Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Determining the Number of Factors in the General Dynamic Factor Model

Marc Hallin & Roman Liška

To cite this article: Marc Hallin & Roman Liška (2007) Determining the Number of Factors in the General Dynamic Factor Model, Journal of the American Statistical Association, 102:478, 603-617, DOI: 10.1198/016214506000001275

To link to this article: https://doi.org/10.1198/016214506000001275

Published online: 01 Jan 2012.

Determining the Number of Factors in the General Dynamic Factor Model

Marc HALLIN and Roman LIŠKA

This article develops an information criterion for determining the number q of common shocks in the *general* dynamic factor model developed by Forni et al., as opposed to the *restricted* dynamic model considered by Bai and Ng and by Amengual and Watson. Our criterion is based on the fact that this number q is also the number of diverging eigenvalues of the spectral density matrix of the observations as the number n of series goes to infinity. We provide sufficient conditions for consistency of the criterion for large n and T (where T is the series length). We show how the method can be implemented and provide simulations and empirics illustrating its very good finite-sample performance. Application to real data adds a new empirical facet to an ongoing debate on the number of factors driving the U.S. economy.

KEY WORDS: Dynamic factor model; Dynamic principal components; Information criterion.

1. INTRODUCTION

Factor models recently have been considered quite successfully in the analysis of large panels of time series data. Under such models, the observation X_{it} (where $i = 1, \dots, n$ represents the cross-sectional index and $t = 1, \dots, T$ denotes time) is decomposed into the sum, $\chi_{it} + \xi_{it}$, of two nonobservable mutually orthogonal (at all leads and lags) variables, the *common component*, χ_{it}, and the *idiosyncratic component*, ξ_{it}.

Under the *general dynamic factor* approach considered in this article, the common component χ_{it} takes the form $\chi_{it} = \sum_{j=1}^{q} b_{ij}(L)u_{jt}$, where the unobservable common shocks u_{jt}—the *dynamic factors*—are loaded through one-sided linear filters $b_{ij}(L)$, $j = 1, \dots, q$ (L, as usual, stands for the lag operator). This approach was first suggested by Sargent and Sims (1977) and Geweke (1977) in a model in which idiosyncratic components were assumed to be mutually orthogonal (*exact* factor model), and developed for large panels with weakly cross-correlated idiosyncratic components (*approximate* factor model) in a series of articles by Forni and Lippi (2001) and Forni, Hallin, Lippi, and Reichlin (2000, 2004). The main theoretical tool in this context is Brillinger's theory of *dynamic principal components* (Brillinger 1981).

A less general model is the *static* (approximate) factor model proposed by Stock and Watson (2002a,b). Under this approach, the common component χ_{it} is expressed as a linear combination $\sum_{j=1}^{r} a_{ij}F_{jt}$ of a small number r of unobserved *static factors* $(F_{1t}, \dots, F_{rt})$. The loadings a_{ij} are real numbers, and all factors are loaded contemporaneously. Classical principal components are the main tool here.

An intermediate model allowing for some dynamics is the *restricted dynamic model*—actually a static model where the r static factors $(F_{1t}, \dots, F_{rt})$ are driven by a number, $q \le r$, of *dynamic factors*. This model reduces to the (strictly) static model through stacking (see, e.g., Forni et al. 2005b; Forni, Giannone, Lippi, and Reichlin 2005a; Bai 2004; Bai and Ng 2005).

The relative merits, in terms of predictive power, of the various approaches (static, restricted dynamic, and general dynamic) have been studied extensively (see, e.g., Boivin and Ng 2005; d'Agostino and Giannone 2006). However, the general dynamic model is the only one supported by a characterization theorem (Forni and Lippi 2001) that matches the empirical evidence (i.e., the divergence, as $n \to \infty$, of a small number of empirical spectral eigenvalues).

Whether static or dynamic, factor models with large cross-sectional dimensions are attracting increasing attention in finance and macroeconometric applications. In these fields, information usually is scattered through a (very) large number n of interrelated time series (n values of the order of several hundreds, or even 1,000, are not uncommon). Classical multivariate time series techniques are totally helpless, and to the best of our knowledge, factor model methods are the only ones that can handle such datasets. In macroeconomics, factor models are used in business cycle analysis (Forni and Reichlin 1998; Giannone, Reichlin, and Sala 2005, 2006), in the identification of economy-wide and global shocks (Forni et al., 2005a; Giannone and Lenza 2004; Giannone and Reichlin 2006), in the construction of indexes and forecasts exploiting the information scattered in a huge number of interrelated series (Altissimo et al. 2001), in the monitoring of economic policy (Giannone et al. 2005), in consumer theory (Lewbel 1991), and in monetary policy applications (Bernanke and Boivin 2003; Favero, Marcellino, and Neglia 2005). In finance, factor models have been applied to the forecast of multivariate volatility (Alessi, Barigozzi, and Capasso 2006). They are at the heart of the extensions proposed by Chamberlain and Rothschild (1983) and Ingersol (1984) of classical arbitrage pricing theory; they also have been considered in performance evaluation and risk measurement (Campbell, Lo, and MacKinlay 1997, chaps. 5 and 6).

In the past few years, factor models generated a huge amount of applied work, including that of Artis, Banerjee, and Marcellino (2002), Bruneau, De Bandt, Flageollet, and Michaux (2003), den Reijer (2005), Dreger and Schumacher (2004), Nieuwenhuyzen (2004), Schneider and Spitzer (2004), and Stock and Watson (2002b) for applications to data from United Kingdom, France, the Netherlands, Germany, Belgium, and the United States; Altissimo et al. (2001), Angelini, Henry, and Mestre (2001), Forni et al. (2003), and Marcellino, Stock, and Watson (2003) for the Euro area; and Aiolfi, Catão, and Timmerman (2005) for South American data, to quote only a few.

Marc Hallin is Professor, Department of Mathematics (E-mail: *mhallin@ulb.ac.be*), and Roman Liška is Doctoral Student (E-mail: *rliska@ulb.ac.be*), ISRO and ECARES, Free University of Brussels, B-1050 Brussels, Belgium. Hallin's research was supported by a PAI contract from the Belgian Federal Government and an Action de Recherche Concertée from the Communauté Française de Belgique. Liška's research was supported by an Action de Recherche Concertée of the Communauté Française de Belgique. The authors thank Mario Forni, Marco Lippi, and Lucrezia Reichlin for their insightful comments on the manuscript, and Domenico Giannone for his decisive contribution to the proof of Lemma A.1.

Journal of the American Statistical Association
June 2007, Vol. 102, No. 478, Theory and Methods
DOI 10.1198/016214506000001275

Dynamic factor models also have entered the practice of a number of economic and financial institutions, including several central banks and national statistical offices, who are using them in their current analysis and prediction of economic activity. A real-time coincident indicator of the Euro area business cycle (EuroCOIN), based on work of Forni et al. (2000), is published monthly by the London-based Center for Economic Policy Research and the Banca d'Italia (see *http://www.cepr.org/data/EuroCOIN/*). A similar index based on these methods was established for the U.S. economy by the Federal Reserve Bank of Chicago.

A critical step in the statistical analysis of all of these factor models is the preliminary identification of the number, q, of common dynamic and/or the number, r, of static factors. This number indeed is needed in the implementation of the various estimation and forecasting algorithms. Moreover, it has a crucial economic interpretation (on this latter point, see Giannone et al. 2006; Forni, Giannone, Lippi, and Reichlin 2005a; Stock and Watson 2005).

A method for identifying r in the static model has been proposed by Bai and Ng (2002). The criterion that they considered is based on an information-theoretical quantity; they established, under appropriate assumptions, the consistency of their method as n, the cross-sectional dimension, and T, the length of the observed series, both tend to infinity. This criterion has been adapted (Bai and Ng 2005; Amengual and Watson 2006) to the restricted dynamic framework. But in the general dynamic case, this approach is likely to overestimate the actual q (as confirmed by our simulation study; see Sec. 5). More recently, alternative criteria based on the theory of random matrices have been developed by Kapetanios (2005) and Onatski (2006), still for the number r of static factors but in a model with iid idiosyncratics (instantaneous idiosyncratic cross-sectional corelation is allowed in Onatski 2006 where, however, Gaussian assumptions are made).

In the general dynamic model, no formal statistical criterion exists; Forni et al. (2000) suggested only a very heuristic eye-inspection rule. The purpose of this article is to propose such a formal criterion and to establish its consistency as n and T approach infinity. From a technical standpoint, due to the spectral techniques involved, the tools that we use in the proofs are entirely different from those used in the restricted dynamic or static frameworks; our criterion builds directly on the (n, T)-asymptotic properties of the eigenvalues of sample spectral density matrices. Simulations indicate that the method performs quite well even in relatively small panels with moderate series lengths and nonnegligible idiosyncratic cross-correlation. When the Bai and Ng (2005) criterion applies—that is, under the assumptions of the restricted dynamic model—the two methods perform equally well. But when those assumptions are not met, simulations indicate that the Bai and Ng criterion significantly overestimates q, whereas ours still does extremely well.

The article is organized as follows. Section 2 briefly describes the general dynamic factor model proposed by Forni et al. (2000). Section 3 introduces the criterion that we are proposing for identifying q and establishes sufficient consistency conditions (as n and T tend to infinity). We recommend a version of our method based on lag window spectral estimation and a cross-validation idea. Section 4 discusses practical implementation in detail. Section 5 presents a simulation study of the small-sample properties of the proposed identification procedure and an application to macroeconomic data, and Section 6 concludes. The Appendix provides proofs.

Throughout, boldface is used for vectors and matrices, primes for transposes, and stars for complex conjugates. A sequence $\{\zeta(n,T,\theta);\ n \in \mathbb{N}, T \in \mathbb{N}, \theta \in [-\pi,\pi]\}$ of real numbers is said to be $o(1)$ [resp. $O(1)$] as $T \to \infty$ uniformly in n and θ if $\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \zeta(n,T,\theta)$ is $o(1)$ [resp. $O(1)$] as $T \to \infty$. A sequence $\{\zeta(n,T,\theta); n \in \mathbb{N}, T \in \mathbb{N}, \theta \in [-\pi,\pi]\}$ of random variables is said to be $o_P(1)$ [resp. $O_P(1)$] as $T \to \infty$ uniformly in n and θ if for all $\epsilon > 0$ and $\eta > 0$, there exists a $T_{\epsilon,\eta}$ such that $\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mathrm{P}[|\zeta(n,T,\theta)| > \eta] < \epsilon$ for all $T > T_{\epsilon,\eta}$ [resp. for all $\epsilon > 0$ there exist $B_\epsilon > 0$ and T_ϵ such that $\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mathrm{P}[|\zeta(n,T,\theta)| > B_\epsilon] < \epsilon$ for all $T > T_\epsilon$]. Note that this concept of uniformity, which is sufficient for our purposes, is weaker than the classical one (where sup's lie within probabilities).

2. THE DYNAMIC FACTOR MODEL

The model that we consider throughout is Forni et al.'s (2000) general dynamic factor model, under which the observation is a finite realization of a double array $\{X_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ of random variables, where

$$X_{it} = \chi_{it} + \xi_{it},$$

$$\text{with } \chi_{it} := \sum_{j=1}^{q} b_{ij}(L)u_{jt} \text{ and } b_{ij}(L) := \sum_{k=1}^{\infty} b_{ijk}L^k, \quad (1)$$

and Assumptions A1–A4 are assumed to hold.

Assumption A1. (a) The q-dimensional vector process $\{\mathbf{u}_t := (u_{1t}\ \ u_{2t}\ \ \cdots\ \ u_{qt})'; t \in \mathbb{Z}\}$ is orthonormal white noise.

(b) The n-dimensional processes $\{\boldsymbol{\xi}_n := (\xi_{1t}\ \ \xi_{2t}\ \ \cdots\ \ \xi_{nt})'; t \in \mathbb{Z}\}$ are mean-0 stationary for any n; moreover, $\xi_{i,t_1} \perp u_{j,t_2}$ for any i, j, t_1, and t_2.

(c) The filters $b_{ij}(L)$ are square summable: $\sum_{k=1}^{\infty} b_{ijk}^2 < \infty$ for all $i \in \mathbb{N}$ and $j = 1, \ldots, q$.

The processes $\{u_{jt}, t \in \mathbb{Z}\}$, $j = 1, \ldots, q$, are called the *common shocks* or *factors*, and the random variables ξ_{it} and χ_{it} are the *idiosyncratic* and *common components* of X_{it}.

Assumption A2. For all n, the vector process $\{\mathbf{X}_{nt} := (X_{1t}\ \ X_{2t}\ \ \cdots\ \ X_{nt})'; t \in \mathbb{Z}\}$ is a linear process with a representation of the form $\mathbf{X}_{nt} = \sum_{k=-\infty}^{\infty} \mathbf{C}_k \mathbf{Z}_{t-k}$, where $\mathbf{Z}_t$ is full-rank n-dimensional white noise with finite fourth-order cumulants, and the $n \times n$ matrices $\mathbf{C}_k = (C_{ij,k})$ are such that $\sum_{k=-\infty}^{\infty} |C_{ij,k}||k|^{1/2} < \infty$ for all $1 \le i, j \le n$.

