.259
100	100	3.001	3.001	2.842	3.001	2.938	2.999	2.688
100	200	3.007	3.007	2.855	3.007	2.940	3.005	2.727

Note: This table is computed based on the correlation matrix of VAR residuals. The entries are the average values over 1000 iterations.

Table 2 : Empirical Analysis

	τ	T	$R^2_{\widehat{r}}$	$R^2_{\widehat{q}}$	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$
mean	0.3	330.500	0.323	0.916	3.121	2.255	2.066
min	0.3	133.000	0.300	0.882	3.000	2.000	1.000
max	0.3	528.000	0.348	1.000	4.000	4.000	3.000
mean	0.4	330.500	0.418	0.906	5.386	3.533	3.101
min	0.4	133.000	0.400	0.869	5.000	3.000	3.000
max	0.4	528.000	0.438	0.967	6.000	4.000	4.000
mean	0.5	330.500	0.511	0.902	8.833	6.068	5.210
min	0.5	133.000	0.500	0.805	8.000	5.000	4.000
max	0.5	528.000	0.524	0.952	10.000	7.000	6.000
mean	0.6	330.500	0.608	0.848	13.811	8.523	7.773
min	0.6	133.000	0.600	0.714	12.000	6.000	6.000
max	0.6	528.000	0.619	0.908	15.000	10.000	9.000
mean	r^*	330.500	0.430	0.910	5.763	3.864	3.119
min	r^*	133.000	0.239	0.819	2.000	2.000	2.000
max	r^*	528.000	0.460	1.000	7.000	4.000	4.000

Note: The table is computed based on correlation matrix method with $m = 1.25$ and 2.25 for $\widehat{q}_3$ and $\widehat{q}_4$, respectively. $R^2_{\widehat{r}}$ is average variation in x_{it} explained by $\widehat{r}$ factors, when $\widehat{r}$ explains at least τ percent of the variation in the data up to time t. $R^2_{\widehat{q}}$ is the percent variation in $\widehat{F}_t$ explained by $\widehat{q}$ primitive shocks. The last three columns report the mean, minimum, and the maximum of $\widehat{r}$, $\widehat{q}_3$, and $\widehat{q}_4$ over the expanding samples with sample sizes from $T = 132$ to $T = 528$.

Table 3 : Explanatory Power of $\widehat{r}$ factors

	r							
Series	1	2	3	4	5	6	7	8
ALL	0.172	0.242	0.296	0.350	0.393	0.429	0.460	0.486
IPS10	0.690	0.723	0.774	0.779	0.869	0.869	0.875	0.911
A0M059	0.060	0.140	0.149	0.152	0.183	0.321	0.321	0.341
A0M057	0.266	0.362	0.390	0.397	0.455	0.557	0.559	0.562
FYFF	0.180	0.332	0.443	0.445	0.478	0.482	0.505	0.515
PUNEW	0.008	0.028	0.081	0.706	0.711	0.723	0.726	0.726
A0M224R	0.066	0.142	0.153	0.158	0.174	0.227	0.231	0.279

Note: IP is industrial production, RTQ (a0m059) is retail trade, MSMTQ (a0m057) is manufacturing trade is , FYFF Federal funds rate, PUNEW is CPI, and GMCQ (a0m224R) is consumption expenditure.

Figure 1: Estimated r, q_3, and q_4

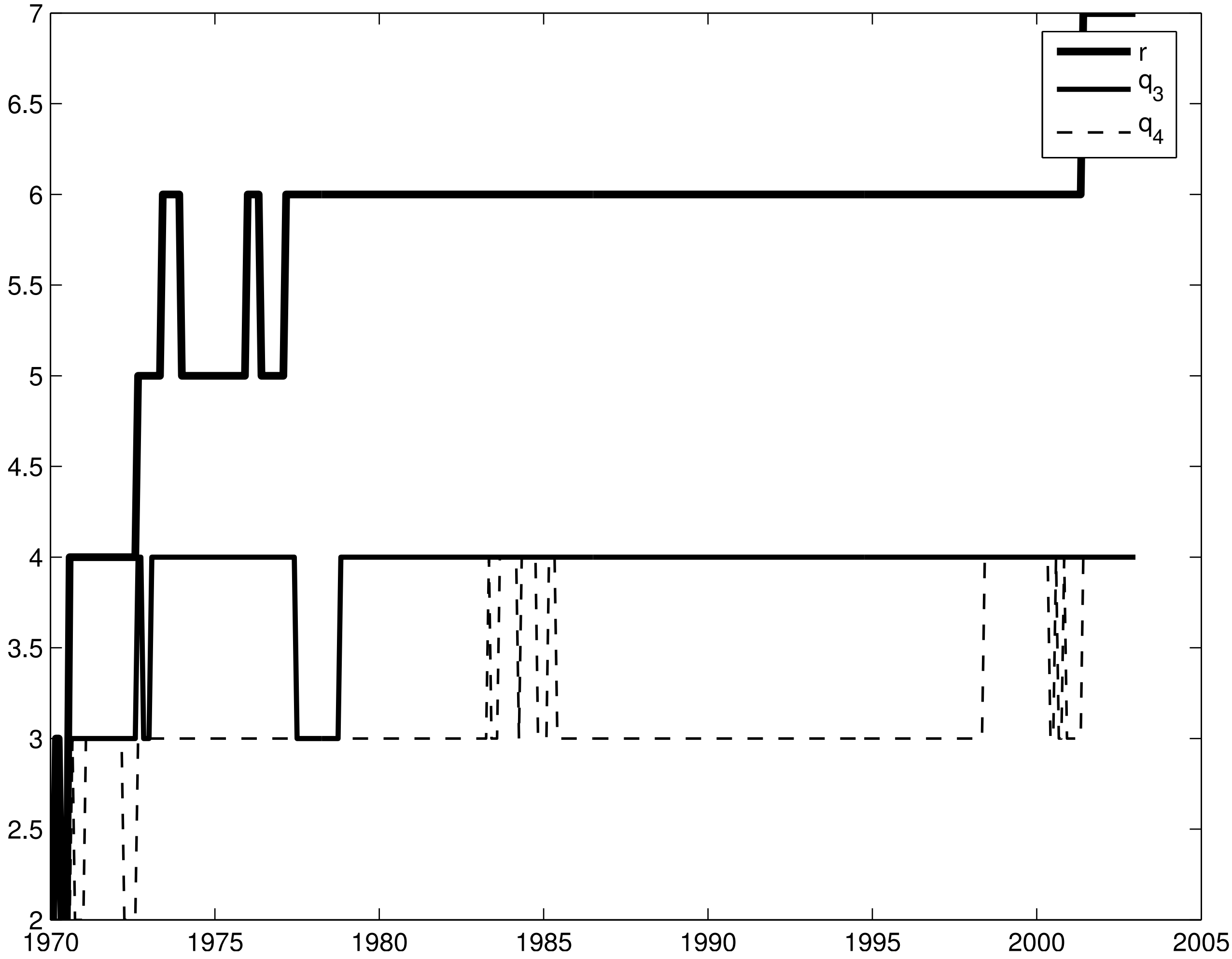

Figure 2: Importance of Common Component