Determining the Number of Primitive Shocks in Factor Models

Jushan Bai* Serena Ng †

August 5, 2005

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. Our analysis is based on the residuals of a VAR in r static factors, where the factors are themselves obtained by applying the method of principal components to a large panel of data. Our algebraic tests are based on a spectral decomposition of the covariance matrix of the residuals of a VAR. The tests are exact if the residuals were observable. Because of sampling variability from having to estimate the VAR, the tests are accurate up to an error that vanishes asymptotically. We show how this error depends on the rate of convergence of the sample covariance to the population covariance of the true innovations. An important aspect of the present analysis is to make precise the relation between the dynamic factors and the static factors, which is a result of independent interest.

*Department of Economics, NYU, 269 Mercer St, New York, NY 10003 Email: Jushan.Bai@nyu.edu.

†Department of Economics, University of Michigan, Ann Arbor, MI 48109 Email: Serena.Ng@umich.edu

The authors acknowledge financial support from the NSF (grants SES-0137084, SES-0136923)

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 q. More precisely, the q that we determine is the rank of the spectral density matrix of 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, if any, 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. (2004) 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. (2004) 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 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.

We begin in Section 2 with a discussion of how q can be determined from r observed time series. In cases of practical interest, the r innovations of interest are not observed. Accordingly, our ability to precisely determine q depends on our ability to obtain consistent estimates of the r innovations. We motivate the tests in the context of a canonical VAR in Section 3. Tests for the number of dynamic factors are formally developed in Sections 4 and 5. Simulations and two applications are considered in Section 6.

2 Preliminaries

We first need to make precise what is the q that we seek to determine.

Definition 1 *Let u_t be a $r \times 1$ vector of* innovations. *We say that u_t is driven by a minimal number of q innovations if there exists a $r \times q$ matrix, R, such that*

$$u_t = R\varepsilon_t \tag{1}$$

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

Our definition of 'primitive' is algebraic. The number of primitive shocks in u_t is simply the q linearly independent shocks that span u_t. Those familiar with the literature on common features, coitegrating vectors, or canonical analysis would notice a resemblance between these concepts and our definition of primitiveness. These analysis seek a reduced rank matrix $R = \gamma\alpha'$, where γ and α are both $r \times q$, satisfying

$$u_t = R\varepsilon_t + v_t$$

and such that u_t and ε_t have time series properties different from v_t. In coitegration analysis of a r vector of variables x_t, $u_t = \Delta x_t$, $\varepsilon_t = x_{t-1}$, and $u_t = R\varepsilon_t + v_t$ is just the error-correction representation. While some shocks in the system have permanent effects while others are transitory, the total number of shocks always equals the number of variables in the system. In a well known paper, King et al. (1991) used cointegration restrictions to isolate one permanent and two transitory shocks in a system of three variables. Their analysis uses the

fact that the population long-run covariance matrix of the differenced data is reduced rank when there are $q < r$ permanent shocks. We do not make use of cointegration restrictions. In fact, we assume that the variables under consideration are (or transformed to be) stationary. The unique number of shocks is simply the rank of R, which can be less than r, because there is no v_t in our analysis.

We make use of the following result from matrix analysis:

Lemma 1 *Let A be a $r \times r$ symmetric, then we can write $A = BCB'$ where B is an $r \times r$ orthogonal matrix, and C is a diagonal matrix of eigenvalues. Let β_j be the j-th column of B and let c_j be the j-eigenvalue of A ordered such that $c_1 > c_2 \ldots > c_r$. The spectral decomposition of A is*

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

The j-th column of B is the eigenvector associated with the j-th largest eigenvalue of A. An implication of Lemma 1 is that if A has $r - q$ eigenvalues that are zero, then

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

Let $d_k = vech(A(k))$, where

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

We will refer to $A(k)$ as the k^{th} pseudo matrix of A. We also define $d_0 = d_r = vech(A)$. Define

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

Under $H_0 : rank(A) = q$, $c_k = 0$ for $k > q$. This implies $d_k = d_q$ for $k \geq q$. Thus if rank$(A) = q$, $D_{1,k} = D_{2,k} = 0$ exactly, for $k \geq q$. These are measures in ratios and are scale free.

In the next two sections, the A matrix whose rank is to be determined is Σ_u, the covariance matrix of a set of innovations. Lemma 1 cannot be applied immediately because Σ_u is not observed, but it can be estimated from the data. Considering its associated pseudo-matrices yields $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$. It will shown that $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ will converge to zero asymptotically at

a rate that depends on the convergence rate of $\widehat{\Sigma}_u$ to Σ_u, where $\widehat{\Sigma}_u$ is a consistent estimate of Σ_u.

Before turning to our main analysis, a remark on existing rank tests is in order. Available tests seek to determine the rank of a $m \times n$ matrix, say, R, when R is consistently estimated from the regression $u_t = R\varepsilon_t + v_t$. First of all, our focus is in problems for which v_t plays no role, but with unobservable u_t. Furthermore, existing tests are built upon the fact that the rank of a matrix is the number of non-zero eigenvalues it possesses. Thus if $\widehat{R} \xrightarrow{p} R$, then by continuity of eigenvalues, the estimated eigenvalues can in theory be used for hypothesis testing. But rank tests tend not to be asymptotically normal, see Anderson (1951), Gill and Lewbel (1992), Cragg and Donald (1996), Cragg and Donald (1997), and Robin and Smith (2000). Recently, Kleibergen and Paap (2003) and Ratsimalahelo (2003) suggest orthogonal rotation of the sample eigenvalues around the origin to restore normality. In doing so, these tests necessitate a consistent estimate of $\text{var}(\widehat{R})$. The matrix whose rank we seek to test is Σ_u. This creates two problems. First, the estimation of $\text{var}(\widehat{\Sigma}_u)$ entails evaluation of a matrix of fourth moments, which tend to be quite imprecisely estimated unless the sample size is extremely large. Second, Σ_u is a variance covariance matrix which only has $r(r+1)/2$ unique elements. Thus, $\text{var}(\Sigma_u)$ and its estimate do not have full rank, an assumption maintained by Kleibergen and Paap (2003). The test of Ratsimalahelo (2003) allows for reduced rank in the variance of apparently matrices that are not symmetric. Our attempts to adopt existing tests have not been successful. This motivates the development of a test aimed specifically at determining q as laid out in Definition 1.

