The Generalized Dynamic Factor Model determining the number of factors *

Marc Hallin[†] and Roman Liška[‡]

E.C.A.R.E.S. and I.S.R.O.
Université Libre de Bruxelles
Brussels, Belgium

Abstract

This paper develops an information criterion for the choice of the number of common shocks for the approximate dynamic factor model developed by Forni, Hallin, Lippi, and Reichlin (2000). In this framework, the number q of common shocks is associated to the number of diverging eigenvalues of the spectral density matrix of the observations as the number n of time series goes to infinity. The criterion exploits this characteristic of the model. We provide sufficient conditions for consistency of the criterion for large n and T (where T is the series length). The paper shows how the method can be implemented and provides simulations and empirics which illustrate its good performance in finite samples.

Key Words: Dynamic factor model; Dynamic principal components; Information criterion.

1 Introduction

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

In the *dynamic factor* approach, the common component results from the action of a small number q of unobserved shocks. More specifically, χ_{it} takes the form $\chi_{it} = \sum_{j=1}^{q} b_{ij}(L)u_{jt}$, where

*The authors are grateful to Lucrezia Reichlin for insightful discussion and critical comments on the manuscript, and to Domenico Giannone for helpful suggestions, and a decisive contribution in the proof of Lemma 2.

†Research supported by by a P.A.I. contract of the Belgian Federal Government and an Action de Recherche Concertée of the Communauté française de Belgique.

‡Research supported by by an Action de Recherche Concertée of the Communauté française de Belgique.
Current affiliation: Institute for the Protection and Security of the Citizen (IPSC), European Commission - Joint Research Centre, Ispra , Italy.

the common shocks u_{jt}—call them the *dynamic factors*—are loaded via linear one-sided filters $b_{ij}(L)$, $j = 1, \ldots, q$ (L, as usual, stands for the lag operator). This approach first was proposed by Sargent and Sims (1977) and Geweke (1977) in a model where the idiosyncratic components are 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 papers by Forni and Lippi (2001) and Forni, Hallin, Lippi, and Reichlin (2000, 2004). The main theoretical tool in the latter papers is Brillingers' theory of *dynamic principal components* (Brillinger 1981).

A similar *approximate* factor model has been proposed by Stock and Watson (2002a and b). In their approach however, 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 common factors $(F_{1t}, \ldots, F_{rt})$—the *static* factors; the loadings a_{ij} are real numbers, and all factors, in this approach, are loaded contemporaneously.

A crucial step in the statistical analysis of these factor models is the preliminary identification of the number q of common shocks or the number r of static factors. A method for the identification of r in the static model has been proposed by Bai and Ng (2002), using an information criterion approach. The criterion they are proposing is shown to be consistent (under appropriate assumptions) as n, the cross-section dimension, and T, the length of the observed series, both tend to infinity. More recently, another criterion, based on the theory of random matrices, has been developed by Oniatski (2005), still for the number r of static factors, but in a model l with iid idiosyncratic components.

For the number q of common shocks in the general dynamic model, Forni et al. 2000 only suggest a heuristic rule based on the number of diverging (as $n \to \infty$) dynamic eigenvalues. The purpose of this paper is to propose a statistical criterion for this identification, and to establish its consistency as n and T approach infinity. This number q indeed plays an essential role in the practical implementation of the generalized dynamic factor method. Moreover, common shocks in a dynamic framework can be given an economic interpretation (on this latter point, see Giannone, Reichlin, and Sala 2005, Forni, Giannone, Lippi, and Reichlin 2005, and Stock and Watson 2005).

As shown by Forni, Hallin, Lippi, and Reichlin (2005) and Forni et al. (2005), for restricted forms of the dynamic structure, one can bring the dynamic factor model back under the traditional umbrella of the static one via stacking. In such a setting, static factors are functions of the number of common shocks and their lags, and the relation between q and r can be exploited to develop an identification criterion for q. Building on this idea, Bai and Ng (2005) recently proposed, in that restricted setting, a criterion for q adapted from Bai and Ng (2002)' s criterion for r. The criterion we are developing in this paper, however, is valid under much more general assumptions on the dynamic structure.

From a technical point of view, due to the spectral techniques involved, the tools we are using in the proofs are entirely different from those used in the static framework; our criterion builds directly on the (n, T)-asymptotic properties of the eigenvalues of sample spectral density matrices, as in Forni et al. (2004). Simulations indicate that the method performs quite well, even in relatively small panels with moderate series lengths.

The paper is organized as follows. In Section 2, the generalized dynamic factor model proposed by Forni et al. (2000) is briefly described, together with the required identifiability assumptions. Section 3 introduces the information criterion we are proposing for the identification of q, and establishes sufficient conditions for consistency as n and T tend to infinity. We recommend a covariogram-smoothing version of our method, the practical implementation of which is carefully discussed in Section 4. A simulation study of the small sample properties of the proposed identification procedure, and an application to macroeconomic data, are presented in Section 5. Section 6 concludes. Proofs are concentrated in an appendix (Section 7).

Boldface are 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 $\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mathrm{P}\left[|\zeta(n, T, \theta)| > \eta\right] < \epsilon$ for all $T > T_{\epsilon,\eta}$ (resp. if for all $\epsilon > 0$ there exist B_ϵ and T_ϵ such that $\sup_{n\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \mathrm{P}\left[|\zeta(n, T, \theta)| > B_\epsilon\right] < \epsilon$ for all $T > T_\epsilon$.

2 The dynamic factor model

The model we are considering throughout is Forni et al. (2000)'s *generalized dynamic factor model*, which we now briefly describe. Let $\{X_{it}, i \in \mathbb{N}, t \in \mathbb{Z}\}$ be a double array of random variables, where

$$X_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \ldots + b_{iq}(L)u_{qt} + \xi_{it}, \tag{2.1}$$

and the following assumptions A1 through A4 are assumed to hold.

Assumption A1.

(i) *The q-dimensional vector process $\{\mathbf{u}_t := (u_{1t}\ u_{2t}\ \ldots\ u_{qt})'; t \in \mathbb{Z}\}$ is orthonormal white noise;*

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

(iii) *the one-sided filters $b_{ij}(L) := \sum_{k=1}^{\infty} b_{ijk}L^k$ have square summable coefficients: $\sum_{k=1}^{\infty} b_{ijk}^2 < \infty$ 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*. The random variables ξ_{it} and χ_{it} are called the *idiosyncratic* and *common components* of X_{it}, respectively.

Assumption A2. *For all n, the vector process $\boldsymbol{X}_{nt} := (x_{1t}\ x_{2t}\ \ldots\ x_{nt})'$ is a linear process, with a representation of the form $\boldsymbol{X}_{nt} = \sum_{k=-\infty}^{\infty} \mathbf{C}_k \mathbf{Z}_{t-k}$, 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 $\sum_{k=-\infty}^{\infty} |C_{ij,k}||k|^{1/2} < \infty$.*

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. Consistent identification of q, as we shall see, is more demanding: denoting by

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

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

Assumption A2′. *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 $1 \leq \ell \leq 4$ and all $1 \leq j < \ell$,*

$$\sup_{i_1,\ldots,i_\ell} \left[\sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{\ell-1}=-\infty}^{\infty} (1+|k_j|)\, |c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1})| \right] < \infty \tag{2.2}$$

This assumption is the uniform version of a condition considered in Section 4.3 of Brillinger (1981) for the consistency of periodogram-based estimation of the spectrum.

Denote by $\boldsymbol{\Sigma}_n(\theta)$, $\theta \in [-\pi, \pi]$, the spectral density matrix of $\mathbf{X}_{nt}$, with elements $\sigma_{ij}(\theta)$, and by $\lambda_{n1}(\theta), \ldots, \lambda_{nn}(\theta)$ the corresponding eigenvalues in decreasing order of magnitude. Similarly, with obvious notation, let $\lambda^{\chi}_{nj}(\theta)$ and $\lambda^{\xi}_{nj}(\theta)$ be the eigenvalues associated with the spectral densities $\boldsymbol{\Sigma}^{\chi}_n(\theta)$ and $\boldsymbol{\Sigma}^{\xi}_n(\theta)$ of $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$, respectively. Such eigenvalues (actually, the functions $\theta \mapsto \lambda(\theta)$) are called *dynamic* eigenvalues.

Assumption A3. *The first idiosyncratic dynamic eigenvalue $\lambda^{\xi}_{n1}(\theta)$ is uniformly (with respect to $\theta \in [-\pi, \pi]$) bounded as $n \to \infty$, i.e. $\sup_{\theta \in [-\pi,\pi]} \lambda^{\xi}_{n1}(\theta) < \infty$ as $n \to \infty$.*

Assumption A4. *The qth common dynamic eigenvalue $\lambda^{\chi}_{nq}(\theta)$ diverges $\theta-$a.e. in $[-\pi, \pi]$ as $n \to \infty$.*

Assumptions A3 and A4 play a key role in the identification of the common and the idiosyncratic components in (2.1). However, only the X_{it}'s are observable, and Assumptions A3 and A4 thus involve the unobserved quantities $\boldsymbol{\chi}_{nt}$ and $\boldsymbol{\xi}_{nt}$. This, at first sight, may seem unrealistic, and the following proposition provides a X_{it}-based counterpart.

PROPOSITION 1 (Forni and Lippi 2001) *Let Assumption A2 (or A2′) hold. Then, Assumptions A1, A3, and A4 are satisfied iff the first q eigenvalues of $\mathbf{\Sigma}_n(\theta)$ diverge as $n \to \infty$, a.e. in $[-\pi, \pi]$, while the $(q+1)$th one is uniformly bounded.*

Forni et al. (2000) show how, under Assumptions A1-A4, the common components χ_{it} and the idiosyncratic components ξ_{it} are asymptotically identified as $n \to \infty$ and are consistently estimated, as both n and $T \to \infty$, by means of the dynamic principal components method. Dynamic principal components are the solutions of an optimization problem, the main features of which we briefly summarize in the following proposition.