Under this form, Assumption A2 is sufficient for a consistent estimation of the model (see Forni et al. 2000), provided that the number q of factors is known. As we show later, consistent identification of q is more demanding. Denoting by

$$c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1}) := \mathrm{cum}\big(X_{i_1}(t+k_1), \ldots, X_{i_{\ell-1}}(t+k_{\ell-1}), X_{i_\ell}(t)\big)$$

the cumulant of order ℓ of $X_{i_1}(t+k_1), \ldots, X_{i_{\ell-1}}(t+k_{\ell-1})$, $X_{i_\ell}(t)$, we also require some uniform decrease, as the lags tend to infinity, of $c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1})$ up to order $\ell = 4$.

Assumption A2′. This is the same as Assumption A2, but the convergence condition on the $C_{ij,k}$'s is uniform: $\sup_{i,j\in\mathbb{N}} \sum_{k=-\infty}^{\infty} |C_{ij,k}||k|^{1/2} < \infty$. Moreover, for all $1 \le \ell \le 4$ and $1 \le j < \ell$, $\sup_{i_1,\dots,i_\ell}[\sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{\ell-1}=-\infty}^{\infty} |c_{i_1,\dots,i_\ell}(k_1, \dots, k_{\ell-1})|] < \infty$.

Denote by $\mathbf{\Sigma}_n(\theta)$, $\theta \in [-\pi,\pi]$, the spectral density matrix of $\mathbf{X}_{nt}$, with elements $\sigma_{ij}(\theta)$, and by $\lambda_{n1}(\theta), \dots, \lambda_{nn}(\theta)$ its eigenvalues in decreasing order of magnitude. Similarly, let $\lambda_{nj}^{\chi}(\theta)$ and $\lambda_{nj}^{\xi}(\theta)$ be the eigenvalues associated with the spectral densities $\mathbf{\Sigma}_n^{\chi}(\theta)$ and $\mathbf{\Sigma}_n^{\xi}(\theta)$ of $\boldsymbol{\chi}_{nt} := (\chi_{1t}, \dots, \chi_{nt})'$ and $\boldsymbol{\xi}_{nt} := (\xi_{1t}, \dots, \xi_{nt})'$. Such eigenvalues [actually, the functions $\theta \mapsto \lambda(\theta)$] are called *dynamic* eigenvalues.

Assumption A3. (a) The first idiosyncratic dynamic eigenvalue $\lambda_{n1}^{\xi}(\theta)$ is uniformly (with respect to $\theta \in [-\pi,\pi]$) bounded as $n \to \infty$; that is, $\sup_{\theta\in[-\pi,\pi]} \lambda_{n1}^{\xi}(\theta) < \infty$ as $n \to \infty$.

(b) The qth common dynamic eigenvalue $\lambda_{nq}^{\chi}(\theta)$ diverges for all θ in $[-\pi,\pi]$ as $n \to \infty$.

Assumption A3 plays a key role in identifying the common and idiosyncratic components in (1), but it involves the unobservable quantities $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$. The following proposition provides a X_{it}-based counterpart.

Proposition 1 (Forni and Lippi 2001). Let Assumption A2 (or A2′) hold. Then Assumptions A1 and A3 are satisfied iff the qth eigenvalue of $\mathbf{\Sigma}_n(\theta)$ diverges as $n \to \infty$, for all θ in $[-\pi,\pi]$, whereas the $(q+1)$th one is bounded uniformly.

Forni et al. (2000) showed how, under Assumptions A1–A3, the common components χ_{it} and the idiosyncratic components ξ_{it} can be consistently reconstructed, as both n and T tend to infinity, by projecting X_{it} onto the space spanned by the q first *dynamic principal components* of the spectral density matrix $\mathbf{\Sigma}^{(n)}(\theta)$ or an adequate estimate of the same. Thus determining q prior to this projection step is absolutely crucial.

3. AN INFORMATION CRITERION

3.1 Population Level

In practice, only finite segments of length T of a finite number n of $\{X_{it}\}$'s are observed, and selection of q must be based on this finite-sample information. As a preparatory step, however, we first prove a consistency result, as $n \to \infty$, at the population level, assuming that the spectral density matrices $\mathbf{\Sigma}_n(\theta)$ are known.

In this setting, the projection $\chi_{it}^{(n)}$ of X_{it} onto the space spanned by $\mathbf{\Sigma}^{(n)}(\theta)$'s first q dynamic principal components provides a consistent reconstruction of χ_{it}. For given q, those dynamic principal components, and hence the $\chi_{it}^{(n)}$'s, can be viewed as solutions of an optimization problem in which the "expected mean of squared residuals," $n^{-1}\mathrm{E}[\sum_{i=1}^{n}(X_{it} - \chi_{it}^{(n)})^2]$, is minimized, yielding a minimum of $n^{-1} \times \sum_{j=k+1}^{n}\{\int_{-\pi}^{\pi} \lambda_{nj}(\theta)\,d\theta\}$. This minimum plays the role of the residual variance classically appearing, in empirical form, in information criterion methods. Accordingly, denoting by $q_{\max}$ some predefined upper bound and by $p(n)$ some adequate penalty, we propose selecting the number of factors as

$$\hat{q}_n := \operatorname*{arg\,min}_{0\le k\le q_{\max}} IC_{0;n}(k),$$

$$\text{with } IC_{0;n}(k) := \frac{1}{n}\sum_{j=k+1}^{n}\int_{-\pi}^{\pi} \lambda_{nj}(\theta)\,d\theta + kp(n). \quad (2)$$

Note that here $\hat{q}_n$ is deterministic, because the spectral density matrices $\mathbf{\Sigma}_n(\theta)$ are supposed to be known.

The intuition behind (2) is clear. For the bounded eigenvalues ($k > q$), the averaged contribution, $\frac{1}{n}\sum_{j=k+1}^{n}\int_{-\pi}^{\pi} \lambda_{nj}(\theta)\,d\theta$, should be "small." The penalty $kp(n)$, as $n \to \infty$, should not be too large, or q will be underestimated, yet should be large enough to avoid overestimation. This delicate balance between overestimation and underestimation is intimately related to the rate of divergence, as $n \to \infty$, of the diverging eigenvalues. To impose consistency conditions on the penalty $p(n)$, an assumption about the divergence rate of the smallest diverging eigenvalue is needed.

Assumption A4. (a) The qth diverging eigenvalue of $\mathbf{\Sigma}_n(\theta)$ diverges at least linearly in n, that is, $\liminf_{n\to\infty} \inf_\theta n^{-1} \times \lambda_{nq}(\theta) > 0$.

(b) $q \le q_{\max}$, and $\lambda_{n(q_{\max}+1)}(\theta)$ is uniformly bounded away from 0; that is, there exists a constant $c_\lambda > 0$ such that for all θ, and $n \in \mathbb{N}$, $\lambda_{n(q_{\max}+1)}(\theta) > c_\lambda$.

The "at least linear" divergence in Assumption A4(a) is sufficient for the consistency proofs given later. However, linear divergence [i.e., diverging eigenvalues being $O(n)$ but not $o(n)$] corresponds to the very natural assumption that the influence of the common shocks is in some sense "stationary along the cross-section," which is a quite sensible assumption. (See Forni et al. 2004 for a discussion.)

The following lemma states a consistency result for $\hat{q}_n$ as $n \to \infty$.

Lemma 1. Let $\hat{q}_n$ be as defined in (2), and let the penalty $p(n)$ be such that

$$\lim_{n\to\infty} p(n) = 0 \qquad \text{and} \qquad \lim_{n\to\infty} np(n) = \infty. \quad (3)$$

Then, under Assumptions A1–A4, $\lim_{n\to\infty} \hat{q}_n = q$.

Proof. See the Appendix.

Examples of penalty functions satisfying (3) are $c/\sqrt{n}$ or $c\log(n)/n$, where c is an arbitrary positive real number. Lemma 1 has few practical consequences, of course. But the pedagogical value of its proof, which is extremely simple, is worth some attention. First, the proof clearly shows that the $\frac{1}{n}$ coefficient in the definition of $IC_{0;n}(k)$ and the second assumption on the penalty $[np(n) \to \infty]$ are directly related to the minimal $O(n)$ divergence rate in Assumption A4. A different divergence rate $[r(n)]$ would result in a different coefficient $[1/r(n)]$ and a different assumption on the penalty $[r(n)p(n) \to \infty]$. A second remark is that a penalty $p(n)$ leads to consistent estimation of q iff $cp(n)$ does, where c is an arbitrary positive constant. Thus multiplying the penalty with an arbitrary constant has no influence on the asymptotic performance of the identification method. But obviously it can affect the actual result for given n quite dramatically. We exploit this later when implementing the method (Sec. 4).

3.2 Sample Level

In real-life situations, the spectral density matrix $\boldsymbol{\Sigma}_n(\theta)$ is unknown and must be estimated from n observed series with finite length T; n and T, as well as the spectral estimation method, quite naturally play a role in the consistency conditions to be satisfied by the penalty. We develop mainly the case of lag window estimation, for which a consistency result is provided in Proposition 2. An alternative method is periodogram smoothing, which is presented briefly in Proposition 3.

Denoting by $\boldsymbol{\Gamma}_{nu}^T$ the sample cross-covariance matrix of $\mathbf{X}_{nt}$ and $\mathbf{X}_{n,t-u}$ based on T observations, the lag window estimator of $\boldsymbol{\Sigma}_n(\theta)$ is defined as

$$\boldsymbol{\Sigma}_n^T(\theta) := \frac{1}{2\pi}\sum_{u=-M_T}^{M_T} w(M_T^{-1}u)\boldsymbol{\Gamma}_{nu}^T e^{-iu\theta}, \tag{4}$$

where $\alpha \mapsto w(\alpha)$ is a positive even weight function and $M_T > 0$ is a truncation parameter; it is consistent for any n, as $T \to \infty$, under the following assumptions on w and M_T.

Assumption B1. (a) $M_T \to \infty$ and $M_T T^{-1} \to 0$, as $T \to \infty$.
(b) $\alpha \mapsto w(\alpha)$ is an even piecewise continuous function, piecewise differentiable up to order three, with bounded first three derivatives, satisfying $w(0) = 1$, $|w(\alpha)| \le 1$ for all α and $w(\alpha) = 0$ for $|\alpha| > 1$.

But such fixed-n consistency is not sufficient here; some uniformity over the cross-section is needed. This uniformity can be obtained by requiring some uniformity in the smoothness of the spectrum and its derivatives.

Assumption B2. The entries $\sigma_{ij}(\theta)$ of $\boldsymbol{\Sigma}_n(\theta)$ are uniformly (in n and θ) bounded and have uniformly (in n and θ) bounded derivatives up to order two; namely, there exists $Q < \infty$ such that $\sup_{i,j\in\mathbb{N}} \sup_\theta |\frac{d^k}{d\theta^k}\sigma_{ij}(\theta)| \le Q$, $k = 0, 1, 2$.

Under Assumptions A2′, B1, and B2, we then have the following uniform consistency result: There exist constants L_1, L_2, and T_0 such that

$$\sup_n \max_{1\le i,j\le n} \sup_\theta \big[\mathrm{E}|\boldsymbol{\Sigma}_n^T(\theta) - \boldsymbol{\Sigma}_n(\theta)|_{ij}^2\big] \le L_1 M_T T^{-1} + L_2 M_T^{-4} \quad \text{for any } T > T_0. \tag{5}$$

The proof of this result is long but easy; it consists mainly of following all of the steps of Parzen's (1957) proof of theorem 5A, taking into account the uniformity of Assumptions A2′, B1, and B2.

Associated with the lag window estimator $\boldsymbol{\Sigma}_n^T(\theta)$, consider the information criterion

$$IC_{1;n}^T(k) := \frac{1}{n}\sum_{i=k+1}^{n} \frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T} \lambda_{ni}^T(\theta_l) + kp(n,T), \quad 0 \le k \le q_{\max}, \tag{6}$$

where $p(n, T)$ is a penalty function, $\theta_l := \pi l/(M_T + 1/2)$ for $l = -M_T, \dots, M_T$, and $q_{\max}$ again is some predetermined upper bound; the eigenvalues $\lambda_{ni}^T(\theta)$ are those of $\boldsymbol{\Sigma}_n^T(\theta)$.

