WORKING PAPER SERIES

Eigenvalue Ratio Estimators for the Number of Dynamic Factors

Maddalena Cavicchioli, Mario Forni, Marco Lippi and Paolo Zaffaroni

Working Paper 123

October 2016

www.recent.unimore.it

RECent: c/o Dipartimento di Economia Politica, Viale Berengario 51, I-41100 Modena, ITALY
Phone +39 059 2056856, Fax +39 059 2056947, E-mail: RECent@unimore.it

EIGENVALUE RATIO ESTIMATORS FOR THE NUMBER OF DYNAMIC FACTORS*

Maddalena Cavicchioli[†]

University of Modena and Reggio Emilia

Mario Forni[‡]

Università di Modena e Reggio Emilia, CEPR and RECent

Marco Lippi

Einaudi Institute for Economics and Finance, Roma

Paolo Zaffaroni

Imperial College London and Università di Roma La Sapienza

Abstract. In this paper we introduce three dynamic eigenvalue ratio estimators for the number of dynamic factors. Two of them, the Dynamic Eigenvalue Ratio (DER) and the Dynamic Growth Ratio (DGR) are dynamic counterparts of the eigenvalue ratio estimators (ER and GR) proposed by Ahn and Horenstein (2013). The third, the Dynamic eigenvalue Difference Ratio (DDR), is a new one but closely related to the test statistic proposed by Onatsky (2009). The advantage of such estimators is that they do not require preliminary determination of discretionary parameters. Finally, a static counterpart of the latter estimator, called eigenvalue Difference Ratio estimator (DR), is also proposed. We prove consistency of such estimators and evaluate their performance under simulation. We conclude that both DDR and DR are valid alternatives to existing criteria. Application to real data gives new insights on the number of factors driving the US economy.

Keywords: Generalized dynamic factor model, dynamic principal components, number of factors, static factor model.

JEL Classification: C01, C13, C38.

---

*The authors thank Alexei Onatski, Marc Hallin, Roman Liška and Alex Horenstein for sharing codes and providing useful suggestions.

[†]Research supported by the PRIN-MIUR Grant 2010J3LZEN and FAR 2014.

[‡]Research supported by the PRIN-MIUR Grant 2010J3LZEN and FAR 2014.

1. Introduction

In the last ten-fifteen years, "approximate" or "generalized" dynamic factor models have been applied quite successfully to the analysis of large panels of time series. Such data-rich environments are a common feature of macroeconomics and finance, where a few common shocks drive the comovements of many variables, so that information is scattered through a large number of interrelated series. Early theoretical contributions are Forni and Reichlin (1998), Forni *et al.* (2000, 2005), Forni and Lippi (2001), Stock and Watson (2002a), Bai and Ng (2002, 2007), Bai (2003). Applications include forecasting (Stock and Watson, 2002a, 2002b, Marcellino *et al.*, 2003, Boivin and Ng 2006, D'Agostino and Giannone, 2012), structural macroeconomic analysis (Bernanke and Boivin, 2003, Bernanke *et al.*, 2005, Giannone *et al.*, 2006, Favero *et al.* 2005, Eickmeier 2007, Forni *et al.*, 2009, Forni and Gambetti, 2010), nowcasting and business cycle indicators (Forni *et al.*, 2001, Cristadoro *et al.* 2005, Giannone *et al.* 2008, Altissimo *et al.*, 2010), the analysis of financial markets (Corielli and Marcellino 2006, Ludvigson and Ng 2007, 2009, Hallin *et al.*, 2011).

In large factor models, each variable, say x_{it}, is decomposed into the sum of two unobservable components, the "*common component*", χ_{it}, and the "*idiosyncratic component*", ξ_{it}. The idiosyncratic components are poorly correlated across sections in a sense that will be specified below; within a macroeconomic context, they can be interpreted as capturing sectoral elements and/or measurement errors. By contrast, the common components are driven by a small number q of unobservable shocks, u_{jt}, $j = 1, \ldots, q$, which are the same for all cross-sectional units. Such shocks, often called "dynamic factors", are loaded through one-sided linear filters, or impulse-response functions. The loadings are quite general, so that the model is flexible enough to accommodate pro-cyclical and counter-cyclical as well as leading, lagging and coincident behaviors. A restriction which we shall not impose in this paper, but is often assumed in the literature, is that the common components are contemporaneous linear combinations of $r \geq q$ unobservable variables f_{kt}, $k = 1, \ldots, r$, often called "static factors". In such a case, we say that the model admits a static factor representation. The dynamic nature of the model comes from the fact that the static factors have a dynamic representation in the common shocks.

The main feature distinguishing large approximate factor model from traditional dynamic factor models (Sargent and Sims, 1977, and Geweke, 1977) is the fact that the idiosyncratic components are not necessarily orthogonal to each other. This important generalization has the consequence that common and idiosyncratic components are no longer conceptually distinguishable to each other if the cross-sectional dimension is finite. This motivates the assumption of an infinite number of variables. The observed data are thought of as an (n, T)-dimensional realization of a double-indexed process x_{it}, $i = 1, \ldots, \infty$, $t = -\infty, \ldots, \infty$. Uniqueness of the common-idiosyncratic decomposition is ensured by appropriate assumptions about the asymptotic behavior, as n gets larger, of the covariance structure of the common and the idiosyncratic components, which, on the one hand, limit the total amount of idiosyncratic correlation and, on the other hand, impose that all common shocks be sufficiently "pervasive" along the cross sectional dimension. A pivotal implication of these assumptions is divergence of the q largest eigenvalues of the *spectral density matrix* of the first n processes, as $n \to \infty$, along with boundedness of the $q+1$-th. Such eigenvalues, call them $\lambda_{nk}(\theta)$, $\theta \in [-\pi, \pi]$, are often called "*dynamic eigenvalues* "[1]. An inverse result (Forni and Lippi, 2001) establishes the existence of a factor representation for the x's when such asymptotic behavior of the dynamic eigenvalues is assumed. Similarly, the existence of a well defined static factor representation is linked to an analoguous μ_{nk} of the *variance-covariance matrix* of the first n variables (Chamberlin and Rothschild, 1983).

A crucial preliminary step in the statistical analysis of large factor models is the estimation of the number q of dynamic factors, which is needed for the implementation of the estimation methods proposed in the literature. If the static factor representation restriction is imposed, both the number q of dynamic factors and the number r of static factors must be estimated. Given the above characterization of large factor models in terms of the asymptotic behavior of eigenvalues, it is not surprising that most existing criteria to determine the number of factors are based on the sample counterparts of the dynamic eigenvalues, say $\lambda^T_{nk}(\theta)$, and the static eigenvalues, say μ^T_{nk}. For n and T large enough, the first q estimated dynamic eigenvalues should be large as compared to the smallest $n-q$. Similarly, for the first r static eigenvalues.

Most of the estimators proposed in the literature for q and r, in analogy with information criteria such as AIC and BIC, entail minimization of a loss function which includes a "penalty" term, increasing in the number of factors (Bai and Ng, 2002, 2007, Amengual and Watson, 2007, Hallin and Liška, 2007, Alessi *et al.* 2010). As noticed by Onatski (2010), this is equivalent to retain a number of factors equal to the number of eigenvalues larger than a threshold value[2], given by the penalty term. A problem with the use of penalty functions is that they are discretionary to a large extent. Several functional forms can in principle satisfy the consistency requirement, and each one of them can be multiplied by an arbitrary constant, which calls for calibration to work properly in small samples. Hallin and Liška (2007) proposes an ingenious and effective method to calibrate the penalty function. The method, however, requires evaluation of the loss function over a grid n_j, T_j, $j = 1, \ldots, J$, and the outcome is sensitive to the choice of such a grid. Moreover, the use of sub-samples entails that the result may depend on the ordering of the

[1] A basic reference about dynamic eigenvalues is Brillinger (1981).

