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 $i = 1, \ldots, n$ represents the cross-sectional index and $t = 1, \ldots, T$ denotes time) is decomposed into the sum $X_{it} = \chi_{it} + \xi_{it}$, of two nonobservable mutually orthogonal (at all leads and lags) variables, the common component, $\chi_{it}$, and the idiosyncratic component, $\xi_{it}$.

Under the general dynamic factor approach considered in this article, the common component $\chi_{it}$ 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 $b_{ij}(L)$, $j = 1, \ldots, q$ ($L$, 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 $\chi_{it}$ is expressed as a linear combination $\sum_{j=1}^{r} a_{ij}F_{jt}$ of a small number $r$ of unobserved static factors $(F_{1t}, \ldots, F_{rt})$. 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 $(F_{1t}, \ldots, F_{rt})$ are driven by a number, $q \leq r$, 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 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 2005). 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 is usually scattered, and large number $n$ of interrelated time series (at many 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 2005a,b), in the identification in the economy-wide and global shocks (Forni et al., 2005a), 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. 2004), in consumer theory (Lewbel 1991), and in monetary policy applications (Bernanke and Boivin 2003; Favero and Marcellino 2001). In finance, factor models 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, Flagelollet, 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), Angeline, 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.

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. 2005b; 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 criterion; 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 2005) 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 (2005), still for the number \( r \) of static factors but in a model with iid idiosyncrasies.

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 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 generalized 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.

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 \( \theta \) if \( \sup_{n \in \mathbb{N}} \sup_{\theta \in [-\pi, \pi]} |\zeta(n, T, \theta)| = o(1) [\text{resp. } O(1)] \) 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 \( \theta \) if for all \( \varepsilon > 0 \) and \( \eta > 0 \), there exists a \( T_{\varepsilon, \eta} \) such that \( \sup_{n \in \mathbb{N}} \sup_{\theta \in [-\pi, \pi]} P[|\zeta(n, T, \theta)| > \eta] < \varepsilon \) for all \( T > T_{\varepsilon, \eta} \) [resp. for all \( \varepsilon > 0 \) there exist \( B_{\varepsilon, T} > 0 \) and \( T_{\varepsilon, \eta} \) such that \( \sup_{n \in \mathbb{N}} \sup_{\theta \in [-\pi, \pi]} P[|\zeta(n, T, \theta)| > B_{\varepsilon, T}] < \varepsilon \) for all \( T > T_{\varepsilon, \eta} \)]. 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

\[ X_{it} = \chi_{it} + \xi_{it}, \]

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

and Assumptions A1–A4 are assumed to hold.

**Assumption A1.** (a) The \( q \)-dimensional vector process \( \{u_{it} := (u_{1t}, u_{2t}, \ldots, u_{qt}); t \in \mathbb{Z}\} \) is orthonormal white noise.

(b) The \( n \)-dimensional processes \( \{\xi_{it} := (\xi_{1it}, \xi_{2it}, \ldots, \xi_{nit}); t \in \mathbb{Z}\} \) are mean-zero stationary for any \( n \); moreover, \( \xi_{it} \perp u_{jt} \) for any \( i, j, t \), 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 \( \xi_{it} \) and \( \chi_{it} \) are the idiosyncratic and common components of \( X_{it} \).

**Assumption A2.** For all \( n \), the vector process \( \{X_{nt} := (X_{1nt}, X_{2nt}, \ldots, X_{qnt}); t \in \mathbb{Z}\} \) is a linear process with a representation of the form \( X_{nt} = \sum_{k=-\infty}^{\infty} C_{it}Z_{t-k} \), where \( Z_{t} \) is full-rank \( n \)-dimensional white noise with finite fourth-order cumulants, and the \( n \times n \) matrices \( C_{it} = (C_{ijk}) \) are such that \( \sum_{k=-\infty}^{\infty} |C_{ijk}|^{1/2} < \infty \) for all \( 1 \leq i, j \leq n \).

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

\[ c_{i_{1}, \ldots, i_{L}}(k_{1}, \ldots, k_{L-1}) := \sum_{l=1}^{L} X_{i_{l}(t+k_{l-1})} \]

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

**Assumption A2’.** This is the same as Assumption A2, but the convergence condition on the \( C_{ij} \)’s is uniform: \( \sup_{i, j \in \mathbb{N}} \sum_{k=-\infty}^{\infty} |C_{ij}(k)|^{1/2} < \infty \). Moreover, for all \( 1 \leq \ell \leq 4 \) and \( 1 \leq j < \ell \), \( \sup_{i_{1}, \ldots, i_{\ell}} |\sum_{k=-\infty}^{\infty} \cdots \sum_{k_{\ell-1}=-\infty}^{\infty} c_{i_{1}, \ldots, i_{\ell}}(k_{1}, \ldots, k_{\ell-1})| < \infty \).
Denote by $\Sigma_n(\theta)$, $\theta \in [-\pi, \pi]$, the spectral density matrix of $X_{nt}$, with elements $\sigma_j(\theta)$, and by $\lambda_1(\theta), \ldots, \lambda_m(\theta)$ its eigenvalues in decreasing order of magnitude. Similarly, let $\lambda^X_j(\theta)$ and $\lambda^X_{nj}(\theta)$ be the eigenvalues associated with the spectral densities $\Sigma^X_n(\theta)$ and $\Sigma^X_{nj}(\theta)$ of $X_{nt} := (x_{1t}, \ldots, x_{nt})'$ and $\xi_{nt} := (\xi_{1t}, \ldots, \xi_{nt})'$. Such eigenvalues [actually, the functions $\theta \mapsto \lambda(\theta)$] are called dynamic eigenvalues.