This criterion has a structure comparable to that of Bai and Ng (2002). In a corollary to their theorem 2, those authors also showed that a logarithmic form of their criterion has similar consistency properties as the original one. Therefore, we also consider

$$IC_{2;n}^T(k) := \log\left[\frac{1}{n}\sum_{i=k+1}^{n} \frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T} \lambda_{ni}^T(\theta_l)\right] + kp(n,T), \quad 0 \le k \le q_{\max}. \tag{7}$$

Depending on the criterion adopted, the estimated factor number, for given n and T, is

$$q_{a;n}^T := \underset{0\le k\le q_{\max}}{\operatorname{argmin}}\ IC_{a;n}^T(k), \qquad a = 1, 2. \tag{8}$$

The following proposition provides sufficient conditions for the consistency of $q_{1;n}^T$ and $q_{2;n}^T$.

Proposition 2. Let Assumptions A1, A2′, A3, A4, B1, and B2 hold. Then $\mathrm{P}(q_{a;n}^T = q) \to 1$ for $a = 1, 2$ provided that n and T both tend to infinity, in such a way that

$$\begin{aligned} &\text{(a)} \quad p(n,T) \to 0 \quad \text{and} \\ &\text{(b)} \quad \min\big(n, M_T^2, M_T^{-1/2}T^{1/2}\big)p(n,T) \to \infty. \end{aligned} \tag{9}$$

Proof. See the Appendix.

Note that if $p(n, T)$ is an appropriate penalty function [i.e., if (9) holds], then $cp(n, T)$, where c is an arbitrary positive real, is appropriate as well. For given n and T, the value of a penalty function $p(n, T)$ satisfying (9) thus can be arbitrarily small or arbitrarily large; the same remark holds for the approaches of Bai and Ng (2002, 2005) and Amengual and Watson (2006).

A variant of this method also can be based on the periodogram-smoothing estimator

$$\boldsymbol{\Sigma}_n^{\dagger T}(\theta) := \frac{2\pi}{T}\sum_{t=1}^{T-1} W^{(T)}\left(\theta - \frac{2\pi t}{T}\right)\mathbf{I}_n^T\left(\frac{2\pi t}{T}\right),$$

where

$$\mathbf{I}_n^T(\alpha) := \frac{1}{2\pi T}\left[\sum_{t=1}^{T-1}\mathbf{X}_{nt}\mathrm{e}^{-i\alpha t}\right]\left[\sum_{t=1}^{T-1}\mathbf{X}_{nt}'\mathrm{e}^{i\alpha t}\right]$$

and

$$W^{(T)}(\alpha) := \sum_{j=-\infty}^{\infty} W(B_T^{-1}(\alpha + 2\pi j)).$$

Here $\alpha \mapsto W(\alpha)$ is a positive even weight function of bounded variation, with bounded derivative, satisfying $\int_{-\infty}^{\infty} W(\alpha)\,d\alpha = 1$ and $\int_{-\infty}^{\infty} |\alpha|^2 W(\alpha)\,d\alpha < \infty$, and $B_T > 0$ is a bandwidth such that $B_T \to 0$ and $B_T T \to \infty$, as $T \to \infty$. The information criterion then takes the form ($\theta_l := 2\pi l/T$ for $l = 1, \dots, T-1$)

$$IC_n^{\dagger T}(k) := \frac{1}{n}\sum_{i=k+1}^{n} \frac{1}{T-1}\sum_{l=1}^{T-1} \lambda_{ni}^{\dagger T}(\theta_l) + kp(n,T), \quad 0 \le k \le k_{\max} < \infty,$$

where $p(n, T)$ is a penalty. Reinforcing the condition on cumulants in Assumption A2′ into

$$\sup_{i_1,\dots,i_\ell} \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{\ell-1}=-\infty}^{\infty} (1 + |k_j|)\big|c_{i_1,\dots,i_\ell}(k_1, \dots, k_{\ell-1})\big| < \infty,$$

and letting $q_n^{\dagger T} := \arg\min_{0\le k\le q_{\max}} IC_n^{\dagger T}(k)$, the following counterpart of Proposition 2 holds.

Proposition 3. Under the assumptions made previously, $\mathrm{P}(q_n^{\dagger T} = q) \to 1$ provided that n and T tend to infinity in such a way that

(a) $p(n,T) \to 0$ and

(b) $\min[n, B_T^{-2}, B_T^{1/2}T^{1/2}]p(n,T) \to \infty$.

The proof, being similar to that of Proposition 2, is left to the reader. Again, if $p(n,T)$ is an appropriate penalty function, then so is $cp(n,T)$, where $c > 0$ is an arbitrary real value.

4. A PRACTICAL GUIDE TO THE SELECTION OF q

As emphasized in the previous section, if our identification procedures are consistent for penalty $p(n,T)$, then they also are consistent for any penalty of the form $cp(n,T)$, where $c \in \mathbb{R}^+$. Important as they are, the consistency results of Section 3 are therefore of limited practical value. In this section we show how this degree of freedom in the choice of c—at first sight, a distressing fact—can be exploited. After some empirical considerations based on two examples, we describe the practical implementation of our method. In Section 5.1 we validate the method through simulation, and in Section 5.2 we apply it to real data.

Denote by $q_{c;1;n}^T$ and $q_{c;2;n}^T$ the number of factors resulting from applying criterion (6) or (7), with penalty $cp(n,T)$. Because both n and T are fixed in practice, the only information that we can obtain on the functions $(n,T) \mapsto q_{c;a;n}^T$ is contained in J-tuples of the form $q_{c;a;n_j}^{T_j}$, $a = 1, 2$, $j = 1, \dots, J$, where $0 < n_1 < \cdots < n_J = n$ and $0 < T_1 \le \cdots \le T_J = T$. For any fixed value of (n_j, T_j), $c \mapsto q_{c;a;n_j}^{T_j}$ is clearly nonincreasing; thus for given a, the curves $[n_j, T_j] \mapsto q_{c;a;n_j}^{T_j}$ never cross each other. The typical situation is as follows (for simplicity, we drop the a subscripts).

Assume that $q > 0$. If we let $c = 0$ (i.e., no penalty at all, thus a nonvalid value of c), then $j \mapsto q_{0;n_j}^{T_j}$ coincides with $q_{\max}$. If $c > 0$ is "very small" (i.e., severe *under*penalization), then, although Proposition 3 applies, the situation for finite (n,T) will not be very different; $j \mapsto q_{c;n_j}^{T_j}$ is increasing until it reaches $q_{\max}$ and will redescend and tend to q (as implied by Prop. 3) only if n and T are allowed to increase without limits. As c grows (and hence also the penalization), this increase of $j \mapsto q_{c;n_j}^{T_j}$ is less and less marked; for c large enough, it eventually decreases, or even may be decreasing from the beginning. A common feature of all of these underpenalized cases, however, is that (except for $c = 0$) the variability among the J values of $q_{c;n_j}^{T_j}$, $j = 1, \dots, J$, is high; this variability is captured by, for instance,

$$S_c^2 := J^{-1}\sum_{j=1}^{J}\left(q_{c,n_j}^{T_j} - J^{-1}\sum_{j=1}^{J} q_{c,n_j}^{T_j}\right)^2. \tag{10}$$

Next consider a "very large" value of c, and hence severely *over*penalized $q_{c;n_j}^{T_j}$'s. If c is large enough, then $q_{c;n_j}^{T_j}$ is identically 0 for all (n_j, T_j)'s, and its eventual convergence to q will not be visible for the values of (n,T) at hand. As c decreases, this convergence is observed for smaller and smaller values of (n,T), yielding horizontal segments at underestimated values of q.

Due to the monotonicity of $c \mapsto q_{c;n_j}^{T_j}$, somewhere between those "small" underpenalizing values of c (with $j \mapsto q_{c;n_j}^{T_j}$ eventually tending to q from above) and the "large" overpenalizing values (with $j \mapsto q_{c;n_j}^{T_j}$ eventually tending to q from below), a range of "moderate" values of c, yielding stable behavior of $j \mapsto q_{c;n_j}^{T_j} \approx q$, typically exists. This stability can be assessed, for given c, through the empirical variance (10) of the $q_{c;n_j}^{T_j}$'s, $j = 1, \dots, J$.

As an illustration, we consider two examples, with $n = T = 300$ and $q = 3$. In both cases, our method has been applied with $M_T = [.75\sqrt{T}]$, $q_{\max} = 19$, penalty $p_1(n,T) = (M_T^{-2} + M_T^{1/2}T^{-1/2} + n^{-1})\log(\min[n, M_T^2, M_T^{-1/2}T^{1/2}])$, and criteria $IC_{1;n}^T$ and $IC_{2;n}^T$, for $(n_j, T_j) = (120, 120), (130, 130), \dots, (300, 300)$ and $c \in \mathcal{C} = \{.001, .002, \dots, 2.000\}$.

Example 1. The common part was modeled with moving average (MA) loadings; see Section 5.1 for details. The graphs of $(n_j, T_j) \mapsto q_{c,n_j}^{T_j}$ and $c \mapsto S_c$ are presented in Figures 1(a1) and 1(a2) for criterion $IC_{1;n}^T$ and in Figures 1(b1) and 1(b2) for $IC_{2;n}^T$. The typical patterns just described are all present in Figure 1(a1) as well as in Figure 1(b1). Inspection of Figure 1(a2) in conjunction with Figure 1(a1) reveals the characteristic fact that S_c vanishes over certain intervals of c values, associated with a stable behavior of the corresponding curves in Figure 1(a1); the $c \mapsto S_c$ curve in Figure 1(a2) yields four "stability intervals," $[0, .005]$, $[.122, .188]$, $[.210, .351]$, $[.388, .547]$, and $[.579, 2.000]$, corresponding to a selection of $q = q_{\max} = 19$, 3, 2, 1, and 0 factors. These "stability intervals" are separated by "instability" intervals, corresponding to more fluctuations in the Figure 1(a1)'s curves. The correct value of q in Figure 1(a1) is obtained for $c = .15$. Note that $q_{.10,n_j}^{T_j}$ converges, as j increases, to $q_{.15,n_j}^{T_j}$ from above, whereas $q_{.20,n_j}^{T_j}$ converges to $q_{.15,n_j}^{T_j}$ from below, and that $\hat{c} = .15$ is the only value of c exhibiting such pattern. Similar comments can be made for $IC_{2;n}^T$ and Figure 1(b1), with $\hat{c} = .35$. Note that both $\hat{c}$ values lie in the second "stability interval" of $c \mapsto S_c$—namely, $[.122, .188]$ in Figure 1(a2) and $[.346, .385]$ in Figure 1(b2)—whereas the first "stability interval"—$[0, .005]$ in Figure 1(a2) and $[0, .178]$ in Figure 1(b2)—is clearly associated with severe underpenalization, yielding the maximal possible value $\hat{q} = q_{\max}$.

This example suggests that irrespective of the choice of $IC_{1;n}^T$ or $IC_{2;n}^T$, the selection of q can be based on an inspection of the family of curves $(n_j, T_j) \mapsto q_{c,n_j}^{T_j}$ indexed by $c \in \mathcal{C}$, trying to spot the value of c (the curve) with "neighbors" $c \pm \delta$ ($\delta > 0$) such that $q_{c+\delta}^{T_j} \uparrow q_{c,n_j}^{T_j}$ whereas $q_{c-\delta}^{T_j} \downarrow q_{c,n_j}^{T_j}$ as $j \uparrow J$. This search is greatly facilitated, and can be made automatic, by also considering the $c \mapsto S_c$ mapping and choosing $\hat{q} = q_{\hat{c},n}^T$, where $\hat{c}$ belongs to the second "stability interval." The relevant figure for this is a joint plot of $c \mapsto S_c$ and $c \mapsto q_{c,n}^T$, as shown in Figure 2.

