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

$$A(L)F_t = u_t$$

where $A(L) = I - A_1L - \ldots - A_pL^p$. 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

$$u_t = R \varepsilon_t$$

where $\varepsilon_t$ 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 $\Sigma_u = E(u_tu'_t)$, 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 $\varepsilon_t$ 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

\[
D_{1,k} = \left( \frac{c_{k+1}^2}{\sum_{j=1}^{r} c_j^2} \right)^{1/2}
\]

\[
D_{2,k} = \left( \frac{\sum_{j=k+1}^{r} c_j^2}{\sum_{j=1}^{r} c_j^2} \right)^{1/2}
\]

Under the assumption of \( \text{rank}(A) = q \), \( c_k = 0 \) 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 \( A \):

\[
A = \sum_{j=1}^{r} c_j \beta_j \beta_j^t
\]

where \( \beta_j \) is the eigenvector corresponding to \( c_j \). Define the \( k^{th} \) pseudo matrix of \( A \) as

\[
A(k) = \sum_{j=1}^{k} c_j \beta_j \beta_j^t
\]

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

\[
D_{1,k} = \frac{\|d_{k+1} - d_k\|}{\|d_0\|}
\]

\[
D_{2,k} = \frac{\|d_k - d_0\|}{\|d_0\|}
\]

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 \( \text{trace}(\beta_j \beta_j^t) = \|\beta_j\|^2 = 1 \) 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 \( \Sigma_u \), the covariance matrix of a set of innovations. However, \( \Sigma_u \) and \( F \) are not observed, but it can be estimated from the data, denoted by \( \hat{\Sigma}_u \). Let \( \hat{D}_{1,k} \) and \( \hat{D}_{2,k} \) be constructed from the eigenvalues of \( \hat{\Sigma}_u \). It will be shown that \( \hat{D}_{1,k} \) and \( \hat{D}_{2,k} \) converge to zero \( (k \geq q) \) asymptotically at a rate that depends on the convergence rate of \( \hat{\Sigma}_u \) to \( \Sigma_u \).

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 \( \text{var}(\hat{R}) \). The matrix whose rank we seek to test is \( \Sigma_u \). This creates two problems. First, the estimation of \( \text{var}(\Sigma_u) \) entails evaluation of a matrix of fourth moments, which tend to be quite imprecisely estimated unless the sample size is extremely large. Second, \( \Sigma_u \) 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 \( b^{D_1,k} \) and \( b^{D_2,k} \). We begin with the intermediate case when \( F_t \) is observed, but \( \Sigma_u \) is not.

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

\[
\mathcal{K}_1 = \{ k : \hat{D}_{1,k} < m/T^{1/2-\delta} \}
\]

\[
\mathcal{K}_2 = \{ k : \hat{D}_{2,k} < m/T^{1/2-\delta} \}.
\]

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

The proposition follows from \( \sqrt{T}(\hat{\Sigma}_u - \Sigma_u) = O_p(1) \). By continuity of eigenvalues \( \hat{D}_{1,k} = D_{1,k} + O_p(T^{-1/2}) \), and likewise for \( \hat{D}_{2,k} \). From \( \hat{D}_{1,k} - D_{1,k} = O_p(T^{-1/2}) \), we have \( \hat{D}_{1,k} = O_p(T^{-1/2}) \) for \( k \geq q \) because \( D_{1,k} = 0 \). Thus, \( \hat{D}_{1,k} < m/T^{1/2-\delta} \) 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 \( \hat{D}_{1,k} > c > 0 \) and thus greater than \( m/T^{1/2-\delta} \) for \( k < q \). This gives the consistency result. Essentially, the cut-off point \( m/T^{1/2-\delta} \) is the tolerated error induced by sampling variability from estimation of \( \Sigma_u \). 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 $i = 1, \ldots, N$, $t = 1, \ldots, T$. In the language of factor analysis, $e_{it}$ is referred to as the idiosyncratic error, $\Lambda_i$ 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 $\varepsilon_t$ 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 $\varepsilon_t$, 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

$$
x_{it} = \lambda_{i0}f_t + \lambda_{i1}f_{t-1} + \cdots + \lambda_{is}f_{t-s} + e_{it}
$$

$$
= \lambda_i(L)f_t + e_{it}
$$

where $f_t$ is $q$ dimensional, and

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

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

$$
x_{it} = \Lambda_i F_t + e_{it}
$$

where

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

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

$$
(I_q - B_1L - \cdots - B_hL^h)f_t = \epsilon_t
$$
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