Assumption A3. (a) The first idiosyncratic dynamic eigenvalue $\lambda^X_1(\theta)$ is uniformly (with respect to $\theta \in [-\pi, \pi]$) bounded as $n \to \infty$; that is, $\sup_{\theta \in [-\pi, \pi]} \lambda^X_1(\theta) < \infty$ as $n \to \infty$.

(b) The qth common dynamic eigenvalue $\lambda^X_q(\theta)$ diverges for all $\theta \in [-\pi, \pi]$ 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 $X_{nt}$ and $\xi_{nt}$. 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 $\Sigma_n(\theta)$ diverges as $n \to \infty$, for all $\theta$ 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 $X_{it}$ and the idiosyncratic components $\xi_{it}$ 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 $\Sigma^{(n)}(\theta)$ or an adequate estimate of the same. Thus determining $q$ prior to this projection step thus is absolutely crucial.

3. AN INFORMATION CRITERION

3.1 Population Level

In practice, only finite segments of length $T$ of a finite number 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 $\Sigma_n(\theta)$ are known.

In this setting, the projection $\hat{X}^{(n)}_{it}$ of $X_{it}$ onto the space spanned by the $\Sigma^{(n)}(\theta)$'s first $q$ dynamic principal components provides a consistent reconstruction of $X_{it}$. For given $q$, those dynamic principal components, and hence the $\chi^{(n)}_{it}$'s, can be viewed as solutions of an optimization problem in which the “expected mean of squared residuals,” $n^{-1}E[\sum_{t=1}^{n} (X_{it} - \hat{X}^{(n)}_{it})^2]$, is minimized, yielding a minimum of $n^{-1} \times \sum_{j=k+1}^{n-q} \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta$. Thus minimum plays the role of the residual variance classically appearing, in empirical form, in information criterion methods. Accordingly, denoting by $q_{\text{max}}$ some predefined upper bound and by $p(n)$ some adequate penalty, we propose selecting the number of factors as

$$\hat{q}_n := \arg \min_{0 \leq k \leq q_{\text{max}}} IC_{0,n}(k),$$

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

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

The intuition behind (2) is clear: For the bounded eigenvalues ($k > q$), the averaged contribution, $\frac{1}{n} \sum_{j=k+1}^{n} \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta$, 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 $\Sigma_n(\theta)$ diverges at least linearly in $n$, that is, $\liminf_{n \to \infty} \inf_{\theta \in \mathbb{R}} n^{-1} \times \lambda_n(\theta) > 0$.

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

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

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

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 $\frac{1}{n} \times \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta$ 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 $c \propto (n/p(n))$, 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 $\Sigma_n(\theta)$ 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 $\mathbf{\Gamma}_{n,\ell}$ the sample cross-covariance matrix of $\mathbf{X}_n$ and $\mathbf{X}_{n,\ell-n}$ based on $T$ observations, the lag window estimator of $\Sigma_n(\theta)$ is defined as

$$\Sigma_T^n(\theta) := \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \sum_{u=-M_T}^{M_T} w(M_T^{-1} u) \mathbf{\Gamma}_{n,\ell} e^{-i\theta\ell}, \quad (4)$$

where $\theta \mapsto w(\theta)$ 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 $M_T$.

\textbf{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 $w(0) = 1$, $|w'(\alpha)| \leq 1$ for all $\alpha$ 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.

\textbf{Assumption B2.} The entries $\sigma_{ij}(\theta)$ of $\Sigma_n(\theta)$ are uniformly (in $n$ and $\theta$) bounded and have uniformly (in $n$ and $\theta$) bounded derivatives up to order two; namely, there exists $Q < \infty$ such that

$$\sup_{i,j \in [n]} \sup_{\theta} \left| \frac{d^k}{d\theta^k} \sigma_{ij}(\theta) \right| \leq Q, \quad k = 0, 1, 2.$$

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

$$\sup_{\theta} \sup_{n} \left( \frac{1}{n} \sum_{i=1}^{n} \left| \Sigma_T^n(\theta) - \Sigma_n(\theta) \right|_{ij}^2 \right) \leq L_4 M_T T^{-1} + L_2 M_T^{-4} \quad \text{for any } T > T_0. \quad (5)$$

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 $\Sigma_T^n(\theta)$, consider the information criterion

$$\text{IC}_{1,n}(k) := \sum_{i=k+1}^{n} \frac{1}{2M_T + 1} \sum_{\ell=-\infty}^{\infty} \lambda_{n,\ell}(\theta) + kp(n,T), \quad 0 \leq k \leq q_{\text{max}}. \quad (6)$$

where $p(n,T)$ is a penalty function, $\theta := \pi l/(M_T + 1/2)$ for $l = -M_T, \ldots, M_T$, and $q_{\text{max}}$ is some predetermined upper bound; the eigenvalues $\lambda_{n,\ell}(\theta)$ are those of $\Sigma_T^n(\theta)$.

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

$$\text{IC}_{3,n}(k) := \log \left[ \sum_{i=k+1}^{n} \frac{1}{2M_T + 1} \sum_{\ell=-\infty}^{\infty} \lambda_{n,\ell}(\theta) \right] + kp(n,T), \quad 0 \leq k \leq q_{\text{max}}. \quad (7)$$

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

$$q_{1,n}^T := \arg \min_{0 \leq k \leq q_{\text{max}}} \text{IC}_{1,n}(k), \quad a = 1, 2. \quad (8)$$

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

\textbf{Proposition 2.} Let Assumptions A1, A2', A3, A4, B1, and B2 hold. Then $P(q_{1,n}^T = q) \to 1$ for $a = 1, 2$ provided that $n$ and $T$ both tend to infinity, in such a way that

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

\textbf{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 (2005).