PROPOSITION 2 (BRILLINGER 1981, THEOREM 9.3.1.) *Let $\{\mathbf{Y}_t,\ t \in \mathbb{Z}\}$ be an n-dimensional stationary process, with zero-mean and rational spectrum $\mathbf{\Sigma}_Y(\theta)$. Denote by $\mathbf{V}_j(\theta)$ the eigenvector associated with the jth largest eigenvalue $\mu_j(\theta)$ of $\mathbf{\Sigma}_Y(\theta)$. Then, the coefficients of the $(q \times n)$ filter $\boldsymbol{b}(L) := \sum_k \mathbf{b}_k L^k$ and the coefficients of the $(n \times q)$ filter $\boldsymbol{c}(L) := \sum_k \mathbf{c}_k L^k$ that minimize*

$$\mathrm{E}\{[\mathbf{Y}_t - \mathbf{c}(L)\mathbf{b}(L)\mathbf{Y}_t]^{*\prime}[\mathbf{Y}_t - \mathbf{c}(L)\mathbf{b}(L)\mathbf{Y}_t]\} \tag{2.3}$$

are

$$\boldsymbol{b}_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{B}(\theta)e^{-ik\theta}\,d\theta \quad \textit{and} \quad \boldsymbol{c}_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{C}(\theta)e^{-ik\theta}\,d\theta, \tag{2.4}$$

respectively, where $\mathbf{C}(\theta) = [\mathbf{V}_1(\theta) \cdots \mathbf{V}_q(\theta)]$ and $\mathbf{B}(\theta) = \mathbf{C}(\theta)^{*\prime}$. The resulting minimum of (2.3) is $\int_{-\pi}^{\pi}\{\sum_{j>q}\mu_j(\theta)\}d\theta$.

The first q *dynamic principal components* are defined as the components of the random q-dimensional vector $\mathbf{c}(L^{-1})'\mathbf{Y}_t$. Under model (2.1) and Assumptions A1-A4, Forni et al. (2000) show that the common component $\boldsymbol{\chi}_{nt}$ can be consistently (as $n \to \infty$) reconstructed by

$$\chi_{it}^{(n)} := (\mathbf{K}(L)\mathbf{X}_{nt})_{it}\,, \qquad \text{with} \qquad \mathbf{K}(L) := \mathbf{c}(L)\mathbf{c}(L^{-1})' \tag{2.5}$$

(with $\mathbf{X}_{nt}$ playing the role of $\mathbf{Y}_t$).

This $\chi_{it}^{(n)}$ of course cannot be computed from the data, since it involves the true spectral density matrix $\mathbf{\Sigma}_n(\theta)$ and the true number of factors q. The spectral density $\mathbf{\Sigma}_n(\theta)$ however can be estimated from the data by means of periodogram or covariogram smoothing methods. Provided that q is known, applying (2.5) to the estimated spectral density yields a consistent estimator of $\boldsymbol{\chi}_{nt}$ (Forni et al. 2000, 2004). Determining q prior to this estimation 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 n of the processes $\{X_{it}\}$ are observed, and the selection of q has to be based on this finite-sample information. As a preparation, however, we first prove a consistency result, as $n \to \infty$, at *population level*, that is, assuming that the processes $\{X_{it}\}$ are observed over $t \in \mathbb{Z}$, so that the spectral density matrices $\mathbf{\Sigma}_n(\theta)$ are known. Only asymptotics in n are of interest here. We define a (deterministic) selection criterion (3.1) and provide sufficient conditions for its consistency as the size n of the panel tends to infinity.

As mentioned in Proposition 2, the estimated common components $\chi_{it}^{(n)}$ in (2.5) can be viewed as solutions of the optimization problem (2.3). For fixed k, this optimization is equivalent to minimizing $n^{-1}\mathrm{E}\{[\mathbf{X}_{nt} - \mathbf{K}(L)\mathbf{X}_{nt}]^{*\prime}[\mathbf{X}_{nt} - \mathbf{K}(L)\mathbf{X}_{nt}]\}$ with respect to $\mathbf{K}(L)$. The corresponding minimum is then $n^{-1}\sum_{j=k+1}^{n}\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)d\theta\}$.

The information criterion approach consists in selecting the number of factors as

$$\hat{q}_n := \mathrm{argmin}_{0\le k\le k_{\max}} L_n(k), \qquad \text{where} \quad L_n(k) := \frac{1}{n}\sum_{j=k+1}^{n}\int_{-\pi}^{\pi}\lambda_{nj}(\theta)d\theta + kp(n), \tag{3.1}$$

where $q_{\max}$ is some predefined upper bound for the actual q, and $p(n)$ is an adequate penalty function. Note that $p(n)$ here is deterministic, and depends only on n since the spectral density matrices $\mathbf{\Sigma}_n(\theta)$ are assumed to be known; the solution $\hat{q}_n$ is deterministic as well, since $\mathbf{\Sigma}_n(\theta)$ is.

The intuition behind (3.1) 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; still, it should be large enough to avoid overestimation. This delicate balance between over- and under-estimation is intimately related to the rate of divergence, as $n \to \infty$, of the diverging eignevalues. In order to impose consistency conditions on the penalty function $p(n)$, an assumption about the behavior of the diverging eigenvalues is needed.

Assumption A5.

(i) All diverging eigenvalues of $\mathbf{\Sigma}_n(\theta)$ diverge linearly in n, θ - a.e., that is, there exist $2q$ constants $0 < c_i^- \le c_i^+$, $i = 1, ..., q$, such that $c_i^+ < c_{i-1}^-$ and

$$c_i^- \le \mathrm{liminf}_{n\to\infty} n^{-1}\lambda_{ni}(\theta) \le \mathrm{limsup}_{n\to\infty} n^{-1}\lambda_{ni}(\theta) \le c_i^+, \quad \theta\text{−a.e}, \ i = 1, ..., q, \ n \in \mathbb{N}.$$

(ii) The non diverging eigenvalues $\lambda_{ni}(\theta)$ ($i > q$) are uniformly bounded away from zero, that is, there exists a constant $c_\lambda > 0$ such that, for all $i > q$ and $n \in \mathbb{N}$, $\lambda_{ni}(\theta) > c_\lambda$, θ - a.e.

The linear divergence in (i) has a natural interpretation: the influence of the common shocks, in some sense, is "stationary along the cross-section". For a detailed discussion of this assumption, we refer to Forni et al (2004).

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

Lemma 1. *Let $\hat{q}_n$ be defined in (3.1), and let the penalty $p(n)$ be such that*

$$\lim_{n\to\infty} p(n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} np(n) = \infty. \tag{3.2}$$

Then, under Assumptions A1-A5, $\lim_{n\to\infty}\hat{q}_n = q$.

PROOF. See Appendix.

Examples of penalty functions satisfying (3.2) are $c/\sqrt{n}$ or $c\log(n)/n$, where c is an arbitrary positive real number. Lemma 1 of course has little practical consequences. But the pedagogical value of its proof, which is extremely simple, is worth some attention. First of all, it very clearly appears from that proof, that the $\frac{1}{n}$ coefficient, in the definition of the criterion $L_n(k)$ and the second assumption on the penalty ($np(n) \to \infty$) are directly related to the $O(n)$ divergence rate in Assumption A5: 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. Multiplying the penalty with an arbitrary constant thus has no influence on the asymptotic performance of the identification method. But it obviously quite dramatically may affect the actual result for given n. This will be exploited later on in the implementation of the criterion (Section 4).

In practical situations, the spectral density matrix has to be estimated from observed series with finite length T; this series length moreover quite naturally has to play a role in the penalty function.

3.2 Sample level : periodogram smoothing estimation

In this section, we derive sufficient conditions for consistent estimation of q as both n and T tend to infinity. As mentioned at the end of Section 2, one possibility consists in using a periodogram-smoothing estimate $\mathbf{\Sigma}_n^T(\theta)$ of $\mathbf{\Sigma}_n(\theta)$. Based on the nT observations $\{X_{it}; t = 1, ..., T,\ i = 1, ..., n\}$, this estimator is defined as

$$\mathbf{\Sigma}_n^T(\theta) := \frac{2\pi}{T}\sum_{t=1}^{T-1} W^{(T)}\Big(\theta - \frac{2\pi t}{T}\Big)\mathbf{I}_n^T\Big(\frac{2\pi t}{T}\Big), \tag{3.3}$$

where

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

and

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

with a positive even weight function $W(\alpha)$, and a bandwidth B_T. This estimator $\mathbf{\Sigma}_n^T(\theta)$ is consistent for any n, as $T \to \infty$, provided that W and B_T satisfy the following assumption.

ASSUMPTION B1.

(i) $B_T > 0$, $B_T \to 0$, *and* $B_T T \to \infty$, *as* $T \to \infty$;

(ii) $\alpha \mapsto W(\alpha)$ *is a differentiable positive even function, of bounded variation, with bounded derivative* W'*, satisfying* $\int_{-\infty}^{\infty} W(\alpha)d\alpha = 1$ *and* $\int_{-\infty}^{\infty} |\alpha|^3 W(\alpha)d\alpha < \infty$.

However, such fixed-n consistency is not sufficient here, and 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* $\mathbf{\Sigma}_n(\theta)$ *are uniformly (in* n *and* θ*) bounded, and have uniformly (in* n *and* θ*) bounded derivatives up to the order three: there exists* $Q < \infty$ *such that* $\sup_{i,j \in \mathbb{N}} \sup_\theta \left| \frac{d^k}{d\theta^k} \sigma_{ij}(\theta) \right| \leq Q, \quad k = 0, 1, 2, 3.$

Assuming that Assumptions A2′, B1, and B2 hold, we then have the following uniform consistency result (see equation (7.4.20) in Brillinger 1981): there exist constants K_1, K_2, and T_0 such that for all θ, $1 \leq i, j \leq n$, and n,

$$\sup_n \max_{1 \leq i,j \leq n} \sup_\theta \left[\mathrm{E} \left| \mathbf{\Sigma}_n^T(\theta) - \mathbf{\Sigma}_n(\theta) \right|_{ij}^2 \right] \leq K_1 B_T^{-1} T^{-1} + K_2 B_T^4 \tag{3.4}$$

for any $T > T_0$.