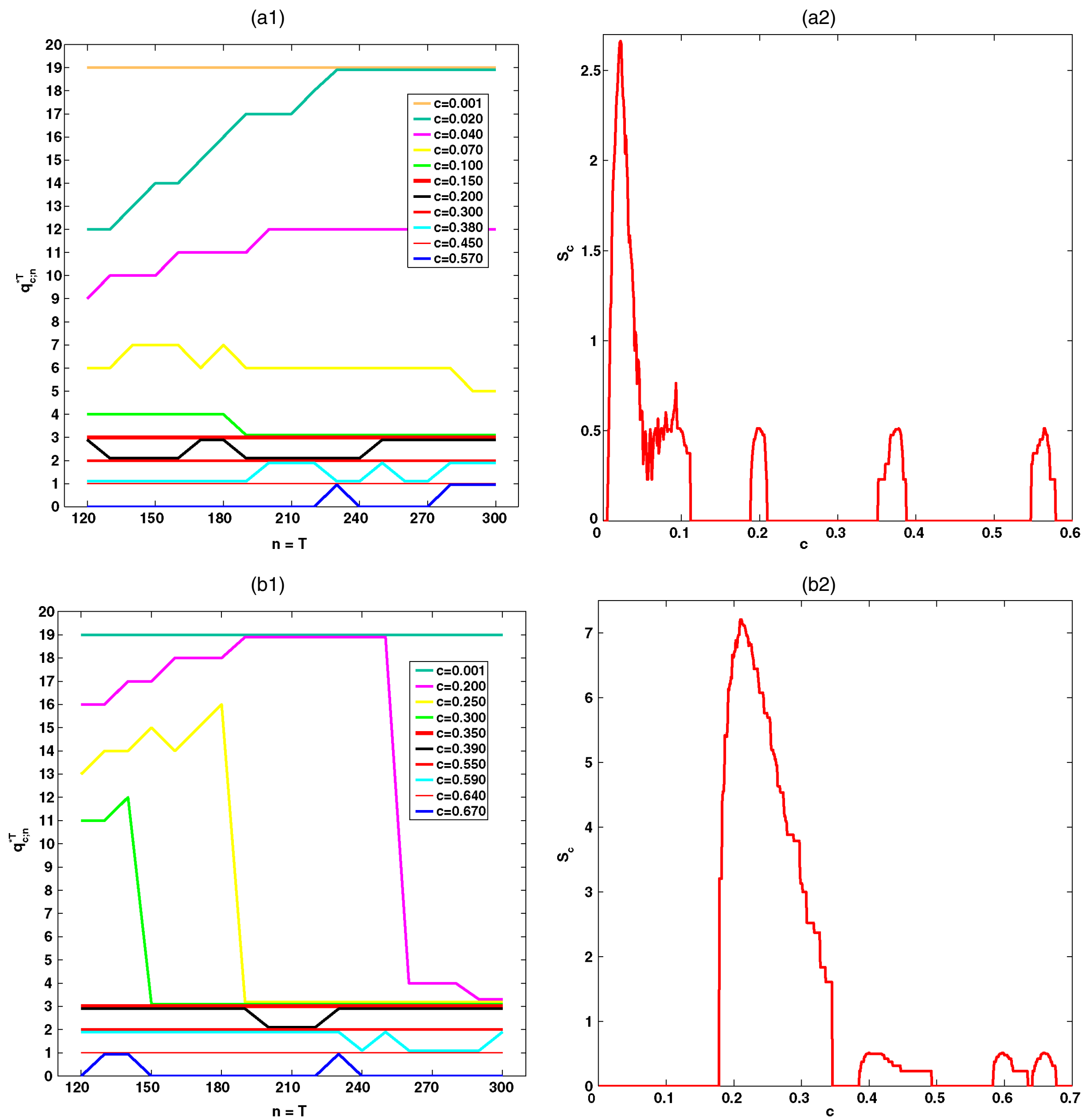

Figure 1. Example 1: MA Loadings, $q = 3$, $n = T = 300$, $M_T = [.75\sqrt{T}]$. Graphs of $(n_j, T_j) \mapsto q^{T_j}_{c;n_j}$ and $c \mapsto S_c$ for $(n_j, T_j) = (120, 120), (130, 130), \ldots, (300, 300)$ and various values of c, using penalty function p_1, $q_{max} = 19$, and [(a1) and (a2)] $IC^T_{1;n}$ criterion, [(b1) and (b2)] $IC^T_{2;n}$ criterion.

Example 2. The common part was modeled with autoregressive (AR) loadings (see Sec. 5.1 for details). Here we apply the automatic selection rule just described, based on the $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$ plots in Figure 3. For the $IC^T_{1;n}$ criterion, the stability intervals in Figure 1(d1) are $[0, .007]$, $[.164, .284]$, $[.336, .464]$, $[.512, .765]$, and $[.867, 2.000]$, yielding $q^T_{c,n} = q_{\max} = 19$, 3 (correct identification), 2, 1, and 0. The situation is even clearer with $IC^T_{2;n}$ in Figure 1(d2), with stability intervals $[0, .242]$, $[.509, .591]$, $[.729, .732]$, $[.878, .894]$, and $[1.03, 2.000]$, yielding $q^T_{c,n} = q_{\max} = 19$, 3 (correct identification), 2, 1, and 0. Thus, in both cases the second stability interval correctly identifies $q = 3$.

When T is small relative to n, which is typically the case in macroeconomic datasets, one may want to look at J-tuples $n_1, \ldots, n_J$ only, keeping T fixed. The monotonicity of $c \mapsto q^T_{c;n_j}$ still holds, and the same discussion as before can apply, although some patterns may not be present. Finally, whenever the actual value of q is 0 (i.e., no common component at all), the same analysis can be made, but the overpenalization part of the picture is not present; typically, no $(n_j, T_j) \mapsto q^{T_j}_{c,n_j}$ curve will tend to any other one from below, and only two stability intervals will appear in the $c \mapsto S_c$ plots, with the second one extending to the maximal possible value of c and corresponding to $q^T_{c,n} = 0$ (corect identification).

This existence of "stability intervals" in the $s \mapsto S_c$ graphs, and their relation to $q_{\max}, q, q-1, \ldots, 0$, is an empirical finding, but it can be explained as follows. For simplicity, consider the "population-level problem" of Section 3.1. There, the se-

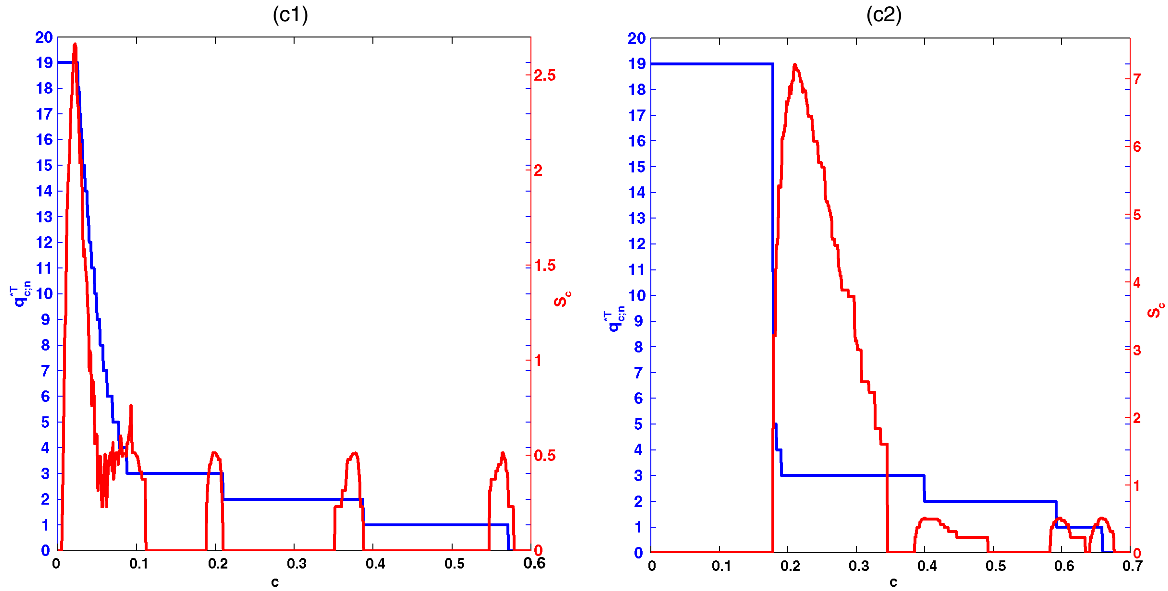

Figure 2. Example 1: MA Loadings, $q = 3$, $n = T = 300$, $M_T = [.75\sqrt{T}]$. Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{max} = 19$, and (c1) $IC^T_{1;n}$ criterion, (c2) $IC^T_{2;n}$ criterion.

lected number of factors $\hat{q} =: \hat{q}^n_c$ associated with penalty $cp(n)$ is characterized by the fact that

$$\begin{cases} \dfrac{1}{(k-\hat{q})}\displaystyle\sum_{j=\hat{q}+1}^{k} \int \lambda_{nj}(\theta)\,d\theta < cnp(n), \\ \qquad\qquad\qquad\qquad k = \hat{q}+1, \hat{q}+2, \ldots \\ \dfrac{1}{(\hat{q}-k)}\displaystyle\sum_{j=k+1}^{\hat{q}} \int \lambda_{nj}(\theta)\,d\theta > cnp(n), \\ \qquad\qquad\qquad\qquad k = \hat{q}-1, \hat{q}-2, \ldots, \end{cases} \tag{11}$$

which, noting that both $k \mapsto \frac{1}{(k-q)}\sum_{j=q+1}^{k} \int \lambda_{nj}(\theta)\,d\theta$ and $k \mapsto \frac{1}{(q-k)}\sum_{j=k+1}^{q} \int \lambda_{nj}(\theta)\,d\theta$ are decreasing, reduces to ($k = \hat{q} \pm 1$),

$$\frac{1}{n}\int \lambda_{n\hat{q}+1}(\theta)\,d\theta < cp(n) < \frac{1}{n}\int \lambda_{n\hat{q}}(\theta)\,d\theta. \tag{12}$$

Thus, once c is chosen, the criterion identifies the number of factors as the (unique, in view of monotonocity in c) $\hat{q}^n_c$ such that $cp(n)$ "separates" $\frac{1}{n}\int \lambda_{n\hat{q}+1}(\theta)\,d\theta$ and $\frac{1}{n}\int \lambda_{n\hat{q}}(\theta)\,d\theta$. Figure 4 illustrates this finding. The red lines are plots of $n \mapsto \frac{1}{n}\int \lambda_{nk}(\theta)\,d\theta$, roughly yielding [in the least favorable case of $O(n)$ diverging eigenvalues] constants for $1 \le k \le q$ and hy-

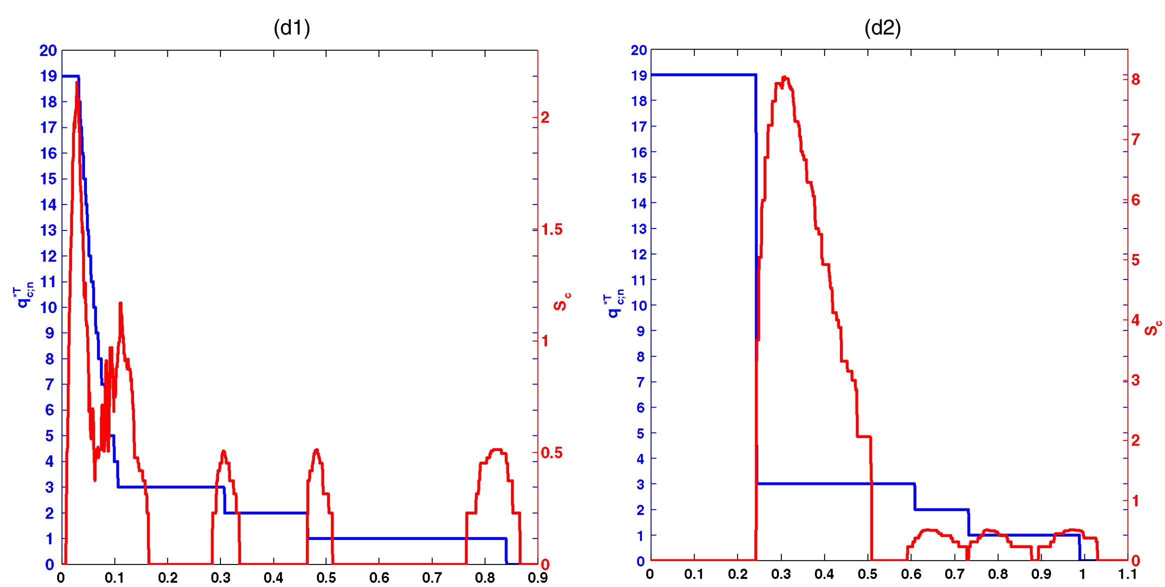

Figure 3. Example 2. AR Loadings, $q = 3$, $n = T = 300$, $M_T = [.75\sqrt{T}]$. Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{max} = 19$, and (d1) $IC^T_{1;n}$ criterion and (d2) $IC^T_{2;n}$ criterion.