3 The Minimal Number of Primitive Shocks in a VAR

Consider a r vector of (observed) stationary time series, x_t, $t = 1, \ldots T$. Assume that the dynamics of x_t are well approximated by a VAR(p), where p is finite. We have

$$A(L)x_t = u_t \tag{2}$$

where $A(L) = I - A_1 L - \ldots - A_p L^p$. The r variable VAR is said to have q minimal primitive shocks, denoted ε_t, if $u_t = R\varepsilon_t$, where R is a $r \times q$ matrix.[1] Consideration of such a VAR structure is useful in developing our main analysis, which concerns distinguishing the dynamic factors from the static factors.

[1] Such a VAR can be motivated by log-linearizing a dynamic optimization under perfect foresight. Allowing for expectational error would lead to an error structure of the form $u_t = R\varepsilon_t + v_t$.

Let $\widehat{A}_i$ be a $\sqrt{T}$ consistent estimate of A_i, with A_i being the $r \cdot p$ vector of coefficients in the i-th equation, for $i = 1, \ldots r$. Let $\widehat{u}_i$ be the corresponding set of $T \times 1$ residuals. Let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t'$ be an estimate of $\Sigma_u = E(u_t u_t')$. Then

$$\sqrt{T}(\widehat{\Sigma}_u - \Sigma_u) = O_p(1).$$

Consider the spectral decomposition $\Sigma_u = BCB'$ and $\widehat{\Sigma}_u = \widehat{B}\widehat{C}\widehat{B}$. By continuity of eigenvalues and the spaces spanned by the eigenvectors, we have $\sqrt{T}(\widehat{\Sigma}_u(k) - \Sigma_u(k)) = O_p(1)$, where $\widehat{\Sigma}_u(k) = \sum_{j=1}^{k} \widehat{c}_j \widehat{\beta}_j \widehat{\beta}_j'$ and $\Sigma_u(k) = \sum_{j=1}^{k} c_j \beta_j \beta_j'$. Let $\widehat{d}_k = \text{vech}(\widehat{\Sigma}_u(k))$, it follows that $\left\| \widehat{d}_k - d_k \right\| = O_p(T^{-1/2})$. Furthermore,

$$\left\| \widehat{d}_{k+1} - \widehat{d}_k \right\| \leq \|d_{k+1} - d_k\| + \left\| d_{k+1} - \widehat{d}_{k+1} \right\| + \left\| d_k - \widehat{d}_k \right\|.$$

Define

$$\begin{aligned} \widehat{D}_{1,k} &= \left\| \widehat{d}_{k+1} - \widehat{d}_k \right\| / \left\| \widehat{d}_0 \right\| \\ \widehat{D}_{2,k} &= \left\| \widehat{d}_{k+1} - \widehat{d}_0 \right\| / \left\| \widehat{d}_0 \right\|. \end{aligned}$$

From $\widehat{D}_{1,k} - D_{1,k} = O_p(T^{-1/2})$, we have $\widehat{D}_{1,k} = O_p(T^{-1/2})$ for $k \geq q$ because $D_{1,k} = 0$. In addition, for $k < q$, $\widehat{D}_{1,k} \geq c > 0$ with large probability for some c because $D_{1,k} > 0$ for $k < q$. Similarly, $\widehat{D}_{2,k} = O_p(T^{-1/2})$ for $k \geq q$ and $\widehat{D}_{2,k} \geq c > 0$ for $k < q$. Define

$$\mathcal{K}_1 = \{k : \widehat{D}_{1,k} < m/T^{1/2-\delta}\} \quad (3)$$

$$\mathcal{K}_2 = \{k : \widehat{D}_{2,k} < m/T^{1/2-\delta}\}. \quad (4)$$

where $m > 0$ and $0 < \delta < 1/2$.

Proposition 1 *Let $\widehat{\Sigma}_u$ be a $\sqrt{T}$ consistent estimator of Σ_u. Let $0 < m < \infty$ be a generic positive constant and $0 < \delta < 1/2$. With $\mathcal{K}_1$ defined by (3) and $\mathcal{K}_2$ defined by (4), let $\widehat{q}_1 = \min\{k \in \mathcal{K}_1\}$ and $\widehat{q}_2 = \min\{k \in \mathcal{K}_2\}$. Then under H_0 that rank(Σ_u)=q, $\widehat{q}_1 \xrightarrow{p} q$ and $\widehat{q}_2 \xrightarrow{p} q$ as $T \to \infty$.*

From $\widehat{D}_k = O_p(T^{-1/2})$ for $k \geq q$, we have $\widehat{D}_k < m/T^{1/2-\delta}$ with probability tending to 1 as $T \to \infty$. This means that $q \in \mathcal{K}$ for large T. But $q - 1$ does not belong to $\mathcal{K}$ because $\widehat{D}_k > c > 0$ and thus greater than $m/T^{1/2-\delta}$ for $k < q$. This gives the consistency result. While $D_k = 0$ for $k \geq q$, $\widehat{D}_k$ will not be exactly zero. The cut-off point $m/T^{\frac{1}{2}-\delta}$ is thus the tolerated error induced by sampling variability from estimation of Σ_u. Throughout, we let

$\delta = 0.1$ so that the tolerance is $m/T^{2/5}$. For $T = 100$, setting m to 0.5 translates into a tolerance of 0.079.

Our estimate of q is the smallest k at which $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ are approximately zero. Both $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ are unit free measure of how far are the unique elements of the k-th pseudo matrix of Σ_k from Σ_u. In large samples, the two tests should arrive at the same conclusion.

Notice that we have circumvented the problem of testing the eigenvalues of $\widehat{\Sigma}_u$ directly. In finite samples, the $q+1$ eigenvalues of the matrix may be small but not numerically zero. Establishing if they are exactly zero will be difficult. We also bypassed the estimation and/or inversion of $\text{var}(\widehat{\Sigma}_u)$, while still accounting for sampling error that arises from having to estimate Σ_u. Instead, we use $\widehat{D}_{1,k}$ and $\widehat{D}_{2,k}$ as metric to judge if the eigenvalues are small enough to render the k-th pseudo matrices indistinguishable from the true Σ_u.

