Determining the Number of Primitive Shocks in Factor Models

Jushan BAI

Department of Economics, New York University, 269 Mercer St., New York, NY 10003, and Department of Economics, Tsinghua University, Beijing, China (*Jushan.Bai@nyu.edu*)

Serena NG

Department of Economics, University of Michigan, Ann Arbor, MI 48109 (*Serena.Ng@umich.edu*)

A widely held but untested assumption underlying macroeconomic analysis is that the number of shocks driving economic fluctuations, q, is small. In this article 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, then compute the eigenvalues of the residual covariance or correlation matrix. We then test whether their eigenvalues satisfy an asymptotically shrinking bound that reflects sampling error. We apply the procedure 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.

KEY WORDS: Common shocks; Dynamic factor model; Number of factors; Principal components analysis; Static factor.

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. In fact, it is not easy to find a business cycle model built from microfoundations with more than four shocks. Macroeconomists have been preoccupied with understanding the transmission and quantifying the importance of three shocks: technology, monetary, and fiscal policy. But what is the exact number of primitive shocks in the data? In this article we propose a simple testing procedure to determine this number, which we 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. We do this without having to estimate a dynamic factor model, however.

Surprisingly, few test exists to formally evaluate what is the exact value of q. Using a dynamic index model to analyze quarterly data for 14 series over the sample 1950:1–1970:1, Sargent and Sims (1977) rejected the one- and two-index models 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 articles Forni, Giannone, Lippi, and Reichlin (2003) and Giannone, Reichlin, and Sala (2005) argued that the number of macroeconomic shocks, which they referred to as the stochastic dimension of the economy, is two. They estimated common factors from quarterly data on 190 series over the sample period 1970–1996 and arrived at the conclusion of two shocks using a reasonable (albeit informal) judgment that two dynamic factors explain about 60% of the variation in 12 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 cutoff 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 $q = 2$ remains very much an assertion.

The analytical framework used by 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 can 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. We formally develop tests for the number of dynamic factors in Sections 3 and 4, and consider simulations and an empirical application in Section 5. We present conclusions in Section 6.

2.   THE MINIMAL NUMBER OF PRIMITIVE SHOCKS IN VECTOR AUTOREGRESSION

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

$$\mathbf{A}(L)\mathbf{F}_t = \mathbf{u}_t, \tag{1}$$

where $\mathbf{A}(L) = \mathbf{I} - \mathbf{A}_1 L - \cdots - \mathbf{A}_p L^p$. Throughout, we assume that the roots of $\mathbf{A}(L) = \mathbf{0}$ all lie outside of the unit circle, and the $\mathbf{u}_t$'s are iid with $E\|\mathbf{u}_t\|^{4+\delta} < M < \infty$ for some $\delta > 0$. We consider the case in which $\mathbf{u}_t$ 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 $\mathbf{F}_t$ is driven by a minimal number of q innovations if there exists a $r \times q$ matrix, $\mathbf{R}$, with rank q such that

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t, \tag{2}$$

where $\boldsymbol{\epsilon}_t$ is a $q \times 1$ vector of innovations that are mutually uncorrelated; that is, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t')$ is diagonal. If we define $\boldsymbol{\Sigma}_{\mathbf{u}} = E(\mathbf{u}_t \mathbf{u}_t')$, then, under (2), $\boldsymbol{\Sigma}_{\mathbf{u}} = \mathbf{R}\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\mathbf{R}'$ has rank $q \le r$.

Bernanke (1986) used $\mathbf{u}_t$ as the residuals of an estimated VAR and assumed $\mathbf{R}$ to be full rank. The number of primitive shocks thus equals the number of variables in the system. We allow the rank of $\mathbf{R}$ 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 $\mathbf{u}_t$ is simply the q linearly independent shocks that span $\mathbf{u}_t$.

Suppose that $\mathbf{B}$ is an arbitrary $r \times r$ matrix with rank q and that $\mathbf{v}_t$ is vector of $r \times 1$ shocks. Let $\mathbf{u}_t = \mathbf{B}\mathbf{v}_t$. Then $\mathbf{u}_t$ can be expressed as $\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t$, where $\mathbf{R}$ is $r \times q$ and $\boldsymbol{\epsilon}_t$ is $q \times 1$. To determine q, we use as starting point that a $r \times r$ semipositive definite matrix $\mathbf{A}$ of rank q has q nonzero eigenvalues. Let $c_1 > c_2 \ge \cdots \ge c_r \ge 0$ be the ordered eigenvalues (at least one nonzero eigenvalue) and define

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

and

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

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

A different interpretation of our test can be obtained using the spectral decomposition of $\mathbf{A}$,

$$\mathbf{A} = \sum_{j=1}^r c_j \boldsymbol{\beta}_j \boldsymbol{\beta}_j',$$

where $\boldsymbol{\beta}_j$ is the eigenvector corresponding to c_j. Define the kth pseudomatrix of $\mathbf{A}$ as

$$\mathbf{A}(k) = \sum_{j=1}^k c_j \boldsymbol{\beta}_j \boldsymbol{\beta}_j'.$$

If $\mathbf{A}$ has $r - q$ eigenvalues that are 0, then $\mathbf{A} = \mathbf{A}(k)$ for $k = q+1, \ldots, r$. If $\mathbf{d}_k = \text{vec}(\mathbf{A}(k))$, with $\mathbf{d}_0 = \text{vec}(\mathbf{A})$, then

$$D_{1,k} \equiv \|\mathbf{d}_{k+1} - \mathbf{d}_k\| / \|\mathbf{d}_0\| \quad \text{and} \quad D_{2,k} \equiv \|\mathbf{d}_k - \mathbf{d}_0\| / \|\mathbf{d}_0\|.$$

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 trace$(\boldsymbol{\beta}_j \boldsymbol{\beta}_j') = \|\boldsymbol{\beta}_j\|^2 = 1$ by construction, and $d_0^2 = \text{vec}(\mathbf{A})' \text{vec}(\mathbf{A}) = \text{tr}(\mathbf{A}'\mathbf{A}) = \|\mathbf{A}^2\|$.