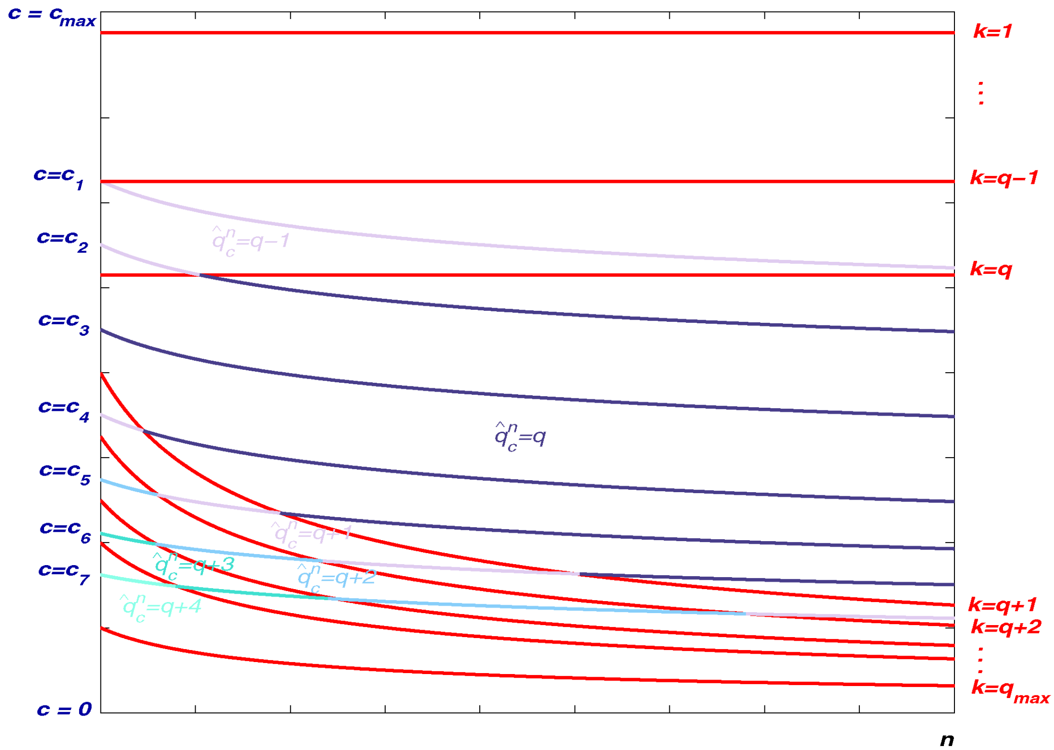

Figure 4. Heuristic Behavior of $c \mapsto S_c$. Graphs of $n \mapsto \frac{1}{n}\int \lambda_{nk}(\theta)\,d\theta$, $k = 1, \ldots, q_{max}$ (in red), and $n \mapsto cp(n)$, $c = 0, \ldots, c_{max}$ (in blue), along with the corresponding $\hat{q}_c^n$'s. Note that $S_{c_1} = S_{c_3} = 0$, whereas $S_{c_2}, S_{c_4}, \ldots, S_{c_7}$ are strictly positive.

perbolically decreasing values for $k \geq q$. The blue lines are the $n \mapsto cp(n)$ curves for various values of c. Each choice of c and n yields a blue line and a point on that line; $\hat{q}_c^n$ is the value of k associated with the first red curve lying above that point; the value of S_c associated with a blue line is 0 iff the blue line does not cross any of the red lines. For instance, $c = c_1$ leads to a "stable" ($S_{c_1} = 0$) underidentification $\hat{q}_{c_1} = q - 1$; S_{c_2} is strictly positive, with a hesitation between $\hat{q}_{c_2} = q - 1$ and $\hat{q}_{c_2} = q$, but S_{c_3} again is 0, with a correct identification of q. From $c = c_4$ on, S_c is strictly positive.

Summing up, in practice our identification method is performed as follows:

1. Preliminary to the analysis, it may be worth choosing a random permutation of the n cross-sectional items, because some irrelevant structure may exist in the initial ordering (although this is completely optional). Fix the upper bound $q_{\max}$.
2. Choose $T \mapsto M_T$ (e.g., $M_T := [.5\sqrt{T}]$ or $M_T := [.7\sqrt{T}]$) and a smoothing function $\mapsto w(\alpha)$ such that Assumptions B1(a) and (b) be satisfied.
3. Choose a penalty, $(n, T) \mapsto p(n, t)$, and a criterion $[IC_{1;n}^T(k)$ or $IC_{2;n}^T(k)]$, and define $p_c^*(n, t) = cp(n, t)$ for $c \in \mathcal{C} \subset [C^-, C^+] \subset \mathbb{R}^+$ (e.g., $\mathcal{C} := \{.01, .02, \ldots, 3\}$).
4. Define sequences $n_1 < n_2 < \cdots < n_J = n$ and $T_1 \leq T_2 \leq \cdots \leq T_J = T$; for example, for $(n, T) = (150, 100)$, set $n_j := 120 + 10j$ and $T := 70 + 10j$, $j = 1, \ldots, 3$.
5. Defining S_c as in (10), identify q as $\hat{q} := q_{\hat{c},n}^T$, where $\hat{c}$ belongs to the second stability interval of $c \mapsto S_c$.

Note that this data-driven selection of c does not affect consistency, because $\mathrm{P}[q_{C^+,n}^T \leq q_{\hat{c},n}^T \leq q_{C^-,n}^T] = 1$ for all n and T, where $q_{C^-,n}^T$ and $q_{C^+,n}^T$ satisfy the assumptions of Proposition 2. Also note that $c = 1$ does a very poor job in both examples.

5. NUMERICAL STUDY

5.1 Simulations

To evaluate the performance of the strategy proposed in the previous section, we conducted the following Monte Carlo experiment. Three datasets ($n = 150$, $T = 120$) were generated as follows, with $q = 1$, 2, and 3 factors from model (1):

1. For each $k = 1, \ldots, q$, the random shocks u_{kt}, $t = 1, \ldots, T$, are iid $\mathcal{N}(0, D_k)$, with $D_1 = 1$, $D_2 = .5$, and $D_3 = 1.5$.
2. The idiosyncratic components are of the form $\xi_{it} = \sum_{j=0}^{4} \sum_{k=0}^{2} g_{i,j;k} v_{i+j,t-k}$, where the v_{it}'s, $i = 1, \ldots, n$, $t = -1, \ldots, T$, are iid $\mathcal{N}(0, 1)$, and, denoting by $U_{[a,b]}$ the uniform distribution over $[a, b]$, the $g_{i,j;k}$'s, $i = 1, \ldots, n$, $j = 1, \ldots, 4$, $k = 0, 1, 2$, are iid $\mathrm{U}_{[1.0,1.5]}$, with the v_{it}'s and $g_{i,j;k}$'s mutually independent and independent of the u_{it}'s; this ensures both autocorrelation and cross-correlation among idiosyncratics.
3. The filters $b_{ik}(L)$ ($i = 1, \ldots, n$, $k = 1, \ldots, q$) are randomly generated (independently from the u_{kt}'s and ξ_{it}'s) by one of the following two devices:
   - MA loadings: $b_{ik}(L) = b_{ik;0} + b_{ik;1}L + b_{ik;2}L^2$ with iid and mutually independent coefficients $(b_{ik;0}, b_{ik;1}, b_{ik;2}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$
   - AR loadings: $b_{ik} = b_{ik;0}(1 - b_{ik;1}L)^{-1}(1 - b_{ik;2}L)^{-1}$, with iid and mutually independent coefficients $b_{ik;0}^0 \sim \mathcal{N}(0, 1)$, $b_{ik;1} \sim \mathrm{U}_{[.8,.9]}$, and $b_{ik;2} \sim \mathrm{U}_{[.5,.6]}$;

for each i, the variance of ξ_{it} and that of the common component $\sum_{k=1}^{q} b_{ik}(L)u_{it}$ are normalized to .5. Note that the MA loadings satisfy the assumptions of Bai and Ng's (2005) restricted dynamic model, but the AR loadings do not.

For each case, 500 replications were generated, and for each of these, q was identified by the following procedure

1. Running the automatic identification procedure described in the previous section, with sequences $n_j := n - 10j, j = 1, \ldots, 3, T_i := T - 10i, i = 1, \ldots, 3$; penalties

$$p_1(n, T) = \left(M_T^{-2} + M_T^{1/2} T^{-1/2} + n^{-1}\right) \times \log\left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right),$$

$$p_2(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)^{-1/2},$$

and

$$p_3(n, T) = \left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)^{-1} \times \log\left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right);$$

$q_{\max} = 19$; and $\mathcal{C} = [.01, .02, \ldots, 3]$. Spectral densities were estimated with a triangular smoothing function $w(v) = 1 - |v|$, $M_T = [.50\sqrt{T}]$, and $M_T = [.75\sqrt{T}]$.

2. Running the method of Bai and Ng (2005) based on covariance and correlation matrices of AR(4) residuals, with penalties

$$p_1^{\mathrm{BNg}}(n, T) = \log(nT/(n + T))/nT,$$
$$p_2^{\mathrm{BNg}}(n, T) = \log(\min(n, T))(n + T)/nT,$$

and

$$p_3^{\mathrm{BNg}}(n, T) = \log(\min(n, T))/\min(n, T);$$

a maximum of 10 static factors; statistics $\hat{D}_{1,k}$ and $\hat{D}_{2,k}$, and various values of m (see Bai and Ng 2005 for precise definitions).

Tables 1, 2, and 3 provide, for each case, the percentages (over the 500 replications) of underidentification and overidentification of q. Inspection of Table 1 shows that correct identification rates for our method are uniformly very good, with a slight deterioration at $q = 3$ for small values of n, due to the substantial amount of idiosyncratic cross-correlation in the data. The logged criterion $IC_{2;n}^T$, with $M_T = [.75\sqrt{T}]$ and p_1, yields the best performance, but this choice of the penalty function and M_T seems to have a limited impact on the results. For AR loadings, all versions of the Bai and Ng method behave very

Table 1. Underidentification and Overidentification Relative Frequencies (in %), Over 500 Replications of the MA and AR Loading Models, for q = 1, 2, and 3, Penalty Functions p_1, p_2, and p_3, of the Number of Factors Identified by Applying the Automatic Procedure of Section 4 With $IC_{1;n}^T$ and $IC_{2;n}^T$, $q_{max} = 19$; $M_T = [.5\sqrt{T}]$ and $M_T = [.75\sqrt{T}]$

			$IC_{1;n}^T$						$IC_{2;n}^T$					
			$M_T = [.50\sqrt{T}]$			$M_T = [.75\sqrt{T}]$			$M_T = [.50\sqrt{T}]$			$M_T = [.75\sqrt{T}]$		
q	n	T	p_1	p_2	p_3	p_1	p_2	p_3	p_1	p_2	p_3	p_1	p_2	p_3
MA loadings														
1	60	100	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	100	100	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	70	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	120	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	150	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
2	60	100	2/0	2/0	2/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	100	100	0/0	1/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	70	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	120	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	150	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3	60	100	29/2	39/0	38/0	31/0	33/0	31/0	50/0	47/0	49/0	27/0	33/0	36/0
	100	100	22/1	30/0	27/0	22/0	22/0	24/0	24/0	21/0	24/0	11/0	11/0	12/0
	70	120	21/0	22/0	23/0	17/0	19/0	18/0	22/0	20/0	23/0	9/0	9/0	11/0
	120	120	14/0	13/0	14/0	11/0	13/0	13/0	9/0	6/0	7/0	4/0	4/0	4/0
	150	120	13/0	12/0	12/0	8/0	10/0	10/0	6/0	5/0	5/0	3/0	3/0	3/0
AR loadings														
1	60	100	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	100	100	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	70	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	120	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	150	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
2	60	100	5/0	5/0	5/0	4/0	4/0	4/0	3/0	3/0	4/0	3/0	3/0	3/0
	100	100	3/0	2/0	3/0	3/0	3/0	3/0	1/0	1/0	1/0	1/0	1/0	1/0
	70	120	2/0	2/0	1/0	1/0	1/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
	120	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	150	120	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
3	60	100	42/0	40/0	40/0	36/0	40/0	37/0	46/0	46/0	46/0	47/0	50/0	50/0
	100	100	32/0	31/0	30/0	31/0	31/0	30/0	25/0	23/0	24/0	30/0	31/0	31/0
	70	120	23/0	23/0	23/0	22/0	23/0	23/0	19/0	19/0	19/0	18/0	20/0	20/0
	120	120	16/0	15/0	16/0	17/0	16/0	17/0	10/0	10/0	9/0	12/0	11/0	12/0
	150	120	15/0	14/0	14/0	15/0	15/0	16/0	9/0	7/0	8/0	10/0	10/0	10/0