4 Dynamic vs. 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, Λ_i is the vector of factor loadings for unit i on the r (static) common factors F_t. The *dynamic* model is written as $x_{it} = \lambda_i'(L) f_t + e_{it}$ where $\lambda_i(L)$ is the dynamic factor loading matrix of order s. In general, $f_t = C(L)\varepsilon_t$, where ε_t are iid vectors and $x_{it} = \lambda_i(L)C(L)\varepsilon_t + e_{it}$. The dimension of f_t, which is the same as the dimension of ε_t, is called the number of dynamic factors; this is denoted by q. Dynamic factor model can be written as static factor models, but then 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. In general, we expect $r \geq q$. 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 be consistently estimated when N and T are both large.[2]

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

[2]See, e.g., Forni et al. (2000), Ding and Hwang (2001), Stock and Watson (2002a), Bai and Ng (2002), and Bai (2003).

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. We will ultimately propose a way to determine q from r estimated static factors. 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 the 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.

4.1 Putting the Dynamic Model into Static Form

Consider the dynamic factor model

$$\begin{aligned} x_{it} &= \lambda_{i1}' f_t + \lambda_{i2}' f_{t-1} + \cdots + \lambda_{is} f_{t-s} + e_{it} \\ &= \lambda_i'(L) f_t + e_{it} \end{aligned} \tag{5}$$

where f_t is q dimensional, and

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

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

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

where

$$\Lambda_i = \begin{bmatrix} \lambda_{i1} \\ \lambda_{i2} \\ \vdots \\ \lambda_{is} \end{bmatrix} \quad \text{and} \quad F_t = \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-s} \end{bmatrix}. \tag{8}$$

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

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

where q is the dimension of f_t. Although the relation between x_{it} and F_t is static, F_t itself can be a dynamic process, depending on the dynamics of f_t. We consider two cases: f_t is a finite order autoregressive process, and f_t has a moving-average structure.

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

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

From $F_t = (f_t', f_{t-1}', \ldots f_{t-s}')'$ and since f_t is q dimensional, the number of static factors is clearly $r = (s+1)q$, which does not depend on h, the order of the dynamic process governing f_t in (9). Unless $s = 0$, the number of static factors is larger than the number of dynamic factors.

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

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

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

$$\begin{aligned} F_t^* &= A F_{t-1}^* + u_t \\ u_t &= R \varepsilon_t \end{aligned} \tag{10}$$

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

$$\begin{aligned} F_t &= A_1 F_{t-1} + A_2 F_{t-1} + \cdots + A_p F_{t-p} + u_t \quad (11) \\ u_t &= (I_q, 0, ..., 0)' \varepsilon_t \quad (12) \end{aligned}$$

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

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

To write $F_t = (f_t', ..., f_{t-s}')'$ in the AR form, define $F_t^* = (\varepsilon_t', \varepsilon_{t-1}', ..., \varepsilon_{t-s-h}')'$. Then

$$\begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-s-h} \end{bmatrix} = \begin{pmatrix} 0 & 0 & . & . & 0 \\ I_q & 0 & . & . & 0 \\ 0 & I_q & 0 & . & 0 \\ . & . & . & 0 & 0 \\ 0. & . & . & I_q & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \\ \varepsilon_{t-s-h-1} \end{pmatrix} + \begin{pmatrix} I_q \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_t.$$

That is,

$$F_t^* = A F_{t-1}^* + R^* \varepsilon_t.$$

From $F_t = (f_t', ..., f_{t-s}')'$, we can write F_t as

$$F_t = D F_t^*$$

where D is a $q(s+1) \times q(s+h+1)$ matrix with the following form

$$D = \begin{pmatrix} I_q & C_1 & . & . & C_h & 0 & . & . \\ 0 & I_q & C_1 & . & . & C_h & . & . \\ . & . & . & . & . & . & . & . \\ 0. & . & 0 & I_q & C_1 & . & . & C_h \end{pmatrix}.$$

Note that F_t is of dimension $r = q(s+1)$, the same as when f_t is an AR(h). However, because F_t is a linear combination of F_t^* when f_t is $MA(h)$, F_t is now a vector ARMA process driven by the q shocks ε_t, even though F_t^* is a vector AR process with no moving average part. But a VARMA process also has an infinite order VAR representation. Thus,

$$\begin{aligned} F_t &= A_1 F_{t-1} + A_2 F_{t-1} + \cdots + \ldots + u_t \quad (14) \\ u_t &= R \varepsilon_t \quad (15) \end{aligned}$$

where $R = DR^*$ has rank q. Although F_t is an infinite order VAR, we can approximate it by a finite order VAR.

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

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

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

Although the dynamics of the static factors are in the same form as the observable VAR system in (2) in the sense that both are driven by shocks with dimension less than the dimension of the variables, Proposition I cannot be used immediately to determine q. This is because F_t is not observable, and the convergence rate of $\widehat{\Sigma}_u$ is not $\sqrt{T}$. These issues are dealt with in the next section.

5 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 $S_F(\omega)$ or $S_f(\omega)$. Such a procedure would necessitate the choice of many auxiliary parameters (such as bandwidth and kernel), and even then, we do not have a formal theory for determining q. We now show how q can be estimated in the time domain, and limiting distributions of the eigenvalues are not necessary.

If F_t were observed, and since it has a VAR representation, Proposition 1 then implies that q can be determined from a spectral decomposition of $\widehat{\Sigma}_u$ provided T is large. What prevents such an analysis is that neither F_t nor its dimension (r) is observed. However, the following holds. Let $\widehat{F}_t^r$ be the r factors obtained by the method of principal components. That is, let $\widehat{\Lambda}$ be a $N \times r$ matrix consisting of the r eigenvectors (multiplied by $\sqrt{N}$) associated with the r largest eigenvalues of the matrix $X'X$ in decreasing order. Then $\widehat{F} = X\widehat{\Lambda}/N$. These

principal component estimates adopt the normalization that $\widehat{\Lambda}'\widehat{\Lambda}/N = I_r$. Then under the assumption that (i) $\Sigma_F = E(F_t F_t')$ and $\Lambda'\Lambda/N$ are both rank r, (ii) moment restrictions are satisfied, and (iii) that the time and cross-section correlation in the idiosyncratic errors is weak, Bai and Ng (2002) and Bai (2003) showed that if the data are generated by the static factor model, then as $N, T \to \infty$, there exists a matrix H, of rank r, such that as $N, T \to \infty$ (jointly),

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

and

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

where

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

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

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

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

Lemma 2 clarifies that when F_t is reduced rank, the method of principal components will estimate the space spanned by the r^* independent factors. The IC will select $r^* \leq r$ factors, since r^* is the rank of F_t.