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

Define $F_t^* = [f'_t, f'_{t-1}, \ldots, f'_{t-h}]'$. We have

$$
F_t^* = AF_{t-1}^* + u_t
$$

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

$$
\begin{align*}
F_t &= A_1 F_{t-1} + A_2 F_{t-2} + \cdots + A_p F_{t-p} + u_t \\
u_t &= (I_q, 0, \ldots, 0)' \varepsilon_t
\end{align*}
$$

7
**Case II: \( f_t \) is \( MA(h) \)**  
Consider

\[
\begin{align*}
    f_t &= \varepsilon_t + C_1\varepsilon_{t-1} + \cdots + C_h\varepsilon_{t-h} 
\end{align*}
\]

(13)

In this case, \( F_t = (f'_t, f'_{t-1}, \ldots, f'_{t-s})' \), 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(\( \infty \)) 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(\( \infty \)) 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 \( \omega \)

\[
S_F(\omega) = A(e^{-i\omega})^{-1}RS_e(\omega)R'A(e^{i\omega})^{-1},
\]

has rank \( q \) if \( S_e(\omega) \) 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 \( \hat{\Sigma}_u \) 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 \( \hat{\Sigma}_u \) 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 \( \hat{\Lambda} \) 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 \( \hat{\Sigma} = X\hat{\Lambda}/N \). These principal component estimates adopt the normalization that \( \hat{\Lambda}/N = I_r \). Then under the assumption that (i) \( \Sigma_F = E(F_t F_t') \) and \( \hat{\Lambda}/N \) 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),

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

and

\[
\text{prob}(\hat{k} = r) \xrightarrow{p} 1
\]

where

\[
\hat{k} = \arg\min_k \text{IC}(k) = \arg\min_k \log(\sigma_k^2) + kC_{NT}
\]

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 \( \Lambda \) and \( F_t \) are rank \( r \). But \( \Sigma_F \) may have rank less than \( r \). For example, if \( F_t = AF_{t-1} + R\xi_t \) is such that \( A = \rho I_r, |\rho| < 1 \), then \( \text{var}(F_t) = R\Sigma_x R'/(1 - \rho^2) \), \( \text{var}(F_t) \) 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 $\Sigma_u$ is only $q$. But existing results do not cover cases when $\text{var}(F_i) < r$, which can arise as in the example above when $F_i$ has very simple dynamics. We now extend our results on determining the number of factors to also cover these special cases.

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

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

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

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

$$\min[\sqrt{N}, \sqrt{T}] (\hat{\Sigma}_u - H^* \hat{\Sigma}_u H'^*) = O_p(1).$$

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

$$\left\| \hat{\Sigma}_u - H^* \hat{\Sigma}_u H'^* \right\| = O_p(1 / \min[\sqrt{N}, \sqrt{T}]).$$

Note that $\Sigma_u$ and $H^* \Sigma_u H'^*$ have the same rank since $H^*$ is of full rank. We have $\hat{D}_{1,k} - D_{1,k} = O_p(1 / \min[\sqrt{N}, \sqrt{T}])$ because $D_{1,k} = 0$ when $k \geq q$, we also have $\hat{D}_{1,k} = O_p(1 / \min[\sqrt{N}, \sqrt{T}])$. A similar result holds for $\hat{D}_{2,k}$. Thus, we have
Proposition 2 Let \( \hat{\Sigma}_u = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' \), where \( \hat{u}_t \) are the residuals from estimation of a VAR in \( \hat{F}_t \), \( \hat{F}_t \) being the principal components estimator for \( F_t \). For some \( 0 < m < \infty \) and \( 0 < \delta < 1/2 \), let

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

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

Let \( \hat{q}_3 = \min \{ k \in \mathcal{K}_3 \} \) and \( \hat{q}_4 = \min \{ k \in \mathcal{K}_4 \} \). Then under \( H_0 \) with \( \text{rank}(\Sigma_u) = q \), we have \( \hat{q}_3 \overset{p}{\to} q \) and \( \hat{q}_4 \overset{p}{\to} q \) 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 \( \tau \) 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 \( \tau \) are determined using well-defined criteria. This is in contrast to Giannone et al. (2005), in which \( q \) is chosen for a subjectively chosen \( \tau \).