[2] Such a threshold level depends on the sample size n, T in such a way as to ensure consistency of the estimator as n and T get larger.

variables in the data set. Two notable exceptions to the use of penalty functions are Ahn and Horenstein (2013) and Onatski (2009). Firstly, Ahn and Horenstein (2013) concerns static factors. The "Eigenvalue Ratio" estimator is given by the number k maximizing the ratio of two adjacent eigenvalues, $\mathrm{ER}(k) = \mu_{nk}^T/\mu_{n,k+1}^T$. When both eigenvalues are either large or small, the ratio should be relatively small, but when $k = r$ the numerator is large and the denominator is small, so that the ratio should be large. The "Growth Rate" criterion GR is based on a similar idea. Such estimators are very effective under simulation, and do not require preliminary determination of nuisance parameters. Both ER and GR have quite natural dynamic counterparts, the "Dynamic Eigenvalue Ratio" (DER) and the "Dynamic Growth Rate" (DGR), which have never been studied so far. In particular, let $\theta_h = 2\pi h/(2M_T + 1)$, $h = -M_T, \dots, M_T$, and $\lambda_{nk}^T = \sum_{h=-M_T}^{M_T} \lambda_{nk}^T(\theta_h)$. Under standard assumption on the behavior of the window size M_T as $T \to \infty$, λ_{nk}^T is a consistent estimator of $\lambda_{nk} = \int_{-\pi}^{\pi} \lambda_{nk}(\theta)\, d\theta$, which is the variance of the k-th "dynamic" principal component of $x_{1t}, \dots, x_{nt}$[3]. Then the DER can be defined as

$$\mathrm{DER}(k) = \lambda_{nk}^T/\lambda_{n,k+1}^T.$$

The definition of DGR will be given below. In this paper we prove consistency of arg max $\mathrm{DER}(k)$, arg max $\mathrm{DGR}(k)$ and arg max $\mathrm{DDR}(k)$ as estimators of q, as $\min(n, T) \to \infty$. Moreover, we perform a few simulation experiments to evaluate the performance of such estimators, as compared to the methods proposed by Hallin and Liška (2007) and Onatski (2009), call them HL and O, respectively, and the DDR method introduced below.

Secondly, Onatski (2009) concerns dynamic factors. It proposes a test for the null of $q = k$ against the alternative of $k < q \le q_{\max}$. The test can be used sequentially as a device to estimate q; however, the procedure proposed in the paper requires preliminary choices which are discretionary to some extent and may affect the final result. An interesting similarity with Ahn and Horenstein (2013) is that the test statistic is based on an eigenvalue ratio, i.e.

$$\frac{\lambda_{nk}^T(\theta) - \lambda_{n,k+1}^T(\theta)}{\lambda_{n,k+1}^T(\theta) - \lambda_{n,k+2}^T(\theta)}.$$

In the present paper we define the "Dynamic eigenvalue Difference Ratio" (DDR) as the ratio above, where $\lambda_{nk}^T(\theta)$ is replaced by λ_{nk}^T:

$$\mathrm{DDR}(k) = \frac{\lambda_{nk}^T - \lambda_{n,k+1}^T}{\lambda_{n,k+1}^T - \lambda_{n,k+2}^T}.$$

In other words, we do not consider the variance of the principal components at a specific frequency, but the overall variance. Since aggregation is performed before computing the

[3]For a basic reference about "principal component series" or dynamic principal components see Brillinger (1981).

ratio, we do not need either to choose a specific frequency or to average in some way the potentially different results obtained at different frequencies. Having the above ratio, we simply use the arg max as our estimator of q (like Ahn and Horestein, 2013), rather than using the maximum to perform a recursive test. Since we do not have a test, we also avoid the need to choose a significance level either. The static equivalent of DDR, the "eigenvalue Difference Ratio" $\text{DR}(k) = (\mu^T_{nk} - \mu^T_{n,k+1})/(\mu^T_{n,k+1} - \mu^T_{n,k+2})$ can in principle be used as an alternative to ER and GR to determine the number of static factors r.
The basic idea behind DDR and DR is that the difference between diverging eigenvalues should be large, whereas the difference between bounded eigenvalues should be small. The illustration of Figure 1 can also be useful to get an intuition of the method. Consider the plot of λ^T_{nk} (or μ^T_{nk}) as a function of k. Then $\text{DDR}(k)$ (or $\text{DR}(k)$) is the ratio of two adjacent slopes of the polyline and therefore represents the percentage variation of slope in $k+1$. Going along the curve from the right to the left, the maximum is $\text{DDR}(k^*)$ (or $\text{DR}(k^*)$) as long as k^*+1 is the point where, so to speak, the climb becomes hard.

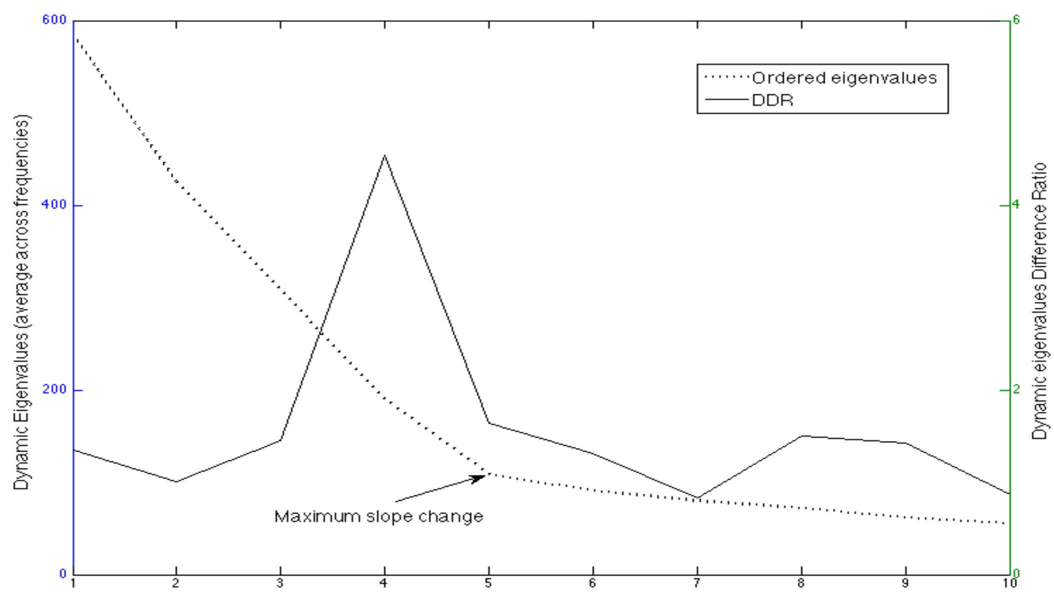

Figure 1: Estimated dynamic eigenvalues λ^T_{nk} (dotted line, left y axis) and $\text{DDR}(k)$ (solid line, right y axis), plotted as functions of i (x axis). Estimates are obtained from data generated with the ARMA specification 5.3 (Section 5), $(n, T) = (100, 200)$, $q = 4$. $\text{DDR}(k)$ reaches its maximum in k^* when the maximum slope change of the dotted line is in k^*+1.

Below, we prove that argmax $\text{DDR}(k)$ is a consistent estimator of q, as $\min(n, T) \to \infty$, and evaluate its performance in some experiments. In addition, we show that arg max

DR(k) is a consistent estimator of r as $\min(n, T) \to \infty$ and evaluate its performance under simulation, in comparison with the estimators ER and GR proposed by Ahn and Horenstein (2013).
Simulation results are the following: (i) DDR dominates DER and DGR in all experiments; (ii) DDR performs comparably or even better than both HL and O in most experiments; (iii) DR overperforms ER and GR when the static factors have different variances.
The paper is organized as follows. Section 2 describes the Generalized Dynamic Factor model along with its fundamental assumptions. Sections 3 and 4 introduce formally our criteria and state our consistency results. Section 5 presents numerical simulations in dynamic settings. Section 6 studies the DR criterion for the number of static factors along with some simulations. Section 7 concludes. Proofs are given in the Appendix.

2. The reference model

Our reference model is the Generalized Dynamic Factor Model (GDFM) introduced by Forni *et al.* (2000). The basic assumptions are reinforced following Hallin and Liška (2007), with slight variations. Precisely, the double-indexed processes x_{it}, χ_{it} and ξ_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$, conform to the following assumptions.