In the next two sections, the matrix $\mathbf{A}$, the rank of which is to be determined, is $\boldsymbol{\Sigma}_{\mathbf{u}}$, the covariance matrix of a set of innovations. Although $\boldsymbol{\Sigma}_{\mathbf{u}}$ and $\mathbf{F}$ are not observed, they can be estimated from the data, denoted by $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$. Let $\hat{D}_{1,k}$ and $\hat{D}_{2,k}$ be constructed from the eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$. We show that $\hat{D}_{1,k}$ and $\hat{D}_{2,k}$ converge to 0 ($k \ge q$) asymptotically at a rate depending on the convergence rate of $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$ to $\boldsymbol{\Sigma}_{\mathbf{u}}$.

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 $\mathbf{R}$, when $\mathbf{R}$ is consistently estimated from the regression $\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t + \mathbf{v}_t$. Our focus is on problems in which $\mathbf{v}_t$ plays no role, but with unobservable $\mathbf{u}_t$. Furthermore, rank tests tend to not be asymptotically normal (see Anderson 1951; Gill and Lewbel 1992; Cragg and Donald 1996, 1997; Robin and Smith 2000). Recently, Kleibergen and Paap (2006) and Ratsimalahelo (2003) suggested orthogonal rotation of the sample eigenvalues around the origin to restore normality. Doing so, these tests necessitate a consistent estimate of var$(\hat{\mathbf{R}})$. We want to test the rank of the matrix $\boldsymbol{\Sigma}_{\mathbf{u}}$. This creates two problems. First, estimating var$(\hat{\boldsymbol{\Sigma}}_{\mathbf{u}})$ entails evaluating a matrix of fourth moments, which tend to be quite imprecisely estimated unless the sample size is extremely large. Second, $\boldsymbol{\Sigma}_{\mathbf{u}}$ is a variance–covariance matrix that has only $r(r+1)/2$ unique elements; thus var$(\boldsymbol{\Sigma}_{\mathbf{u}})$ and its estimate do not have full rank, an assumption maintained by Kleibergen and Paap (2006). 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 $\hat{D}_{1,k}$ and $\hat{D}_{2,k}$. We begin with the intermediate case when $\mathbf{F}_t$ is observed but $\boldsymbol{\Sigma}_{\mathbf{u}}$ is not.

Proposition 1. Let $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t'$, where $\hat{\mathbf{u}}_t$ are the residuals from estimation of a VAR in $\mathbf{F}_t$, where $\mathbf{F}_t$ 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 $\hat{\boldsymbol{\Sigma}}_{\mathbf{u}}$. For some $0 < m < \infty$ and $0 < \delta < 1/2$, define

$$\mathcal{K}_1 = \left\{ k : \hat{D}_{1,k} < m/T^{1/2-\delta} \right\} \tag{3}$$

and

$$\mathcal{K}_2 = \left\{ k : \hat{D}_{2,k} < m/T^{1/2-\delta} \right\}. \tag{4}$$

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$(\boldsymbol{\Sigma}_{\mathbf{u}}) = q$, $\hat{q}_1 \xrightarrow{p} q$ and $\hat{q}_2 \xrightarrow{p} q$ as $T \to \infty$.

This proposition follows from $\sqrt{T}(\hat{\boldsymbol{\Sigma}}_{\mathbf{u}} - \boldsymbol{\Sigma}_{\mathbf{u}}) = O_p(1)$. 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 $k \ge q$ 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 is $> m/T^{1/2-\delta}$ for $k < q$. This gives the consistency result. Essentially, the cutoff point $m/T^{1/2-\delta}$ is the tolerated error induced by sampling variability from estimation of $\boldsymbol{\Sigma}_{\mathbf{u}}$. 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 $\mathbf{F}_t$ is assumed to be observed. In the next two sections, $\mathbf{F}_t$ is a vector of unobserved common factors 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 $x_{it} = \boldsymbol{\Lambda}_i' \mathbf{F}_t + e_{it}$, where $i = 1, \ldots, N, t = 1, \ldots, T$. In the language of factor analysis,

e_{it} is referred to as the idiosyncratic error, and $\mathbf{\Lambda}_i$ is a vector of factor loadings for unit i on the r (static) common factors $\mathbf{F}_t$. The term *static factor* model refers to the static relationship between x_{it} and $\mathbf{F}_t$, but $\mathbf{F}_t$ itself can be a dynamic process. The *dynamic factor* model is written as $x_{it} = \mathbf{\lambda}_i'(L)\mathbf{f}_t + e_{it}$, where $\mathbf{\lambda}_i(L)$ 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, $\mathbf{f}_t = \mathbf{C}(L)\boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t$ are iid vectors and $x_{it} = \mathbf{\lambda}_i(L)\mathbf{C}(L)\boldsymbol{\epsilon}_t + e_{it}$. In this article we consider dynamic factor models (finite s). The dimension of $\mathbf{f}_t$, which is the same as the dimension of $\boldsymbol{\epsilon}_t$, is called the number of dynamic factors; and is denoted by q. Dynamic factor models with s finite can be written as static factor models with r finite, but the dimension of $\mathbf{F}_t$ is in general different from the dimension of $\mathbf{f}_t$, because $\mathbf{F}_t$ includes the leads and lags of $\mathbf{f}_t$ with $r \geq q$. In practice, $\mathbf{F}_t$ is estimated using an eigenvalue–eigenvector decomposition of the sample covariance matrix of the data, whereas the dynamic estimates are based on a eigenvalue decomposition of the spectrum smoothed over various frequencies. Recent research has shown that the space spanned by the static and dynamic factors can be consistently estimated when N and T are both large (see, e.g., Forni, Hallin, Lippi, and Reichlin 2000, 2005; Ding and Hwang 2001; Stock and Watson 2002a; Forni and Lippi 2001; Bai and Ng 2002; 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 to be useful in forecasting exercises and in understanding the conduct of monetary policy (see, e.g., Stock and Watson 2002b; Bernanke and Boivin 2003). Although for forecasting purposes, little is to be gained from a clear distinction between the static factors and the dynamic factors, many economic analyses hinge on the ability to isolate the primitive shocks or, in other words, the number of dynamic factors.