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 \( x_{it} = \Lambda_i F_i + \rho_i(L) x_{i,t-1} + e_{it} \) 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 \( x_{it} = \Lambda_i A^+(L) F_{i,t-1} + \Lambda_i' R \hat{\varepsilon}_t + \rho_i(L) x_{i,t-1} + e_{it} \). The composite residuals of a VAR in \( X_t \) and \( F_t \) is of the form \( \Lambda_i' R \hat{\varepsilon}_t + e_{it} \), which has \( q \) common factors \( \hat{\varepsilon}_t \). Stock and Watson exploit this factor representation to determine \( q \) using the criteria developed in Bai and Ng (2002).

As written, \( \hat{q}_3 \) and \( \hat{q}_4 \) are the estimated rank of \( \hat{\Sigma}_u \), the sample covariance matrix of \( \hat{u}_t \). But the number of non-zero eigenvalues of \( \hat{\Sigma}_u \) is the same as the number of non-zero eigenvalues of \( \hat{\Sigma}_u \), the sample correlation matrix of \( \hat{u}_t \). In our experience, using one or the other matters only for the choice of \( m \). We found \( m = 1 \) works for both \( \hat{q}_3 \) and \( \hat{q}_4 \) when \( \hat{\Sigma}_u \) is used. As we will see in the simulations, the preferred values of \( m \) are different for \( \hat{q}_3 \) and \( \hat{q}_4 \) when correlation matrix \( \hat{\Sigma}_u \) 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. \( x_{it} = (\lambda_0 + \lambda_1 L + \lambda_2 L^2) f_t + e_{it} \), where \( f_t = C(L)\varepsilon_t \) and \( f_t \) is \( q \) dimensional;

2. \( x_{it} = (\lambda_0 + \lambda_1 L) f_t + e_{it} \), 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 \( A_1 = \rho I_r \), \( u_t = R\varepsilon_t \), \( \text{rank}(R) = q \).

4. \( x_{it} = \lambda_i' F_t + e_{it} \) where \( F_t = A_1 F_{t-1} + u_t \), where \( A_1 = \text{diag}(.2, .375, .55, .725, .9) \) 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 \( \Gamma \) 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 \( \Gamma \) 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 DGP 3. 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 \( \lambda_{ij}, e_{it}, \text{and} \varepsilon_t \) are iid standard normal.

For all four DGPs, the testing proceeds as follows. Given the data \( x_{it}, i = 1, \ldots, N, t = 1, \ldots, T \), 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,

\[
\hat{r}^* = \arg\min_{k \in [0, 2r]} \log(\hat{\sigma}_k^2) + k \log(\min[N, T]) \frac{\log(NT/(N + T))}{NT/(N + T)},
\]
where $\hat{\sigma}_k^2 = \frac{1}{NT} \sum_i \sum_t (x_{it} - \hat{\gamma}_k \hat{F}_t^k)^2$, $\hat{F}_t^k$ is $k \times 1$. Given $\hat{F}_t$, a $\hat{r}^*$ dimensional VAR in $\hat{F}_t$ is estimated to obtain $\hat{u}_t$. Selecting too few lags will be problematic as $\hat{u}_t$ will not be innovations. We report results for VAR(2). Results for higher lags are similar. Given $\hat{u}_t$, its $\hat{r}^* \times \hat{r}^*$ covariance matrix is constructed. Then $\hat{q}_3$ and $\hat{q}_4$ are obtained with $\delta = .1$ so that $m^* = \frac{m}{\min[N/2, T/2]}$. The number of simulations is 1000.

We report the average values for $\hat{r}$ and $\hat{q}$ in Table 1a estimated based on the covariance matrix of $\hat{u}_t$. The four panels correspond to four different data generating processes. For all cases, the number of static factors $\hat{r}$ is determined by minimizing the IC(k) for $k$ between 0 and 2$r$. For small values of $\min[N, T]$, the IC tends to select a large number of static factors. However, even when $\hat{r}^*$ is overestimated, $\hat{q}$ 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, $\hat{q}$ is larger than $q$. For all four DGPs we find that $m = 1.25$ works well for $\hat{q}_3$, and $m = 2.25$ works well for $\hat{q}_4$ (when using correlation matrices). Between $\hat{q}_3$ and $\hat{q}_4$, 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 $\hat{q}$ 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, $\hat{q} = r$ when $r = 2, 3, \hat{q} = 3$, $r = 4, 5, 6$, $\hat{q} = 4$ 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 $\hat{q} = 2$ 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 \( \hat{r}_t(\tau) \) static factors, where \( \hat{r}_t(\tau) \) explains closest \( \tau \) percent of the variation in the data up to time \( t \), and then determine \( \hat{q}_t(\tau) \) given \( \hat{r}_t(\tau) \) factors. Note that \( \hat{r}_t(\tau) \) 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 \( \hat{r}_t^* \), and the corresponding number of primitive factors is \( \hat{q}_t^* \). 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 \( T = 528 \)). \( \hat{R}^2_r \) is the average explanatory power of \( \hat{r} \) factors, when \( \hat{r} \) is chosen with cut-off of \( .3, .4, .5 \) or \( .6 \), and also optimally. The column \( \hat{R}^2_q \) is the average explanatory power of the \( \hat{q} \) shocks given \( \hat{r} \) 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 \( \hat{r}_t^* \), \( \hat{q}_t \), and \( \hat{u}_t \). As we can see, \( \hat{r} \) jumped from 4 to 6 around 1973, and has remained roughly at 6 till 2000. On the other hand, \( \hat{q} \) jumped from 3 to 4 and stayed at 4 most of the time. If we had ended the estimation in 2000, we would have \( \hat{r} = 6 \) and \( \hat{q} = 4 \). However, in the past few years, \( \hat{r} \) seemed to have taken another jump from 6 to 7, though \( \hat{q} \) seems to have stayed at 4.