Table 2. Bai and Ng (2005) Criterion, MA Loadings. Underidentification and overidentification relative frequencies (in %), over 500 replications of the MA loading model, for $q = 1$, 2, and 3, of the number of factors identified by applying the Bai and Ng (2005) criterion based on (a) the covariance and (b) the correlation matrices of VAR(4) residuals, with penalty functions p_1^{BNg}, p_2^{BNg}, p_3^{BNg}, $m = 2$, 1, and .5, and an upper bound on the number of static factors set to 10

			p_1^{BNg}						p_2^{BNg}						p_3^{BNg}					
			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$		
q	n	T	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$
(a) covariance																				
1	60	100	0/0	0/0	0/67	0/0	0/37	0/94	0/0	0/0	0/23	0/0	0/2	0/49	0/0	0/0	0/82	0/0	0/63	0/100
	100	100	0/0	0/0	0/84	0/0	0/60	0/100	0/0	0/0	0/27	0/0	0/1	0/48	0/0	0/0	0/86	0/0	0/62	0/100
	70	120	0/0	0/0	0/64	0/0	0/36	0/95	0/0	0/0	0/23	0/0	0/1	0/45	0/0	0/0	0/76	0/0	0/63	0/100
	120	120	0/0	0/0	0/77	0/0	0/65	0/100	0/0	0/0	0/27	0/0	0/4	0/56	0/0	0/0	0/77	0/0	0/66	0/100
	150	120	0/0	0/0	0/66	0/0	0/47	0/100	0/0	0/0	0/47	0/0	0/21	0/89	0/0	0/0	0/66	0/0	0/47	0/100
2	60	100	14/0	0/2	0/93	4/0	0/72	0/99	13/0	0/2	0/81	7/0	0/31	0/99	16/0	0/2	0/97	5/0	0/85	0/100
	100	100	1/0	0/2	0/98	0/0	0/84	0/100	0/0	0/0	0/70	0/0	0/24	0/96	1/0	0/2	0/98	0/0	0/84	0/100
	70	120	7/0	0/0	0/93	1/0	0/75	0/100	7/0	0/0	0/75	3/0	0/27	0/99	7/0	0/0	0/97	1/0	0/88	0/100
	120	120	0/0	0/0	0/98	0/0	0/88	0/100	0/0	0/0	0/72	0/0	0/28	0/97	0/0	0/0	0/98	0/0	0/88	0/100
	150	120	0/0	0/0	0/95	0/0	0/69	0/100	0/0	0/0	0/88	0/0	0/49	0/97	0/0	0/0	0/95	0/0	0/69	0/100
3	60	100	90/0	6/11	0/99	52/0	0/89	0/99	94/0	21/10	1/87	77/0	7/44	1/91	91/0	3/11	0/99	47/0	0/95	0/100
	100	100	65/0	0/10	0/99	28/0	0/93	0/100	76/0	15/7	0/92	61/0	3/48	0/98	65/0	0/10	0/99	27/0	0/93	0/100
	70	120	83/0	2/8	0/99	38/0	0/94	0/100	88/0	14/7	0/96	62/0	1/57	0/99	82/0	1/7	0/99	36/0	0/97	0/100
	120	120	49/0	0/5	0/99	15/0	0/91	0/100	57/0	6/3	0/94	33/0	1/62	0/99	49/0	0/5	0/99	15/0	0/91	0/100
	150	120	47/0	0/2	0/96	20/0	0/68	0/100	48/0	1/2	0/93	24/0	0/59	0/99	47/0	0/2	0/96	20/0	0/68	0/100
(b) correlation			$m = 1.25$	$m = 1$	$m = 1.5$	$m = 2.25$	$m = 2$	$m = 2.5$	$m = 1.25$	$m = 1$	$m = 1.5$	$m = 2.25$	$m = 2$	$m = 2.5$	$m = 1.25$	$m = 1$	$m = 1.5$	$m = 2.25$	$m = 2$	$m = 2.5$
1	60	100	0/97	0/97	0/96	0/88	0/91	0/85	0/67	0/72	0/62	0/34	0/42	0/24	0/100	0/100	0/100	0/100	0/100	0/100
	100	100	0/100	0/100	0/100	0/100	0/100	0/100	0/72	0/76	0/70	0/55	0/65	0/46	0/100	0/100	0/100	0/100	0/100	0/100
	70	120	0/97	0/98	0/97	0/92	0/93	0/91	0/65	0/70	0/60	0/40	0/49	0/31	0/100	0/100	0/100	0/100	0/100	0/100
	120	120	0/100	0/100	0/100	0/100	0/100	0/100	0/81	0/83	0/81	0/72	0/77	0/60	0/100	0/100	0/100	0/100	0/100	0/100
	150	120	0/100	0/100	0/100	0/100	0/100	0/100	0/96	0/97	0/96	0/92	0/96	0/89	0/100	0/100	0/100	0/100	0/100	0/100
2	60	100	0/94	0/98	0/62	0/76	0/88	0/51	0/62	0/79	0/30	0/17	0/32	1/7	0/98	0/100	0/72	0/95	0/99	0/71
	100	100	0/100	0/100	0/98	0/99	0/100	0/96	0/75	0/85	0/62	0/37	0/52	0/25	0/100	0/100	0/98	0/99	0/100	0/96
	70	120	0/97	0/99	0/74	0/86	0/92	0/72	0/71	0/86	0/43	0/25	0/39	0/13	0/100	0/100	0/80	0/99	0/100	0/92
	120	120	0/100	0/100	0/100	0/100	0/100	0/100	0/85	0/90	0/78	0/58	0/68	0/44	0/100	0/100	0/100	0/100	0/100	0/100
	150	120	0/100	0/100	0/99	0/100	0/100	0/99	0/97	0/99	0/95	0/88	0/92	0/81	0/100	0/100	0/99	0/100	0/100	0/99
3	60	100	1/60	0/92	6/13	6/16	3/44	17/3	10/27	4/56	25/8	52/1	32/7	70/0	0/68	0/97	2/16	0/23	0/61	4/5
	100	100	0/82	0/97	0/42	0/40	0/69	0/16	4/43	2/67	12/19	26/7	15/17	36/1	0/82	0/97	0/43	0/41	0/70	0/16
	70	120	0/71	0/97	1/21	1/29	0/62	5/7	5/40	1/73	15/13	29/5	15/15	48/1	0/75	0/99	0/23	0/38	0/74	1/10
	120	120	0/88	0/99	0/53	0/48	0/77	0/22	1/63	0/83	3/37	7/23	4/41	14/7	0/88	0/99	0/53	0/48	0/77	0/22
	150	120	0/82	0/98	0/42	0/31	0/60	0/14	0/75	0/93	0/38	2/25	1/51	3/12	0/82	0/98	0/42	0/31	0/60	0/14

Table 3. Bai and Ng (2005) Criterion, AR Loadings. Underidentification and overidentification relative frequencies (in %), over 500 replications of the AR loading model, for $q = 1, 2$, and 3, of the number of factors identified by applying the Bai and Ng (2005) criterion based on (a) the covariance and (b) the correlation matrices of VAR(4) residuals, with penalty functions $p_1^{BNg}, p_2^{BNg}, p_3^{BNg}$, $m = 2, 1$, and .5 and an upper bound on the number of static factors set to 10

			p_1^{BNg}						p_2^{BNg}						p_3^{BNg}					
			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$		
q	n	T	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$	$m = 2$	$m = 1$	$m = .5$
(a) covariance																				
1	60	100	0/20	0/90	0/94	0/72	0/92	0/94	7/0	0/33	0/35	0/13	0/33	0/35	0/21	0/98	0/100	0/97	0/100	0/100
	100	100	0/49	0/98	0/100	0/97	0/100	0/100	0/17	0/50	0/54	0/25	0/51	0/54	0/51	0/98	0/100	0/98	0/100	0/100
	70	120	0/18	0/91	0/96	0/76	0/94	0/96	0/8	0/35	0/38	0/14	0/36	0/38	0/18	0/98	0/100	0/97	0/100	0/100
	120	120	0/43	0/99	0/100	0/98	0/100	0/100	0/22	0/61	0/65	0/34	0/62	0/65	0/43	0/99	0/100	0/98	0/100	0/100
	150	120	0/37	0/98	0/100	0/98	0/100	0/100	0/31	0/90	0/95	0/74	0/93	0/95	0/37	0/98	0/100	0/98	0/100	0/100
2	60	100	8/18	0/95	0/95	0/82	0/95	0/95	4/12	0/44	0/44	1/21	0/44	0/44	11/10	0/100	0/100	0/99	0/100	0/100
	100	100	1/59	0/100	0/100	0/100	0/100	0/100	0/26	0/56	0/56	0/36	0/56	0/56	1/60	0/100	0/100	0/100	0/100	0/100
	70	120	4/20	0/96	0/96	0/86	0/96	0/96	2/15	0/48	0/49	1/23	0/49	0/49	6/14	0/99	0/100	0/100	0/100	0/100
	120	120	1/63	0/100	0/100	0/100	0/100	0/100	0/38	0/71	0/71	0/52	0/71	0/71	1/63	0/100	0/100	0/100	0/100	0/100
	150	120	1/56	0/100	0/100	0/100	0/100	0/100	1/51	0/96	0/96	0/89	0/96	0/96	1/56	0/100	0/100	0/100	0/100	0/100
3	60	100	68/0	0/95	0/95	2/82	0/95	0/95	50/1	0/47	0/47	28/12	0/47	0/47	74/0	0/99	0/100	0/99	0/100	0/100
	100	100	11/29	0/100	0/100	0/100	0/100	0/100	13/16	0/66	0/66	6/37	0/66	0/66	12/29	0/100	0/100	0/100	0/100	0/100
	70	120	47/1	0/98	0/98	0/89	0/98	0/98	34/2	0/51	0/51	19/17	0/51	0/51	55/0	0/100	0/100	0/100	0/100	0/100
	120	120	4/52	0/100	0/100	0/100	0/100	0/100	5/32	0/76	0/76	1/53	0/76	0/76	4/51	0/100	0/100	0/100	0/100	0/100
	150	120	5/45	0/100	0/100	0/100	0/100	0/100	6/44	0/97	0/97	0/92	0/97	0/97	5/45	0/100	0/100	0/100	0/100	0/100
(b) correlation			$m = 1.25$	$m = 1$	$m = 1.5$	$m = 2.25$	$m = 2$	$m = 2.5$	$m = 1.25$	$m = 1$	$m = 1.5$	$m = 2.25$	$m = 2$	$m = 2.5$	$m = 1.25$	$m = 1$	$m = 1.5$	$m = 2.25$	$m = 2$	$m = 2.5$
1	60	100	0/94	0/94	0/94	0/94	0/94	0/94	0/35	0/35	0/35	0/35	0/35	0/35	0/100	0/100	0/100	0/100	0/100	0/100
	100	100	0/100	0/100	0/100	0/100	0/100	0/100	0/54	0/54	0/54	0/54	0/54	0/54	0/100	0/100	0/100	0/100	0/100	0/100
	70	120	0/96	0/96	0/96	0/96	0/96	0/96	0/38	0/38	0/38	0/38	0/38	0/37	0/100	0/100	0/100	0/100	0/100	0/100
	120	120	0/100	0/100	0/100	0/100	0/100	0/100	0/65	0/65	0/65	0/65	0/65	0/65	0/100	0/100	0/100	0/100	0/100	0/100
	150	120	0/100	0/100	0/100	0/100	0/100	0/100	0/95	0/95	0/95	0/95	0/95	0/95	0/100	0/100	0/100	0/100	0/100	0/100
2	60	100	0/95	0/95	0/95	0/93	0/95	0/89	0/44	0/44	0/44	1/35	0/41	3/26	0/100	0/100	0/100	0/100	0/100	0/100
	100	100	0/100	0/100	0/100	0/100	0/100	0/100	0/56	0/56	0/56	0/54	0/55	0/51	0/100	0/100	0/100	0/100	0/100	0/100
	70	120	0/96	0/96	0/96	0/94	0/96	0/92	0/49	0/49	0/49	1/42	0/47	2/36	0/100	0/100	0/100	0/100	0/100	0/100
	120	120	0/100	0/100	0/100	0/100	0/100	0/100	0/71	0/71	0/71	0/70	0/71	0/69	0/100	0/100	0/100	0/100	0/100	0/100
	150	120	0/100	0/100	0/100	0/100	0/100	0/100	0/96	0/96	0/96	0/96	0/96	0/95	0/100	0/100	0/100	0/100	0/100	0/100
3	60	100	0/95	0/95	0/94	2/86	1/90	4/80	0/46	0/47	0/41	24/13	11/21	43/9	0/100	0/100	0/98	0/100	0/100	0/100
	100	100	0/100	0/100	0/100	0/100	0/100	0/100	0/66	0/66	0/64	4/44	1/54	10/34	0/100	0/100	0/100	0/100	0/100	0/100
	70	120	0/98	0/98	0/98	0/92	0/94	1/88	0/51	0/51	0/49	16/19	5/29	29/13	0/100	0/100	0/100	0/100	0/100	0/100
	120	120	0/100	0/100	0/100	0/100	0/100	0/100	0/76	0/76	0/76	1/64	0/71	3/54	0/100	0/100	0/100	0/100	0/100	0/100
	150	120	0/100	0/100	0/100	0/100	0/100	0/100	0/97	0/97	0/97	0/95	0/96	0/94	0/100	0/100	0/100	0/100	0/100	0/100

poorly (Table 3), which is not surprising, because their consistency assumptions are not satisfied. Penalty p_2^{BNg}, with $\hat{D}_{1,k}$ and $m = 2$, seems to be least misleading here, but still yields a percentage of incorrect identification (mainly overestimation) uniformly greater than 50%. But the same version behaves rather poorly under MA loadings (Table 2), with severe underidentification of $q = 3$ even under large values of n and T. The best results here are obtained for p_1^{BNg}, with $\hat{D}_{1,k}$ and $m = 1$, and are quite similar to those of our method; but, contrary to our method, its performance under AR loadings is very poor (90% overestimation in the best case, increasing to 100% under large n and T). Correlation-based versions are uniformly worse than their covariance-based counterparts.