In earlier work (Bai and Ng 2002), we showed that under certain conditions, information criteria with appropriately chosen penalties will consistently estimate r, where r is 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 precisely state the relationship between the dynamic and static factors, treating $\mathbf{F}_t$ and $\mathbf{f}_t$ as though they are observed. In the remainder of this section, we show that a dynamic factor model always has a static factor representation in which the dynamics of $\mathbf{F}_t$ is characterized by a VAR with order depending on the dynamics of $\mathbf{f}_t$. We will see from the VAR representation that the spectrum of the static factors has rank q.

3.1 Putting the Dynamic Model Into Static Form

Consider the dynamic factor model

$$x_{it} = \mathbf{\lambda}_{i0}'\mathbf{f}_t + \mathbf{\lambda}_{i1}'\mathbf{f}_{t-1} + \cdots + \mathbf{\lambda}_{is}'\mathbf{f}_{t-s} + e_{it}$$

$$= \mathbf{\lambda}_i'(L)\mathbf{f}_t + e_{it}, \tag{5}$$

where $\mathbf{f}_t$ is q-dimensional and

$$\mathbf{\lambda}_i(L) = \mathbf{\lambda}_{i1} + \mathbf{\lambda}_{i2}L + \cdots + \mathbf{\lambda}_{is}L^s. \tag{6}$$

Clearly we can rewrite (5) in the static form,

$$x_{it} = \mathbf{\Lambda}_i'\mathbf{F}_t + e_{it}, \tag{7}$$

where

$$\mathbf{\Lambda}_i = \begin{bmatrix} \mathbf{\lambda}_{i0} \\ \mathbf{\lambda}_{i1} \\ \vdots \\ \mathbf{\lambda}_{is} \end{bmatrix} \quad \text{and} \quad \mathbf{F}_t = \begin{bmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \vdots \\ \mathbf{f}_{t-s} \end{bmatrix}. \tag{8}$$

The foregoing is a simple mathematical identity and is true whether $\mathbf{f}_t$ itself is an AR process or an MA process. The dimension of $\mathbf{F}_t$ is always equal to

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

where q is the dimension of $\mathbf{f}_t$. As mentioned earlier, whereas the relation between x_{it} and $\mathbf{F}_t$ is static, $\mathbf{F}_t$ itself can be a dynamic process with precise characterization depending on the dynamics of $\mathbf{f}_t$. We consider two cases, one case in which $\mathbf{f}_t$ is a finite-order AR process, and the other in which $\mathbf{f}_t$ has an MA structure.

Case I. $\mathbf{f}_t$ is AR(h), h is finite. In this case,

$$(\mathbf{I}_q - \mathbf{B}_1 L - \cdots - \mathbf{B}_h L^h)\mathbf{f}_t = \boldsymbol{\epsilon}_t. \tag{9}$$

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

We now want to establish that $\mathbf{F}_t$ is a VAR with order depending on h and s. To see the VAR representation of $\mathbf{F}_t$, we put $\mathbf{f}_t$ into a state-space form. Let $\kappa = \max(h, s)$ and define $\mathbf{B}_{h+1} = \cdots = \mathbf{B}_\kappa = \mathbf{0}$; then

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ . \\ . \\ \mathbf{f}_{t-\kappa} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 & . & . & . & \mathbf{B}_{\kappa+1} \\ \mathbf{I}_q & 0 & 0 & . & . & 0 \\ 0 & \mathbf{I}_q & & . & . & 0 \\ . & . & . & . & . & . \\ 0 & . & . & \mathbf{I}_q & . & 0 \end{pmatrix} \begin{pmatrix} \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ . \\ . \\ \mathbf{f}_{t-\kappa-1} \end{pmatrix}$$
$$+ \begin{pmatrix} \mathbf{I}_q \\ 0 \\ . \\ . \\ 0 \end{pmatrix} \boldsymbol{\epsilon}_t.$$

Define $\mathbf{F}_t^* = [\mathbf{f}_t'\ \mathbf{f}_{t-1}'\ \cdots\ \mathbf{f}_{t-\kappa}']'$. We have

$$\mathbf{F}_t^* = \mathbf{A}\mathbf{F}_{t-1}^* + \mathbf{u}_t, \tag{10}$$

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t,$$

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

$$\mathbf{F}_t \equiv \mathbf{F}_t^*,$$

so $\mathbf{F}_t$ also has a VAR(1) representation. When $s < h$, $\mathbf{F}_t$ is a subvector of $\mathbf{F}_t^*$. In general, any subvector of a VAR is a vector ARMA process, not necessarily VAR. However, due to the special structure of $\mathbf{F}_t^*$, the subvector $\mathbf{F}_t$ itself is a VAR. This point

can be easily made clear through an illustration. Suppose that $h = 3$ and $s = 1$, and consider

$$\mathbf{f}_t = \mathbf{B}_1 \mathbf{f}_{t-1} + \mathbf{B}_2 \mathbf{f}_{t-2} + \mathbf{B}_3 \mathbf{f}_{t-3} + \boldsymbol{\epsilon}_t,$$

so that $\mathbf{f}_t$ is VAR(3). Let $\mathbf{F}_t = (\mathbf{f}_t', \mathbf{f}_{t-1}')'$. Clearly,

$$\begin{bmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{B}_3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f}_{t-2} \\ \mathbf{f}_{t-3} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix} \boldsymbol{\epsilon}_t.$$

This implies that $\mathbf{F}_t$ is VAR(2). In fact, we can show that for the general situation, $\mathbf{F}_t$ is VAR(p) with $p = \max(1, h - s)$. Therefore, the dynamic factor model defined by (5)–(6) when $\mathbf{f}_t$ is a VAR(h) can be written as a static factor model as

$$\mathbf{F}_t = \mathbf{A}_1 \mathbf{F}_{t-1} + \mathbf{A}_2 \mathbf{F}_{t-2} + \cdots + \mathbf{A}_p \mathbf{F}_{t-p} + \mathbf{u}_t \qquad (11)$$

and

$$\mathbf{u}_t = (\mathbf{I}_q, 0, \ldots, 0)' \boldsymbol{\epsilon}_t \qquad (12)$$