The proof of this results is long but easy; it mainly consists in going through all the steps of Brillinger's proofs (Section 7.4), and taking into account the uniformity of Assumption A2′ and B2.

The stochastic information criterion we are proposing is defined, in terms of the eigenvalues $\lambda_{ni}^T(\theta)$ of the estimated spectral density matrices $\mathbf{\Sigma}_n^T(\theta)$, as

$$IC_n^T(k) := \frac{1}{n} \sum_{i=k+1}^{n} \frac{1}{T-1} \sum_{l=1}^{T-1} \lambda_{ni}^T(\theta_l) + kp(n,T), \quad 0 \leq k \leq k_{\max} < \infty, \tag{3.5}$$

where $p(n,T)$ is a penalty function, $\theta_l := 2\pi l/T$ for $l = 1, \ldots, T-1$, and $q_{\max}$ is some predetermined upper bound. For given n and T, the number of factors q is estimated as

$$q_n^T := \mathrm{argmin}_{0 \leq k \leq q_{\max}} IC_n^T(k). \tag{3.6}$$

The following consistency property is the first main result of this paper.

Proposition 3. *Let Assumptions A1, A2′, A3 through A5, B1, and B2 hold. Then,* $\mathrm{P}(q_n^T = q) \to 1$ *as* n *and* T *both tend to infinity in such a way that*

$$(i)\ p(n,T) \to 0, \quad (ii)\ \min\left[n, B_T^{-2}, B_T^{1/2} T^{1/2}\right] p(n,T) \to \infty. \tag{3.7}$$

Proof. See the Appendix.

Observe that if $p(n,T)$ is an appropriate penalty function, that is, if (3.7) holds, then $cp(n,T)$, where c is an arbitrary positive real, also is an appropriate penalty function.

3.3 Sample level : covariogram smoothing estimation

The consistency conditions in Proposition 3 are derived for the periodogram smoothing estimator (3.3) of $\mathbf{\Sigma}_n(\theta)$. For computational convenience, however, covariogram smoothing estimation is preferable in the practical implementation of the Forni et al (2000) method. The covariogram smoothing estimator of $\mathbf{\Sigma}_n(\theta)$ is defined as

$$\mathbf{\Sigma}_n^{*T}(\theta) := \frac{1}{2\pi} \sum_{u=-M_T}^{M_T} w(M_T^{-1}u)\mathbf{\Gamma}_{nu}^T e^{-iu\theta} \tag{3.8}$$

where $\mathbf{\Gamma}_{nu}^T$ denotes the sample cross-covariance matrix of $\mathbf{X}_{nt}$ and $\mathbf{X}_{n,t-u}$ based on T observations, $w(\alpha)$ is a positive even weight function, and M_T is a truncation parameter. The estimator $\mathbf{\Sigma}_n^{*T}(\theta)$ is consistent for any n, as $T \to \infty$, provided that w and M_T satisfy the following assumption.

Assumption B1$'$.

(i) $M_T > 0$, $M_T \to \infty$, *and* $M_T T^{-1} \to 0$, *as* $T \to \infty$;

(ii) $\alpha \mapsto w(\alpha)$ *is an even, piecewise continuous function, differentiable with bounded first three derivatives, satisfying* $w(0) = 1$, $|w(\alpha)| \leq 1$ *for all* α *and* $w(\alpha) = 0$ *for* $|\alpha| > 1$.

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 for all θ, $1 \leq i, j \leq n$, and n,

$$\sup_n \max_{1\leq i,j\leq n} \sup_\theta \left[\mathrm{E} \left| \mathbf{\Sigma}_n^{*T}(\theta) - \mathbf{\Sigma}_n(\theta) \right|_{ij}^2 \right] \leq L_1 M_T T^{-1} + L_2 M_T^{-4} \tag{3.9}$$

for any $T > T_0$.

As in the periodogram smoothing case, the proof of this result is long but easy; it mainly consists in going through all the steps of Parzen's Theorem 5A proof (Parzen 1957), and taking into account the uniformity of Assumption A2$'$, B1$'$, and B2.

Associated with the covariogram smoothing estimator $\mathbf{\Sigma}_n^{*T}(\theta)$, consider the following information criterion

$$IC_{1;n}^{*T}(k) := \frac{1}{n} \sum_{i=k+1}^{n} \frac{1}{2M_T + 1} \sum_{l=-M_T}^{M_T} \lambda_{ni}^{*T}(\theta_l) + kp(n,T), \quad 0 \leq k \leq q_{\max} \tag{3.10}$$

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

This criterion has a structure comparable to that of Bai and Ng (2002). In a Corollary to their Theorem 2, these authors also show that a logarithmic form of their criterion has similar consistency properties as the original one. Experience seems to indicate, moreover, that this logarithmic form has better finite sample performances. We therefore also consider the criterion

$$IC_{2;n}^{*T}(k) := \log\left[\frac{1}{n}\sum_{i=k+1}^{n}\frac{1}{2M_T+1}\sum_{l=-M_T}^{M_T}\lambda_{ni}^{*T}(\theta_l)\right] + kp(n,T), \quad 0 \leq k \leq q_{\max}, \qquad (3.11)$$

Depending on the criterion adopted, the resulting estimated number of factors, for given n and T, is,

$$q_{a;n}^{*T} := \mathrm{argmin}_{0\leq k\leq q_{\max}} IC_{a;n}^{*T}(k), \quad a = 1, 2. \qquad (3.12)$$

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

Proposition 4. *Let Assumptions A1, A2′, A3 through A5, B1′, and B2 hold. Then,* $\mathrm{P}(q_{a;n}^{*T} = q) \to 1$ *for* $a = 1, 2$ *as* n *and* T *both tend to infinity, in such a way that*

$$(i)\ p(n,T) \to 0, \quad and \quad (ii)\ \min\left(n, M_T^2, M_T^{-1/2}T^{1/2}\right) p(n,T) \to \infty. \qquad (3.13)$$

Proof. See the Appendix.

Here again, if $p(n,T)$ is an appropriate penalty function, then $cp(n,T)$, where c is an arbitrary positive real, also is; for given n and T, a penalty function $p(n,T)$, although satisfying (3.13), can be arbitrary bad (the same remark holds for all information criteria developed in the literature).

4 A practical guide to the selection of q

As emphasized in the previous section, if our identification procedures are consistent for penalty $p(n,T)$, they also are for any penalty of the form $cp(n,T)$, where $c \in \mathbb{R}^+$. Important as they are, the above consistency results thus are of limited value for practical implementation. In this section, we show how this degree of freedom in the choice of c can be exploited. We first give some theoretical considerations, which we check on two examples (Examples 1 and 2) before describing a practical implementation of our method. In Section 5.1, we validate the method through simulation; in Section 5.2, we apply it to a dataset of quarterly macroeconomic indicators.

Denote by $q_{c;1;n}^{*T}$ and $q_{c;2;n}^{*T}$ the number of factors resulting from applying (3.10) or (3.11), respectively, with penalty $cp(n,T)$: as both n and T in practice are fixed, the only information we can obtain on the functions $(n,T) \mapsto q_{c;a;n}^{*T}$ is to be obtained from J-tuples of the form $q_{c;a;n_j}^{*T_j}$, $a = 1, 2$, $j = 1, \ldots J$, where $0 < n_1 < n_2 < \ldots < n_J = n$, and $0 < T_1 < T_2 < \ldots < T_J = T$. For any fixed value of (n_j, T_j), $q_{c;a;n_j}^{*T_j}$ clearly is a nonincreasing function of c: for given a, the curves $[n_j, T_j] \mapsto q_{c;a;n_j}^{*T_j}$ thus never cross each other. The typical situation is as follows (for simplicity, we drop a subscripts).

Assume that $q > 0$. If we let $c = 0$ (no penalty at all—this is thus a non-valid value of c), $q_{0;n_j}^{*T_j}$ is increasing with j, and would tend to infinity if n and T would. If $c > 0$ is "very small" (severe *under*penalization), although Proposition 4 applies, the situation for finite (n, T) will not be very different: $q_{c;n_j}^{*T_j}$ is still an increasing function of j, and only would redescend and tend to q (as implied by Proposition 4) if n and T were allowed to increase without limits. As c grows, 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 these *under*penalized cases however is that the variability among the J values of $q_{c;n_j}^{*T_j}$, $j = 1, \ldots, J$, is high; this variability can be captured, for instance, by the mean squared deviation $J^{-1} \sum_{j=1}^{J} \left(q_{c,n_j}^{*T_j} - J^{-1} \sum_{j=1}^{J} q_{c,n_j}^{*T_j} \right)^2$ or its square root.

Let us now consider, quite on the contrary, a "very large" value of c, hence severely *over*penalized $q_{c;n_j}^{*T_j}$'s. If c is large enough, $q_{c;n_j}^{*T_j}$ will be identically zero for all $[n_j, T_j]$'s, and convergence to q will not be visible for the values of n and 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.

In view of the monotonicity of $c \mapsto q_{c;n_j}^{*T_j}$, somewhere between those "small" underpenalizing values of c (with $j \mapsto q_{c;n_j}^{*T_j}$ curves eventually tending to q from above) and the "large" overpenalizing ones (with $j \mapsto q_{c;n_j}^{*T_j}$ curves tending to q from below), a range of "moderate" values of c, yielding a stable behavior of $j \mapsto q_{c;n_j}^{*T_j} \approx q$, typically exists. This stability can be assessed, for instance, via the empirical standard error, for given c, of the $q_{c;n_j}^{*T_j}$'s, $j = 1, \ldots, J$ (see (4.14) below).