Thus, it seems that our method significantly outperforms that of Bai and Ng (2005), certainly when the possibility of AR loadings cannot be excluded.

5.2 A Real Data Application

The proposed criteria seems to work rather well in simulated data. We now consider a real case study, with data comprising $n = 132$ monthly time series observed from January 1960 through December 2003 ($T = 528$). These series are considered by economists to be a representative summary of the U.S. economy and have been studied by Stock and Watson (2005), Giannone et al. (2005, 2006), and Bai and Ng (2005) with surprisingly divergent conclusions. Stock and Watson found seven static factors and seven dynamic ones; thus $\hat{q} = 7$. Based on a different methodology (restricting to a carefully selected subset of 12 series), Giannone et al. (2005, 2006) arrived at $\hat{q} = 2$ dynamic factors. Bai and Ng did not give a clear-cut conclusion; in an early version of their article, they mentioned up to 10 static factors and 7 dynamic ones ($\hat{q} = 7$), but in the final version, they concluded in favor of 4 dynamic factors ($\hat{q} = 4$) spanning 7 static factors.

On the other hand, economists seem to agree that a change in the structure of the U.S. economy occurred around 1982–1983. Therefore, we applied our own methods to the same dataset (downloadable at *http://www.princeton.edu/~mwatson*), first over the full period ($T = 528$), then over each of the subperiods 1960–1982 ($T = 276$) and 1983–2003 ($T = 252$). In view of the simulation results, we chose the $IC_{2;n}^T$ criterion based on p_1, with $M_T = [.75\sqrt{T}]$. Our results are shown in Figures 5(a), 5(b), and 5(c). These pictures clearly show three factors ($\hat{q} = 3$) for 1960–1982, and one factor ($\hat{q} = 1$) for the more recent 1983–2003 period; for the full period, the conclusion is less clear, with four but perhaps only one factor (identification is borderline), which confirms the hypothesis of a changepoint.

6. CONCLUDING REMARKS

This article attempts to fill a gap on dynamic factor models in the literature by providing an efficient, yet flexible tool for identifying the number, q, of factors. We have established the consistency, as both n and T approach infinity in an appropriate way, of two versions of a method based on penalized information criterion ideas. We have also shown how to take advantage of the fact that penalty functions in such criteria are defined up to a multiplicative constant. We evaluated the performance of the method through simulation and found that it outperforms existing methods. Application to real data suggest that the number of factors driving the U.S. economy may in fact be quite low—much lower than suggested by an application of static or restricted dynamic factor models.

APPENDIX: PROOFS

Proof of Lemma 1.

We must show that $\lim_{n\to\infty}[IC_{0;n}(k) - IC_{0;n}(q)] > 0$ for all $k \neq q$, $k \leq q_{\max} < \infty$. This inequality holds provided that there exists a finite n_0 such that for all $n > n_0$ and $k \neq q$,

$$\frac{1}{n}\sum_{j=k+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)\,d\theta\right\} + kp(n) > \frac{1}{n}\sum_{j=q+1}^{n}\left\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)\,d\theta\right\} + qp(n).$$

Two cases are possible: either $k > q$, in which case, for n sufficiently large, $(k-q)p(n) > \frac{1}{n}\sum_{j=q+1}^{k}\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)\,d\theta\}$, because $np(n) \to \infty$ as $n \to \infty$, or $k < q$ and, for n sufficiently large, $\frac{1}{n}\sum_{j=k+1}^{q}\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)\,d\theta\} > (q-k)p(n)$, because $p(n) \to 0$ as $n \to \infty$ and $\lambda_{nj}(\theta), j \leq q$, under Assumption A4, is $O(n)$ but not $o(n)$. The result follows.

Before turning to the proof of Proposition 2, we prove a general result (Lemma A.1) on the asymptotic behavior of eigenvalues of (n, T)-indexed sequences of $n \times n$ random matrices as both n and T tend to infinity. This result relies on a matrix inequality of Weyl (1912), the importance of which in the context of factor models was first recognized by Giannone (2004) (see also lemma 1 of Forni et al. 2005a). Corollary A.1 translates Lemma A.1 into the form that we need in the problem at hand.

Let $\{\zeta_{ij}; i, j \in \mathbb{N}\}$ denote a collection of complex numbers such that for all n, the $n \times n$ matrices $\boldsymbol{\zeta}_n$ with entries $(\zeta_{ij}; 1 \leq i, j \leq n)$ are Hermitian. Denote by $\{\zeta_{n,ij}^T; 1 \leq i, j \leq n,\ n \in \mathbb{N},\ T \in \mathbb{N}\}$ a collection of complex-valued random variables such that, similarly, for all n and T, the $n \times n$ matrices $\boldsymbol{\zeta}_n^T$ with entries $(\zeta_{n,ij}^T; 1 \leq i, j \leq n)$ be Hermitian. Write $\lambda_{ni}(\boldsymbol{\zeta})$ and $\lambda_{ni}^T(\boldsymbol{\zeta})$ for the eigenvalues of $\boldsymbol{\zeta}_n$ and $\boldsymbol{\zeta}_n^T$ in decreasing order of magnitude. The following lemma characterizes the asymptotic behavior of $\lambda_{ni}(\boldsymbol{\zeta}) - \lambda_{ni}^T(\boldsymbol{\zeta})$ when $\boldsymbol{\zeta}_n - \boldsymbol{\zeta}_n^T$ converges to 0 in a sense made precise in (A.1).

Lemma A.1. Assume that for all $1 \leq i, j \leq n$, $n \in \mathbb{N}$ and $T \in \mathbb{N}$, there exist a positive constant K that does not depend on n, T, i or j, and a sequence of positive constants M_T depending on T only, such that $M_T \to \infty$ as $T \to \infty$ and

$$\mathrm{E}[|\zeta_{n,ij}^T - \zeta_{ij}|^2] \leq KM_T^{-1}. \tag{A.1}$$

Then for any $\epsilon > 0$, there exist B_ϵ and T_ϵ such that for any fixed $q_{\max}$, n, and $T > T_\epsilon$,

$$\max_{1\leq k\leq q_{\max}} \mathrm{P}\left[M_T^{1/2}\frac{1}{n}|\lambda_{nk}(\boldsymbol{\zeta}) - \lambda_{nk}^T(\boldsymbol{\zeta})| > B_\epsilon\right] \leq \epsilon. \tag{A.2}$$

Corollary A.1. Let Assumptions A1, A2′, B1, and B2 hold. Then for any $\epsilon > 0$, there exist M_ϵ and T_ϵ such that for any fixed $q_{\max}$, n, and $T > T_\epsilon$,

$$\max_{1\leq k\leq q_{\max}} \sup_{\theta} \mathrm{P}\left[\min\left(M_T^2, M_T^{-1/2}T^{1/2}\right) \times \left|\frac{\lambda_{nk}^T(\theta)}{n} - \frac{\lambda_{nk}(\theta)}{n}\right| > M_\epsilon\right] \leq \epsilon.$$

Proof of Lemma A.1. Weyl's inequality implies that for any Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$ with eigenvalues $\lambda_j(\mathbf{A})$ and $\lambda_j(\mathbf{B})$,

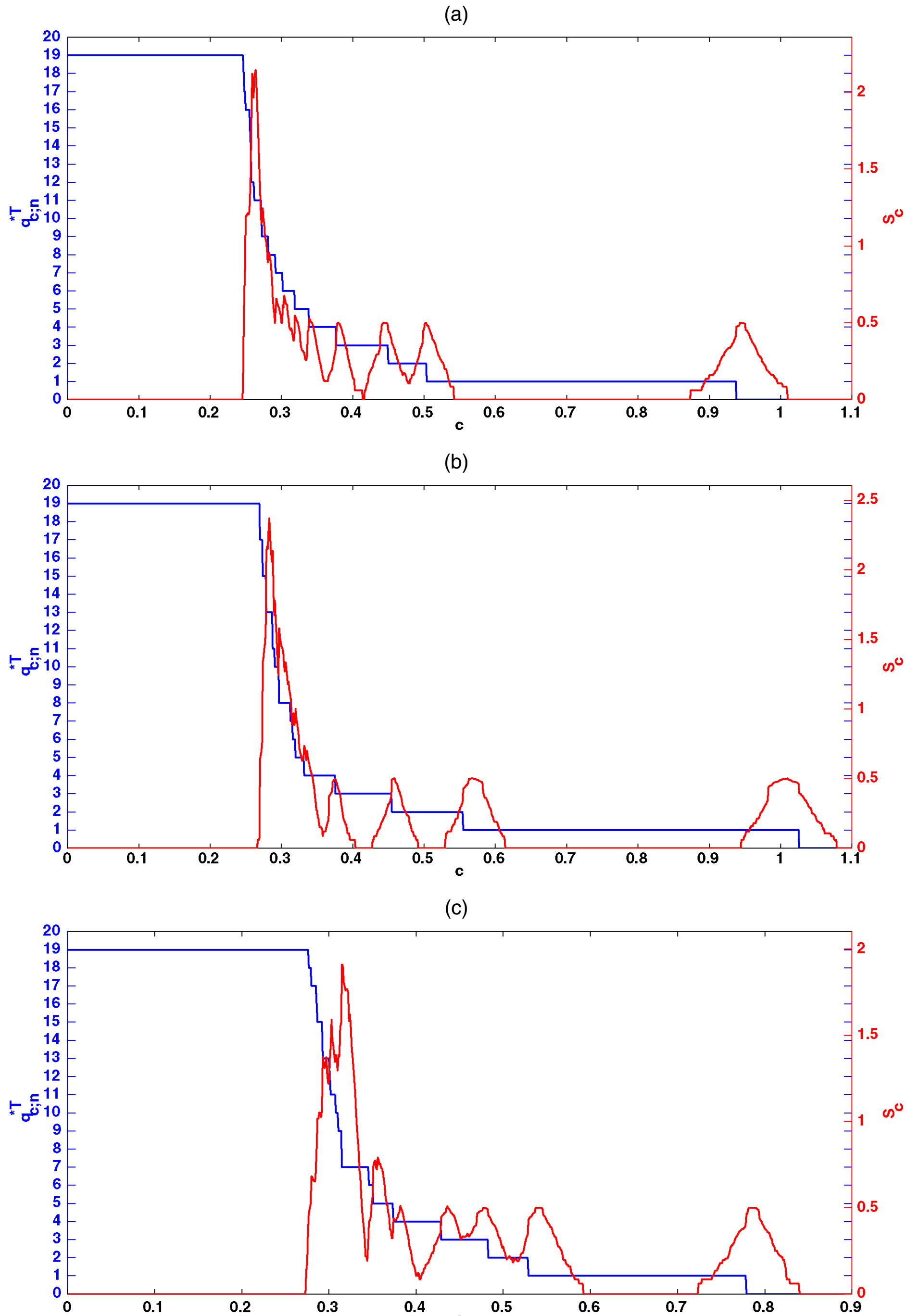

Figure 5. An Analysis of the U.S. Economy Dataset (1960–2003). Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^T_{c,n}$, using penalty function p_1, $q_{max} = 19$, and $IC^T_{2;n}$ criterion, over the periods (a) 1960–2003, (b) 1960–1982, and (c) 1983–2003.

$\max_j |\lambda_j(\mathbf{B}) - \lambda_j(\mathbf{A})|^2 \le \mathrm{tr}((\mathbf{B}-\mathbf{A})(\mathbf{B}-\mathbf{A})')$. It follows that for all n, T, and k,

$$|\lambda^T_{nk}(\boldsymbol{\zeta}) - \lambda_{nk}(\boldsymbol{\zeta})|^2 \le \mathrm{tr}\big((\boldsymbol{\zeta}^T_n - \boldsymbol{\zeta}_n)(\boldsymbol{\zeta}^T_n - \boldsymbol{\zeta}_n)'\big) = \sum_{i=1}^n \sum_{j=1}^n |\zeta^T_{n,ij} - \zeta_{ij}|^2.$$

Taking expectations, we thus have, in view of (A.1),

$$\mathrm{E}\big[|\lambda^T_{nk}(\boldsymbol{\zeta}) - \lambda_{nk}(\boldsymbol{\zeta})|^2\big] \le \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}[|\zeta^T_{n,ij} - \zeta_{ij}|^2] \le n^2 K M_T^{-1}$$

for all n, T, and k. The Markov inequality completes the proof.

Proof of Corollary A.1. From (5), there exist constants L_1 and L_2 such that $\boldsymbol{\Sigma}^T_n(\theta)$ and $\boldsymbol{\Sigma}_n(\theta)$ for all θ satisfy (A.1) in Lemma A.1, with a constant $K = \max[L_1, L_2]$ and a rate $M_T = \max[M_T T^{-1}, M_T^{-4}]$ that do not depend on θ. The corollary follows.

We now turn to the proof of Proposition 2.