Case II. $\mathbf{f}_t$ is *MA(h)*. Consider

$$\mathbf{f}_t = \boldsymbol{\epsilon}_t + \mathbf{C}_1 \boldsymbol{\epsilon}_{t-1} + \cdots + \mathbf{C}_h \boldsymbol{\epsilon}_{t-h}. \qquad (13)$$

In this case $\mathbf{F}_t = (\mathbf{f}_t', \mathbf{f}_{t-1}', \ldots, \mathbf{f}_{t-s}')'$, and $\mathbf{F}_t$ is $q(s+1) \times 1$. That is, the number of static factors is $r = q(s+1)$ irrespective of h, the order of $\mathbf{f}_t$.

Under the invertibility assumption, the MA process $\mathbf{f}_t$ can be expressed as an VAR(∞) process, which can be approximated by a finite-order VAR. This implies that $\mathbf{F}_t$ can also be approximated by a finite-order VAR as in (11) and (12) (see Berk 1974; Kuersteiner 2004 for theoretical results), particularly when the coefficients of the AR process decay quickly to 0. In general, the lag length should be chosen by a data-dependent method and is an increasing function of T. Although this procedure works well in practice, some theoretical issues are left unexplored when the VAR order increases with T. Proposition 2 is stated only for VARs of fixed orders, and in effect it does not cover case II. Nevertheless, we note that it is possible to extend the theory to include VAR(∞) processes with rapidly decaying coefficients. In particular, the theory can be extended to $\mathbf{F}_t$ as vector ARMA processes. For simplicity, this extension is not considered here. Monte Carlo simulations show that the procedure works well for $\mathbf{f}_t$ as MA processes.

The point that we highlight is that data generated by the dynamic model can always be mapped into a static model of the form $x_{it} = \boldsymbol{\Lambda}_i' \mathbf{F}_t + e_{it}$ by suitably defining a $\mathbf{F}_t$ that evolves according to a VAR with order depending on the dynamics of $\mathbf{f}_t$. The dimension of $\mathbf{F}_t$ is always $r = q(s+1)$ irrespective of the order of the VAR. Given that $\mathbf{A}(L)\mathbf{F}_t = \mathbf{R}\boldsymbol{\epsilon}_t$, the spectrum of $\mathbf{F}$ at frequency ω,

$$\mathbf{S}_\mathbf{F}(\omega) = \mathbf{A}(e^{-i\omega})^{-1} \mathbf{R} \mathbf{S}_\boldsymbol{\epsilon}(\omega) \mathbf{R}' \mathbf{A}(e^{i\omega})^{-1},$$

has rank q if $\mathbf{S}_\boldsymbol{\epsilon}(\omega)$ has rank q for all $|\omega| \leq \pi$. Accordingly, the spectrum of the static factors $\mathbf{S}_\mathbf{F}(\omega)$ will also have q nonzero eigenvalues. Therefore, we 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 1 cannot be used immediately to determine q. This is because $\mathbf{F}_t$ is not observable, and the convergence rate of $\hat{\boldsymbol{\Sigma}}_\mathbf{u}$ is not $\sqrt{T}$. We deal with these issues in the next section.

4. DETERMINING q

Let $\mathbf{S}_\mathbf{x}(\omega)$ be the population spectrum of the N cross-sectional units. The static model implies that

$$\mathbf{S}_\mathbf{x}(\omega) = \boldsymbol{\Lambda} \mathbf{S}_\mathbf{F}(\omega) \boldsymbol{\Lambda}' + \mathbf{S}_\mathbf{e}(\omega), \qquad -\pi \leq \omega \leq \pi.$$

Because $\mathbf{S}_\mathbf{e}(\omega)$ has rank N, $\mathbf{S}_\mathbf{x}(\omega)$ is also rank N. This would seem to suggest that q cannot be determined without working on $\mathbf{S}_\mathbf{F}(\omega)$. Such a procedure would necessitate choosing 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, with limiting distributions of the eigenvalues unnecessary.

If $\mathbf{F}_t$ is observed, and because it has a VAR representation, Proposition 1 then implies that q can be determined from a spectral decomposition of $\hat{\boldsymbol{\Sigma}}_\mathbf{u}$ provided that T is large. What prevents such an analysis is that neither $\mathbf{F}_t$ nor its dimension (r) is observed. However, the following holds. Let $\hat{\mathbf{F}}_t^r$ be the r factors obtained by the method of principal components; that is, let $\hat{\boldsymbol{\Lambda}}$ be a $N \times r$ matrix consisting of the r eigenvectors (multiplied by $\sqrt{N}$) associated with the r largest eigenvalues of the matrix $\mathbf{X}'\mathbf{X}$ in decreasing order. Then $\hat{\mathbf{F}} = \mathbf{X}\hat{\boldsymbol{\Lambda}}/N$. These principal component estimates adopt the normalization that $\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Lambda}}/N = \mathbf{I}_r$. Then, under the assumption that (a) $\boldsymbol{\Sigma}_\mathbf{F} = E(\mathbf{F}_t \mathbf{F}_t')$ and $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N$ are both rank r, (b) moment restrictions are satisfied, and (c) the time and cross-sectional 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 $\mathbf{H}$ of rank r, such that as $N, T \to \infty$ (ointly),