Assumption A1 (Forni *et al.*, 2000). For all $i \in \mathbb{N}$, $t \in \mathbb{Z}$,

$$
\begin{aligned}
x_{it} &= \chi_{it} + \xi_{it} \\
\chi_{it} &= \sum_{j=1}^{q} b_{ij}(L) u_{jt}
\end{aligned}
\tag{1}
$$

where

(i) the q-dimensional vector process $\mathbf{u}_t = (u_{1t} \ \cdots \ u_{qt})'$ is orthonormal white noise;

(ii) the n-dimensional vector process $\boldsymbol{\xi}_{nt} = (\xi_{1t} \ \cdots \ \xi_{nt})'$ is zero mean stationary for any $n \in \mathbb{N}$ and ξ_{it} is orthogonal to $u_{j\tau}$ for any $i, j \in \mathbb{N}$ and $t, \tau \in \mathbb{Z}$;

(iii) the filters $b_{ij}(L)$, $i \in \mathbb{N}$, $j = 1, \ldots, q$ are one-sided in the lag operator L and their coefficients are square summable, i.e., $\sum_{k=1}^{\infty} b_{ijk}^2 < \infty$ for all $i \in \mathbb{N}$ and $j = 1, \ldots, q$.

In matrix notation, we can write

$$
\mathbf{x}_{nt} = \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt}, \qquad \boldsymbol{\chi}_{nt} = \mathbf{B}_n(L)\mathbf{u}_t \tag{2}
$$

where $\mathbf{x}_{nt} = (x_{1t} \ \cdots \ x_{nt})'$, $\boldsymbol{\chi}_{nt} = (\chi_{1t} \ \cdots \ \chi_{nt})'$ and $\mathbf{B}_n(L)$ is the $n \times q$ matrix $(b_{ij}(L))_{\substack{i=1,\ldots,n \\ j=1,\ldots,q}}$.
The variables u_{jt} are the *common shocks* or *dynamic factors*. The variables χ_{it} and ξ_{it} are the *common components* and the *idiosyncratic components* of x_{it}, respectively.

Assumption A2 (Hallin and Liška, 2007). For all $n \in \mathbb{N}$,

$$\mathbf{x}_{nt} = \sum_{k=-\infty}^{\infty} \mathbf{C}_k \, \mathbf{z}_{t-k}$$

where $\mathbf{z}_t$ is an n-dimensional white noise with non-singular covariance matrix and finite fourth-order cumulants, and the $n \times n$ matrices $\mathbf{C}_k = (c_{ij,k})$ are such that

$$\sup_{i,j \in \mathbb{N}} \sum_{k=-\infty}^{\infty} |c_{ij,k}| \, |k|^{1/2} < \infty$$

for all $i, j = 1, \ldots, n$. Moreover, let $c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1})$ denote the cumulant of order ℓ of $x_{i_1,t+k_1}, \ldots, x_{i_\ell,t+k_{\ell-1}}$. For $\ell = 1, \ldots, 4$, we assume

$$\sup_{i_1,\ldots,i_\ell} \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_{\ell-1}=-\infty}^{\infty} |c_{i_1,\ldots,i_\ell}(k_1, \ldots, k_{\ell-1})| < \infty.$$

Let $\mathbf{\Sigma}_n(\theta)$, $\mathbf{\Sigma}^{\chi}_n(\theta)$, $\mathbf{\Sigma}^{\xi}_n(\theta)$, $\theta \in [-\pi, \pi]$, denote the spectral density matrices of $\mathbf{x}_{nt}$, $\boldsymbol{\chi}_{nt}$, $\boldsymbol{\xi}_{nt}$, respectively, and $\lambda_{ni}(\theta)$, $\lambda^{\chi}_{ni}(\theta)$, $\lambda^{\xi}_{ni}(\theta)$, $i = 1, \ldots, n$, denote their respective eigenvalues, in decreasing order of magnitude. Such eigenvalues are sometimes called *dynamic eigenvalues* to avoid confusion with the eigenvalues of the variance-covariance matrices of the corresponding processes.

Assumption A3 (Forni *et al.*, 2000). The first idiosyncratic dynamic eigenvalue $\lambda^{\xi}_{n1}(\theta)$ is uniformly bounded, i.e. there exists a real number Λ such that $\lambda^{\xi}_{n1}(\theta) \leq \Lambda$ for any $\theta \in [-\pi, \pi]$ and for any $n \in \mathbb{N}$.

Assumption A4. The q-th common dynamic eigenvalue $\lambda^{\chi}_{nq}(\theta)$ diverges almost everywhere, that is,

$$\lim_{n \to \infty} \lambda^{\chi}_{nq}(\theta) = \infty$$

a.e. in $[-\pi, \pi]$.

Assumptions A3 and A4 ensure uniqueness of the common and the idiosyncratic components appearing in decomposition (1), see Forni *et al.* (2000). Assumption A3 implies that the idiosyncratic components are, so to speak, "weakly" correlated across sections. It is substantially weaker than the orthogonality assumption which is typical of "exact" (as opposite to "approximate") factor models, in that, for instance, each one of the ξ_{it}'s can be correlated with ξ_{jt}, $j \in S$, S being a finite set of natural numbers. Notice that A4 does not rule out the case $\lambda^{\chi}_{nk}(0) = 0$, for any n and some $k \leq q$. Such a case may be economically interesting for macroeconomic data sets, since it holds when some of the common shocks (e.g. monetary policy shocks) have transitory effects on all variables.

The orthogonality Assumption A1(ii) implies that $\mathbf{\Sigma}_n(\theta) = \mathbf{\Sigma}_n^{\chi}(\theta) + \mathbf{\Sigma}_n^{\xi}(\theta)$. Weyl's inequality for Hermitian matrices implies

$$\lambda_{nk}(\theta) \geq \lambda_{nk}^{\chi}(\theta) \qquad \text{and} \qquad \lambda_{nk}(\theta) \leq \lambda_{nk}^{\chi}(\theta) + \lambda_{n1}^{\xi}(\theta)$$

for $k = 1, \dots, n$. Moreover, $\lambda_{nk}^{\chi}(\theta) = 0$ for $k > q$. Hence A1, A3 and A4 imply

Property A3′. $\lim_{n\to\infty} \lambda_{nq}(\theta) = \infty$ a.e. in $[-\pi, \pi]$.

Property A4′. $\lambda_{nq+1}(\theta) \leq \Lambda$ for any $\theta \in [-\pi, \pi]$ and for any $n \in \mathbb{N}$.

Forni and Lippi (2001) establish a remarkable converse result, i.e., if A2 holds, Properties A3′ and A4′ imply A1, A3 and A4[4]. In other words, under the maintained Assumption A2, divergence of the first q eigenvalues of the x's, along with boundedness of the $q+1$-th, ensure that the x's conform to the dynamic factor structure A1, A3 and A4, so that A1-A4 are equivalent to A2, A3′ and A4′. This equivalence is the dynamic factor analogue of the static factor representation result of Chamberlain and Rothschild (1983).

3. The estimators for finite n and infinite T

Clearly the selection of the number q of dynamic factors must be based on finite-sample information. Before turning to this task in Section 4, it is convenient to define our estimators in population and obtain consistency results as n approaches infinity, assuming that the spectral density matrices $\mathbf{\Sigma}_n(\theta)$ are known.

The first two criteria that we introduce are dynamic generalizations of the Eigenvalue Ratio (ER) criterion and the Growth Ratio (GR) criterion proposed by Ahn and Horestein (2013) for the estimation of the number of static factors within a static model. The *Dynamic Eigenvalue Ratio* is simply the ratio of the variances of two adjacent dynamic principal components, i.e.

$$\mathrm{DER}_n(k) = \frac{\lambda_{nk}}{\lambda_{n,k+1}},$$

where $\lambda_{nk} = \int_{-\pi}^{\pi} \lambda_{nk}(\theta)\, d\theta$. The *Dynamic Growth Ratio* is

$$\mathrm{DGR}_n(k) = \frac{\ln[V_n(k-1)/V_n(k)]}{\ln[V_n(k)/V_n(k+1)]} = \frac{\ln(1+\widetilde{\lambda}_{nk})}{\ln(1+\widetilde{\lambda}_{n,k+1})},$$

where

$$V_n(k) = \sum_{j=k+1}^{n} \lambda_{nj} \qquad \text{and} \qquad \widetilde{\lambda}_{nk} = \frac{\lambda_{nk}}{V_n(k)}.$$

[4]Indeed, in Forni and Lippi (2001) the maintained assumption is milder than Assumption A2.