As an illustration, let us consider two examples:

– in Example 1, a panel of size $n = 200$ and length $T = 200$ was generated. The common part was modelled with $q = 3$ factors and MA loadings, see Section 5.1 for details. The truncation parameter was set as $M_T = [0.7\sqrt{T}]$;

– in Example 2 a panel of size $n = 150$ and length $T = 120$ was generated. The common part was modelled with $q = 2$ factors and AR loadings, see Section 5.1 for details. The truncation parameter was set as $M_T = [0.5\sqrt{T}]$.

In both cases, a triangular window was used, $q_{\max}$ was set to 19, and the penalty function

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

was chosen. The values of c in the interval $[0, 2]$ were explored with a grid step of size 0.01.

– Example 1. The graphs of $(n_j, T_j) \mapsto q_{c,n_j}^{*T_j}$ and

$$c \mapsto S_c := \left[J^{-1} \sum_{j=1}^{J} \left(q_{c,n_j}^{*T_j} - J^{-1} \sum_{j=1}^{J} q_{c,n_j}^{*T_j} \right)^2 \right]^{1/2} \tag{4.14}$$

are presented for $n_j = T_j = 50, 60, \ldots, 200$ and various values of c in Figure 1, based on criterion $IC^{*T}_{1;n}(k)$ in (a1) and (a2), on criterion $IC^{*T}_{2;n}(k)$ in (b1) and (b2). The typical patterns described are all present in (a1) as well as in (b1). Inspection of (a2) in conjunction with (a1) reveals the very characteristic fact that S_c vanishes over certain intervals, corresponding with a stable behavior of the corresponding graphs in (a1): (a2) yields four "stability intervals", $(0, 0.02]$, $[0.20, 0.29]$, $[0.36, 0.48]$ and $[0.54, 0.70]$, corresponding to a selection of $q = q_{\max} = 19$, 3, 2, and 1 factors, respectively. Those "stability intervals" are separated by "instability" intervals, corresponding to more fluctuations in (a1) curves. The correct value of q, in (a1), is obtained for $c = 0.25$. Note that $q^{*T_j}_{0.15,n_j}$, as $j \uparrow$, converges to $q^{*T_j}_{0.25,n_j}$ from above, while $q^{*T_j}_{0.35,n_j}$ converges to $q^{*T_j}_{0.25,n_j}$ from below and that $c = 0.25$ is the only value $\hat{c}$ of c in (a1) exhibiting that pattern. The same comments can be made for the logarithmic version of the criterion: see (b1). Moreover, this $\hat{c}$ corresponds to the second "stability interval" in the $c \mapsto S_c$ graphs (a2) and (b2), while the first "stability interval" (namely, $(0, 0.02]$ in (a2), and (0, .24] in (b2)) clearly is associated with severe underpenalization, hence the maximal possible number of factors $q_{\max}$; Figure (b2) in this respect provides a somewhat clearer picture than (a2).

This example suggests that, irrespective of the choice of $IC^{*T}_{1;n}$ or $IC^{*T}_{2;n}$, the selection of q should be based on an inspection of the family of curves $(n_j, T_j) \mapsto q^{*T_j}_{c,n_j}$, trying to spot (as in Figure 1(a1)) the curve (and the associated value of c) the neighbors of which (corresponding to $c \pm \delta$) tend to, both from above (for $\delta < 0$) as from below (for $\delta > 0$). This search is greatly facilitated, and can be made automatic, by considering also the $c \mapsto S_c$ mapping, and choosing $q^{*T}_{\hat{c},n}$, where $\hat{c}$ belongs to the second "stability interval". The relevant figure then is a joint plot of $c \mapsto S_c$ and $c \mapsto q^{*T}_{c,n}$: see Figure 2 (c1) and (c2).

– Example 2. Here we apply the automatic selection rule just described, but with $n_j = 80 + 10j$, $j = 1, \ldots, J = 7$, $T_i = 60 + 10i$, $i = 1, \ldots, I = 6$, and

$$S_c := \left[(IJ)^{-1} \sum_{i,j} \left(q^{*T_i}_{c,n_j} - (IJ)^{-1} \sum_{i,j} q^{*T_i}_{c,n_j} \right)^2 \right]^{1/2}; \qquad (4.15)$$

the relevant plots of $c \mapsto S_c$ and $c \mapsto q^{*T}_{c,n}$ are given in Figure 3. For the $IC^{*T}_{1;n}(k)$ criterion, the stability intervals (in Figure (d1)) are (0, 0.02], [0.17, 0.54], [0.63, 0.93], and [1, 2], yielding $q^{*T}_{c,n} = 19$, 2 (correct identification), 1, and 0, respectively. The situation again is rather clearer with $IC^{*T}_{2;n}(k)$ (Figure (d2)), with stability intervals (0, 0.26], [0.39, 0.95], [1.06, 1.12], and [1.2, 2], yielding $q^{*T}_{c,n} = 19$, 2 (correct identification), 1, and 0, respectively. In both cases, thus, the second stability interval identifies the correct value $q = 2$.

When T is small relative to n, which is typically the case in macroeconomic data sets, one may like to look at J-tuples $n_1, \ldots, n_J$ only, keeping T fixed. The monotoniocity of $c \mapsto q^{*T}_{c;n_j}$ still holds, and the same discussion as above can be made, though all patterns may not be present (typically, the "redescending" to q of $j \mapsto q^{*T}_{c;n_j}$ may not be observed). Finally, whenever the actual value of q is zero (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 $c\S_c$ plots, the second one extending to the maximal possible value of c, and corresponding to $q^{*T}_{c,n} = 0$.

Summing up, our identification method in practice is to be performed as follows.

(0) Preliminary to the analysis, it may be worth choosing a random permutation of the n cross-sectional items, as some irrelevant structure may exist in the initial ordering of the panel;

(i) fix the upper bound $q_{\max}$ on the number of factors;

(ii) choose a covariogram smoothing function $w(\alpha)$ satisfying Assumption B1*(ii);

(iii) choose $T \mapsto M_T$ so that Assumption B1′(i) be satisfied, e.g., $M_T := [0.5\sqrt{T}]$ or $M_T := [0.7\sqrt{T}]$;

(iv) choose a penalty function $(n, T) \mapsto p(n, t)$ and a criterion ($IC^{*T}_{1;n}(k)$ or $IC^{*T}_{2;n}(k)$), and define $p^*_c(n, t) = cp(n, t)$ for a suitable set $\mathcal{C} \subset \mathbb{R}^+$ of values of c (e.g. $\mathcal{C} := [0.01, 0.02, \ldots, 3]$);

(v) define sequences $n_1 < n_2 < \ldots < n_J = n$ and $T_1 < T_2 < \ldots < T_I = T$ (e.g., for $n = 150$, set $n_j := 40 + 10j$, $j = 1, \ldots, 11$, for $T = 100$, set $T_i := 70 + 10i$, $i = 1, \ldots, 3$); if T is too small, let $I = 1$, that is, keep T fixed;

(vi) defining S_c as in (4.14) or (4.15), identify the number of factors as $\hat{q} := q^{*T}_{\hat{c},n}$, where $\hat{c}$ is selected as explained above (see Examples 1-2), either by inspecting the $(n_j, T_j) \mapsto q^{*T_j}_{c,n_j}$ curves, or by selecting the second stability interval of $c \mapsto S_c$.

5 Numerical study

5.1 Simulations

In order to evaluate the performance of the selection strategy proposed in the previous section, the following Monte-Carlo experiment has been conducted. Three datasets were generated, with $q = 1$, 2, and 3 factors, respectively, from the model

$$\mathbf{x}_{nt} = \mathbf{B}_{nq}(L)\mathbf{u}_t + \mathbf{e}_{nt}, \quad (n, T) = (70, 60), (90, 90), (150, 120),$$

where

- the random shocks $\mathbf{u}_t = (u_{1t}, ..., u_{qt})'$, the idiosyncratic components $\mathbf{e}_{nt} = (e_{1t}, ..., e_{nt})'$, and the loading filters $[\mathbf{B}_{nq}(L)]_{rs} = b_{rs}(L)$, $r = 1, ..., n$, $s = 1, ..., q$ are randomly generated as follows:
  - the vectors $\mathbf{u}_t$ and $\mathbf{e}_t$ are i.i.d., with $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{I_q})$,
  - the e_{it}'s are of the form $e_{i,t} = d_i f_{i,t}$, $f_{it} = y_{i,t} + 0.1 y_{i,t-1} + 0.1 y_{i+1,t}$, with $y_{it} \sim N(0, 1)$ and $d_i \sim U(0.9, 1.1)$ mutually independent and independent of the $\mathbf{u}_t$'s;
  - the filters $b_{ik}(L)$ $(i = 1, ..., n, \quad k = 1, ..., q)$ are randomly generated (independent from the $\mathbf{u}_t$'s and $\mathbf{e}_t$'s) by one of the following two devices :
    (MA loadings): $b_{ik}(L) = b^0_{ik} + b^1_{ik} L + b^2_{ik} L^2$ with $(b^0_{ik},\ b^1_{ik},\ b^2_{ik}) \sim N(\mathbf{0}, \mathbf{I})$;
    (AR loadings): $b_{ik} = b^0_{ik}(1 + b^1_{ik} L)^{-1}$, with $b^0_{ik} \sim N(0, 1)$ and $b^1_{ik} \sim U(-0.8, 0.8)$, mutually independent;
- for each i, the variance of $f_{i,t}$ and that of the common component $\sum_{k=1}^{q} b_{ik}(L) u_{it}$ were normalized to 0.5.

In each case, the number of replications was set to 500, the upper bound $q_{\max}$ to 19. Spectral density matrices were estimated with a triangular smoothing function $w(v) = 1 - |v|$ and two different values of M_T, $M_T = [0.5\sqrt{T}]$ and $M_T = [0.7\sqrt{T}]$. For each pair (n, T), the automatic identification rule described in the previous section was performed with sequences $n_j := n - 10j$, $j = 1, \dots, 3$, $T_i := T - 10i$, $i = 1, \dots, 3$, $\mathcal{C} := [0.01, 0.02, \dots, 3]$, and penalty functions

$$
\begin{aligned}
p_1(n,T) &= \left(M_T^{-2} + M_T^{1/2} T^{-1/2} + n^{-1}\right) \log\left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right) \\
p_2(n,T) &= \left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)^{-1/2} \\
p_3(n,T) &= \left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2} T^{1/2}\right]\right),
\end{aligned}
$$

respectively.