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

and

$$\Pr(\hat{k} = r) \xrightarrow{p} 1,$$

where

$$\hat{k} = \arg \min_k IC(k) = \arg \min_k \log(\sigma_k^2) + k C_{NT}$$

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

Importantly, the foregoing large-sample results assume that the second moment matrix of $\boldsymbol{\Lambda}_i$ and $\mathbf{F}_t$ is rank r. But $\boldsymbol{\Sigma}_\mathbf{F}$ may have rank less than r. For example, if $\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{R}\boldsymbol{\epsilon}_t$ is such that $\mathbf{A} = \rho \mathbf{I}_r$, $|\rho| < 1$, then $\mathrm{var}(\mathbf{F}_t) = \mathbf{R}\boldsymbol{\Sigma}_\boldsymbol{\epsilon}\mathbf{R}'/(1 - \rho^2)$, $\mathrm{var}(\mathbf{F}_t)$ has only rank $r^* = q$. In general, when the dynamics of $\mathbf{F}_t$ is sufficiently rich, $\mathrm{var}(\mathbf{F}_t)$ is of rank r even though the rank of $\boldsymbol{\Sigma}_\mathbf{u}$ is only q. But existing results do not cover cases when $\mathrm{var}(\mathbf{F}_t) < r$, which can arise as in the foregoing example when $\mathbf{F}_t$ has very simple dynamics. We now extend our results on determining the number of factors to also cover these special cases.

Lemma 1. Let $\mathbf{F}_t$ be a $r \times 1$ vector of factors generated by q primitive common shocks $\boldsymbol{\epsilon}_t$. Let $q \leq r^* \leq r$. Let $\boldsymbol{\Sigma}_\mathbf{F} = E(\mathbf{F}_t \mathbf{F}_t')$ and $\boldsymbol{\Sigma}_\boldsymbol{\Lambda} = \mathrm{plim}\, \boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N$. Suppose that the $r \times r$ matrix $\boldsymbol{\Sigma}_\mathbf{F} \cdot \boldsymbol{\Sigma}_\boldsymbol{\Lambda}$ has rank r^* and the remaining assumptions of Bai and Ng (2002) hold. Let $\hat{\mathbf{F}}_t^{r*}$ be the $r^* \times 1$ vector of factor estimates obtained by the method of principle components. There exists a matrix $\mathbf{H}^*$ with rank r^* such that (a) $\min[N, T](\frac{1}{T} \sum_{t=1}^{T} \|\hat{\mathbf{F}}_t^{r*} - \mathbf{H}^*\mathbf{F}_t\|^2) = O_p(1)$ and (b) $\Pr(\hat{k} = r^*) = 1$ if $\hat{k} = \arg \min_k IC(k)$.

Lemma 1 clarifies that when $\mathbf{F}_t$ is reduced rank, the method of principal components will estimate the space spanned by the r^* independent factors. The IC will select $r^* \leq r$ factors, because r^* is the rank of $\mathbf{\Sigma_F}$ (assuming that $\mathbf{\Sigma_\Lambda}$ is of full rank).

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

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

$$\min[\sqrt{N}, \sqrt{T}](\hat{\mathbf{\Sigma}}_{\mathbf{u}} - \mathbf{H}^* \tilde{\mathbf{\Sigma}}_{\mathbf{u}} \mathbf{H}^{*\prime}) = O_p(1).$$

The ultimate interest is in the rank of $\mathbf{\Sigma_u} = E(\mathbf{u}_t \mathbf{u}_t')$. But $\sqrt{T}(\tilde{\mathbf{\Sigma}}_{\mathbf{u}} - \mathbf{\Sigma_u}) = O_p(1)$. Lemma 2 thus implies that

$$\|\hat{\mathbf{\Sigma}}_{\mathbf{u}} - \mathbf{H}^* \mathbf{\Sigma_u} \mathbf{H}^{*\prime}\| = O_p(1/\min[\sqrt{N}, \sqrt{T}]).$$

Note that $\mathbf{\Sigma_u}$ and $\mathbf{H}^* \mathbf{\Sigma_u} \mathbf{H}^{*\prime}$ have the same rank because $\mathbf{H}^*$ 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 the following result.

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

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

and

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

Let $\hat{q}_3 = \min\{k \in \mathcal{K}_3\}$ and $\hat{q}_4 = \min\{k \in \mathcal{K}_4\}$. Then, under $\mathbf{H}_0$ with rank$(\mathbf{\Sigma_u}) = q$, we have $\hat{q}_3 \xrightarrow{p} q$ and $\hat{q}_4 \xrightarrow{p} q$ as $N, T \to \infty$.

Our main insight is to exploit the relationship between the dynamic and the static factors so that estimating the dynamic factors is not necessary to determine q. In our setup, if r^* factors explain τ percent of the variation in the data, then 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 criteria. This is in contrast to the approach of 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 $\mathbf{S_F}(\omega)$ and is a useful cross-check of the informal method used by Giannone et al. (2005). In independent

work completed the same time the first draft of this article was written, Stock and Watson (2005) also developed a test for q that uses a rather different approach. Instead of the rank of $\mathbf{S_F}(\omega)$, they estimated the rank of the restricted residuals of a $(N + r)$-dimensional VAR. In our notation, Stock and Watson started with $x_{it} = \mathbf{\Lambda}_i' \mathbf{F}_t + \rho_i(L)x_{it-1} + e_{it}$ to allow serial correlation in the idiosyncratic errors. The factor dynamics $\mathbf{A}(L)\mathbf{F}_t = \mathbf{R}\mathbf{u}_t$, where $\mathbf{A}(L) = \mathbf{I} - \mathbf{A}^+(L)L$ implies that $x_{it} = \mathbf{\Lambda}_i' \mathbf{A}^+(L)\mathbf{F}_{t-1} + \mathbf{\Lambda}_i' \mathbf{R}\boldsymbol{\epsilon}_t + \rho_i(L)x_{it-1} + e_{it}$. The composite residuals of a VAR in $\mathbf{X}_t$ and $\mathbf{F}_t$ is of the form $\mathbf{\Lambda}_i' \mathbf{R}\boldsymbol{\epsilon}_t + e_{it}$, which has q common factors $\boldsymbol{\epsilon}_t$. Stock and Watson exploited this factor representation to determine q using the criteria developed by Bai and Ng (2002).