Proof of Proposition 2

We have to prove that $\mathrm{P}[IC^T_{a;n}(q) < IC^T_{a;n}(k)] \to 1$ for all $k \neq q$, $k \le q_{\max}$, $a = 1, 2$, as $\min(n, T) \to \infty$ in such a way that (9) holds.

Let $V_n^T(k) := \sum_{i=k+1}^n \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \lambda_{ni}^T(\theta_l)/n$. For all $k < q$,

$$IC_{1;n}^T(q) < IC_{1;n}^T(k) \tag{A.3}$$

if and only if

$$\sum_{i=k+1}^{q} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \frac{\lambda_{ni}^T(\theta_l)}{n} > (q-k)p(n,T), \tag{A.4}$$

that is, in view of Corollary A.1, if and only if

$$\sum_{i=k+1}^{q} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{1n}(T)\right] > (q-k)p(n,T), \tag{A.5}$$

where $K_{1n}(T)$ is $O_P(\max[M_T^{-2}, M_T^{1/2}T^{-1/2}])$ uniformly in n and θ. By Assumption A4, the first q eigenvalues $\lambda_{ni}(\theta)$ diverge linearly in n, which implies that $\sup_\theta \frac{\lambda_{ni}(\theta)}{n} = O(1)$ and $\liminf_{n\to\infty} \sup_\theta \frac{\lambda_{ni}(\theta)}{n} > 0$, for $i = k+1, \dots, q$. Because $K_{1n}(T)$ converges to 0, a sufficient condition for (A.3) to hold with probability tending to 1 as $\min(n,T) \to \infty$ is that $p(n,T) \to 0$ as $\min(n,T) \to \infty$.

Similarly, for the logarithmic version of the criterion,

$$IC_{2;n}^T(q) < IC_{2;n}^T(k) \tag{A.6}$$

for $k < q$ if and only if [note that under Assumption A4(b), $V_n^T(q) > 0$]

$$\log[V_n^T(k)/V_n^T(q)] > (q-k)p(n,T). \tag{A.7}$$

In view of Corollary A.1, we have, for $k \le q$,

$$V_n^T(k) = \sum_{i=k+1}^{n} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{2n}(T)\right], \tag{A.8}$$

where $K_{2n}(T)$ is $O_P(\max[M_T^{-2}, M_T^{1/2}T^{-1/2}])$ uniformly in n and θ. By Assumption A4, the eigenvalues $\lambda_{ni}(\theta)$, $i > q$, are, uniformly in n and θ in $[-\pi, \pi]$, bounded and bounded away from 0. Thus there exist positive constants c_0 and c_1 such that $\mathrm{P}[c_0 > V_n^T(q) > c_1] \to 1$ as $\min(n,T) \to \infty$. For $k < q$, we have

$$V_n^T(k) - V_n^T(q) = \sum_{i=k+1}^{q} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{3n}(T)\right], \tag{A.9}$$

where $K_{3n}(T)$ is $O_P(\max[M_T^{-2}, M_T^{1/2}T^{-1/2}])$ uniformly in n and θ in $[-\pi, \pi]$. As (A.9) coincides with the left side of inequality (A.5), there exists a positive constant c_2 such that $\mathrm{P}[V_n^T(k) - V_n^T(q) > c_2] \to 1$, and hence a constant $c_3 > 0$ such that

$$\mathrm{P}\big[\log\big[\big(V_n^T(k) - V_n^T(q)\big)/V_n^T(q) + 1\big] > c_3\big] = \mathrm{P}\big[\log[V_n^T(k)/V_n^T(q)] > c_3\big] \to 1$$

as $\min(n,T) \to \infty$. The same condition $p(n,T) \to 0$ is thus sufficient for both (A.4) and (A.7) to hold with probability tending to 1 as $\min(n,T) \to \infty$.

Next, for any $k > q$, (A.3) holds if and only if

$$\sum_{i=q+1}^{k} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \frac{\lambda_{ni}^T(\theta_l)}{n} < (k-q)p(n,T),$$

that is, in view of Corollary A.1, if and only if

$$\sum_{i=q+1}^{k} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{4n}(T)\right] < (k-q)p(n,T), \tag{A.10}$$

where $K_{4n}(T)$ is $O_P(\max[M_T^{-2}, M_T^{1/2}T^{-1/2}])$ uniformly in n and θ. As $\lambda_{nq+1}(\theta)$, $\lambda_{nq+2}(\theta), \dots$ are bounded uniformly in n and θ, $\sup_\theta \frac{\lambda_{ni}(\theta)}{n} = O(n^{-1})$ as $n \to \infty$ for $i = q+1, \dots, k$. Thus, when $k > q$, it is sufficient for inequality (A.3) to hold with probability arbitrarily close to 1 as $\min(n,T) \to \infty$, that

$$np(n,T) \to \infty \quad \text{and} \quad \min\big[M_T^2, M_T^{-1/2}T^{1/2}\big]p(n,T) \to \infty \quad \text{as } \min(n,T) \to \infty. \tag{A.11}$$

Turning to the logarithmic criterion, (A.6) holds for $k > q$ if and only if

$$\log[V_n^T(q)/V_n^T(k)] < (k-q)p(n,T). \tag{A.12}$$

By the same arguments as in (A.8), there exist positive constants c_4 and c_5 such that $\mathrm{P}[c_4 < V_n^T(k) < c_5] \to 1$ as $\min(n,T) \to \infty$. Similarly,

$$V_n^T(q) - V_n^T(k) = \sum_{i=q+1}^{k} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{5n}(T)\right],$$

where $K_{5n}(T)$ is $O_P(\max[M_T^{-2}, M_T^{1/2}T^{-1/2}])$ uniformly in n and θ. This term coincides with the left side of (A.10), and the same arguments imply that $V_n^T(q) - V_n^T(k)$, and hence $(V_n^T(q) - V_n^T(k))/V_n^T(k)$, are $O_P(\max[n^{-1}, M_T^{-2}, M_T^{1/2}T^{-1/2}])$ as $\min(n,T) \to \infty$. Therefore, $\log[(V_n^T(q) - V_n^T(k))/V_n^T(k) + 1] = \log[V_n^T(q)/V_n^T(k)]$, which, as $\min(n,T) \to \infty$, is $O_P(\max[n^{-1}, M_T^{-2}, M_T^{1/2}T^{-1/2}])$, so that (A.12), under (A.11), holds with probability arbitrarily close to 1 as $\min(n,T) \to \infty$. This completes the proof.

[Received September 2005. Revised July 2006.]

REFERENCES

Aiolfi, M., Catão, L., and Timmerman, A. (2005), "Common Factors in Latin America's Business Cycle," working paper, International Monetary Fund, Research Dept.

Alessi, L., Barigozzi, M., and Capasso, M. (2006), "Dynamic Factor GARCH: Multivariate Volatility Forecast for a Large Number of Series," LEM Working Paper 2006/25, Sant'Anna School of Advanced Studies, Pisa, Italy.

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2001), "A Real Time Coincident Indicator of the Euro Area Business Cycle," Discussion Paper 3108, Center for Economic Policy Research.

Amengual, D., and Watson, M. W. (2006), "Consistent Estimation of the Number of Dynamic Factors in a Large *N* and *T* Panel," *Journal of Business & Economic Statistics*, to appear.

Angelini, E., Henry, J., and Mestre, R. (2001), "Diffusion Index–Based Inflation Forecasts for the Euro Area," Working Paper 61, European Central Bank.

Artis, M., Banerjee, A., and Marcellino, M. (2002), "Factor Forecasts for the U.K.," Discussion Paper 3119, Center for Economic Policy Research.

Bai, J. (2004), "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panell Data," *Journal of Economics*, 122, 137–183.

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

——— (2005), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, to appear.

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

Boivin, J., and Ng, S. (2005), "Understanding and Comparing Factor-Based Forecasts," *International Journal of Central Banking*, 1, 117–151.

Brillinger, D. R. (1981), *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day.

Bruneau, C., De Bandt, A., Flageollet, A., and Michaux, E. (2003), "Forecasting Inflation Using Economic Indicators: The Case of France," Working Paper 101, Banque de France.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 70, 191–221.

d'Agostino, A., and Giannone, D. (2006), "Comparing Alternative Predictors Based on Large-Panel Dynamic Factor Models," Working Paper 600, European Central Bank.

den Reijer, A. (2005), "Forecasting Dutch GDP Using Large Scale Factor Models," DNB Working Paper 28/2005, Bank of the Netherlands.

Dreger, C., and Schumacher, C. (2004), “Estimating Large Scale Factor Models for Economic Activity in Germany,” in *Jahrbücher für Nationalökonomie und Statistik*, ed. A. Wagner, Stuttgart: Lucius & Lucius Verlagsgesellschaft, pp. 731–750.

Favero, C., Marcellino, M., and Neglia, F. (2005), “Principal Components at Work: Empirical Analysis of Monetary Policy With Large Data Sets,” *Journal of Applied Econometrics*, 20, 603–620.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2005a), “Opening the Black Box: Identifying Shocks and Propagation Mechanisms in VAR and Factor Models,” IAP Technical Report TR 0482, Catholic University of Louvain.

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.

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

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

——— (2005b), “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 Factor Model: Representation Theory,” *Econometric Theory*, 17, 1113–1141.

Forni, M., and Reichlin, L. (1998), “Let’s Get Real: A Factor-Analytical Approach to Disaggregated Business Cycle Dynamics,” *Review of Economic Studies*, 65, 453–473.

Geweke, J. (1977), “The Dynamic Factor Analysis of Economic Time Series,” in *Latent Variables in Socio-Economic Models*, eds. D. J. Aigner and A. S. Goldberg, Amsterdam: North-Holland, pp. 365–383.

Giannone, D. (2004), “Essays on Dynamic Factor Models of Large Cross-Sections,” unpublished doctoral dissertation, ISRO, Free University of Brussels.

Giannone, D., and Lenza, M. (2004), “The Feldstein–Horioka Fact,” Discussion Paper 4610, Center for Economic Policy Research.

Giannone, D., and Reichlin, L. (2006), “Does Information Help Recovering Structural Shocks From Part Observations?” *Journal of the European Economic Association*, 4, 455–465.

Giannone, D., Reichlin, L., and Sala, L. (2005), “Monetary Policy in Real Time,” in *NBER Macroeconomic Annual 2004*, eds. M. Gertler and K. Rogoff, Cambridge, MA: MIT Press, pp. 161–200.

——— (2006), “VARs, Factor Models and the Empirical Validation of Equilibrium Business Cycle Models,” *Journal of Econometrics*, 127, 257–279.

Ingersol, J. (1984), “Some Results in the Theory of Arbitrage Pricing,” *Journal of Finance*, 39, 1021–1039.

Kapetanios, G. (2005), “A New Method for Determining the Number of Factors in Factor Models With Large Datasets,” Working Paper 525, Queen Mary University of London.

Lewbel, A. (1991), “The Rank of Demand Systems: Theory and Nonparametric Estimation,” *Econometrica*, 59, 711–730.

Marcellino, M., Stock, J., and Watson, M. (2003), “Macroeconomic Forecasting in the Euro Area: Country-Specific versus Area-Wide Information,” *European Economic Review*, 47, 1–18.

Nieuwenhuyzen, C. (2004), “A Generalized Dynamic Factor Model for the Belgian Economy,” *Journal of Business Cycle Measurement and Analysis*, 2, 213–248.

Onatski, A. (2005), “Determining the Number of Factors From Empirical Distribution of Eigenvalues,” preprint, Columbia University, Economics Dept.

——— (2006), “A Formal Statistical Test for the Number of Factors in Approximate Factor Models,” preprint, Columbia University, Economics Dept.

Parzen, E. (1957), “On Consistent Estimates of the Spectrum of a Stationary Time Series,” *The Annals of Mathematical Statistics*, 28, 329–348.

Sargent, T. J., and Sims, C. A. (1977), “Business Cycle Modelling Without Pretending to Have Too Much *a priori* Economic Theory,” in *New Methods in Business Research*, ed. C. A. Sims, Minneapolis: Federal Reserve Bank of Minneapolis, pp. 45–109.

Schneider, M., and Spitzer, M. (2004), “Forecasting Austrian GDP Using the Generalized Dynamic Factor Model,” Working Paper 89, Österreichische Nationalbank.

Stewart, G. W., and Sun, J. (1990), *Matrix Perturbation Theory*, San Diego: Academic Press.

Stock, J. H., and Watson, M. W. (1989), “New Indices of Coincident and Leading Indicators,” in *NBER Macroeconomics Annual 1989*, eds. O. J. Blanchard and S. Fischer, Cambridge, MA: MIT Press, pp. 351–393.

——— (2002a), “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business & Economic Statistics*, 20, 147–162.

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

——— (2005), “Implications of Dynamic Factor Models for VAR Analysis,” Working Paper 11467, National Bureau of Economic Research.

Weyl, H. (1912), “Das Asymptotische Verteilungsgesetz der Eigenwerte Linearer Partieller Differentialgleichungen,” *Mathematische Annalen*, 71, 441–479.