Tables 1 and 2 provide, for each case, the percentages (over the 500 replications) of under-, correct, and over-identification of the number of factors. Inspection of the results show that identification is uniformly very good. The choice of M_T, the penalty function and the criterion ($IC^{*T}_{1;n}(k)$ or $IC^{*T}_{2;n}(k)$) apparently have very little impact when n and T are large; larger values of q ($q = 3$) and the MA loadings in this respect are "more difficult" than smaller $q$values ($q = 1$) and AR loadings.

5.2 A real data application

The proposed criteria thus seems to work rather well in simulated data. We now consider a real case study. We build a panel of $n = 62$, $T = 40$ by pooling seven quarterly macroeconomic indicators for all countries of the Eurozone, excluding Luxembourg and Ireland, from 1995 to first quarter 2005 (source: Eurostat). For all those countries, the panel includes seasonally adjusted series of imports of goods and services (millions of euros, at 1995 prices and exchange rates), exports of goods and services (millions of euros, at 1995 prices and exchange rates), harmonized consumer price indices (3rd, 6th, 9th, 12th months values), quarterly production index, total industry (excluding construction), gross domestic product at market prices (constant prices, millions of euros, at 1995 prices and exchange rates), final consumption expenditure of households (millions of euros, at 1995 prices and exchange rates), gross fixed capital formation (millions of euros, at 1995 prices and exchange rates). Only the Austrian quarterly production index is missing. Data are taken in log-differenced and then normalized by their sample standard deviations. The truncation parameter is $M_T = [0.5\sqrt{T}] = 3$ and $n_j = 47 + j$, $j = 1, \ldots, 15$. A triangular window was used, and the penalty function $p_1(n, T)$ was chosen. The automatic identification based on $IC^{*T}_{2;n}(k)$ yields the stability interval $[0.44,\ 0.63]$, and the number of factors $q = 1$ is identified: see Figure 5.

6 Concluding remarks

This paper is an attemps to fill a gap in the literature on dynamic factor models, by providing an efficient yet flexible tool for identification of the number q of factors. We establish the consistency, as both n and T approach infinity in an appropriate way, of two methods, based on periodogram and covariogram smoothing, logged and non-logged criteria, respectively. We also show how to take advantage of the fact that penalty functions are defined up to a multiplicative constant. The performance of the method is evaluated through simulation, and appears to be surprisingly good.

7 Appendix.

Proof of Lemma 1. We have to show that $\lim_{n\to\infty}[L_n(k) - L_n(q)] > 0$ for all $k \neq q$, $k \leq q_{\max} < \infty$. This inequality holds true 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.

(a) Either $k > q$; then, for n sufficiently large, $(k-q)p(n) > \frac{1}{n}\sum_{j=q+1}^{k}\left\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)d\theta\right\}$, since $np(n) \to \infty$ as $n \to \infty$; or

(b) $k < q$ and , for n sufficiently large, $\frac{1}{n}\sum_{j=k+1}^{q}\left\{\int_{-\pi}^{\pi}\lambda_{nj}(\theta)d\theta\right\} > (q-k)p(n)$, since $p(n)\to 0$ as $n\to\infty$ and $\lambda_{nj}(\theta)$, $j\le q$, under Assumption E, is $O(n)$ but not $o(n)$.

The result follows. *Q.E.D.*

Before turning to the proof of Proposition 3, we prove a general result 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. 2005). Lemma 2 and Corollary 1 below collect the statements under the form we need in the sequel; the ideas and the arguments of the proof, however, essentially belong to Domenico Giannone.

Denote by $\{\zeta_{ij}; i,j\in\mathbb{N}\}$ a collection of complex numbers such that for all n the $n\times n$ matrices $\boldsymbol{\zeta}_n$ with entries $(\zeta_{ij}; 1\le i,j\le n)$ be hermitian. Denote by $\left\{\zeta_{n,ij}^T; 1\le i,j\le n, n\in\mathbb{N}, T\in\mathbb{N}\right\}$ 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 $\left(\zeta_{n,ij}^T; 1\le i,j\le n\right)$ be hermitian. Write $\lambda_{ni}(\boldsymbol{\zeta})$ and $\lambda_{ni}^T(\boldsymbol{\zeta})$, respectively, for $\boldsymbol{\zeta}_n$ and $\boldsymbol{\zeta}_n^T$'s eigenvalues 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 zero in a sense to be made precise in (7.16) below.

Lemma 2. *Assume that, for all $1\le i,j\le n$, $n\in\mathbb{N}$ and $T\in\mathbb{N}$, there exist a positive constant K that does not depend on n,T, i nor j, and a sequence of positive constants M_T depending on T only such that $M_T\to\infty$ as $T\to\infty$ and*

$$\mathrm{E}\left[\left|\zeta_{n,ij}^T-\zeta_{ij}\right|^2\right]\le KM_T^{-1}. \tag{7.16}$$

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

$$\max_{1\le k\le q_{max}}\mathrm{P}\left[M_T^{1/2}\frac{1}{n}\left|\lambda_{nk}(\boldsymbol{\zeta})-\lambda_{nk}^T(\boldsymbol{\zeta})\right|>B_\epsilon\right]\le\epsilon. \tag{7.17}$$

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

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

Proof of Lemma 2. Weyl's inequality implies that, for any hermitian matrices $\mathbf{A}$ and $\mathbf{B}$, with eigenvalues $\lambda_j(\mathbf{A})$ and $\lambda_j(\mathbf{B})$, respectively, $\max_j|\lambda_j(\mathbf{B})-\lambda_j(\mathbf{A})|^2\le\mathrm{tr}\left((\mathbf{B}-\mathbf{A})(\mathbf{B}-\mathbf{A})'\right)$. It follows that, for all n, T, and k,

$$\left|\lambda_{nk}^T(\boldsymbol{\zeta})-\lambda_{nk}(\boldsymbol{\zeta})\right|^2\le\mathrm{tr}\left((\boldsymbol{\zeta}_n^T-\boldsymbol{\zeta}_n)(\boldsymbol{\zeta}_n^T-\boldsymbol{\zeta}_n)'\right)=\sum_{i=1}^n\sum_{j=1}^n\left|\zeta_{n,ij}^T-\zeta_{ij}\right|^2.$$

Taking expectations, we thus have, in view of (7.16),

$$\mathrm{E}\left[\left|\lambda_{nk}^{T}(\boldsymbol{\zeta})-\lambda_{nk}(\boldsymbol{\zeta})\right|^{2}\right]\leq\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}\left[\left|\zeta_{n,ij}^{T}-\zeta_{ij}\right|^{2}\right]\leq n^{2}KM_{T}^{-1}$$

for all n, T, and k. The Markov inequality completes the proof. *Q.E.D.*

PROOF OF COROLLARY 1. From (3.4) there exist constants K_1, K_2, and T_0 such that

$$\sup_{n}\max_{1\leq i,j\leq n}\sup_{\theta}\left[\mathrm{E}\left|\boldsymbol{\Sigma}_{n}^{T}(\theta)-\boldsymbol{\Sigma}_{n}(\theta)\right|_{ij}^{2}\right]\leq K_{1}B_{T}^{-1}T^{-1}+K_{2}B_{T}^{4}$$

for any $T>T_0$. Therefore, $\boldsymbol{\Sigma}_n^T(\theta)$ and $\boldsymbol{\Sigma}_n(\theta)$ for all θ satisfy the assumption (7.16) of Lemma 2, with a constant $K=\max\left[K_1,K_2\right]$ and a rate $M_T=\max\left[B_T^{-1}T^{-1},B_T^4\right]$ that do not depend on θ. The corollary follows.

PROOF OF PROPOSITION 3: We will prove that, under (3.7), $\mathrm{P}\left[IC_n^T(q)\ <\ IC_n^T(k)\right]\to 1$ for all $k\neq q$, $k\ \leq\ q_{\max}$, as $\min(n,T)\to\infty$. For all $k<q$,

$$IC_{n}^{T}(q)\ <\ IC_{n}^{T}(k) \tag{7.18}$$

if and only if

$$\sum_{i=k+1}^{q}\frac{1}{T-1}\sum_{l=1}^{T-1}\frac{\lambda_{ni}^{T}(\theta_{l})}{n}>(q-k)p(n,T),$$

that is, in view of Corollary 2, if and only if

$$\sum_{i=k+1}^{q}\frac{1}{T-1}\sum_{l=1}^{T-1}\left[\frac{\lambda_{ni}(\theta_{l})}{n}+K_{1n}(T)\right]>(q-k)p(n,T), \tag{7.19}$$

where $K_{1n}(T)$ is $O_P\left(\max\left[B_T^2,B_T^{-1/2}T^{-1/2}\right]\right)$ uniformly in n and θ. By Assumption E, the first q eigenvalues $\lambda_{ni}(\theta)$ diverge linearly in n, which implies that there exists a dense set Ω in $[-\pi,\pi]$ such that

$$\sup_{\theta}\frac{\lambda_{ni}(\theta)}{n}=O(1)\qquad\text{and}\qquad\liminf_{n\to\infty}\sup_{\theta\in\Omega}\frac{\lambda_{ni}(\theta)}{n}>0,$$

for $i=k+1,...,q$. Since $K_{1n}(T)$ converges to 0, a sufficient condition for (7.18) to hold with probability one as $\min(n,T)\to\infty$ is that $p(n,T)\to 0$ as $\min(n,T)\to\infty$.