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

$$\Sigma_{n}^{T}(\theta) := \frac{2\pi}{T} \sum_{t=1}^{T-1} W(T)(\theta - \frac{2\pi t}{T}) \mathbf{I}_{n}^{T}(\theta) \left( \pi \frac{2\pi t}{T} \right), \quad (10)$$

where

$$\mathbf{I}_{n}^{T}(\alpha) := \frac{1}{2\pi T} \sum_{t=1}^{T-1} \sum_{i=1}^{n} \mathbf{X}_{n,i} e^{-i\alpha t} \left[ \sum_{s=1}^{t-1} \mathbf{X}_{n,s} e^{i\alpha t} \right]$$

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 $\int_{-\infty}^{\infty} |\alpha|^2 W(\alpha) d\alpha = 1$ 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 $l = 1, \ldots, T-1$)

$$\text{IC}_{1,n}^{T}(k) := \sum_{i=k+1}^{n} \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^{n} \lambda_{n,\ell}(\theta_l) + kp(n,T), \quad 0 \leq k \leq k_{\text{max}} < \infty, \quad (11)$$

where $p(n,T)$ is a penalty. Reinforce the condition on cumulants in Assumption A2' into

$$\sup_{i_1, \ldots, i_{k-1}} \sum_{k=0}^{\infty} \sum_{k=1}^{\infty} \cdots \sum_{k=1}^{\infty} (1 + |k|) \left| c_{i_1, \ldots, i_{k-1}, k_{l-1}} \right| < \infty, \quad (12)$$

and let $q_{1,n}^{T} := \arg \min_{0 \leq k \leq q_{\text{max}}} \text{IC}_{1,n}(k)$. The following counter-part of Proposition 2 holds.

\textbf{Proposition 3.} Under the assumptions made previously, $P(q_{1,n}^{T} = q) \to 1$ provided that $n$ and $T$ tend to infinity in such a way that (a) $p(n,T) \to 0$ and (ii) $\min[n, B_T^{-2}, B_T^{-1/2} T^{1/2}] \times p(n,T) \to \infty$.\]
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 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 (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 $(n, T) \mapsto q_{c,1:n}^T$ is obtained from $J$-tuples of the form $q_{c,a,n}^T$, $a = 1, 2, j = 1, \ldots, J$, where $0 < n_1 < \cdots < n_j = n$ and $0 < T_1 \leq \cdots \leq T_j = T$. For any fixed value of $(n_j, T_j)$, $c \mapsto q_{c,a,n}^T$ is clearly nonincreasing; thus for given $a$, the curves $(n_j, T_j) \mapsto q_{c,a,n}^T$ 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 $j \mapsto q_{c,n}^T$ is increasing and would tend to infinity if $n$ would do so. If $c > 0$ is “very small” (i.e., severe underpenalization), then, although Proposition 3 applies, the situation for finite $(n, T)$ will not be very different; $j \mapsto q_{c,n}^T$ is still increasing 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 $j \mapsto q_{c,n}^T$ 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 the variability among the $J$ values of $q_{c,n}^T$, $j = 1, \ldots, J$, is high; this variability is captured by, for instance,

$$S_c^2 := J^{-1} \sum_{j=1}^J \left( q_{c,n}^T - J^{-1} \sum_{j=1}^J q_{c,n}^T \right)^2. \quad (10)$$

Next consider a “very large” value of $c$, and hence severely overpenalized $q_{c,n}^T$’s. If $c$ is large enough, then $q_{c,n}^T$ is identically 0 for all $(n_j, T_j)$’s, and 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 $c \mapsto q_{c,n}^T$, somewhere between those “small” underpenalizing values of $c$ (with $j \mapsto q_{c,n}^T$ eventually tending to $q$ from above) and the “large” overpenalizing values (with $j \mapsto q_{c,n}^T$ tending to $q$ from below), a range of “moderate” values of $c$, yielding stable behavior of $j \mapsto q_{c,n}^T \approx q$, typically exists. This stability can be assessed, for given $c$, through the empirical variance (10) of the $q_{c,n}^T$’s, $j = 1, \ldots, J$.

As an illustration, we consider two examples, with $n = T = 300$ and $q = 3$. In both cases, our method has been applied with $M_T = [\sqrt{75}, T]$, penalty $p_1(n, T) = (M_T^2 + M_T^{-1} - T^{-1} - n^{-1}) \log(\min(n, M_T, M_T^{-1} - T^{-1}/2))$, and criteria $I_{c,1:n}^T$ and $I_{c,2:n}^T$ for $(n_j, T_j) = (120, 120), (130, 130), \ldots, (300, 300)$, and $c \in \mathcal{C} = \{0.001, 0.002, \ldots, 2.000\}$.

**Example 1.** The common part was modeled with ?? (MA) loadings; see Section 5.1 for details. The graphs of $(n_j, T_j) \mapsto q_{c,n}^T$ and $c \mapsto S_c$ are presented in Figures 1(a1) and 1(a2) for criterion $I_{c,1:n}^T$ and in Figures 1(b1) and 1(b2) for $I_{c,2:n}^T$. 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, 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,” $[0, 0.005], [0.122, .188], [0.210, .351], [0.388, .547]$, and $[0.579, 2.000]$, corresponding to a selection of $q = q_{max} = 19, 3, 2, 1, 0$ 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 $q_{0.10,n}^T$ converges, as $j$ increases, to $q_{0.15,n}^T$ from below, whereas $q_{20.0}^T$ converges to $q_{15.0}^T$ from below, and that $\hat{c} = .15$ is the only value of $c$ exhibiting such pattern. Similar comments can be made for $I_{c,2:n}^T$ 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, $[0.122, .188]$ in Figure 1(a2) and $[0.346, .385]$ 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; hence the maximal possible value $\hat{c} = q_{max}$.