Consider now the determination of q given $\widehat{F}_t$, where $\widehat{F}_t$ are the $\widehat{r}^*$ factors obtained by the static method of principal components. For notation simplicity, the dimension of $\widehat{F}_t$ is suppressed, but is understood to be of dimension $\widehat{r}^*$, where $\widehat{r}^*$ is determined by the IC. Let

$\widehat{u}_t$ be the residuals from estimating a VAR(p) in $\widehat{F}_t$, and let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t\widehat{u}_t$. It should be remarked that since we can only estimate the space spanned by the factors, we need the residuals from estimation of a VAR in $\widehat{F}_t$. Performing $\widehat{r}^*$ univariate autoregressions for each of the $\widehat{F}_t$ will not be appropriate.

Lemma 3 *Consider the model $x_{it} = \lambda_i' F_t + e_{it}$ with $A(L)F_t = u_t$, where $A(L)$ is possibly of infinite order. Let $\widehat{F}_t$ be the $r^* \times 1$ factors estimated by the method of principal components under the normalization that $\Lambda'\Lambda/N = I_{r^*}$, where $q \leq r^* \leq r$. Let $\widehat{u}_t$ be the residuals obtained by least squares estimation of a VAR(p) in $\widehat{F}_t$. Let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=}^{T} \widehat{u}_t'\widehat{u}_t$, and let $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} u_t u_t'$. Then*

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

Now $\widetilde{\Sigma}_u$ is the sample covariance matrix that would have obtained if F_t were observable. The ultimate interest is the rank of $\Sigma_u = E(u_t u_t')$. But $\sqrt{T}(\widetilde{\Sigma}_u - \Sigma_u) = O_p(1)$. Lemma 3 thus implies that

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

Note that Σ_u and $H^*\Sigma_u H^{*\prime}$ have the same rank since H^* is of full rank. Let $\widehat{d}_k = \text{vech}(\widehat{\Sigma}_u(k))$ and $d_k = \text{vech}(H^*\Sigma_u H^{*\prime}(k))$. It follows that $\left\|\widehat{d}_k - d_k\right\| = O_p(1/\min[\sqrt{N}, \sqrt{T}])$. Define

$$\begin{aligned}
\widehat{D}_{3,k} &= \left\|\widehat{d}_{k+1} - \widehat{d}_k\right\| / \left\|\widehat{d}_0\right\| \\
\widehat{D}_{4,k} &= \left\|\widehat{d}_k - \widehat{d}_0\right\| / \left\|\widehat{d}_0\right\|.
\end{aligned}$$

Let $0 < m < \infty$ be a generic positive constant. We have $\widehat{D}_{3,k} - D_{3,k} = O_p(1/\min[\sqrt{N}, \sqrt{T}])$ and $\widehat{D}_{4,k} - D_{4,k} = O_p(1/\min[\sqrt{N}, \sqrt{T}])$. For $k \geq q$, because $D_{3,k} = 0$, we have $\widehat{D}_{3,k} = O_p(1/\min[\sqrt{N}, \sqrt{T}])$. Similarly, for $k \geq q$, $\widehat{D}_{4,k} = O_p(1/\min[\sqrt{N}, \sqrt{T}])$. Let

$$\mathcal{K}_3 = \{k : \widehat{D}_{3,k} < m/\min[N^{2/5}, T^{2/5}]\} \quad (16)$$

$$\mathcal{K}_4 = \{k : \widehat{D}_{4,k} < m/\min[N^{2/5}, T^{2/5}]\}. \quad (17)$$

Proposition 2 *Let $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t\widehat{u}_t'$, where $\widehat{u}_t$ are the residuals from estimation of a VAR in $\widehat{F}_t$, $\widehat{F}_t$ being the principal components estimator for F_t. With $\mathcal{K}_3$ defined by (16) and $\mathcal{K}_4$ defined by (17), let $\widehat{q}_3 = \min\{k \in \mathcal{K}_3\}$ and $\widehat{q}_4 = \min\{k \in \mathcal{K}_4\}$. Then under H_0 with rank(Σ_u)=q, we have $\widehat{q}_3 \xrightarrow{p} q$ and $\widehat{q}_4 \xrightarrow{p} q$ as $N, T \to \infty$.*

Our main insight is to exploit the relation between the dynamic and the static factors so that estimation of the dynamic factors is not necessary to determine q. In our setup, if r^* factors explain θ 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 θ are jointly optimally determined. This is in contrast to Giannone et al. (2004), in which q is chosen for a subjectively chosen θ.

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. (2004). In independent work that was completed the same time as the first draft of this paper was written, Stock and Watson (2005) also developed a test for q that uses a rather different approach. Instead of the rank of $S_F(\omega)$, they estimate the rank of the restricted residuals of a $N+r$-dimensional VAR. In our notation, Stock and Watson starts with $x_{it} = \Lambda_i' F_t + \rho_i(L)x_{it-1} + e_{it}$ to allow serial correlation in the idiosyncratic errors. The factor dynamics $A(L)F_t = Ru_t$ where $A(L) = I - A^+(L)L$ implies that $x_{it} = \Lambda^{i\prime} A^+(L)F_{t-1} + \Lambda_i' R\varepsilon_t + \rho_i(L)x_{it-1} + e_{it}$. The composite residuals of a VAR in X_t and F_t is of the form $\Lambda_i' R\varepsilon_t + e_{it}$. Stock and Watson exploit this factor representation to determine q using the criteria developed in Bai and Ng (2002). As we will see, our method arrives at the same conclusion with the Stock and Watson procedure for the application being considered.

6 Simulations

We consider two sets of simulation exercises. Subsection 1 considers situations when $\widehat{\Sigma}_u$ is $\sqrt{T}$ consistent for Σ_u. Subsection 2 is aimed specifically at factor models.

6.1 Two Simple Examples

We begin with the simple exercise of determining the rank of a $r \times r$ covariance matrix associated with some $T \times r$ data matrix, u. For each $t = 1, \ldots T$, data are generated as $u_t = R\varepsilon_t$ where R is the Choleski decomposition of BCB', B is orthonormal, and C is a diagonal matrix with q non-zero eigenvalues and $r - q$ eigenvalues set to 1e-12. Although these last eigenvalues are zero for practical purposes, they are non-zero in terms of machine precision to permit Choleski decomposition to be performed.

We suppose that we do not observe u_t, but that $\sqrt{T}$ consistent estimates of u_t are available. Thus in each replication, we generate

$$\widehat{u}_t = u_t + error/\sqrt{T}$$

so that $\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t\widehat{u}_t'$ is $\sqrt{T}$ consistent for $\Sigma_u = \frac{1}{T}\sum_{t=1}^{T} E(u_t u_t')$. In the simulations, the perturbation errors are standard normal variates. It should be made clear that $\widehat{u}_t$ are constructed to deviate from u_t. No estimation is involved to obtain $\widehat{u}_t$.