Next, for any $k>q$, (7.18) holds if and only if

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

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

$$\sum_{i=q+1}^{k}\frac{1}{T-1}\sum_{l=1}^{T-1}\left[\frac{\lambda_{ni}(\theta_{l})}{n}+K_{2n}(T)\right]\ <\ (k-q)p(n,T), \tag{7.20}$$

where $K_{2n}(T)$ is $O_P\left(\max\left[B_T^2, B_T^{-1/2}T^{-1/2}\right]\right)$ uniformly in n and θ. By Assumption E, $\lambda_{n\,q+1}(\theta)$, $\lambda_{n\,q+2}(\theta), \ldots$ are bounded uniformly in n and θ. Hence, $\sup_\theta \frac{\lambda_{ni}(\theta)}{n} = O(n^{-1})$ as $n \to \infty$ for $i = q+1, \ldots, k$. It is sufficient, for inequality (7.18) to hold with probability arbitrarily close to one as $\min(n,T) \to \infty$, that

$$np(n,T) \to \infty \quad \text{and} \quad \min\left[B_T^{-2}, B_T^{1/2}T^{1/2}\right] p(n,T) \to \infty,$$

as $\min(n,T) \to \infty$. The result follows. *Q.E.D.*

Turning to covariogram estimation, the proof of Proposition 4 relies on the following counterpart of Corollary 1.

Corollary 2. *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 $T > T_\epsilon$,*

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

Proof of Corollary 2. From (3.9) there exist constants L_1, L_2, and T_0 such that

$$\sup_n \max_{1\le i,j\le n} \sup_\theta \left[\mathrm{E}\left|\mathbf{\Sigma}_n^{*T}(\theta) - \mathbf{\Sigma}_n(\theta)\right|_{ij}^2\right] \le L_1 M_T T^{-1} + L_2 M_T^{-4}$$

for any $T > T_0$.

Therefore, $\mathbf{\Sigma}_n^{*T}(\theta)$ and $\mathbf{\Sigma}_n(\theta)$ for all θ satisfy the assumption (7.16) of Lemma 2, with a constant $K = \max[L_1, L_2]$ and a rate $M_T = \max\left[M_T T^{-1}, M_T^{-4}\right]$ that do not depend on θ. The corollary follows. *Q.E.D.*

Proof of Proposition 4: We will prove that $\mathrm{P}\left[IC_{a;n}^{*T}(q) < IC_{a;n}^{*T}(k)\right] \to 1$ for all $k \ne q$, $k \le q_{\max}$, $a = 1,2$, as $\min(n,T) \to \infty$ in such a way that (3.13) 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 $k < q$,

$$IC_{1;n}^{*T}(q) < IC_{1;n}^{*T}(k) \tag{7.21}$$

if and only if

$$\sum_{i=k+1}^{q} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \frac{\lambda_{ni}^{*T}(\theta_l)}{n} > (q-k)p(n,T), \tag{7.22}$$

that is, in view of Corollary 2, if and only if

$$\sum_{i=k+1}^{q} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{1n}(T)\right] > (q-k)p(n,T), \tag{7.23}$$

where $K_{1n}(T)$ is $O_P\left(\max\left[M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ uniformly in n and θ. By Assumption A5, the first q eigenvalues $\lambda_{ni}(\theta)$ diverge linearly in n, which implies that there exists a dense set Ω in

$[-\pi, \pi]$ such that

$$\sup_{\theta} \frac{\lambda_{ni}(\theta)}{n} = O(1) \quad \text{and} \quad \liminf_{n\to\infty} \sup_{\theta\in\Omega} \frac{\lambda_{ni}(\theta)}{n} > 0, \tag{7.24}$$

for $i = k+1, ..., q$. Since $K_{1n}(T)$ converges to 0, a sufficient condition for (**??**) to hold with probability tending to one 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,

$$IC^{*T}_{2;n}(q) < IC^{*T}_{2;n}(k) \tag{7.25}$$

for $k < q$ if and only if

$$\log \left[V_n^T(k)/V_n^T(q)\right] > (q-k)p(n,T), \tag{7.26}$$

where $V_n^T(k) := \sum_{i=k+1}^{n} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \lambda^{*T}_{ni}(\theta_l)/n$. In view of Corollary 2, we have, for $k = q$,

$$V_n^T(q) = \sum_{i=q+1}^{n} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{2n}(T)\right] \tag{7.27}$$

where $K_{2n}(T)$ is $O_P\left(\max\left[M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ uniformly in n and θ. By Assumption A5, the eigenvalues $\lambda_{ni}(\theta)$, $i > q$ are, uniformly in n and θ-a.e. in $[-\pi, \pi]$, bounded and bounded away from zero. Thus there exist positive constants c_0 and c_1 such that $\mathrm{P}\left[c_0 > V_n^T(q) > c_1\right] \to 1$ as $\min(n,T) \to \infty$. For $k < q$, we have

$$V_n^T(k) - V_n^T(q) = \sum_{i=k+1}^{q} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{3n}(T)\right] \tag{7.28}$$

where $K_{3n}(T)$ is $O_P\left(\max\left[M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ uniformly in n and θ-a.e. in $[-\pi, \pi]$. As (7.28) coincides with the left-hand side of (7.23), by the same argument as above, there exists a constant $c_2 > 0$ such that $\mathrm{P}\left[V_n^T(k) - V_n^T(q) > c_2\right] \to 1$, hence a a constant $c_3 > 0$ such that

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

as $\min(n,T) \to \infty$. The same condition that $p(n,T) \to 0$ is thus sufficient for both (7.22) and (7.26) to hold with probability tending to one as $\min(n,T) \to \infty$.

Next, for any $k > q$, (7.21) holds if and only if

$$\sum_{i=q+1}^{k} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \frac{\lambda^{*T}_{ni}(\theta_l)}{n} < (k-q)p(n,T), \tag{7.29}$$

that is, in view of Corollary 2, if and only if

$$\sum_{i=q+1}^{k} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{4n}(T)\right] < (k-q)p(n,T), \tag{7.30}$$

where $K_{4n}(T)$ is $O_P\left(\max\left[M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ uniformly in n, θ-a.e. in $[-\pi,\pi]$. As, $\lambda_{n\,q+1}(\theta)$, $\lambda_{n\,q+2}(\theta), \ldots$ are bounded uniformly in n and θ, $\sup_\theta \frac{\lambda_{ni}(\theta)}{n} = O(n^{-1})$ as $n \to \infty$ for $i = q+1, \ldots, k$. For $k > q$, it is thus sufficient for inequality (7.21) to hold with probability arbitrarily close to one as $\min(n,T) \to \infty$ that

$$np(n,T) \to \infty \quad \text{and} \quad \min\left[M_T^2, M_T^{-1/2}T^{1/2}\right]p(n,T) \to \infty,$$

as $\min(n,T) \to \infty$.

Turning to the logarithmic criterion, (7.25) holds for $k > q$ if and only if

$$\log\left[V_n^T(q)/V_n^T(k)\right] < (k-q)p(n,T). \tag{7.31}$$

Still in view of Corollary 2,

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

where $K_{5n}(T)$ is $O_P\left(\max\left[M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ uniformly in n and θ. By the same arguments as in (7.27), there exist positive constants c_4 and c_5 such that $\mathrm{P}\left[c_4 > V_n^T(k) > c_5\right] \to 1$ as $\min(n,T) \to \infty$. Similarly,

$$V_n^T(q) - V_n^T(k) = \sum_{i=q+1}^{k} \frac{1}{2M_T+1} \sum_{l=-M_T}^{M_T} \left[\frac{\lambda_{ni}(\theta_l)}{n} + K_{6n}(T)\right]$$

where $K_{6n}(T)$ is $O_P\left(\max\left[M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ uniformly in n and θ. This term coincides with the left-hand side of (7.30), and the same arguments imply that

$$V_n^T(q) - V_n^T(k) = O_P\left(\max\left[n^{-1}, M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$$

as $\min(n,T) \to \infty$. Thus, $\left(V_n^T(q) - V_n^T(k)\right)/V_n^T(k)$ and, therefore,

$$\log\left[\left(V_n^T(q) - V_n^T(k)\right)/V_n^T(k) + 1\right] = \log\left[V_n^T(q)/V_n^T(k)\right]$$

are also $O_P\left(\max\left[n^{-1}, M_T^{-2}, M_T^{1/2}T^{-1/2}\right]\right)$ as $\min(n,T) \to \infty$. Consequently, it is sufficient, for inequality (7.31) to hold with probability arbitrarily close to one as $\min(n,T) \to \infty$, that

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

as $\min(n,T) \to \infty$. This completes the proof. *Q.E.D.*

References

[1] Bai, J., and S. Ng (2002), "Determining the number of factors in approximate factor models," *Econometrica* 70, 191-221.

[2] Bai, J., and S. Ng (2005), "Determining the Number of Primitive Shocks in Factor Models", manuscript, Department of Economics, Harvard University.

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

[4] Brockwell, P.J., and R.A. Davis (1987), *Time Series: Theory and Methods,* New York: Springer-Verlag.

[5] Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2005), "Opening the black box", IAP Preprint, Université catholique de Louvain.

[6] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The generalized dynamic factor model: identification and estimation," *The Review of Economics and Statistics* 82, 540-554.

[7] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004), "The generalized dynamic factor model: consistency and rates," *Journal of Econometrics* 119, 231-255.

[8] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005), "The generalized dynamic factor model: one-sided estimation and forecasting," *Journal of the American Statistical Association,* forthcoming.

[9] Forni, M., and M. Lippi (2001), "The generalized factor model: representation theory," *Econometric Theory* 17, 1113-1141.

[10] Geweke, J. (1977), "The dynamic factor analysis of economic time series," in Dennis J. Aigner and Arthur S. Goldberg (eds.) *Latent Variable in Socio-Economic Models* (Amsterdam, North-Holland).

[11] Giannone, D. (2004), *Essays on Dynamic Factor Models of Large Cross-Sections*, Ph. D. Dissertation, ISRO, Université libre de Bruxelles, Brussels, Belgium.

[12] Giannone, D., L. Reichlin, and L. Sala (2005), "Monetary policy in real time", in M. Gertler and K. Rogoff, Eds, *NBER Macroeconomic Annual 2004*, MIT Press, 161-200.