This example suggests that irrespective of the choice of $I_{c,1:n}^T$ or $I_{c,2:n}^T$, the selection of $q$ can be based on an inspection of the family of curves $(n_j, T_j) \mapsto q_{c,n}^T$ indexed by $c \in \mathcal{C}$, trying to spot the value of $c$ (the curve) with “neighbors” $c \pm \delta (\delta > 0)$ such that $q_{c+\delta}^T \uparrow q_{c,n}^T$ whereas $q_{c-\delta}^T \downarrow q_{c,n}^T$ 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 $\hat{c} = q_{c,n}^T$ 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.

**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 $c \mapsto q_{c,n}^T$ plots in Figure 3. For the $I_{c,2:n}^T$ criterion, the stability intervals Figure 1(d1) are $[0.007], [0.164, .284], [0.336, .464], [0.512, .765]$, and $[0.867, 2.000]$, yielding $q_{c,n}^T = 19, 3$ (correct identification), 2, 1, and 0. The situation is even clearer with $I_{c,2:n}^T$ in Figure 1(d2), with stability intervals $[0, .242], [0.509, .591], [0.729, .732], [0.878, .894]$, and $[1.03, 2.000]$, yielding $q_{c,n}^T = 19, 3$ (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 $(n_j, T_j) \mapsto q^T_{c; n_j}$ 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 $q^T_{c, n} = 0$.

Figure 1. Example 1: MA Loadings, $q = 3$, $n = T = 300$, $M_T = [1.75, \sqrt{T}]$. Graphs of $(n_j, T_j) \mapsto q^T_{c; n_j}$ 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 $p_1$, $q_{\text{max}} = 19$, and [(a1) and (a2)] $IC_{1,n}^T$ criterion, [(b1) and (b2)] $IC_{2,n}^T$ criterion.

This existence of “stability intervals” in the $s \mapsto S_c$ graphs, and their relation to $q_{\text{max}}, q, q - 1, \ldots, 1$, is an empirical finding, but it can be explained as follows. For simplicity, consider the “population-level problem” of Section 3.1. There, the selected number of factors $\hat{q} = \hat{q}^n_c$ associated with penalty $cp(n)$ is characterized by the fact that

$$\begin{align*}
\frac{1}{(k - \hat{q})} \sum_{j=\hat{q}+1}^{k} \int \lambda^j_{c}(\theta) \, d\theta &< cnp(n), \\
\frac{1}{(\hat{q} - k)} \sum_{j=k+1}^{\hat{q}} \int \lambda^j_{c}(\theta) \, d\theta &> cnp(n),
\end{align*}$$

$$k = \hat{q} + 1, \hat{q} + 2, \ldots \ (11)$$
which, noting that both \( k \mapsto \frac{1}{(q-k)} \sum_{j=k+1}^{q} \int \lambda_{j}^{n}(\theta) \, d\theta \) and \( k \mapsto \frac{1}{n} \sum_{j=q+1}^{n} \int \lambda_{j}^{n}(\theta) \, d\theta \) are decreasing, reduces to \( (k = \hat{q} \pm 1) \),

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

Thus, once \( c \) is chosen, the criterion identifies the number of factors as the (unique, in view of monotonocity in \( c \)) \( \hat{q}_{c} \) such that \( cp(n) \) “separates” \( \frac{1}{n} \int \lambda_{\hat{q}+1}^{n}(\theta) \, d\theta \) and \( \frac{1}{n} \int \lambda_{\hat{q}}^{n}(\theta) \, d\theta \). Figure 4 illustrates this finding. The red lines are plots of \( n \mapsto \frac{1}{n} \int \lambda_{q}^{n}(\theta) \, d\theta \), roughly yielding [in the least favorable case of \( O(n) \) diverging eigenvalues] constants for \( 1 \leq k \leq q \) and hyperbolically 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, and on that line; \( \hat{q}_{c} \) 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:
5.1 Simulations

Also note that with the corresponding \( q_T \)'s. Note that \( S_{c1} = S_{c3} = 0 \), whereas \( S_{c2}, S_{c4}, \ldots, S_{c7} \) are strictly positive.

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 \( q_{\text{max}} \).

2. Choose \( T \mapsto M_T \) (e.g., \( M_T := \lfloor 0.5 \sqrt{T} \rfloor \) or \( M_T := \lfloor 0.7 \sqrt{T} \rfloor \)) and a smoothing function \( w(a) \) such that Assumptions B1(a) and (b) be satisfied.

3. Choose a penalty, \((n, T) \mapsto p(n, t)\), and a criterion \([IC_1(n, k) \text{ or } IC_2(n, k)]\), and define \( p^*_c(n, t) = cp(n, t) \) for \( c \in C \subset [C^-, C^+] \subset \mathbb{R}^+ \) (e.g., \( C := \{0.01, 0.02, \ldots, 3\} \)).

4. Define sequences \( n_1 < n_2 < \cdots < n_j = n \) and \( T_1 \leq T_2 \leq \cdots < T_J = T \), for example, \((n, T) = (150, 100)\), set \( n_j := 120 + 10j \) and \( T_j := 70 + 10j, j = 1, \ldots, 3 \).

5. Defining \( S_c \) as in (10), identify \( q \) as \( \hat{q} := q_{c, \hat{n}}^T \), 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 \( P[q_{c,n}^T \leq \hat{q}_c^{T, \hat{n}} \leq q_{c,n}^{-T}] = 1 \) 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

To evaluate the performance of the strategy proposed in the previous section, we conducted the following Monte Carlo experiments. 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 \( u_{ik}, i = 1, \ldots, T \), are iid \( \mathcal{N}(0, D_k) \), with \( D_1 = 1, D_2 = 0.5 \), and \( D_3 = 1.5 \).