Both the numerator and the denominator of DGR can be interpreted as the growth rates of the variances of the dynamic principal components up to index k. Lastly, the *Dynamic Difference eigenvalue Ratio* is closely related to test statistic proposed by Onatski (2009).

$$\mathrm{DDR}_n(k) = \frac{\lambda_{nk} - \lambda_{n,k+1}}{\lambda_{n,k+1} - \lambda_{n,k+2}}.$$

The population counterpart of our estimators are the values of k which maximize $\mathrm{DER}_n(k)$, $\mathrm{DGR}_n(k)$ and $\mathrm{DDR}_n(k)$, for $1 \le k \le q_{\max}$, where $q_{\max}$ is an upper bound for q chosen by the researcher:

$$\begin{aligned}
\widehat{q}_{n\mathrm{DER}} &= \underset{1\le k\le q_{\max}}{\arg\max}\, \mathrm{DER}_n(k) \\
\widehat{q}_{n\mathrm{DGR}} &= \underset{1\le k\le q_{\max}}{\arg\max}\, \mathrm{DGR}_n(k) \\
\widehat{q}_{n\mathrm{DDR}} &= \underset{1\le k\le q_{\max}}{\arg\max}\, \mathrm{DDR}_n(k).
\end{aligned}$$

Note that $\widehat{q}_{n\mathrm{DER}}$, $\widehat{q}_{n\mathrm{DGR}}$ and $\widehat{q}_{n\mathrm{DDR}}$ are deterministic; their sample analogues will be introduced in the next section.

Since we use eigenvalue ratios and eigenvalue difference ratios, we need to ensure that denominators are bounded away from zero. In addition, we need to rule out the case in which the diverging eigenvalues (or eigenvalue differences) diverge at different rates. Properties A3′ and A4′, which are already implied by A1-A4, are not sufficient to such purposes. Hence we make the following additional assumption.

Assumption A5.

(i) There exist positive constants c_k^+, c_k^- and a natural number N_1 such that

$$c_1^+ > \frac{\lambda_{n1}}{n} > c_1^- > \cdots > c_q^+ > \frac{\lambda_{nq}}{n} > c_q^-$$

and

$$c_{q+1}^+ > \lambda_{nq+1} > c_{q+1}^- > \cdots > c_{q_{\max}+2}^+ > \lambda_{n,q_{\max}+2} > c_{q_{\max}+2}^-$$

for all $n > N_1$.

(ii) Let $V_n(q)/n = (\sum_{k=q+1}^{n} \lambda_{nk})/n$. Then there exists a positive constant c and natural number N_2 such that $V_n(q)/n > c$ for all $n > N_2$.

(iii) The first $q_{\max}+2$ eigenvalues are distinct, i.e., $\lambda_{nk}(\theta) > \lambda_{n,k+1}(\theta)$, a.e. in $[-\pi, \pi]$, $k = 1, \dots, q_{\max}+2$.

A5(i) essentially says that the integrals of the diverging eigenvalues, along with the differences between adjacent eigenvalue integrals, diverge linearly with n. Moreover, the non-diverging eigenvalue integrals up to index $q_{\max} + 2$, along with their differences, are bounded away from zero. Linear divergence is a quite natural assumption in the present setting. Below, we shall assume a uniform upper bound for the spectral density functions of the x's (Assumption B2). This implies that the trace of $\mathbf{\Sigma}_n(\theta)$, and therefore the "large" eigenvalues, cannot diverge faster than n. On the other hand, a uniform lower bound for the variances of the x's is sufficient to ensure that λ_{n1} cannot diverge slower than n. Notice however that such very mild regularity conditions do not rule out slow divergence of λ_{nk} for $k > 1$. A5(i) excludes this possibility.

A5 (ii) is needed for the DGR criterion. It amounts to assuming that the average variance of the idiosyncratic components ξ_{it}'s is bounded away from 0 as $n \to \infty$; it rules out the (somewhat bizarre) case in which the ξ_{it}'s vanish as i increases.

A5(iii) is needed to ensure differentiability of the eigenvalues. See Theorem 9, Lancaster (1964).

Notice that A5 does not rule out the case $\lim_{n\to\infty} \lambda_{nk}(0)/n = 0$ for some $k \leq q$, which is necessarily verified when $\lambda^{\chi}_{nk}(0) = 0$ for all n, a case which, as argued above, may be of interest for macroeconomic data sets. In addition, we may have $\lim_{n\to\infty} \lambda_{n,q+1}(0) = 0$. This case can also be of interest when the x's are the first differences of I(1) variables, since it holds true when the idiosyncratic components of the original I(1) variables are already stationary.

Theorem 1. *Let Assumptions A1-A5 hold and $q \geq 1$. Then there is a natural number N such that, for any $n > N$,*

$$\widehat{q}_{n\mathrm{DER}} = \widehat{q}_{n\mathrm{DGR}} = \widehat{q}_{n\mathrm{DDR}} = q.$$

The proof is provided in the Appendix.[5]

4. The estimators for finite n and T

Coming to the sample level, we have to replace the population quantities $\widehat{q}_{n\mathrm{DER}}$, $\widehat{q}_{n\mathrm{DGR}}$ and $\widehat{q}_{n\mathrm{DDR}}$ with the feasible estimators $\widehat{q}^{T}_{n\mathrm{DER}}$, $\widehat{q}^{T}_{n\mathrm{DGR}}$ and $\widehat{q}^{T}_{n\mathrm{DDR}}$, based on finite realizations of the x's. To this end, in Subsection 4.1 we replace the population spectral density matrix $\mathbf{\Sigma}_n(\theta)$ with the lag-window estimator $\mathbf{\Sigma}^T_n(\theta)$ (similar results are obtained in Subsection 4.2 by using a periodogram smoothing estimator). Correspondingly, the eigenvalues $\lambda_{nk}(\theta)$ are replaced by the eigenvalues $\lambda^T_{nk}(\theta)$ of $\mathbf{\Sigma}^T_n(\theta)$. In addition, the integrals of the eigenvalue functions must be approximated numerically. We use simple Riemann sums over a finite

[5] The possibility $q = 0$ (i.e. $x_{it} = \xi_{it}$) could be allowed for, following Ahn and Horenstein (2013), by using a mock eigenvalue integral $\lambda_{n0} = d\,n$, where d is an arbitrary positive constant. The constant d however, would call for calibration.

number, increasing with T, of equally spaced points (different choices would not change the theoretical results). The sample eigenvalue ratios $\mathrm{DER}_n^T(k)$, $\mathrm{DGR}_n^T(k)$ and $\mathrm{DDR}_n^T(k)$ are then defined as the respective population ratios, where the integrals of the population eigenvalues are replaced with their estimators, i.e., the Riemann sums of the sample eigenvalues.

Under standard assumption (see Assumption B1 below), as T gets larger, all such estimators approach in probability their population counterparts for any given n. Therefore, by virtue of Theorem 1, they tend to q in probability for any $n > N$. Such fixed-n consistency, however, is not sufficient for our purposes, since we have to deal with a diverging cross-sectional dimension. An additional condition, borrowed from Hallin and Liška (2007), ensures that consistency is uniform in n (Assumption B2). As a consequence, the estimators approach q in probability as $\min(n, T) \to \infty$.