Figure 2 plots \( \hat{R}^2_r \), 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 $\hat{r}$ 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


Table 1a: Estimated number of dynamic factors

DGP 1: \( x_{it} = (\lambda_0 + \lambda_1 i + \lambda_2 L^2)f_t + \epsilon_{it}, f_t = \varepsilon_t + C_{1e_{t-1}}, q = 2, s = 2, r = 6 \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>10.332</td>
<td>1.137</td>
<td>1.307</td>
<td>1.975</td>
<td>2.090</td>
<td>2.355</td>
<td>3.925</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>10.443</td>
<td>1.164</td>
<td>1.339</td>
<td>1.981</td>
<td>2.088</td>
<td>2.283</td>
<td>4.169</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>6.078</td>
<td>1.963</td>
<td>1.969</td>
<td>2.000</td>
<td>2.000</td>
<td>2.030</td>
<td>2.068</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>6.146</td>
<td>1.978</td>
<td>1.982</td>
<td>2.000</td>
<td>2.000</td>
<td>2.026</td>
<td>2.091</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>6.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>6.011</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.001</td>
</tr>
</tbody>
</table>

DGP 2: \( x_{it} = (\lambda_0 + \lambda_1 i + \lambda_2 L^2)f_t + \epsilon_{it}, f_t = A_1f_{t-1} + u_t, q = 2, s = 1, r = 4 \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>6.820</td>
<td>1.122</td>
<td>1.327</td>
<td>1.995</td>
<td>2.223</td>
<td>2.769</td>
<td>3.859</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>7.231</td>
<td>1.131</td>
<td>1.321</td>
<td>1.978</td>
<td>2.213</td>
<td>2.703</td>
<td>3.996</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>4.356</td>
<td>1.937</td>
<td>1.947</td>
<td>2.000</td>
<td>2.014</td>
<td>2.193</td>
<td>2.300</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>4.795</td>
<td>1.955</td>
<td>1.966</td>
<td>2.000</td>
<td>2.026</td>
<td>2.228</td>
<td>2.537</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>4.003</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.002</td>
<td>2.002</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>4.075</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
<td>2.007</td>
<td>2.032</td>
</tr>
</tbody>
</table>

DGP 3: \( x_{it} = \lambda_1 F_t + \epsilon_{it}, F_t = A_1F_{t-1} + u_t, A_1 = \rho L_t, q = 3, r = 5 (r^* = 3) \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>4.499</td>
<td>1.122</td>
<td>1.686</td>
<td>2.693</td>
<td>2.787</td>
<td>3.043</td>
<td>3.375</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>4.903</td>
<td>1.121</td>
<td>1.782</td>
<td>2.762</td>
<td>2.842</td>
<td>3.020</td>
<td>3.490</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>3.000</td>
<td>2.338</td>
<td>2.356</td>
<td>2.999</td>
<td>2.999</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>3.012</td>
<td>2.405</td>
<td>2.415</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>3.000</td>
<td>2.963</td>
<td>2.963</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>3.000</td>
<td>2.987</td>
<td>2.987</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