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

5. SIMULATIONS

We consider the following data-generating processes (DGPs):

1. $x_{it} = (\boldsymbol{\lambda}_{i0} + \boldsymbol{\lambda}_{i1}L + \boldsymbol{\lambda}_{i2}L^2)' \mathbf{f}_t + e_{it}$, where $\mathbf{f}_t = \mathbf{C}(L)\boldsymbol{\epsilon}_t$ and $\mathbf{f}_t$ is q-dimensional
2. $x_{it} = (\boldsymbol{\lambda}_{i0} + \boldsymbol{\lambda}_{i1}L)' \mathbf{f}_t + e_{it}$, where $\mathbf{A}(L)\mathbf{f}_t = \boldsymbol{\epsilon}_t$ and $\mathbf{f}_t$ is q-dimensional
3. $x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + e_{it}$, where $\mathbf{F}_t = \mathbf{A}_1 \mathbf{F}_{t-1} + \mathbf{u}_t$ and $\mathbf{A}_1 = \rho \mathbf{I}_r$; $\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t$, rank$(\mathbf{R}) = q$
4. $x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + e_{it}$, where $\mathbf{F}_t = \mathbf{A}_1 \mathbf{F}_{t-1} + \mathbf{u}_t$, where $\mathbf{A}_1 = \text{diag}(.2, .375, .55, .725, .9)$ and $\mathbf{u}_t$ is the same as in DGP 3.

DGPs 1 and 2 are dynamic factor models considered by Forni et al. (2000) with $q = 2$ dynamic factors. DGP 1 assumes that $\mathbf{f}_t$ is a bivariate MA process with MA(1) parameters of .2 for f_{1t} and .9 for f_{2t}. DGP 2 assumes that $\mathbf{f}_t$ is a bivariate first-order AR process with AR(1) parameters of .2 for f_{1t} and .9 for f_{2t}. 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 was used by 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 that $r^* = q = 3$. There are many ways to generate $\mathbf{u}_t$ of the form $\mathbf{u}_t = \mathbf{R}\boldsymbol{\epsilon}_t$. Our particular method is as follows. Let $\mathbf{S}$ be a $r \times r$ diagonal matrix of rank q with nonzero elements drawn from uniform U(.8, 1.2) distribution. Let $\mathbf{\Gamma}$ be an arbitrary orthonormal matrix $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}_r$, obtained in Matlab through "orth(rand(r, r))." Then $\mathbf{u}_t$ is generated as $\mathbf{u}_t = \mathbf{\Gamma}\mathbf{S}\mathbf{\Gamma}'\mathbf{v}_t$, where $\mathbf{v}_t$ is an $r \times 1$ vector of iid normal variables. Note that $\mathbf{\Gamma}$ and $\mathbf{S}$ do not vary over t and i. The variance of $\mathbf{u}_t$ is $\mathbf{\Gamma}\mathbf{S}^2\mathbf{\Gamma}'$, with rank q. For DGP 4, $\mathbf{A}_1$ is a diagonal matrix with values .2, .375, .55, .725, and .9; the dynamics of $\mathbf{F}_t$ are thus richer than those of DGP 3. In this case, $q = 3$ and $r = 5$. In

Table 1. Estimated Number of Dynamic Factors Based on the Covariance Matrix of VAR Residuals

N	T	$\hat{r} \setminus m$	$\hat{q}_3$ 2	$\hat{q}_4$ 2	$\hat{q}_3$ 1.0	$\hat{q}_4$ 1.0	$\hat{q}_3$ .5	$\hat{q}_4$ .5
DGP 1: $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)'\mathbf{f}_t + e_{it}$, $\mathbf{f}_t = \boldsymbol{\epsilon}_t + \mathbf{C}_1\boldsymbol{\epsilon}_{t-1}$,								
		$q = 2$, $s = 2$, $r = 6$						
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: $x_{it} = (\lambda_{i0} + \lambda_{i1}L)'\mathbf{f}_t + e_{it}$, $\mathbf{f}_t = \mathbf{A}_1\mathbf{f}_{t-1} + \mathbf{u}_t$,								
		$q = 2$, $s = 1$, $r = 4$						
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: $x_{it} = \lambda_i'\mathbf{F}_t + e_{it}$, $\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \mathbf{u}_t$, $\mathbf{A}_1 = \rho\mathbf{I}_r$,								
		$q = 3$, $r = 5 (r^* = 3)$						
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: $x_{it} = \lambda_i'\mathbf{F}_t + e_{it}$, $\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \mathbf{R}\mathbf{u}_t$,								
		$q = 3$, $r = 5$						
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 entries are the average values over 1,000 iterations.

Table 2. Estimated Number of Dynamic Factors Based on the Correlation Matrix of VAR Residuals

N	T	$\hat{r} \setminus m$	$\hat{q}_3$ 1.25	$\hat{q}_4$ 2.25	$\hat{q}_3$ 1.0	$\hat{q}_4$ 2.0	$\hat{q}_3$ 1.5	$\hat{q}_4$ 2.5
DGP 1: $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)'\mathbf{f}_t + e_{it}$, $\mathbf{f}_t = \boldsymbol{\epsilon}_t + \mathbf{C}_1\boldsymbol{\epsilon}_{t-1}$,								
		$q = 2$, $s = 2$, $r = 6$						
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: $x_{it} = (\lambda_{i0} + \lambda_{i1}L)'\mathbf{f}_t + e_{it}$, $\mathbf{f}_t = \mathbf{A}_1\mathbf{f}_{t-1} + \mathbf{u}_t$,								
		$q = 2$, $s = 1$, $r = 4$						
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: $x_{it} = \lambda_i'\mathbf{F}_t + e_{it}$, $\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \mathbf{u}_t$, $\mathbf{A}_1 = \rho\mathbf{I}_r$,								
		$q = 3$, $r = 5 (r^* = 3)$						
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: $x_{it} = \lambda_i'\mathbf{F}_t + e_{it}$, $\mathbf{F}_t = \mathbf{A}_1\mathbf{F}_{t-1} + \mathbf{R}\mathbf{u}_t$,								
		$q = 3$, $r = 5$						
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: The entries are the average values over 1,000 iterations.