[13] Onatski, A. (2005), "Determining the Number of Factors from Empirical Distribution of Eigenvalues", preprint, Economics Department, Columbia University.

[14] Parzen, E. (1957), "On consistent estimates of the spectrum of a stationary time series," *The Annals of Mathematical Statistics* 28, 329-348.

[15] Sargent, T.J. and C.A. Sims (1977), "Business cycle modelling without pretending to have too much *a priori* economic theory," in C. A. Sims (ed.) *New Methods in Business Research* (Minneapolis: Federal Reserve Bank of Minneapolis).

[16] Stewart, G.W. and J. Sun (1990), *Matrix Perturbation Theory*, San Diego: Academy Press, Inc.

[17] Stock, J. H. and M. W. Watson (2002a), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics* 20, 147-162.

[18] Stock, J. H. and M. W. Watson (2002b), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association* 97, 1167-1179.

[19] Stock, J. H. and M. W. Watson (2005), "Implications of dynamic factor models for VAR analysis", NBER Working Paper 11467.

[20] Weyl, H. (1912), "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen," *Mathematische Annalen* 71, 441-479.

Marc Hallin
Département de Mathématique, I.S.R.O., and E.C.A.R.E.S.
Université Libre de Bruxelles
Campus de la Plaine CP 210
B-1050 Bruxelles Belgium
mhallin@ulb.ac.be

Roman Liška
I.S.R.O. and E.C.A.R.E.S.
Université Libre de Bruxelles
Avenue Jeanne 44 CP 114
B-1050 Bruxelles Belgium
rliska@ulb.ac.be

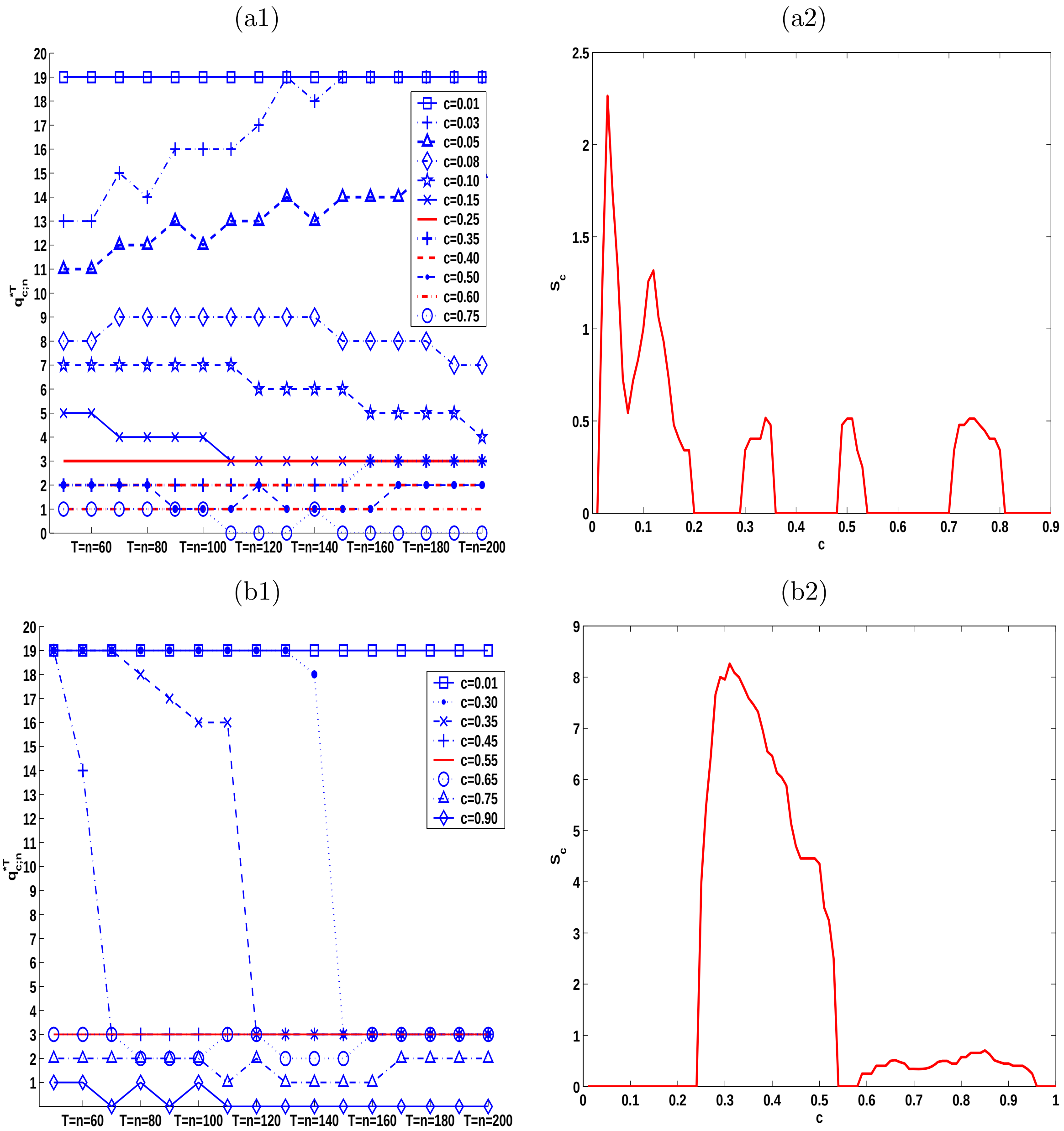

Figure 1: Example 1. MA loadings, $q = 3$, $n = T = 200$; $M_T = [0.7\sqrt{T}]$. Graphs of $(n_j, T_j) \mapsto q^{*T_j}_{c;n_j}$ and $c \mapsto S_c$ for $(n_j, T_j) = (50, 50), (60, 60), \ldots, (200, 200)$ and various values of c, using penalty function $p_3(n, T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$, $q_{\max} = 19$, and ((a1), (a2)) $IC^{*T}_{1;n}(k)$ criterion, ((b1), (b2)) $IC^{*T}_{2;n}(k)$ criterion, respectively.

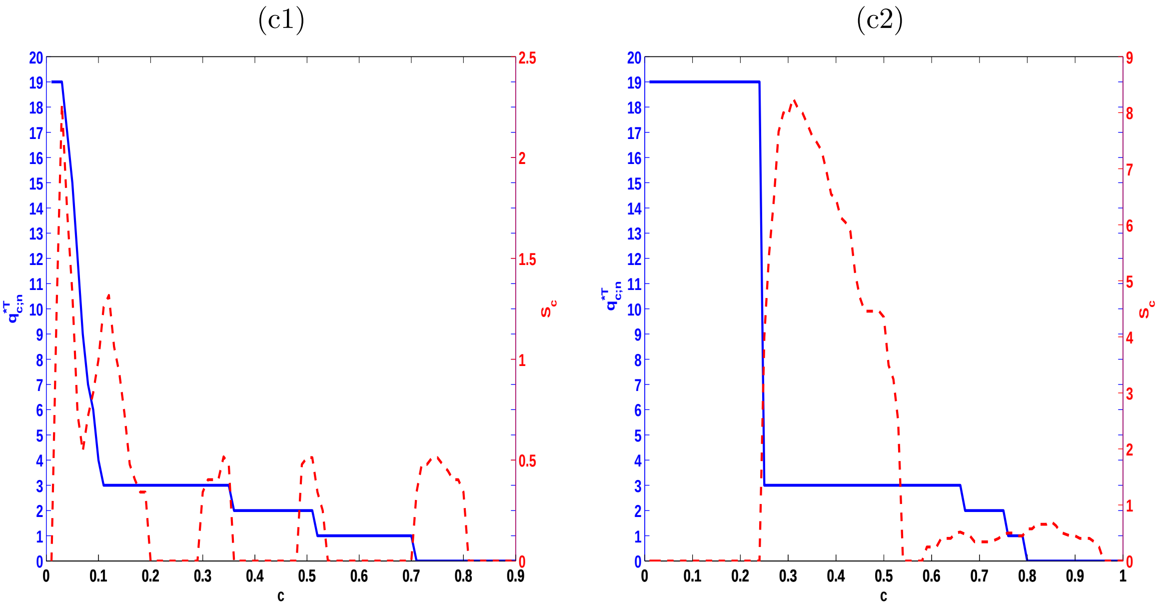

Figure 2: Example 1. MA loadings, $q = 3$, $n = T = 200$; $M_T = [0.7\sqrt{T}]$. Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^{*T}_{c,n}$, using penalty function $p_3(n,T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$, $q_{\max} = 19$, and (c1) $IC^{*T}_{1;n}(k)$ criterion, (c2) $IC^{*T}_{2;n}(k)$ criterion, respectively.