DGP 4: \( x_{it} = \lambda_1 F_t + \epsilon_{it}, F_t = A_1F_{t-1} + Ru_t, q = 3, r = 5 \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>4.596</td>
<td>1.098</td>
<td>1.620</td>
<td>2.573</td>
<td>2.753</td>
<td>3.254</td>
<td>3.570</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>4.943</td>
<td>1.096</td>
<td>1.636</td>
<td>2.566</td>
<td>2.777</td>
<td>3.272</td>
<td>3.686</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>3.097</td>
<td>2.087</td>
<td>2.148</td>
<td>2.990</td>
<td>2.991</td>
<td>3.013</td>
<td>3.023</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>3.289</td>
<td>2.126</td>
<td>2.168</td>
<td>2.992</td>
<td>2.994</td>
<td>3.007</td>
<td>3.061</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>3.001</td>
<td>2.768</td>
<td>2.768</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>3.007</td>
<td>2.776</td>
<td>2.776</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

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: \( x_{it} = (\lambda_{0i} + \lambda_{1i} L + \lambda_{2i} L^2) f_t + \epsilon_{it}, f_t = \epsilon_{t} + C_{1\epsilon_{t-1}}, q = 2, s = 2, r = 6 \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( q_3 )</th>
<th>( q_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>10.332</td>
<td>1.955</td>
<td>1.866</td>
<td>2.448</td>
<td>2.510</td>
<td>1.642</td>
<td>1.529</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>10.443</td>
<td>1.944</td>
<td>1.938</td>
<td>2.259</td>
<td>2.649</td>
<td>1.625</td>
<td>1.578</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>6.078</td>
<td>2.062</td>
<td>1.919</td>
<td>2.179</td>
<td>1.977</td>
<td>2.002</td>
<td>1.794</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>6.146</td>
<td>2.086</td>
<td>1.944</td>
<td>2.253</td>
<td>1.988</td>
<td>1.999</td>
<td>1.827</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>6.000</td>
<td>2.037</td>
<td>1.990</td>
<td>2.076</td>
<td>1.996</td>
<td>2.015</td>
<td>1.978</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>6.011</td>
<td>2.047</td>
<td>1.994</td>
<td>2.088</td>
<td>1.998</td>
<td>2.022</td>
<td>1.978</td>
</tr>
</tbody>
</table>

DGP 2: \( x_{it} = (\lambda_{0i} + \lambda_{1i} L) f_t + \epsilon_{it}, f_t = A_1 f_{t-1} + u_t, q = 2, s = 1, r = 4 \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( q_3 )</th>
<th>( q_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>6.820</td>
<td>1.999</td>
<td>1.842</td>
<td>3.960</td>
<td>2.480</td>
<td>1.529</td>
<td>1.464</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>7.231</td>
<td>1.880</td>
<td>1.934</td>
<td>4.624</td>
<td>2.693</td>
<td>1.440</td>
<td>1.539</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>4.356</td>
<td>2.391</td>
<td>1.936</td>
<td>2.416</td>
<td>2.064</td>
<td>2.316</td>
<td>1.784</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>4.795</td>
<td>2.828</td>
<td>2.131</td>
<td>2.857</td>
<td>2.318</td>
<td>2.599</td>
<td>1.948</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>4.003</td>
<td>2.032</td>
<td>1.994</td>
<td>2.047</td>
<td>2.012</td>
<td>2.021</td>
<td>1.954</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>4.075</td>
<td>2.110</td>
<td>2.018</td>
<td>2.123</td>
<td>2.077</td>
<td>2.093</td>
<td>1.992</td>
</tr>
</tbody>
</table>

DGP 3: \( x_{it} = \lambda_i F_t + \epsilon_{it}, F_t = A_1 f_{t-1} + u_t, A_1 = \rho L, q = 3, r = 5 (r^* = 3) \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( q_3 )</th>
<th>( q_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>4.499</td>
<td>3.124</td>
<td>2.468</td>
<td>3.908</td>
<td>2.791</td>
<td>2.500</td>
<td>2.087</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>4.903</td>
<td>2.991</td>
<td>2.762</td>
<td>4.279</td>
<td>3.167</td>
<td>2.474</td>
<td>2.436</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>3.000</td>
<td>3.000</td>
<td>2.852</td>
<td>3.000</td>
<td>2.987</td>
<td>3.000</td>
<td>2.303</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>3.012</td>
<td>3.012</td>
<td>2.984</td>
<td>3.012</td>
<td>3.011</td>
<td>3.012</td>
<td>2.609</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>2.996</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

DGP 4: \( x_{it} = \lambda_i F_t + \epsilon_{it}, F_t = A_1 f_{t-1} + Ru_t, q = 3, r = 5 \)