4.1. Lag window estimators

Let $\mathbf{\Gamma}_n^T(j)$ be the sample covariance matrix of $\mathbf{x}_{nt}$ and $\mathbf{x}_{n,t-j}$, i.e.

$$\mathbf{\Gamma}_n^T(j) = \frac{1}{T-j}\sum_{t=1}^{T} \mathbf{x}_{nt}\,\mathbf{x}'_{n,t-j}. \tag{3}$$

A possible estimator of the spectral density matrix is the $(2M_T + 1)$-point discrete Fourier transform of the truncated two-sided sequence $\mathbf{\Gamma}_n^T(-M_T)\omega_{-M_T}^T, \ldots, \mathbf{\Gamma}_n^T(0)\omega_0^T, \ldots, \mathbf{\Gamma}_n^T(M_T)\omega_{M_T}^T$, i.e.,

$$\mathbf{\Sigma}_n^T(\theta_h) = \frac{1}{2\pi}\sum_{j=-M_T}^{M_T} \mathbf{\Gamma}_n^T(j)\,\omega_j^T\,e^{-ij\theta_h},$$

where $\theta_h = 2\pi h/(2M_T + 1)$, $h = -M_T, \ldots, M_T$, and $\omega_j^T = \omega(j\,M_T^{-1})$, $\alpha \mapsto \omega(\alpha)$ is a positive even weight function satisfying condition B1 below. In our Monte Carlo exercises we use the triangular window given by

$$\omega_j^T = 1 - \frac{|j|}{M_T}.$$

Let $\lambda_{nk}^T(\theta_h)$ denote the k-th eigenvalue of $\mathbf{\Sigma}_n^T(\theta_h)$ in decreasing order of magnitude. We define

$$\mathrm{DER}_n^T(k) = \frac{\sum_{h=-M_T}^{M_T}\lambda_{nk}^T(\theta_h)}{\sum_{h=-M_T}^{M_T}\lambda_{n,k+1}^T(\theta_h)}$$

$$\mathrm{DGR}_n^T(k) = \frac{\ln[V_n^T(k-1)/V_n^T(k)]}{\ln[V_n^T(k)/V_n^T(k+1)]} = \frac{\ln(1+\widetilde{\lambda}_{nk}^T)}{\ln(1+\widetilde{\lambda}_{n,k+1}^T)}$$

where

$$V_n^T(k) = \sum_{j=k+1}^{n} \sum_{h=-M_T}^{M_T} \lambda_{nj}^T(\theta_h) \qquad\qquad \widetilde{\lambda}_{nk}^T = \frac{\sum_{h=-M_T}^{M_T} \lambda_{nk}^T(\theta_h)}{V_n^T(k)}.$$

$$\mathrm{DDR}_n^T(k) = \frac{\sum_{h=-M_T}^{M_T} \lambda_{nk}^T(\theta_h) - \sum_{h=-M_T}^{M_T} \lambda_{n,k+1}^T(\theta_h)}{\sum_{h=-M_T}^{M_T} \lambda_{n,k+1}^T(\theta_h) - \sum_{h=-M_T}^{M_T} \lambda_{n,k+2}^T(\theta_h)}.$$

Then our estimators of the number q of dynamic factors are

$$\begin{aligned}
\widehat{q}_{n\mathrm{DER}}^T &= \operatorname*{arg\,max}_{1\le k\le q_{\max}} \mathrm{DER}_n^T(k)\\
\widehat{q}_{n\mathrm{DGR}}^T &= \operatorname*{arg\,max}_{1\le k\le q_{\max}} \mathrm{DGR}_n^T(k)\\
\widehat{q}_{n\mathrm{DDR}}^T &= \operatorname*{arg\,max}_{1\le k\le q_{\max}} \mathrm{DDR}_n^T(k).
\end{aligned}$$

The additional assumptions we need for our consistency result are the following.

Assumption B1.

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

(ii) $\alpha \mapsto \omega(\alpha)$ is an even piecewise continuous function, piecewise differentiable up to order three, with bounded first three derivatives, satisfying $\omega(0) = 1$, $|\omega(\alpha)| \le 1$ for all α and $\omega(\alpha) = 0$ for $|\alpha| > 1$.

Assumption B2 (Hallin and Liška, 2007).
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 order two; namely, there exists $Q < \infty$ such that

$$\sup_{i,j\in\mathbb{N}} \sup_{\theta\in[-\pi,\pi]} \left| \frac{d^r \sigma_{ij}(\theta)}{d\theta^r} \right| \le Q$$

for $r = 0, 1, 2$.

Assumption B1 is a standard assumption which is needed to ensure consistency of the estimator of the spectral density matrix. Assumption B2 ensures that consistency is uniform in n (see Hallin and Liška, 2007, equation (5)).

Let us now state our main result, which is proven in the Appendix.

Theorem 2. *Let Assumptions A1 to A5, B1 and B2 hold and $q \ge 1$. Further assume that $\lim nM_T^{*-1} = 0$ as n and T go to infinity, with $M_T^* = \max(M_T^{-2}, M_T^{1/2}T^{-1/2})$. Then*

$$\mathrm{plim}_{m\to\infty}\, \widehat{q}_{n\mathrm{DER}}^T = \mathrm{plim}_{m\to\infty}\, \widehat{q}_{n\mathrm{DGR}}^T = \mathrm{plim}_{m\to\infty}\, \widehat{q}_{n\mathrm{DDR}}^T = q$$

where $m = \min(n, T)$.

4.2. Periodogram smoothing estimators

A variant of the lag window estimator is given by the periodogram smoothing estimator

$$\boldsymbol{\Sigma}_n^{T*}(\theta) = \frac{2\pi}{T} \sum_{t=1}^{T-1} W_T\left(\theta - \frac{2\pi t}{T}\right) \mathbf{I}_n^T\left(\frac{2\pi t}{T}\right) \tag{4}$$

where $\mathbf{I}_n^T(\alpha)$ is the periodogram matrix defined as

$$\mathbf{I}_n^T(\alpha) = \frac{1}{2\pi T} \sum_{t=1}^{T} \mathbf{x}_{nt} e^{-i\alpha t} \sum_{t=1}^{T} \mathbf{x}_{nt}' e^{i\alpha t}$$

and $W^T(\alpha)$, $-\infty < \alpha < \infty$, $T \in \mathbb{N}$ is a family of periodic functions with period 2π, which on $(-\pi, \pi]$ are defined as

$$W_T(\alpha) = \frac{1}{B_T} \sum_{j=-\infty}^{\infty} W\left(\frac{\alpha + 2\pi j}{B_T}\right).$$

Here $W(\beta)$ is a positive even function, independent of T, with bounded derivative, satisfying $\int_{-\infty}^{\infty} W(\beta)d\beta = 1$, $\int_{-\infty}^{\infty} |\beta| W(\beta)d\beta < \infty$ and B_T is a bandwidth satisfying $B_T \to 0$, $B_T T \to \infty$ as $T \to \infty$.

It is known that, irrespective of the precise form of $W(\beta)$, the periodogram smoothing estimator $\boldsymbol{\Sigma}_n^{T*}(\theta)$ is a consistent estimator of $\boldsymbol{\Sigma}_n(\theta)$, as $T \to \infty$ (see Brillinger (1981), Section 7.4).

In our Monte Carlo exercise below we use a simple average of the periodogram across the frequencies in the relevant band. This estimator is obtained by assuming the Daniell window

$$W(\beta) = \begin{cases} 1/(2\pi) & \text{if } |\beta| \le \pi \\ 0 & \text{if } |\beta| > \pi, \end{cases} \tag{5}$$

which gives

$$W_T(\alpha) = \begin{cases} 1/(2\pi B_T) & \text{if } |\alpha| \le B_T\pi \\ 0 & \text{if } |\alpha| > B_T\pi. \end{cases}$$

Let us now consider a frequency grid θ_h, $h = -M_T, \dots, M_T$. Let $\lambda_{nk}^{T*}(\theta_h)$ denote the k-th eigenvalue of $\boldsymbol{\Sigma}_n^{T*}(\theta_h)$ in decreasing order of magnitude. The corresponding information criteria are

$$\mathrm{DER}_n^{T*}(k) = \frac{\sum_{h=-M_T}^{M_T} \lambda_{nk}^{T*}(\theta_h)}{\sum_{h=-M_T}^{M_T} \lambda_{n,k+1}^{T*}(\theta_h)}$$

$$\mathrm{DGR}_n^{T*}(k) = \frac{\ln[V_n^{T*}(k-1)/V_n^{T*}(k)]}{\ln[V_n^{T*}(k)/V_n^{T*}(k+1)]} = \frac{\ln(1+\widetilde{\lambda}_{nk}^{T*})}{\ln(1+\widetilde{\lambda}_{n,k+1}^{T*})}$$

where

$$V_n^{T*}(k) = \sum_{j=k+1}^{n} \sum_{h=-M_T}^{M_T} \lambda_{nj}^{T*}(\theta_h) \qquad \widetilde{\lambda}_{nk}^{T*} = \frac{\sum_{h=-M_T}^{M_T} \lambda_{nk}^{T*}(\theta_h)}{V_n^{T*}(k)}$$

$$\mathrm{DDR}_n^{T*}(k) = \frac{\sum_{h=-M_T}^{M_T} \lambda_{nk}^{T*}(\theta_h) - \sum_{h=-M_T}^{M_T} \lambda_{n,k+1}^{T*}(\theta_h)}{\sum_{h=-M_T}^{M_T} \lambda_{n,k+1}^{T*}(\theta_h) - \sum_{h=-M_T}^{M_T} \lambda_{n,k+2}^{T*}(\theta_h)}.$$

The estimators $\widehat{q}_{n\mathrm{DER}}^{T*}$, $\widehat{q}_{n\mathrm{DGR}}^{T*}$ and $\widehat{q}_{n\mathrm{DDR}}^{T*}$ are defined as the arguments of the maxima of the respective criteria.