We consider $r = 4$ series with $T = 25, 50, 100, 200$ time series observations. In this exercise (as in estimation of VAR when r is the number of variables) , r is assumed known. Table 1a reports results for $\widehat{q}_1$ and $\widehat{q}_2$, averaged over 1000 replications. We consider $m = 1.0, 0.5$, and 0.1. The corresponding values for $m^* = \frac{m}{T^{2/5}}$ are also reported in Table 1.

The larger we tolerate the difference between $\widehat{d}_k$ and $\widehat{d}_0$ as specified by m, the smaller will be $\widehat{q}$. We had expected the results to be sensitive to the choice of m. Table 1a shows surprisingly that this is not the case, except when $T \leq 50$. When $m = 1$ and T is small, the tolerance is too high, causing $\widehat{q}_1$ and $\widehat{q}_2$ to underestimate q. However, when $T \geq 100$, all choices of m give precise estimates of q.

Next, we let $\widehat{u}_t$ be residuals from estimation of vector-autoregressions. We generate data as

$$X_t = AX_{t-1} + R\varepsilon_t$$

where X_t is r dimensional, A is a diagonal matrix with elements $A_i \sim U(0,1)$, and R is a $r \times q$ matrix with elements drawn from the uniform distribution. Results are presented in Table 1b for $r = 4, 6, 8$, with $q = 4$ in every case. Again, we find that the tolerance of $m = 1$ is too high when $T \leq 50$; in fact, even $m = .5$ is too high when $T = 25$. Additional results reveal that for small T, the choice of m becomes a rather delicate matter. But in empirical analysis when we typically have more than 100 observations at our disposal, $m = 0.5$ is an effective tolerance limit.

6.2 Factor Models

We consider four data generating processes:

1. $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)'f_t + e_{it}$, where $f_t = C(L)\varepsilon_t$ and f_t is q dimensional;
2. $x_{it} = (\lambda_{i0} + \lambda_{i1}L)'f_t + e_{it}$, where $A(L)f_t = \varepsilon_t$ and f_t is q dimensional;
3. $x_{it} = \lambda_i'F_t + e_{it}$ where $F_t = A_1F_{t-1} + u_t$, and $A_1 = \rho I_r$; $u_t = R\varepsilon_t$, $rank(R) = q$.
4. $x_{it} = \lambda_i'F_t + e_{it}$ where $F_t = A_1F_{t-1} + u_t$, where $A_1 = diag(.2, .375, .55, .725, .9)$ and u_t is the same as in 3.

DGP 1 and 2 are dynamic factor models considered by Forni et al. (2000). DGP 1 assumes the dynamic factors are moving average processes. We generate f_t as MA(2). DGP 2 assumes that the dynamic factors are autoregressive processes with f_t being AR(1). In both cases, there are $r = q(s+1)$ static factors. DGPs 3 and 4 are static factor models. The static factors are driven by q 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 $r^* = q = 3$. For DGP 4, A_1 is a diagonal matrix with values .2,.375,.55,.725, .9 and the dynamics of F_t is richer than that of DGP3. In this case, $r = 5$ and $q = 3$.

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

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

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

The results are reported in Tables 2a for DGP 1 with $q = 2, s = 2$, Table 2b for DGP 2 with $q = 2$, $s = 1$. The true values of r are 6 and 4, respectively. Table 2c are results for DGP 3 with $r = 5$ and $r^* = q = 3$. Table 2d has results for DGP 4 with $r = 5, q = 3$. Then $\widehat{r}$ is determinined by minimizing the IC(k) for k between 0 and $2r$. For small values of $\min[N, T]$, the IC tends to select a large number of static factors. However, even when $\widehat{r}^*$ is overestimated, $\widehat{q}$ can be very close to q for suitable choice of m. For all four DGPs we find that let $m = 0.5$ gives good estimates of q.

An overview of Tables 1 and 2 suggest that setting $m = 0.5$ in both VAR and factor analysis yields rather robust results when $\min[N, T] \geq 50$, giving m^* in the neighborhood of .1.

6.3 Empirical Analysis: Shocks in the U.S.

To illustrate, we take data used in Stock and Watson (2002b).[3] There are 215 monthly time series, of which 155 are available from 1960:1 to 2003:12. The data are transformed (by

[3]The data are taken from Mark Watson's web site http://www.princeton.edu/ mwatson.

taking logs, first or second difference) as in Stock and Watson. The objective is to determine the number of primitive, or dynamic factors, in this panel of data.

To get a sense of the importance of the factors in the data, we begin by determining $\widehat{q}$ for $r = 2, 3, \ldots 10$. In this exercise, we do not take a stand on what is the optimal number of static factors in the data. We find that for the full sample of 528 observations, $\widehat{q} = r$ when $r = 2, 3, 4$, $\widehat{q} = r - 1$ for $r \in [5, 9]$, and $\widehat{q} = 8$ when $r = 10$. Some of the static factors are linearly dependent in a dynamic sense. It is well known that the first two static factors in the data being analyzed are real factors. The finding that $\widehat{q} = 2$ given $r = 2$ indicates that the first two static factors are dynamically distinct.

Next, we allow the number of static factors, r_t, to be determined optimally for each t. We use two concepts of optimality. We first estimate $\widehat{r}_{jt}$ static factors, where $\widehat{r}_{jt}$ explains j percent of the variation in the data up to time t, and then determine $\widehat{q}_{jt}$ given $\widehat{r}_{jt}$ factors. Note that $\widehat{r}_{jt}$ is not optimal from a statistical point of view. However, the result of Giannone et al. (2004) 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 $j = .6$. We also determine the number of static factors using the IC. This is denoted $\widehat{r}_t^*$, and the corresponding number of primitive factors is q_t^*. All these statistics are computed for t ranging from 133 to 528, corresponding to estimation ending in 1970:12 and 2003:12, respectively. We thus have 396 statistics, one for every t.

Reported in Table 3 are the mean of these statistics over the sample. $R^2_{\widehat{r}}$ is the average explanatory power of $\widehat{r}$ factors, when $\widehat{r}$ is chosen with cut-off of .3,.4,.5 or .6, and also optimally. The column $R^2_{\widehat{q}}$ is the average explanatory power of the $\widehat{q}$ shocks given $\widehat{r}$ innovations. Because $\widehat{r}_t$ and $\widehat{q}_t$ changes as t increases, we also report the minimum and maximum of these statistics over time, along with the frequency that $\widehat{q}_t$ is less than or equal to the mean, minimum, or maximum of $\widehat{q}$ in the simulations.