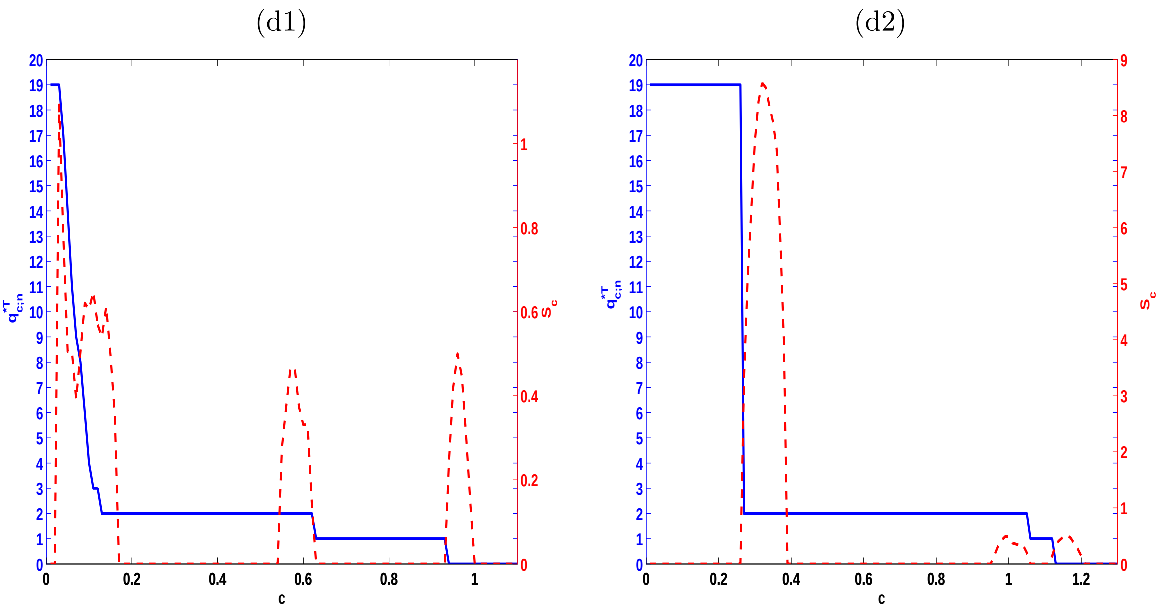

Figure 3: Example 2. AR loadings, $q = 2$, $n = 150$, $T = 120$; $M_T = [0.5\sqrt{T}]$. Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^{*T}_{c,n}$ for $n_j = 80 + 10j$, $j = 1, \ldots, 7$, $T_j = 60 + 10j$, $j = 1, \ldots, 6$, for penalty function $p_3(n,T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1} \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$, $q_{\max} = 19$, and (d1) $IC^{*T}_{1;n}(k)$ criterion, (d2) $IC^{*T}_{2;n}(k)$ criterion, respectively.

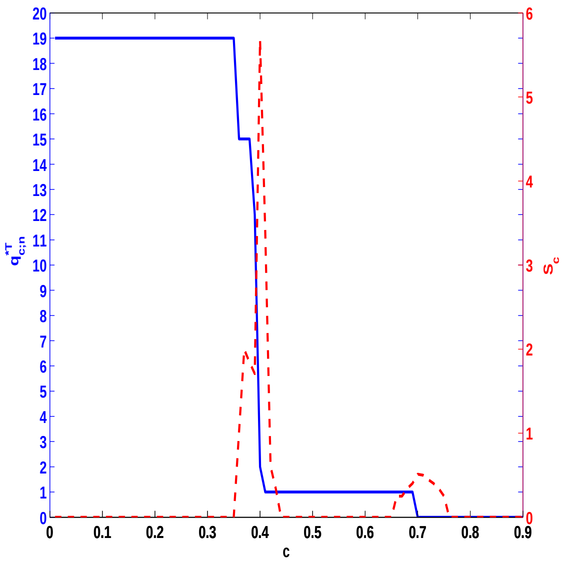

Figure 4: Eurozone macroeconomic indicators ($n = 62$, $T = 40$). Simultaneous plots of $c \mapsto S_c$ and $c \mapsto q^{*T}_{c,n}$ for $n_j = 47 + j$, $j = 1, \dots, 12$, $M_T = [0.5\sqrt{T}] = 3$, penalty function $p_1(n,T) = \left(M_T^{-2} + M_T^{1/2}T^{-1/2} + n^{-1}\right) \log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$, and the $IC^{*T}_{2;n}(k)$ criterion, respectively.

AR Model	T	n	under-identification						correct identification						over-identification					
$M_T = [0.5\sqrt{T}]$			(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)
$q = 1$	60	70	0	0	0	0	0	0	100	100	82	100	99	100	0	0	18	0	1	0
$q = 1$	90	90	0	0	0	0	0	0	100	100	84	100	99	100	0	0	16	0	1	0
$q = 1$	120	150	0	0	0	0	0	0	100	100	76	100	99	100	0	0	24	0	1	0
$q = 2$	60	70	0	0	0	0	0	0	100	99	97	99	100	99	0	1	3	1	0	1
$q = 2$	90	90	0	0	0	0	0	0	100	100	100	100	100	100	0	0	0	0	0	0
$q = 2$	120	150	0	0	0	0	0	0	100	100	100	100	100	100	0	0	0	0	0	0
$q = 3$	60	70	14	83	0	25	1	31	84	15	96	73	98	42	2	2	4	2	1	27
$q = 3$	90	90	0	4	0	0	0	0	99	95	100	100	99	100	1	1	0	0	1	0
$q = 3$	120	150	0	0	0	0	0	0	100	100	100	100	99	100	0	0	0	0	1	0
$M_T = [0.5\sqrt{T}]$			(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)
$q = 1$	60	70	0	0	0	0	0	0	99	100	85	100	99	100	1	0	13	0	1	0
$q = 1$	90	90	0	0	0	0	0	0	99	100	98	100	99	100	1	0	2	0	1	0
$q = 1$	120	150	0	0	0	0	0	0	100	100	96	100	100	100	0	0	4	0	0	0
$q = 2$	60	70	0	0	0	0	0	0	100	99	83	100	99	99	0	1	17	0	1	1
$q = 2$	90	90	0	0	0	0	0	0	99	100	99	100	100	100	1	0	1	0	0	0
$q = 2$	120	150	0	0	0	0	0	0	99	100	98	100	98	100	1	0	2	0	2	0
$q = 3$	60	70	8	37	0	61	2	85	90	14	71	37	96	13	2	49	29	2	2	2
$q = 3$	90	90	0	1	0	0	0	1	99	97	98	99	99	98	1	2	2	1	1	1
$q = 3$	120	150	0	0	0	0	0	0	100	100	97	100	100	100	0	0	3	0	0	0

Table 1: Relative frequencies of under-, correct and over-identification, frequencies (in %), over 500 replications, for $q = 1, 2$, and 3, AR generating device, $M_T = [0.5\sqrt{T}]$ and $M_T = [0.7\sqrt{T}]$, applying the automatic procedure of Section 4 with (a) penalty function $p_1(n,T) := \left(M_T^{-2} + M_T^{1/2}T^{-1/2} + n^{-1}\right)\log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$ and $IC_{1;n}^{*T}$ criterion; (b) penalty function $p_1(n,T)$ and $IC_{2;n}^{*T}$ criterion; (c) penalty function $p_2(n,T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1/2}$ and $IC_{1;n}^{*T}$ criterion; (d) penalty function $p_2(n,T)$ and $IC_{2;n}^{*T}$ criterion; (e) penalty function $p_3(n,T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1}\log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$ and $IC_{1;n}^{*T}$ criterion; (f) penalty function $p_3(n,T)$ and $IC_{2;n}^{*T}$ criterion, respectively.

MA Model	T	n	under-identification						correct identification						over-identification					
$M_T = [0.5\sqrt{T}]$			(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)
$q=1$	60	70	0	0	0	0	0	0	44	97	99	71	9	91	56	3	1	29	91	9
$q=1$	90	90	0	0	0	0	0	0	17	63	100	12	3	41	83	37	0	88	97	59
$q=1$	120	150	0	0	0	0	0	0	5	17	99	0	0	3	95	83	1	100	100	97
$q=2$	60	70	0	10	0	0	0	2	100	90	30	99	96	96	0	0	70	1	4	2
$q=2$	90	90	0	0	0	0	0	0	99	100	90	100	99	100	1	0	10	0	1	0
$q=2$	120	150	0	0	0	0	0	0	99	100	93	100	99	100	1	0	7	0	1	0
$q=3$	60	70	53	100	0	87	9	100	47	0	72	11	90	0	0	0	28	2	1	0
$q=3$	90	90	11	44	0	0	0	3	89	56	99	99	99	97	0	0	1	1	1	0
$q=3$	120	150	0	0	0	0	0	0	100	100	100	100	100	100	0	0	0	0	0	0
$M_T = [0.7\sqrt{T}]$			(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)	(a)	(b)	(c)	(d)	(e)	(f)
$q=1$	60	70	0	0	0	0	0	0	92	100	46	100	93	100	8	0	54	0	7	0
$q=1$	90	90	0	0	0	0	0	0	99	100	63	100	99	100	1	0	37	0	1	0
$q=1$	120	150	0	0	0	0	0	0	99	100	31	100	99	100	1	0	69	0	1	0
$q=2$	60	70	0	5	0	0	0	3	100	95	85	100	100	97	0	0	15	0	0	0
$q=2$	90	90	0	0	0	0	0	0	99	100	99	100	99	100	1	0	1	0	1	0
$q=2$	120	150	0	0	0	0	0	0	99	100	99	100	99	100	1	0	1	0	1	0
$q=3$	60	70	23	99	0	89	8	97	77	1	78	11	91	1	0	0	22	0	1	2
$q=3$	90	90	0	12	0	0	0	9	99	88	99	98	99	89	1	0	1	2	1	2
$q=3$	120	150	0	0	0	0	0	0	100	100	96	100	98	100	0	0	4	0	2	0

Table 2: Relative frequencies of under-, correct and over-identification, frequencies (in %), over 500 replications, for $q = 1, 2$, and 3, MA generating device, $M_T = [0.5\sqrt{T}]$ and $M_T = [0.7\sqrt{T}]$, of the number $\hat{q}$ of factors identified by applying the automatic procedure of Section 4 with (a) penalty function $p_1(n,T) := \left(M_T^{-2} + M_T^{1/2}T^{-1/2} + n^{-1}\right)\log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$ and $IC_{1;n}^{*T}$ criterion; (b) penalty function $p_1(n,T)$ and $IC_{2;n}^{*T}$ criterion; (c) penalty function $p_2(n,T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1/2}$ and $IC_{1;n}^{*T}$ criterion; (d) penalty function $p_2(n,T)$ and $IC_{2;n}^{*T}$ criterion; (e) penalty function $p_3(n,T) := \left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)^{-1}\log\left(\min\left[n, M_T^2, M_T^{-1/2}T^{1/2}\right]\right)$ and $IC_{1;n}^{*T}$ criterion; (f) penalty function $p_3(n,T)$ and $IC_{2;n}^{*T}$ criterion, respectively.