5. Numerical simulations

To evaluate the performance of the criteria defined in the previous section, we conduct three Monte Carlo experiments with three different specifications of Model (1).

First experiment. The first DGP follows the one proposed by Hallin and Liška (2007), Section 5. Precisely:

i. The common shocks u_{kt}, $k = 1, \ldots, q$, $t = 1, \ldots, T$, are $iid \sim \mathcal{N}(0, D_k)$, with $D_1 = 1$, $D_2 = .5$ and $D_3 = 1.5$.

ii. The idiosyncratic components are of form $\xi_{it} = \sum_{j=0}^{4} \sum_{k=0}^{2} g_{i,j,k}\, v_{i+j,t-k}$, where the v_{it}'s are $iid \sim \mathcal{N}(0,1)$, and the $g_{i,j,k}$'s are $iid \sim \mathcal{U}_{[1,1.5]}$, where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 1, \ldots, 4$, $k = 0, 1, 2$. To ensure both autocorrelation and cross-correlation among idiosyncratic, the v_{it}'s and the $g_{i,j,k}$'s are mutually independent and independent of the u_{it}'s.

iii. the filters $b_{ik}(L)$, $i = 1, \ldots, n$, $k = 1, \ldots, q$, are randomly generated (independently from the u_{kt}'s and ξ_{it}'s) by one of the following devices: (1) MA loadings: $b_{ik}(L) = b_{ik,0} + b_{ik,1}L + b_{ik,2}L^2$ with iid and mutually independent coefficients $(b_{ik,0}, b_{ik,1}, b_{ik,2}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$; (2) AR loadings: $b_{ik}(L) = b_{ik,0}(1 - b_{ik,1}L)^{-1}(1 - b_{ik,2}L)^{-1}$ with iid and mutually independent coefficients $b_{ik,0} \sim \mathcal{N}(0,1)$, $b_{ik,1} \sim \mathcal{U}_{[.8,.9]}$ and $b_{ik,2} \sim \mathcal{U}_{[.5,.6]}$.

Finally, for each i, the variance of ξ_{it} and that of the common component $\sum_{k=1}^{q} b_{ik}(L)$ are normalized to 0.5.

The artificial samples were generated with $q = 2, 3$ and (n, T)= $(60, 100)$, $(100, 100)$, $(70, 120)$, $(120, 120)$, $(150, 120)$.

Second experiment. The second DGP is the one studied by Onatski (2009), Sections 5.1 and 5.3. Precisely:

i. The common shocks u_{kt}, $k = 1, \dots, q$, $t = 1, \dots, T$, are $iid \sim \mathcal{N}(0, I_k)$.

ii. The idiosyncratic components follow AR(1) processes both cross-sectionally and over time: $\xi_{it} = \rho_i \xi_{i,t-1} + v_{it}$, $v_{it} = \rho v_{i-1,t} + \epsilon_{it}$, where $\rho_i \sim iid\mathcal{U}_{[-.8,.8]}$, $\rho = .2$ and $\epsilon_{it} \sim iid\mathcal{N}(0, 1)$.

iii. The filters $b_{ik}(L)$, $i = 1, \dots, n$ and $k = 1, \dots, q$, are randomly generated (independently from the u_{kt}'s and ξ_{it}'s) by one of the following devices: (1) MA loadings: $b_{ik}(L) = b_{ik,0}(1 + b_{ik,1}L)(1 + b_{ik,2}L)$ with iid and mutually independent coefficients $b_{ik,0} \sim \mathcal{N}(0, 1), b_{ik,1} \sim \mathcal{U}_{[0,1]}$ and $b_{ik,2}) \sim \mathcal{U}_{[0,1]}$; (2) AR loadings: same as in the first experiment; $b_{ik}(L) = b_{ik,0}(1 - b_{ik,1}L)^{-1}(1 - b_{ik,2}L)^{-1}$ with iid and mutually independent coefficients $b_{ik,0} \sim \mathcal{N}(0, 1), b_{ik,1} \sim \mathcal{U}_{[.8,.9]}$ and $b_{ik,2} \sim \mathcal{U}_{[.5,.6]}$.

For each i, the idiosyncratic component ξ_{it} and the common component $\chi_{it} = \sum_{k=1}^{q} b_{ik}(L)$ are normalized so that their variances equal $\sigma^2[1 - (0.4 + 0.05q)]$ and $0.4 + 0.05q$, respectively. Following Onatski (2009), we set $q = 2$ and (n, T, σ^2) equal to $(70, 70, 1)$, $(70, 70, 2)$, $(70, 70, 4)$, $(100, 120, 1)$, $(100, 120, 2)$, $(100, 120, 6)$, $(150, 500, 1)$, $(150, 500, 8)$, $(150, 500, 16)$.

Third experiment. The third GDP is the following.

i. The common shocks u_{kt}, $k = 1, \dots, q$, $t = 1, \dots, T$, are $iid \sim \mathcal{N}(0, I_k)$ for case (a) and normal iid with random variance between 1 and 1.5 for case (b).

ii. Same as in the second experiment. The idiosyncratic components follow AR(1) processes both cross-sectionally and over time: $\xi_{it} = \rho_i \xi_{it-1} + v_{it}$, $v_{it} = \rho v_{i-1t} + \epsilon_{it}$, where $\rho_i \sim iid\mathcal{U}_{[-.8,.8]}$, $\rho = 0.2$ and $\epsilon_{it} \sim iid\mathcal{N}(0, 1)$.

iii. the filters $b_{ik}(L)$, $i = 1, \dots, n$ and $k = 1, \dots, q$, are randomly generated (independently from the u_{kt}'s and ξ_{it}'s) with ARMA loadings: $b_{ik}(L) = (m_{ik,0} + m_{ik,1}L + m_{ik,2}L^2)(a_{ik,0}(1 - a_{ik,1}L)^{-1})$, where coefficients are iid and mutually independent and $m_{ik,s} \sim \mathcal{U}_{[-1,1]}$, $s = 0, 1, 2$, and $a_{ik,r} \sim \mathcal{U}_{[-0.8,0.8]}$, $r = 0, 1$.

In this experiment we want to control for the common to idiosyncratic variance ratio without forcing all variables in the cross section to have the same ratio. To this end, having computed the common components from drawing j, we compute the square root of the average sample variance, say $\tau(j)$, and use this overall measure of common volatility to normalize the idiosyncratic components in two different ways: (1) all idiosyncratic components are multiplied by $\sqrt{0.5}\,\tau(j)$ (large idiosyncratic components); (2) all idiosyncratic components are multiplied by $\sqrt{0.2}\,\tau(j)$ (small idiosyncratic components). Since the variance of the idiosyncratic components produced with device (ii) above is on average 1.42,[6]

[6]We computed the average sample variance over 10000 replications.

the idiosyncratic to common variance ratio is on average 0.72 in the large idiosyncratic case and 0.28 in the small idiosyncratic case, irrespective of q.

We set $q = 2, 4, 6$ and (n, T)= $(50, 80)$, $(120, 80)$, $(50, 240)$, $(120, 240)$, $(240, 480)$.