The results indicate that three static factors explain .3 of the variation in the data, while five factors explain .4. To explain .6 of the variation in the data would require, on average, 13 factors. When determined optimally by the IC, the data suggest that 8 static factors explain, on average, .5 of the variation in the data, and that there seven dynamic factors spanning the eight static factors.[4] This is the same number of dynamic factors as determined by the alternative method developed in Stock and Watson (2005). The evidence is thus very

[4]In an earlier version of this paper, the tests were applied to a dataset for the sample 1960:1-1998:12. We found an average of seven dynamic factors in ten static factors. We now use exactly the same data as Stock and Watson. There are differences in data construction between the old and new data (at source), but the number of dynamic factors is apparently robust in spite of the data revisions.

compelling that the number of dynamic factors larger than two.

We stress once again that there is substantial variation over the sample. Figure 1 depicts the time series plot of $\widehat{r}_t^*, \widehat{q}_{1t}$ and $\widehat{q}_{2t}$. As we can see, r jumped from 4 to 6 around 1975, and has remained roughly at 6 till 2000 except for a brief increase in the early 1990s. If we had ended the estimation in 2000, we would have $\widehat{r} = 6$ and $\widehat{q} = 5$. However, in the past few years, r seemed to have taken another jump from 6 to 8. Correspondingly, $\widehat{q}$ increased from 5 to 7.

Figure 2 plots R_t^2, the fraction of variance in the data explained by the factors up to time t, along with , $.1\bar{x}_t/\sigma_{xt}$ the coefficient of variation at time t (divided by 10). 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. Simultaneously, the coefficient of variation of the data falls. 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 larger than two.

It remains to reconcile our finding that q exceeds two with the result of Giannone et al. (2004). Our analysis gives the optimal number of factors for the panel of data from which the factors are extracted. Given that $N = 155$ series, our finding suggests seven dynamic factors is optimal in explaining the average variation in the data. In contrast, Giannone et al. (2004) first estimated the factors from close to 200 series. They then restrict their attention to only twelve series when arriving at the conclusion that q is two. Their conclusion should not be taken to mean that two dynamic factors best explain the variation in the panel of data from which the factors are extracted.

To highlight the difference, we calculate the explanatory power of the common factors in the full sample for a selected number of series: IPS10 (industrial production), A0M059 (retail trade), A0M057 (manufacturing trade), FYFF (Federal funds drate), PUNEW (CPI), and A0M224r (consumption expenditure). Reported are the R^2 from a regression of x_{it} on a constant and $\widehat{r}$ static factors, where x_{it} is log first difference of IPS10, A0M059, A0M057, and A0M224r, the first difference FYFF, and second difference of the logarithm of PUNEW.[5] Exact replication of the results of Giannone et al. (2004) is difficult, as these authors have averaged the monthly data to quarterly levels.[6] However, it is evident from Table 3 that

[5]The size of our panels differ in both the T and the N dimensions.

[6]This makes an important difference for the PUNEW and GMCQ. The R^2 for the averaged data are much higher. Our R^2 corresponds to the x_{it} as used in the factor analysis

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, six or fewer static factors would have been necessary, which would have implied three dynamic factors. The full sample analysis also obscures the fact that the optimal number of static and dynamic factors have changed over time. As indicated in Figure 2, the average number of dynamic factors over the sample when r is optimally determined is around seven, not two.

6.4 Empirical Analysis: Shocks in G7

We consider quarterly GDP growth for the G7 countries taken from Stock and Watson (2003). Per capita data from 1960:1-2002:4 are available for the U.S., Canada, France, Italy, Germany, Japan, and the U.K. We then regress output growth for each country on four of its own lag, plus one lag of output growth of each of the remaining six countries to filter out the spillover effects, as in Stock and Watson.

Stock and Watson assumed that the filtered residuals $\widehat{u}_t$ have a two factor structure. This means that in a seven country model, they entertain a total of nine shocks. On the other hand, if all shocks are global in nature, we should find that $\widehat{\Sigma}_u$ is reduced rank. Applying the spectral decomposition analysis, we find that $\widehat{\Sigma}_u$ is full rank since $\widehat{q}_3$ and $\widehat{q}_4$ are both seven at $m = 0.5$, or equivalently, an m^* of .07. This suggests that there are at least seven shocks in the system, and is consistent with an error structure of the form $u_t = R\varepsilon_t + v_t$. While our results do not contradict Stock and Watson's assumption of nine shocks, we also cannot rule out the possibility that seven or eight shocks better characterize the data. How to identify the minimum number of idiosyncratic shocks is work for future research.

7 Conclusion

This paper proposes a procedure to determine the number of dynamically independent shocks, given time series data on a set of innovations. 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 algebraic and is easy to compute. The tests can also be used in VAR analysis.

Table 1a: $\widehat{u}_t = u_t + O_p(1/\sqrt{T})$, $r = 4$

T	q	$\widehat{q}_1$	$\widehat{q}_2$	$\frac{m}{T^{2/5}}$	$\widehat{q}_1$	$\widehat{q}_2$	$\frac{m}{T^{2/5}}$	$\widehat{q}_1$	$\widehat{q}_2$	$\frac{m}{T^{2/5}}$
		$m = 1$			$m = .5$			$m = .1$		
25	1	1.000	1.000	0.276	1.000	1.000	0.138	1.000	1.000	0.028
25	2	1.000	1.000	0.276	2.000	2.000	0.138	2.000	2.000	0.028
25	3	2.000	2.000	0.276	3.000	3.000	0.138	3.000	3.000	0.028
50	1	1.000	1.000	0.209	1.000	1.000	0.105	1.000	1.000	0.021
50	2	2.000	2.000	0.209	2.000	2.000	0.105	2.000	2.000	0.021
50	3	2.000	2.000	0.209	3.000	3.000	0.105	3.000	3.000	0.021
100	1	1.000	1.000	0.158	1.000	1.000	0.079	1.000	1.000	0.016
100	2	1.000	1.000	0.158	2.000	2.000	0.079	2.000	2.000	0.016
100	3	3.000	3.000	0.158	3.000	3.000	0.079	3.000	3.000	0.016
200	1	1.000	1.000	0.120	1.000	1.000	0.060	1.000	1.000	0.012
200	2	2.000	2.000	0.120	2.000	2.000	0.060	2.000	2.000	0.012
200	3	3.000	3.000	0.120	3.000	3.000	0.060	3.000	3.000	0.012