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>( \hat{r} )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
<th>( q_3 )</th>
<th>( q_4 )</th>
<th>( \hat{q}_3 )</th>
<th>( \hat{q}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>4.596</td>
<td>3.364</td>
<td>2.244</td>
<td>4.067</td>
<td>2.621</td>
<td>1.920</td>
<td>1.662</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>4.943</td>
<td>3.739</td>
<td>2.423</td>
<td>4.405</td>
<td>2.902</td>
<td>1.671</td>
<td>1.841</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>3.097</td>
<td>3.067</td>
<td>2.291</td>
<td>3.091</td>
<td>2.524</td>
<td>2.970</td>
<td>2.116</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>3.001</td>
<td>3.001</td>
<td>2.842</td>
<td>3.001</td>
<td>2.938</td>
<td>2.999</td>
<td>2.688</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>3.007</td>
<td>3.007</td>
<td>2.855</td>
<td>3.007</td>
<td>2.940</td>
<td>3.005</td>
<td>2.727</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Series</th>
<th>(\tau)</th>
<th>(T)</th>
<th>(R^2_{R^*})</th>
<th>(R^2_{\hat{q}})</th>
<th>(\hat{\tau})</th>
<th>(\hat{q}_3)</th>
<th>(\hat{q}_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.3</td>
<td>330.50</td>
<td>0.323</td>
<td>0.916</td>
<td>3.121</td>
<td>2.255</td>
<td>2.066</td>
</tr>
<tr>
<td>min</td>
<td>0.3</td>
<td>133.00</td>
<td>0.300</td>
<td>0.882</td>
<td>3.000</td>
<td>2.000</td>
<td>1.000</td>
</tr>
<tr>
<td>max</td>
<td>0.3</td>
<td>528.00</td>
<td>0.348</td>
<td>1.000</td>
<td>4.000</td>
<td>4.000</td>
<td>3.000</td>
</tr>
<tr>
<td>mean</td>
<td>0.4</td>
<td>330.50</td>
<td>0.418</td>
<td>0.906</td>
<td>5.386</td>
<td>3.533</td>
<td>3.101</td>
</tr>
<tr>
<td>min</td>
<td>0.4</td>
<td>133.00</td>
<td>0.400</td>
<td>0.869</td>
<td>5.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>max</td>
<td>0.4</td>
<td>528.00</td>
<td>0.438</td>
<td>0.967</td>
<td>6.000</td>
<td>4.000</td>
<td>4.000</td>
</tr>
<tr>
<td>mean</td>
<td>0.5</td>
<td>330.50</td>
<td>0.511</td>
<td>0.920</td>
<td>8.833</td>
<td>6.068</td>
<td>5.210</td>
</tr>
<tr>
<td>min</td>
<td>0.5</td>
<td>133.00</td>
<td>0.500</td>
<td>0.869</td>
<td>8.000</td>
<td>5.000</td>
<td>4.000</td>
</tr>
<tr>
<td>max</td>
<td>0.5</td>
<td>528.00</td>
<td>0.524</td>
<td>0.952</td>
<td>10.000</td>
<td>7.000</td>
<td>6.000</td>
</tr>
<tr>
<td>mean</td>
<td>0.6</td>
<td>330.50</td>
<td>0.608</td>
<td>0.848</td>
<td>13.811</td>
<td>8.523</td>
<td>7.773</td>
</tr>
<tr>
<td>min</td>
<td>0.6</td>
<td>133.00</td>
<td>0.600</td>
<td>0.714</td>
<td>12.000</td>
<td>6.000</td>
<td>6.000</td>
</tr>
<tr>
<td>max</td>
<td>0.6</td>
<td>528.00</td>
<td>0.619</td>
<td>0.908</td>
<td>15.000</td>
<td>10.000</td>
<td>9.000</td>
</tr>
<tr>
<td>mean</td>
<td>(r^*)</td>
<td>330.50</td>
<td>0.430</td>
<td>0.910</td>
<td>5.763</td>
<td>3.864</td>
<td>3.119</td>
</tr>
<tr>
<td>min</td>
<td>(r^*)</td>
<td>133.00</td>
<td>0.239</td>
<td>0.819</td>
<td>2.000</td>
<td>2.000</td>
<td>2.000</td>
</tr>
<tr>
<td>max</td>
<td>(r^*)</td>
<td>528.00</td>
<td>0.460</td>
<td>1.000</td>
<td>7.000</td>
<td>4.000</td>
<td>4.000</td>
</tr>
</tbody>
</table>

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

Table 3: Explanatory Power of \(\hat{\tau}\) factors

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

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

$R^2$