To compute DER, DGR and DDR, we use the lag window estimator (3) with the triangular smoothing function $\omega_j = 1 - |j|M_T^{-1}$ and truncation parameter $M_T = [0.75\sqrt{T}]$. To compute DER*, DGR* and DDR* we use the periodogram smoothing estimator (4) with the Daniell window (5) and bandwidth $B_T = (2M_T + 1)/T$, $M_T = \lceil\sqrt{T}\rceil$. For both methods, we evaluate the eigenvalues in the frequency grid $\theta_h = 2\pi h/T$, $h = -M_T, \dots, M_T$.

We compare our criteria with the methods proposed by Hallin and Liška (2007) and Onatski (2009). With regard to Hallin and Liška estimator (HL), we use the log information criterion $IC_{2;n}^T$ with penalty $p_1(n, T)$ and the Bartlett lag window with truncation parameter $M_T = \lceil 0.75\sqrt{T}\rceil$, which yield the best performance in the simulations shown by the authors. The method requires evaluation of the loss function over a grid n_j, T_j, $j = 1, \dots, J$. We stick to the one proposed by the authors, i.e. $n_j = n - 10j$, $T_j = T - 10_j$, $j = 0, 1, 2, 3$.

When dealing with Onatski's method (O), we use the procedure described in Section 5.3 of the quoted paper. We found that the results are sensitive to the choice of the parameter m (Onatski, 2009, footnote 7). For the second experiment, we stick to Onatski's choice, which is very effective. For the first DGP, we use $m = 15$. For the third experiment, we use $m = 15, 20, 30$ for $T = 80, 240, 480$, respectively. These values produce better results than the ones suggested in Onatski's paper.

For all experiments and all estimators we set $q_{\max} = 8$. For all experiments we generate 500 artificial data sets. We evaluate the results by using the percentage of correct answers.

Table 1 reports results for the first experiment. Boldface numbers denote the estimator(s) which perform best for each q, n, T configuration. Results for HL are very close to those reported in Hallin and Liška (2007). HL has the best overall performance with MA loadings, but is clearly beaten by DDR and DDR* with AR loadings, case $q = 3$. DDR and DDR* perform similarly to one another and dominate DER, DGR, DER* and DGR* and O.

Table 2 reports results for the second experiment. Results for O are close to those reported in Onatski (2009). With MA loadings, O is the best method for all n, T, σ^2 configurations. DDR and DDR* are close to each other and O. HL does not work well for large σ^2. With AR loadings, DDR and DDR* perform the best for all n, T, σ^2 configurations. Again HL has problems with large σ^2, but for the case $n = 500$. For both MA and AR loadings, DER, DGR, DER* and DGR* have a good performance but are dominated by DDR and DDR*.

Table 3 reports results for the third experiment, case (a), in which all common shocks have the same variance. DDR dominates all other methods for all but two rows in the upper panel (Large idiosyncratic components): the configuration $(q, n, T) = (6, 50, 80)$ (best result O) and $(q, n, T) = (6, 120, 240)$ (best result HL). DDR and DDR* almost uniformly dominate DER, DGR, DER* and DGR* (but DDR performs slightly better than DDR*).

Table 4 reports results for the third experiment, case (b), in which common shocks have different variances. Again, DDR dominates all other methods in almost all cases. The only one notable exception is the configuration $(q, n, T) = (6, 120, 240)$, large idiosyncratic component (best result HL).

Overall, DDR has the best score in 78 out of 98 rows of Tables 1-4. In the remaining 20 raws, it ranks second (16 times) or third (4 times). When DDR is not the best method, the difference with respect to the best method is very small in almost all cases, the largest percentage deviation being DDR/HL=83/95 in the configuration $(q, n, T) = (6, 120, 240)$, large idiosyncratic component, Table 4.

We conclude that DDR is a valid alternative to existing methods.

6. The static factor representation and the DR estimator

As explained in the introduction, Ahn and Horestein (2013) proposes two estimators, called ER (Eigenvalue Ratio) and GR (Growth Ratio), for determining the number r of factors in static factor models. In this section we propose a new criterion, the DR (*eigenvalue Difference Ratio*), which is the static eigenvalue equivalent of the DDR discussed above. Notice that the DR estimator is closely related to Onatski's (2010) estimator, which is the maximum k satisfying $\mu_k^T - \mu_{k+1}^T \geq \delta$, μ_k^T being the kth eigenvalue of the variance-covariance matrix of the data and δ a given threshold. Intuitively, using ratios in place of differences enables us to avoid the calibration of δ. Following Forni *et al.* (2009), we first introduce assumptions under which the generalized dynamic factor model admits a static factor representation. Then we prove consistency of DR under suitable condition as $\min(n, T) \to \infty$. Finally we perform a simulation exercise showing that DR performs well, particularly when some factors have small explanatory power.

Assumption C1. There exist an integer $r \geq q$, a nested sequence of $n \times r$ matrices $\mathbf{\Lambda}_n$, and a one-sided $r \times q$ matrix polynomial $\mathbf{N}(L)$, independent of n, such that $\mathbf{B}_n(L) = \mathbf{\Lambda}_n \mathbf{N}(L)$, $n = 1, \ldots, \infty$.

Let $\mathbf{f}_t = \mathbf{N}(L)\mathbf{u}_t$. Then equation (2) can be rewritten as

$$\mathbf{x}_{nt} = \mathbf{\Lambda}_n \mathbf{f}_t + \boldsymbol{\xi}_{nt}. \tag{6}$$

The above representation is sometimes called "static", since the "static factors", i.e., the r entries of $\mathbf{f}_t$, are loaded contemporaneously by the x's. Assumption C1 amounts to assuming that the common components χ_{it}, $i = 1, \ldots, n$ span a finite-dimensional space for each t. It is easily seen that C1 is fulfilled when $\mathbf{B}_n(L)$ is a finite moving average for all n, since in this case χ_{it} can be written in the form (6) with $\mathbf{f}_t$ is a linear combination of $\mathbf{u}_t, \cdots, \mathbf{u}_{t-p}$, p being the order of $\mathbf{B}_n(L)$. By contrast, a simple example of a dynamic model which does not admit a static factor representation is

$$x_{it} = \frac{1}{1-\alpha_i L} u_t + \xi_{it}$$

(see Forni and Lippi, 2011), since in this case the linear space spanned by χ_{it}, $i = 1, \ldots, n$, is infinite-dimensional (unless the α_i's satisfy special restrictions).

Assumption C1 is not sufficient to guarantee uniqueness of the number of static factors r. A standard assumption ensuring uniqueness is the following (Chamberlain and Rothschild, 1983). Let Γ_n^χ be the variance-covariance matrix of $\boldsymbol{\chi}_{nt}$ and μ_{nk}^χ be its k-th eigenvalue in descending order of magnitude.

Assumption C2. $\mu_{nr}^\chi \to \infty$ as $n \to \infty$.

Let $\boldsymbol{\Gamma}_n$ and $\boldsymbol{\Gamma}_n^\xi$ be the variance-covariance matrices of $\mathbf{x}_{nt}$ and $\boldsymbol{\xi}_{nt}$, respectively. Moreover, let μ_{nk} and μ_{nk}^ξ be their k-th eigenvalues in decreasing order of magnitude. Mutual orthogonality of the common and the idiosyncratic components implies that $\boldsymbol{\Gamma}_n = \boldsymbol{\Gamma}_n^\chi + \boldsymbol{\Gamma}_n^\xi$. By Weyl's inequalities we have

$$\mu_{nk}^\chi + \mu_{nn}^\xi \le \mu_{nk} \le \mu_{nk}^\chi + \mu_{n1}^\xi, \qquad k = 1, \ldots, n.$$

Hence $\mu_{nr}^\chi + \Lambda \ge \mu_{nr} \ge \mu_{nr}^\chi$, so that C2 is equivalent to divergence of μ_{nr} as $n \to \infty$. In addition, $\mu_{n,r+1} \le \mu_{n1}^\xi$, since $\mu_{n,r+1}^\chi = 0$. But obviously μ_{n1}^ξ cannot be larger than λ_{n1}^ξ, so that by Assumption A3 both μ_{n1}^ξ and $\mu_{n,r+1}$ are bounded above by Λ. However, divergence of μ_{nr} along with boundedness of $\mu_{n,r+1}$ are not sufficient for our purposes. We then reinforce C2 by replacing it with the following assumption, which is the "static" analogue of A5(i).