2. The idiosyncratic components are of the form \( \xi_{it} = \sum_{j=0}^{4} \sum_{k=0}^{2} g_{ij,k} v_{i+j-k}, \) 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 \( g_{ij,k} \)'s, \( i = 1, \ldots, n, j = 1, \ldots, 4, k = 0, 1, 2 \), are iid \( U_{[1,0.15]} \) with the \( v_{it} \)'s and \( g_{ij,k} \)'s mutually independent and independent of the \( u_{it} \)'s; this ensures both autocorrelation and cross-correlation among idiosyncratic.

3. Randomly generate the filters \( b_{ik}(L) \) \((i = 1, \ldots, n, k = 1, \ldots, q)\) (independently from the \( u_{it} \)'s and \( \xi_{it} \)'s) by one of the following two devices:

- MA loadings: \( b_{ik}(L) = b_{ik0} + b_{ik1} L + b_{ik2} L^2 \) with iid and mutually independent coefficients \( b_{ik0}; b_{ik1}; b_{ik2} \sim \mathcal{U}(0, 1) \)
- AR loadings: \( b_{ik} = b_{ik0}(1 - b_{ik1} L)^{-1}(1 - b_{ik2} L)^{-1} \), with iid and mutually independent coefficients \( b_{ik0}; b_{ik1}; b_{ik2} \sim \mathcal{U}(0, 1) \)

For each \( i \), the variance of \( \xi_{it} \) 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 (2005)'s 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 \( n_j := n - 10j, j = 1, \ldots, 3, T_j := T - 10i, i = 1, \ldots, 3 \); penalties

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

Figure 4. Heuristic Behavior of \( c \mapsto S_c \). Graphs of \( n \mapsto \frac{1}{n} \int_0^1 f(\theta) d\theta, k = 1, \ldots, q_{\text{max}} \) (in red), and \( n \mapsto cp(n), c = 0, \ldots, c_{\text{max}} \) (in blue), along with the corresponding \( q_T \)'s. Note that \( S_{c1} = S_{c3} = 0 \), whereas \( S_{c2}, S_{c4}, \ldots, S_{c7} \) are strictly positive.
Hallin and Liška: General Dynamic Factor Model

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

and

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

\[ q_{\text{max}} = 19; \] and \( C = [0.01, 0.02, \ldots, 3]. \) Spectral densities were estimated with a triangular smoothing function \( w(v) = 1 - |v|, M_T = [0.50\sqrt{T}], \) and \( M_T = [0.75\sqrt{T}]. \)

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

\[ p_1^{\text{BN}}(n, T) = \log(nT/(n + T))/nT, \]

\[ p_2^{\text{BN}}(n, T) = \log(\min(n, T))(n + T)/nT, \]

and

\[ p_3^{\text{BN}}(n, T) = \log(\min(n, T))/\min(n, T); \]

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 \( IC_n^{\text{I,n}} \), with \( M_T = [0.50\sqrt{T}] \) and \( p_1 \), yields the best performance (red figures), 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 poorly (Table 3), which is not surprising, because their consistency assumptions are not satisfied. Penalty \( p_3^{\text{BN}} \), with \( \hat{D}_{1,k} \) and \( m = 2, \) seems to be least misleading here (blue figures), but still yields a percentage of incorrect identification (mainly overestimation) uniformly >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 (green figures) are obtained for \( p_2^{\text{BN}}, \) 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.

\begin{table}
\centering
\begin{tabular}{cccccccccccc}
\hline
\textbf{q} & \textbf{n} & \textbf{T} & \multicolumn{3}{c}{\textbf{MA loadings}} & & \multicolumn{3}{c}{\textbf{AR loadings}} \\
\hline
 & & & \multicolumn{3}{c}{\textbf{IC}_n^{\text{I,n}}} & & \multicolumn{3}{c}{\textbf{IC}_n^{\text{I,n}}} \\
 & & & \textbf{M}_T = [0.50\sqrt{T}] & \textbf{M}_T = [0.75\sqrt{T}] & \textbf{M}_T = [0.50\sqrt{T}] & & \textbf{M}_T = [0.75\sqrt{T}] & \textbf{M}_T = [0.50\sqrt{T}] & \textbf{M}_T = [0.75\sqrt{T}] \\
\hline
1 & 60 & 100 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
2 & 100 & 100 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
3 & 70 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
4 & 120 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
5 & 150 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
6 & 200 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
7 & 250 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
8 & 300 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
9 & 350 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
10 & 400 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
11 & 450 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
12 & 500 & 120 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 & 0/0 \\
\hline
\end{tabular}
\end{table}
Table 2. Bai and Ng (2005) Criterion, MA Loadings