addition, $r^* = r$ for large T, but r^* can be less than r for finite T. In all four DGPs, we assume that λ_{ij}, e_{it}, and $\boldsymbol{\epsilon}_t$ 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 $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N = \mathbf{I}_{r^*}$. The number of factors is estimated by the IC as done by Bai and Ng (2002). Specifically,

$$\hat{r}^* = \arg \min_{k \in [0, 2r]} \log(\hat{\sigma}_k^2) + k \frac{\log(\min[N, T])}{NT/(N+T)},$$

where $\hat{\sigma}_k^2 = \frac{1}{NT} \sum_i \sum_t (x_{it} - \hat{\boldsymbol{\lambda}}_i^k \hat{\mathbf{F}}_t^k)^2$, $\hat{\mathbf{F}}_t^k$ is $k \times 1$. Given $\hat{\mathbf{F}}_t$, a $\hat{r}^*$-dimensional VAR in $\hat{\mathbf{F}}_t$ is estimated to obtain $\hat{\mathbf{u}}_t$. Selecting too few lags will be problematic, because $\hat{\mathbf{u}}_t$ will not be innovations. We report results for VAR(2); results for higher lags are similar. Given $\hat{\mathbf{u}}_t$, 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 $m^* = m / \min[N^{2/5}, T^{2/5}]$. The number of simulations is 1,000.

Table 1 reports the average values for $\hat{r}$ and $\hat{q}$ estimated based on the covariance matrix of $\hat{\mathbf{u}}_t$. Table 2 reports the corresponding results based on the correlation matrix of $\hat{\mathbf{u}}_t$. The four panels correspond to four different DGPs. For all cases, the number of static factors $\hat{r}$ 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 $\hat{r}^*$ is overestimated, $\hat{q}$ can be very close to q for suitable choice of m.

As noted earlier, when the 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 United States

To illustrate, we take data used by Stock and Watson (2005), which can be downloaded at *http://www.princeton.edu/~mwatson*. A total of 132 monthly time series are available from 1960:1 to 2003:12. The data are transformed (by taking logs, first or second difference) as was done by 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 a covariance matrix is used. But when discrepancy exists, the estimated q from the latter method tends to be higher than that from the former method. Thus to assert that q is larger than 2, 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 $r = 2, 3$; $\hat{q} = 3$, when $r = 4, 5, 6$; and $\hat{q} = 4$ 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 the closest τ 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 standpoint. 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 12 variables. It is thus useful to consider results for cutoffs other than $\tau = .6$. We also determine the number of static factors using the IC. We denoted this by $\hat{r}_t^*$, and the corresponding number of primitive factors by q_t^*. We compute all of these statistics for t ranging from 133 to 528, corresponding to estimation ending in 1970:12 and 2003:12. We thus have 396 statistics, one statistic for every t.

Table 3 reports 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 $T = 528$). $R_{\hat{r}}^2$ is the average explanatory power of $\hat{r}$ factors, when $\hat{r}$ is chosen with cutoff of .3, .4, .5, or .6, and also optimally. The column $R_{\hat{q}}^2$ 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 of the variation. To explain .619 of the variation in the data would require, on average, 15 factors. When determined optimally by IC, the data suggest that seven static factors explain

Table 3. Empirical Analysis

	τ	T	$R_{\hat{r}}^2$	$R_{\hat{q}}^2$	$\hat{r}$	$\hat{q}_3$	$\hat{q}_4$
Mean	.3	330.500	.323	.916	3.121	2.255	2.066
Min	.3	133.000	.300	.882	3.000	2.000	1.000
Max	.3	528.000	.348	1.000	4.000	4.000	3.000
Mean	.4	330.500	.418	.906	5.386	3.533	3.101
Min	.4	133.000	.400	.869	5.000	3.000	3.000
Max	.4	528.000	.438	.967	6.000	4.000	4.000
Mean	.5	330.500	.511	.902	8.833	6.068	5.210
Min	.5	133.000	.500	.805	8.000	5.000	4.000
Max	.5	528.000	.524	.952	10.000	7.000	6.000
Mean	.6	330.500	.608	.848	13.811	8.523	7.773
Min	.6	133.000	.600	.714	12.000	6.000	6.000
Max	.6	528.000	.619	.908	15.000	10.000	9.000
Mean	r^*	330.500	.430	.910	5.763	3.864	3.119
Min	r^*	133.000	.239	.819	2.000	2.000	2.000
Max	r^*	528.000	.460	1.000	7.000	4.000	4.000