Table 1b: $X_t = AX_{t-1} + u_t$, $u_t = R\varepsilon_t$, $q = 4$

T	r	$\widehat{q}_1$	$\widehat{q}_2$	$\frac{m}{T^{2/5}}$	$\widehat{q}_1$	$\widehat{q}_2$	$\frac{m}{T^{2/5}}$	$\widehat{q}_1$	$\widehat{q}_2$	$\frac{m}{T^{2/5}}$
		$m = 1$			$m = .5$			$m = .1$		
25	4	2.190	2.420	0.276	3.118	3.163	0.138	3.979	3.979	0.028
50	4	3.052	3.152	0.209	3.776	3.780	0.105	4.000	4.000	0.021
100	4	3.522	3.552	0.158	3.950	3.950	0.079	4.000	4.000	0.016
200	4	3.808	3.812	0.120	3.998	3.998	0.060	4.000	4.000	0.012
25	6	2.074	2.244	0.276	2.895	2.979	0.138	3.932	3.932	0.028
50	6	3.050	3.140	0.209	3.775	3.778	0.105	4.000	4.000	0.021
100	6	3.567	3.589	0.158	3.962	3.962	0.079	4.000	4.000	0.016
200	6	3.793	3.799	0.120	3.998	3.998	0.060	4.000	4.000	0.012
25	8	1.871	1.980	0.276	2.510	2.572	0.138	3.550	3.552	0.028
50	8	3.035	3.124	0.209	3.766	3.774	0.105	4.000	4.000	0.021
100	8	3.576	3.598	0.158	3.965	3.965	0.079	4.000	4.000	0.016
200	8	3.860	3.863	0.120	3.997	3.997	0.060	4.000	4.000	0.012

Table 2: Factor Models

DGP 1: $x_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)'f_t + e_{it}$, $f_t = \varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2}$, $q = 2, s = 2, r = 6$

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*
			$m = 1$			$m = .5$			$m = .1$		
20	100	10.755	1.953	2.157	0.302	2.267	4.163	0.151	8.804	9.630	0.030
20	200	10.762	1.968	2.132	0.302	2.187	4.294	0.151	9.319	9.955	0.030
50	100	6.150	2.000	2.000	0.209	2.066	2.159	0.105	4.975	5.324	0.021
50	200	6.235	2.000	2.001	0.209	2.049	2.183	0.105	5.234	5.538	0.021
100	100	6.006	2.000	2.000	0.158	2.006	2.006	0.079	3.566	4.535	0.016
100	200	6.026	2.000	2.000	0.158	2.002	2.009	0.079	3.600	4.697	0.016

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

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*
			$m = 1$			$m = .5$			$m = .1$		
20	100	5.869	1.985	2.157	0.302	2.450	3.321	0.151	5.737	5.741	0.030
20	200	6.208	1.970	2.138	0.302	2.391	3.441	0.151	6.108	6.115	0.030
50	100	4.279	2.001	2.035	0.209	2.163	2.232	0.105	4.139	4.140	0.021
50	200	4.513	2.003	2.050	0.209	2.178	2.380	0.105	4.434	4.436	0.021
100	100	4.041	2.000	2.007	0.158	2.034	2.040	0.079	3.364	3.463	0.016
100	200	4.164	2.000	2.009	0.158	2.067	2.128	0.079	3.583	3.662	0.016

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

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*
			$m = 1$			$m = .5$			$m = .1$		
20	100	2.822	2.305	2.329	0.302	2.668	2.719	0.151	2.827	2.827	0.030
20	200	2.955	2.342	2.375	0.302	2.713	2.807	0.151	2.958	2.958	0.030
50	100	2.761	2.671	2.673	0.209	2.765	2.765	0.105	2.765	2.765	0.021
50	200	2.826	2.700	2.703	0.209	2.826	2.826	0.105	2.828	2.828	0.021
100	100	2.814	2.764	2.764	0.158	2.815	2.815	0.079	2.815	2.815	0.016
100	200	2.875	2.785	2.786	0.158	2.876	2.876	0.079	2.876	2.876	0.016

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

N	T	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*	$\widehat{q}_3$	$\widehat{q}_4$	m^*
			$m = 1$			$m = .5$			$m = .1$		
20	100	4.149	2.143	2.384	0.302	2.892	3.215	0.151	4.122	4.123	0.030
20	200	4.652	2.197	2.492	0.302	2.952	3.521	0.151	4.638	4.639	0.030
50	100	3.389	2.579	2.593	0.209	2.808	2.810	0.105	3.382	3.382	0.021
50	200	3.595	2.607	2.636	0.209	2.845	2.851	0.105	3.589	3.589	0.021
100	100	3.500	2.687	2.691	0.158	2.847	2.848	0.079	3.398	3.400	0.016
100	200	3.760	2.722	2.734	0.158	2.885	2.887	0.079	3.637	3.638	0.016

Note: $m^* = m/\min[N^{2/5}, T^{2/5}]$ for $m = 1, .5$ and $.1$.

Table 3: Empirical Analysis

	τ	T	$R^2_{\widehat{r}}$	$R^2_{\widehat{q}}$	$\widehat{r}$	$\widehat{q}_3$	$\widehat{q}_4$	P_1	P_2
mean	0.3	294.500	0.321	0.958	3.133	2.444	2.457	0.556	0.543
min	0.3	133.000	0.300	0.904	3.000	2.000	2.000	0.556	0.543
max	0.3	456.000	0.352	1.000	4.000	3.000	3.000	1.000	1.000
mean	0.4	294.500	0.418	0.970	5.352	4.170	4.281	0.765	0.719
min	0.4	133.000	0.400	0.912	5.000	3.000	4.000	0.065	0.719
max	0.4	456.000	0.437	0.993	6.000	5.000	5.000	1.000	1.000
mean	0.5	294.500	0.515	0.948	8.435	6.571	6.769	0.444	0.386
min	0.5	133.000	0.500	0.917	8.000	5.000	6.000	0.111	0.386
max	0.5	456.000	0.529	0.982	9.000	8.000	8.000	1.000	1.000
mean	0.6	294.500	0.609	0.958	12.679	9.605	9.738	0.398	0.336
min	0.6	133.000	0.600	0.918	12.000	9.000	9.000	0.398	0.336
max	0.6	456.000	0.621	0.974	14.000	11.000	11.000	1.000	1.000
mean	r^*	294.500	0.538	0.947	9.898	7.182	7.562	0.556	0.401
min	r^*	133.000	0.250	0.917	2.000	1.000	1.000	0.006	0.006
max	r^*	456.000	0.640	0.984	15.000	11.000	11.000	1.000	1.000