<table>
<thead>
<tr>
<th>q</th>
<th>n</th>
<th>T</th>
<th>( D_{1,k}^{p_{1N}} )</th>
<th>( D_{2,k} )</th>
<th>( D_{1,k}^{p_{2N}} )</th>
<th>( D_{2,k} )</th>
<th>( D_{1,k}^{p_{3N}} )</th>
<th>( D_{2,k} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
<td>100</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>100</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>70</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>120</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>150</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>100</td>
<td>14/0</td>
<td>0/2</td>
<td>0/93</td>
<td>4/0</td>
<td>0/72</td>
<td>0/99</td>
</tr>
<tr>
<td>100</td>
<td>1/0</td>
<td>0/2</td>
<td>0/98</td>
<td>0/0</td>
<td>0/84</td>
<td>0/0</td>
<td>0/70</td>
<td>0/24</td>
</tr>
<tr>
<td>70</td>
<td>7/0</td>
<td>0/2</td>
<td>0/93</td>
<td>1/0</td>
<td>0/75</td>
<td>0/0</td>
<td>0/75</td>
<td>0/27</td>
</tr>
<tr>
<td>120</td>
<td>0/0</td>
<td>0/2</td>
<td>0/98</td>
<td>0/0</td>
<td>0/88</td>
<td>0/0</td>
<td>0/72</td>
<td>0/28</td>
</tr>
<tr>
<td>150</td>
<td>0/0</td>
<td>0/0</td>
<td>0/95</td>
<td>0/0</td>
<td>0/69</td>
<td>0/0</td>
<td>0/88</td>
<td>0/49</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>100</td>
<td>90/0</td>
<td>6/11</td>
<td>0/99</td>
<td>52/0</td>
<td>0/89</td>
<td>0/99</td>
</tr>
<tr>
<td>100</td>
<td>65/0</td>
<td>10/0</td>
<td>0/99</td>
<td>28/0</td>
<td>0/93</td>
<td>0/100</td>
<td>0/99</td>
<td>76/0</td>
</tr>
<tr>
<td>70</td>
<td>120</td>
<td>38/0</td>
<td>2/8</td>
<td>0/99</td>
<td>0/0</td>
<td>0/94</td>
<td>0/100</td>
<td>88/0</td>
</tr>
<tr>
<td>120</td>
<td>49/0</td>
<td>0/5</td>
<td>0/99</td>
<td>1/0</td>
<td>0/91</td>
<td>0/0</td>
<td>0/94</td>
<td>0/33</td>
</tr>
<tr>
<td>150</td>
<td>47/0</td>
<td>0/2</td>
<td>0/96</td>
<td>0/0</td>
<td>0/68</td>
<td>0/0</td>
<td>0/12</td>
<td>0/59</td>
</tr>
</tbody>
</table>

(b) correlation

<table>
<thead>
<tr>
<th>q</th>
<th>n</th>
<th>T</th>
<th>( m_{1} )</th>
<th>( m_{2} )</th>
<th>( m_{1.5} )</th>
<th>( m_{2.5} )</th>
<th>( m_{1.5} )</th>
<th>( m_{2.5} )</th>
<th>( m_{1.5} )</th>
<th>( m_{2.5} )</th>
<th>( m_{1.5} )</th>
<th>( m_{2.5} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
<td>100</td>
<td>0/97</td>
<td>0/97</td>
<td>0/96</td>
<td>0/88</td>
<td>0/91</td>
<td>0/85</td>
<td>0/67</td>
<td>0/72</td>
<td>0/62</td>
<td>0/34</td>
</tr>
<tr>
<td>100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>70</td>
<td>0/97</td>
<td>0/98</td>
<td>0/97</td>
<td>0/92</td>
<td>0/93</td>
<td>0/91</td>
<td>0/65</td>
<td>0/70</td>
<td>0/60</td>
<td>0/40</td>
<td>0/49</td>
<td>0/31</td>
</tr>
<tr>
<td>120</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>150</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>100</td>
<td>19/4</td>
<td>0/98</td>
<td>0/62</td>
<td>0/76</td>
<td>0/88</td>
<td>0/51</td>
<td>0/62</td>
<td>0/79</td>
<td>0/30</td>
<td>0/17</td>
</tr>
<tr>
<td>100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>70</td>
<td>0/97</td>
<td>0/99</td>
<td>0/97</td>
<td>0/92</td>
<td>0/92</td>
<td>0/72</td>
<td>0/71</td>
<td>0/85</td>
<td>0/43</td>
<td>0/25</td>
<td>0/39</td>
<td>0/13</td>
</tr>
<tr>
<td>120</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>150</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
<td>0/100</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>100</td>
<td>1/60</td>
<td>0/92</td>
<td>6/13</td>
<td>6/16</td>
<td>3/44</td>
<td>17/3</td>
<td>10/27</td>
<td>4/56</td>
<td>25/8</td>
<td>52/1</td>
</tr>
<tr>
<td>100</td>
<td>0/82</td>
<td>0/97</td>
<td>4/0</td>
<td>0/40</td>
<td>0/69</td>
<td>0/16</td>
<td>4/43</td>
<td>2/67</td>
<td>12/19</td>
<td>26/7</td>
<td>15/17</td>
<td>36/1</td>
</tr>
<tr>
<td>70</td>
<td>0/71</td>
<td>0/97</td>
<td>1/21</td>
<td>1/29</td>
<td>0/62</td>
<td>5/7</td>
<td>5/40</td>
<td>1/73</td>
<td>15/13</td>
<td>29/5</td>
<td>15/15</td>
<td>48/1</td>
</tr>
<tr>
<td>120</td>
<td>0/88</td>
<td>0/99</td>
<td>0/53</td>
<td>0/48</td>
<td>0/77</td>
<td>0/22</td>
<td>1/63</td>
<td>0/83</td>
<td>3/37</td>
<td>7/23</td>
<td>4/41</td>
<td>14/7</td>
</tr>
</tbody>
</table>

NOTE: Underidentification and overidentification relative frequencies (\%) 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 \( p_{1N}, p_{2N}, p_{3N}, m=2, 1.5, 5 \), and an upper bound on the number of factors set to 10.
Table 3. Bai and Ng (2005) Criterion, AR Loadings

<table>
<thead>
<tr>
<th>$\hat{D}_{k}$</th>
<th>$\rho_{1}^{\text{mg}}$</th>
<th>$\rho_{2}^{\text{mg}}$</th>
<th>$\rho_{3}^{\text{mg}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 1$</td>
<td>$m = 1$</td>
<td>$m = 1$</td>
<td>$m = 1$</td>
</tr>
<tr>
<td>$m = 2$</td>
<td>$m = 2$</td>
<td>$m = 2$</td>
<td>$m = 2$</td>
</tr>
<tr>
<td>$m = 3$</td>
<td>$m = 3$</td>
<td>$m = 3$</td>
<td>$m = 3$</td>
</tr>
</tbody>
</table>