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

on average .460 of the variation in the data over the full sample, and that four dynamic factors span 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 article, we applied the tests to a different dataset for the sample 1960:1–1998:12. We found an average of 7 dynamic factors in 10 static factors. Thus the evidence is 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 time series plot of $\hat{r}_t^*$, $\hat{q}_{3t}$, and $\hat{q}_{4t}$. As we can see, $\hat{r}$ jumped from 4 to 6 around 1973 and has remained roughly at 6 up to 2000. On the other hand, $\hat{q}$ jumped from 3 to 4 and stayed at 4 most of the time. If we had ended
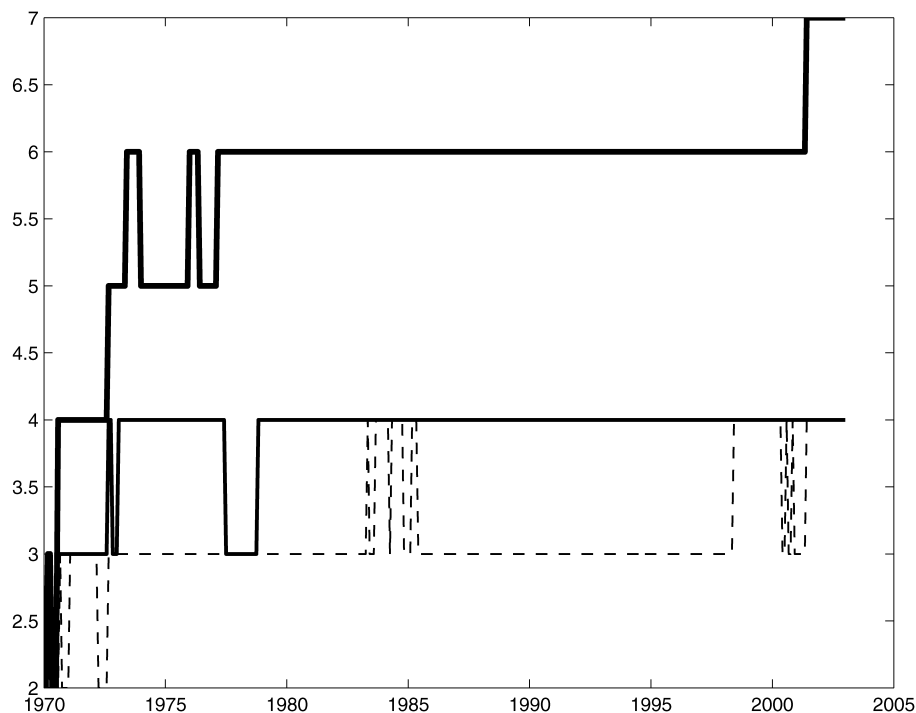
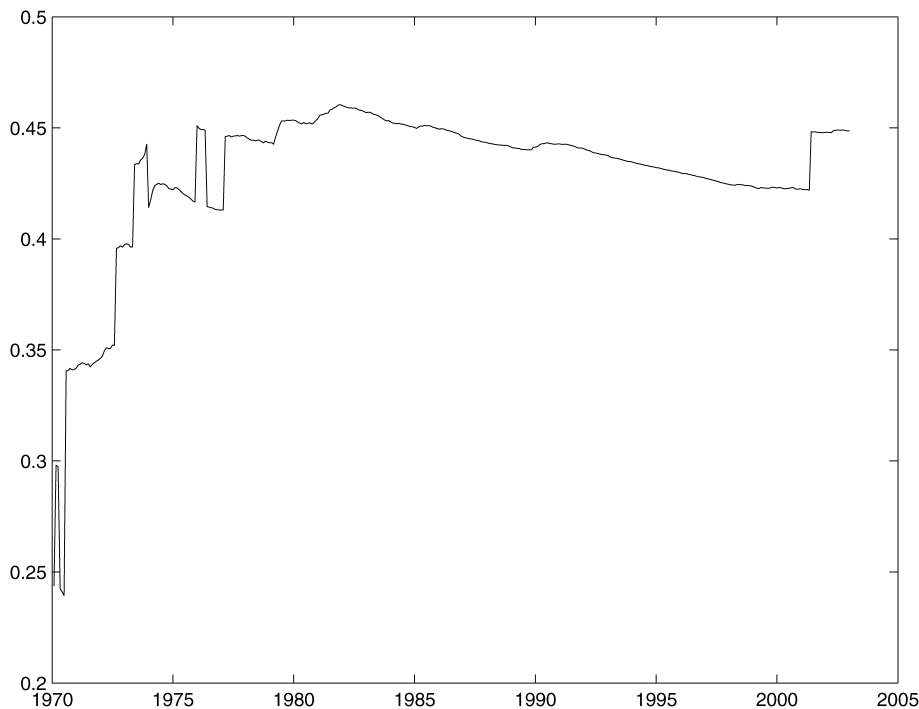*Figure 1. Estimated r, q_3, and q_4 (━━ r; ━━ q_3; - - - q_4).*

Figure 2. Importance of the Common Component (——— R^2).

the estimation in 2000, then we would have $\hat{r} = 6$ and $\hat{q} = 4$. However, in the past few years, $\hat{r}$ seems to have taken another jump from 6 to 7, although $\hat{q}$ seems to have stayed at 4.

Figure 2 plots R_t^2, 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 1970s to peak at about .5 in the early 1990s. 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. Although 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 2 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 findings suggest that a total of 6 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 restricted their attention to only 12 series when arriving at the conclusion that q is 2. 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 rate), PUNEW (CPI), and A0M224r (consumption expenditure). Reported are the R^2 from a regression of x_{it} on 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 4 that the explanatory power of the factors tends to be higher for the selected series than for the panel as a whole. If we had focused on these series, then fewer static factors would have been necessary. It is conceivable that four or five static factors adequately explain the selected series, which implies two or three dynamic factors.

6. CONCLUSION

This article has proposed a procedure for determining the number of primitive common shocks in a large number of series. By making the link between the dynamic and the static factors precise, we arrived 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 estimating 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 4.

Table 4. Explanatory Power of $\hat{r}$ Factors

Series	r							
	1	2	3	4	5	6	7	8
ALL	.172	.242	.296	.350	.393	.429	.460	.486
IPS10	.690	.723	.774	.779	.869	.869	.875	.911
A0M059	.060	.140	.149	.152	.183	.321	.321	.341
A0M057	.266	.362	.390	.397	.455	.557	.559	.562
FYFF	.180	.332	.443	.445	.478	.482	.505	.515
PUNEW	.008	.028	.081	.706	.711	.723	.726	.726
A0M224R	.066	.142	.153	.158	.174	.227	.231	.279

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

REFERENCES

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

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

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

――――― (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., Giannone, D., Lippi, M., and Reichlin, L. (2003), "Opening the Black Box: Identifying Shocks and Propagation Mechanisms in VAR and Factor Models," mimeo, Center for Economic Policy Research (CEPR).

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

――――― (2005), "The Generalized Dynamic Factor Model, One-Sided Estimation and Forecasting," *Journal of the American Statistical Association, 100, 830–840.*

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

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. (2006), "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

Kuersteiner, G. (2004), "Automatic Inference for Infinite-Order Vector Autoregressions," *Econometric Theory*, 21, 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 *New Methods in Business Cycle Research*, ed. C. Sims, Minneapolis, MN: Federal Reserve Bank of Minneapolis, pp. 45–109.

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.

――――― (2002b), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics, 20, 147–162.*

――――― (2005), "Implications of Dynamic Factor Models for VAR Analysis," Working Paper 11467, National Bureau of Economic Statistics.