Note: $m^* = m/\min[N^{2/5}, T^{2/5}]$. $R^2_{\widehat{r}}$ is average variation in x_{it} explained by $\widehat{r}$ factors, when $\widehat{r}$ explains at least τ percent of the variation in the data up to time t. $R^2_{\widehat{q}}$ is the percent variation in $\widehat{F}_t$ explained by $\widehat{q}$ primitive shocks. P_1 is the frequency of $\widehat{q}_3$, where $\widehat{q}_3$ is either the mean, minimum, or maximum of $\widehat{q}$ in the recursive estimations.

Table 4: Explanatory Power of $\widehat{r}$ factors

	r			
Series	3	6	8	$\widehat{r}^*$
ALL	0.298	0.421	0.486	0.643
IP	0.722	0.814	0.887	0.932
PMEMP	0.764	0.781	0.785	0.864
MSMTQ	0.654	0.760	0.782	0.945
FYFF	0.369	0.429	0.461	0.540
PUNEW	0.117	0.676	0.689	0.732
GMCQ	0.231	0.331	0.633	0.881

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

References

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

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

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

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

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

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

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

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

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

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

Giannone, D., Reichlin, L. and Sala, L. 2004, Monetary Policy in Real Time, *Macroeconomic Annual*.

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.

King, R. G., Plosser, C., Stock, J. and Watson, M. 1991, Stochastic Trends and Economic Fluctuations, *American Economic Review* **81:4**, 819–40.

Kleibergen, F. and Paap, R. 2003, Generalized Reduced Rank Tests using the Singular Value Decomposition, University of Amsterdam.

Ratsimalahelo, Z. 2003, Strongly Consistent Determination of the Rank of a Matrix, University of Franche-Comté.

Robin, J. M. and Smith, R. J. 2000, Tests of Rank, **16**, 151–175.

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

Stock, J. H. and Watson, M. W. 2002a, Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association* **97**, 1167–1179.

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

Stock, J. H. and Watson, M. W. 2003, Understanding Changes in International Business Cycle Dyanmics, NBER WP 9859.

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

Figure 1: Estimated r, q_1, and q_2

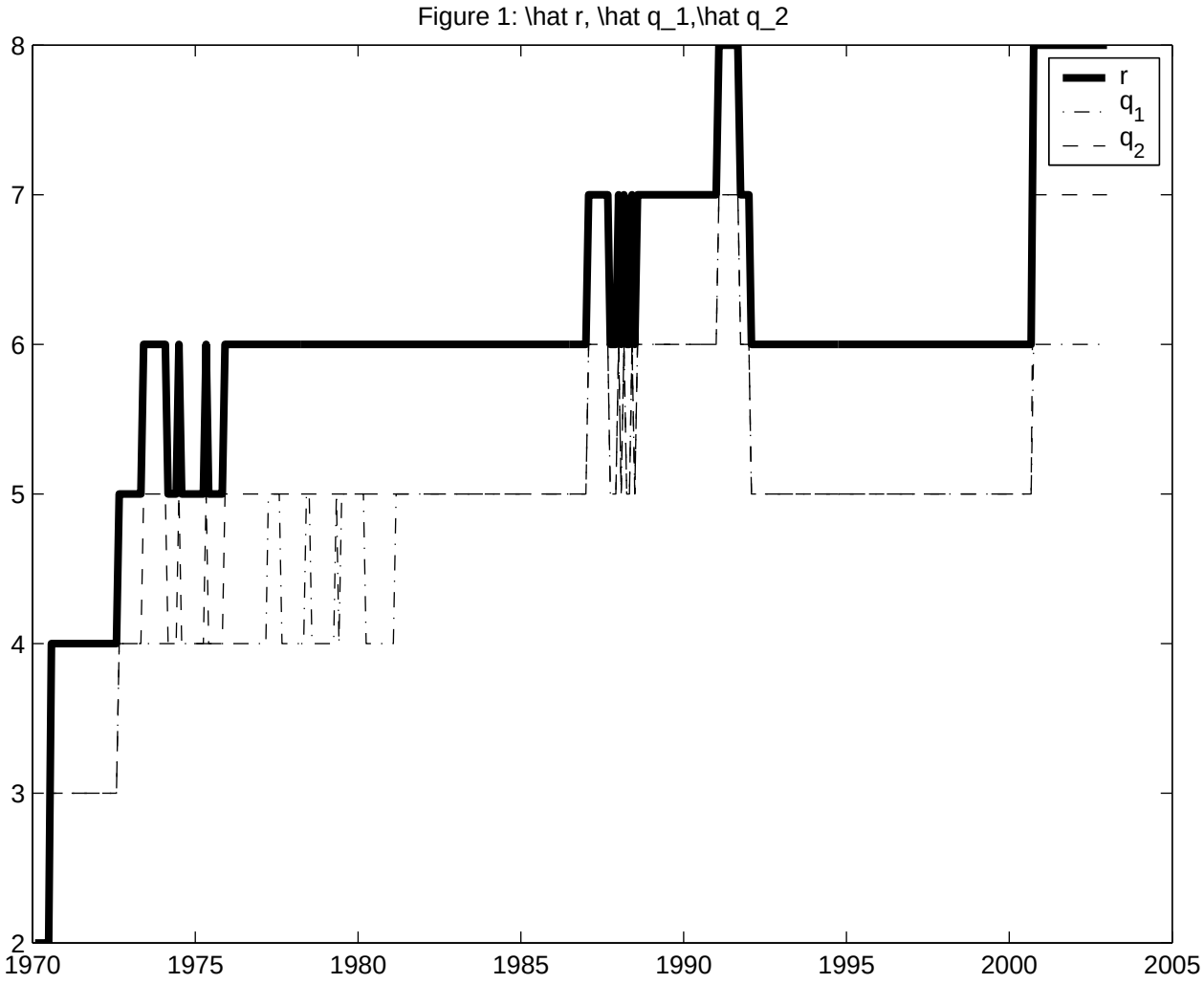

Figure 2: Importance of Common Component