NOTE: Underidentification and overidentification relative frequencies (in %), over 500 replications of the AR loading model, for $q = 1, 2, 3$, and of the number correlation matrices of VAR(4) residuals, with penalty functions $3, m$. Overidentification and underidentification are based on the condition of the number of autoregressive factors set to 10.
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. (2005a,b), 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. (2005a,b) 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 $MT_T = [.75 \sqrt{T}]$. 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 has attempted 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 $\lim_{n \to \infty} |IC_{0,q}(k) - IC_{0,q}(q)| > 0$ for all $k \neq q$. Let $s < q_n$ small. This inequality holds provided that there exists a finite $n_0$ such that for all $n > n_0$ and $k \neq q$, $\frac{1}{n} \sum_{j=q+1}^{n} \{\int_{\mathbb{R}} \lambda_{nj}(\theta) \, d\theta\} \geq k p(n) > \frac{1}{n} \sum_{j=q+1}^{n} \{\int_{\mathbb{R}} \lambda_{nj}(\theta) \, d\theta\} \geq q p(n)$. This holds for any $n$ sufficiently large, $(k < q)$. Therefore, for $n$ sufficiently large, $(k < q)$, $\frac{1}{n} \sum_{j=q+1}^{n} \{\int_{\mathbb{R}} \lambda_{nj}(\theta) \, d\theta\} > \frac{k p(n)}{n}$, because $p(n) \to 0$ as $n \to \infty$ and $\lambda_{nq}(\theta), j < q$, 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 $[\xi_{ij}; i,j \in \mathbb{N}]$ denote a collection of complex numbers such that for all $n$, the $n \times n$ matrices $\xi_n$ with entries $(\xi_{ij}; 1 \leq i,j \leq n)$ are Hermitian. Let $\{\xi_{ij}; 1 \leq i,j \leq n, n \in \mathbb{N}, T \in \mathbb{N}\}$ a collection of complex-valued random variables such that, similarly for all $n$ and $T$, $\xi_{ij}$ are Hermitian. Denote by $\{\xi_{ij}; 1 \leq i,j \leq n, n \in \mathbb{N}, T \in \mathbb{N}\}$ a collection of complex-valued random variables such that, similarly for all $n$ and $T$, $\xi_{ij}$ are Hermitian. Write $\lambda_{n}(\xi)$ and $\lambda_{n}^{T}(\xi)$ for the eigenvalues of $\xi_n$ and $\xi_n^T$'s in decreasing order of magnitude. The following lemma characterizes the asymptotic behavior of $\lambda_{n}(\xi) - \lambda_{n}^{T}(\xi)$ when $\xi_n - \xi_n^{T}$ converges to 0 in a sense made precise in, for example, (A.1).

**Lemma A.1.** Assume that for all $1 \leq i,j \leq n, n \in \mathbb{N}$, 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 $T$ such that $M_T \to \infty$ as $T \to \infty$ and

$$E[|\xi_{ij}^{T} - \xi_{ij}|^2] \leq k M_T^{-1}. \quad \text{(A.1)}$$

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

$$\max_{1 \leq i,j \leq q_{\max}} \mathbb{P} \left[ \left| \frac{M_T^{1/2} \lambda_{nk}(\xi) - \lambda_{nk}^{T}(\xi)}{n} \right| > B_T \right] \leq \epsilon. \quad \text{(A.2)}$$

**Corollary A.1.** Let Assumptions A1, A2', B1, and B2 hold. Then for any $\epsilon > 0$, there exist $M_T$ and $T_\epsilon$ such that for any fixed $q_{\max}$, $n,$ and $T > T_\epsilon$,

$$\max_{1 \leq i,j \leq q_{\max}} \mathbb{P} \left[ \left| \frac{\lambda_{nk}(\xi) - \lambda_{nk}^{T}(\xi)}{n} \right| > M_T^{-1/2} \right] \leq \epsilon. \quad \text{(A.3)}$$

**Proof of Lemma A.1.** Weyl's inequality implies that for any Hermitian matrices $A$ and $B$ with eigenvalues $\lambda_j(A)$ and $\lambda_j(B)$, $\max_j |\lambda_j(A) - \lambda_j(B)|^2 \leq \text{tr}(A - B)^2$. It follows that for all $n, T,$ and $k$,

$$|\lambda_{nk}(\xi) - \lambda_{nk}^{T}(\xi)|^2 \leq \text{tr}(\xi_{nk} - \xi_{nk}^{T}) \times (\xi_{nk}^{T} - \xi_{nk}) = \sum_{i=1}^{n} \sum_{j=1}^{n} |\xi_{ij}^{T} - \xi_{ij}|^2. \quad \text{(A.4)}$$

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

$$E[|\lambda_{nk}^{T}(\xi) - \lambda_{nk}(\xi)|^2] \leq \sum_{i=1}^{n} \sum_{j=1}^{n} E[|\xi_{ij}^{T} - \xi_{ij}|^2] \leq n^2 k M_T^{-1} \quad \text{for all } n, T, \text{ and } k. \quad \text{(A.5)}$$

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

We now turn to the proof of Proposition 2.
Figure 5. An Analysis of the U.S. Economy Dataset (1960–2003). Simultaneous plots of $c \mapsto Sc$ and $c \mapsto qTc$, $n$, using penalty function $p_1$, $q_{\text{max}} = 19$, and $IC^T_{2,n}$ criterion, over the periods (a) 1960–2003, (b) 1960–1982, and (c) 1983–2003.

Proof of Proposition 2

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

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

$$IC^T_{\text{min}}(q) < IC^T_{\text{max}}(k)$$  \hspace{1cm} (A.3)

if and only if

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

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

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

where $K_{1n}(T)$ is $O_p(\max[M_T^{-2}, M_T^{1/2}T^{-1/2}])$ uniformly in $n$ and $\theta$. By Assumption A4, the first $q$ eigenvalues $\lambda_{\text{max}}(\theta_l)$ diverge linearly in $n$, which implies that $\sup_{\theta} \frac{\lambda_{\text{max}}(\theta_l)}{n} = O(1)$ and $\liminf_{n \to \infty} \sup_{\theta} \frac{\lambda_{\text{max}}(\theta_l)}{n} > 0$, for $l = k+1, \ldots, q$. 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,
\[ IC^T_{2,n}(q) < IC^T_{n}(k) \]
for \( k < q \) if and only if [note that under Assumption A4(b), \( V^T_{q}(q) > 0 \)]
\[ \log(V^T_{q}(k)/V^T_{q}(q)) > (q - k)p(n,T). \]  
(A.7)

In view of Corollary A.1, we have, for \( k \leq q \),
\[ V^T_{q}(k) = \sum_{i=k+1}^{q} \frac{1}{2M_T + 1} \sum_{l=-M_T}^{M_T} \left( \frac{\lambda_{n}(\theta)}{n} + K_{2n}(T) \right), \]  
(A.8)

where \( K_{2n}(T) = Op(max[M_T^2, (M_T^2/2-T^{-1/2})]) \) uniformly in \( n \) and \( \theta \).
By Assumption A4, the eigenvalues \( \lambda_{n}(\theta), i > q \), are, uniformly in \( n \) and \( \theta \) in \([-\pi, \pi]\), bounded and bounded away from 0. Thus there exist positive constants \( c_0 \) and \( c_1 \) such that \( P[c_0 > V^T_{q}(q) > c_1] \rightarrow 1 \)
as \( min(n,T) \rightarrow \infty \). For \( k < q \), we have
\[ V^T_{q}(k) - V^T_{q}(q) = \sum_{i=q+1}^{q} \frac{1}{2M_T + 1} \sum_{l=-M_T}^{M_T} \left( \frac{\lambda_{n}(\theta)}{n} + K_{3n}(T) \right). \]
(A.9)

where \( K_{3n}(T) = Op(max[M_T^2, (M_T^2/2-T^{-1/2})]) \) uniformly in \( n \) and \( \theta \) in \([-\pi, \pi]\).  
As (A.9) coincides with the left side of Inequality (A.5), there exists a positive constant \( c_2 \) such that \( P[V^T_{q}(k) - V^T_{q}(q) > c_2] \rightarrow 1 \), and hence a constant \( c_3 > 0 \) such that
\[ P[\log\left( \frac{V^T_{q}(k) - V^T_{q}(q)}{V^T_{q}(q) + 1} \right) > c_3] \]
\[ = P[\log[V^T_{q}(k)/V^T_{q}(q)] > c_3] \rightarrow 1 \]
as \( min(n,T) \rightarrow \infty \). Thus same condition \( p(n,T) \rightarrow 0 \) is sufficient for both (A.4) and (A.7) to hold with probability tending to 1 as \( min(n,T) \rightarrow \infty \).

Next, for any \( k > q \), (A.3) holds if and only if \( \lambda_{q+1}(\theta), \lambda_{q+2}(\theta), \ldots \) are bounded uniformly in \( n \) and \( \theta \),
\[ \sup_{\theta} \left( \frac{\lambda_{q+1}(\theta)}{n} \right) = O(n^{-1}) \]  
as \( n \rightarrow \infty \) for \( i = q + 1, \ldots k \). Thus, \( k > q \), it is sufficient for inequality (A.3) to hold with probability arbitrarily close to 1 as \( min(n,T) \rightarrow \infty \) so that
\[ np(n,T) \rightarrow \infty \]  
and \( \min(M_T^2, M_T^2/2-T^{-1/2})p(n,T) \rightarrow \infty \]
as \( min(n,T) \rightarrow \infty \). (A.11)

Turning to the logarithmic criterion, (A.6) holds for \( k > q \) if and only if
\[ \log(V^T_{q}(q)/V^T_{q}(k)) < (k - q)p(n,T). \]  
(A.12)

By the same arguments as in (A.8), there exist positive constants \( c_4 \) and \( c_5 \) such that \( P[c_4 > V^T_{q}(q) > c_5] \rightarrow 1 \) as \( min(n,T) \rightarrow \infty \). Similarly, \( V^T_{q}(k) - V^T_{q}(q) = \sum_{i=q+1}^{k} \frac{1}{2M_T + 1} \sum_{l=-M_T}^{M_T} \left( \frac{\lambda_{n}(\theta)}{n} + K_{5n}(T) \right), \]  
where \( K_{5n}(T) = Op(max[M_T^2, (M_T^2/2-T^{-1/2})]) \) uniformly in \( n \) and \( \theta \).
This term coincides with the left side of (A.10), and the same arguments imply that \( V^T_{q}(k) - V^T_{q}(q) \), and hence \( (V^T_{q}(q) - V^T_{n}(q))/(V^T_{n}(k)), \)
are \( Op(max[n^{-1}, M_T^2, (M_T^2/2-T^{-1/2})]) \) as \( min(n,T) \rightarrow \infty \). Therefore, \( \log(V^T_{q}(q) - V^T_{q}(k))/V^T_{q}(q) \rightarrow (q - k)p(n,T) \), which, as \( min(n,T) \rightarrow \infty \), is \( Op(max[n^{-1}, M_T^2, (M_T^2/2-T^{-1/2})]) \), so that (A.12), under (A.11), holds with probability arbitrarily close to 1 as \( min(n,T) \rightarrow \infty \). This completes the proof.

References
Faveri, C., and Marcellino, M. (2001), ??