Assumption C2$'$.
There exist positive constants d_k^+, d_k^- and natural numbers N_3 and $r_{\max} > r$ such that

$$d_1^+ > \frac{\mu_{n1}}{n} > d_1^- > \cdots > d_r^+ > \frac{\mu_{nr}}{n} > d_r^-$$

and

$$d_{r+1}^+ > \mu_{nr+1} > d_{r+1}^- > \cdots > d_{r_{\max}+2}^+ > \mu_{n,r_{\max}+2} > d_{r_{\max}+2}^-$$

for all $n > N_3$.

C2$'$ is not particularly restrictive; however, it should be noticed that linear divergence of the divergent eigenvalues rules out "weak" factors in the sense of Onatski (2010).

We are now ready to introduce our estimator. To simplify notation, from now on we write $\mathbf{\Gamma}_n^T$ in place of $\mathbf{\Gamma}_n^T(0)$ (see equation (3)) to denote the sample variance-covariance matrix of $\mathbf{x}_{nt}$, i.e., $\mathbf{\Gamma}_n^T = (T)^{-1}\sum_{t=1}^T \mathbf{x}_{nt}\mathbf{x}_{nt}^{\prime}$. Moreover, let μ_{nk}^T be the k-th eigenvalue of $\mathbf{\Gamma}_n^T$. Our DR criterion function is

$$\mathrm{DR}_n^T(k) = \frac{\mu_{nk}^T - \mu_{n,k+1}^T}{\mu_{n,k+1}^T - \mu_{n,k+2}^T}.$$

The estimator $\widehat{r}_{n\mathrm{DR}}^T$ is defined as the value of k which maximizes $\mathrm{DR}_n^T(k)$, for $1 \leq k \leq r_{\max}$, i.e.,

$$\widehat{r}_{n\mathrm{DR}}^T = \operatorname*{arg\,max}_{1\leq k\leq r_{\max}} \mathrm{DR}_n^T(k).$$

In the Appendix we prove the following consistency result.

Theorem 3. *Let n and T go to infinity so that $\lim nT^{-1/2} = 0$. Then, under Assumptions A1-A5, B2, C1 and C2', if $r \geq 1$, we have*

$$\mathrm{plim}_{m\to\infty}\, \widehat{r}_{n\mathrm{DR}}^T = r$$

for $m = \min(n, T)$.

Consistency of Ahn and Horenstein criteria ER and GR can be proved for the present setting along the lines of Theorem 3. The assumption of T growing faster than n can be found, for example, in Onatski (2009) even if it is not in the spirit of the literature. However, we show in the simulations that our statistics work well even when n is much larger than T. Hence, to evaluate the performance of the DR criterion we run a simulation exercise. We use the following four DGPs.

DGP1. Same as Ahn and Horenstein (2013), first part, case (d). With reference to equation (6), we set $r = q$, $\mathbf{N}(L) = I_r$, so that $\mathbf{f}_t = \mathbf{u}_t$, with entries $f_{jt} \sim \mathcal{N}(0, 1)$. The entries of $\mathbf{\Lambda}$ are *iid* $\mathcal{N}(0, 1)$. The idiosyncratic components are generated as

$$\xi_{it} = \sqrt{\frac{1-\rho^2}{1+2J\beta^2}}\, e_{it},$$

where $e_{it} = \rho e_{i,t-1} + v_{it} + \sum_{h=\max(i-J,1)}^{i-1} \beta v_{ht} + \sum_{h=i+1}^{\min(i+J,n)} \beta v_{ht}$, with v_{ht} $h = 1, \ldots, n$ standard normal i.i.d., $\beta = 0.2$, $J = \min(10, N/20)$, $\rho = 0.5$.

DGP2. As in DGP1, $r = q$, $\mathbf{N}(L) = I_r$, but the factors f_{jt}, $j = 1, \ldots, r$ are independent gaussian white noises with standard deviations $\sigma_j \sim \mathcal{U}_{[0.2,1.2]}$. The idiosyncratic components are independent, unit variance, gaussian white noises.

DGP3. As in GDP1 and GDP2, we set $r = q$, but here $\mathbf{N}(L)$ is a diagonal matrix having on the diagonal the AR(1) filter

$$\frac{\sigma_j\sqrt{1-\rho_j^2}}{1-\rho_j L},$$

so that $(1-\rho_j L)f_{jt} = \sigma_j\sqrt{1-\rho_j^2}u_{jt}$. Here $\rho_j \sim iid\,\mathcal{U}_{[-0.8,0.8]}$ and $\sigma_j \sim \mathcal{U}_{[1,1.4]}$. In such a way the factor f_{jt} has standard deviation σ_j. The loadings in $\mathbf{\Lambda}$ are $iid\,\mathcal{U}_{[-1,1]}$.

The idiosyncratic components are as in the second and third experiments of Section 5, i.e. $\xi_{it} = \rho_i\,\xi_{it-1}+v_{it}$, $v_{it} = \rho v_{i-1t}+\epsilon_{it}$, where $\rho_i \sim iid\,\mathcal{U}_{[-.8,.8]}$, $\rho = 0.2$ and $\epsilon_{it} \sim iid\,\mathcal{N}(0,1)$.

DGP4. Same as DGP3, but $\sigma_j \sim \mathcal{U}_{[0.6,1.8]}$. In this model, unlike the previous one, the f_{it}'s may have very different variance, so that there are both "strong" and "weak" factors.

We produce data sets with $r = 2, 4$ and 6 and (n,T)= $(50,80)$, $(120,80)$, $(50,240)$, $(120,240)$, $(240,480)$. We set $r_{\max} = 10$ and compute Ahn and Horenstein estimators (ER and GR), along with the DR estimator.

Table 5 shows the percentage of correct estimates over 1000 replications. The maximal difference between ER and GR is 0.3% so that, to save space, we report results only for ER.

With DGP2 and DGP4, DR outperforms ER (and GR). This is because DR is more able than ER and GR in detecting weak factors, which characterize DGP2 and DGP4. Our intuition is that the bounded eigenvalues are usually very close to each other, compared to the diverging ones: ER and GR do not exploit this fact, whereas DR does. On the other hand, with DGP1, where the factors have the same variance, and DGP3, where the factors have similar variances, ER (like GR) outperforms DR, so that the static version of the difference ratio estimator, unlike the dynamic version, does not dominate the competing criteria.

Our conclusion is that the DR estimator may be a valid alternative to ER and GR when the researcher suspects that there are weak common factors driving the data.

7. An empirical illustration

In this section we present an empirical illustration based on US macroeconomic data.

The literature is controversial about the number of dynamic factors driving the macroeconomy. With regard to US macro data, the analysis of Stock and Watson (2005) finds seven static factors and seven dynamic ones. Giannone et al. (2005) finds evidence supporting the existence of only two dynamic factors. Hallin and Liška (2007) base their analysis on the IC_2 criterion (with penalty p_1 and $M_T = [.75\sqrt{T}]$) and find four factors. Onatski (2009) performs his test on Stock and Watson (2002a) macroeconomic data and does not reject the null hypothesis of two dynamic factors.

Clearly, a very small number of dynamic factors, like one or two, is at odds with both modern DSGE models and empirical macroeconomic analysis based on structural VARs, FAVARs or factor models, which usually assume several different structural sources of variation hitting the economy (demand, supply, technology, policy and so on).

We consider the macro data set for estimating factors available in the FRED website. The final sample turns out to be composed of $n = 111$ monthly time series from January

1973 to December 2011 (T = 468). The full list of variables along with the corresponding transformations (first differences of levels, logs or first differences of logs) is reported in Appendix B.

Figure 2, left column, plots the lag-window estimators DER(k), DGR(k) and DDR(k) obtained with $q_{\max} = 10$, Bartlett window, 9 lags (upper panel) and 16 lags (lower panel). DDR peaks at 4 dynamic factors, while DER and DGR estimate either 1 or 2 dynamic factors, respectively. Similar conclusions hold for the periodogram-smoothing versions of the three criteria, Daniell window, 16 points (upper panel) and 32 points (lower panel). With the same $q_{\max}$, the HL criterion (same version used in the simulation exercise) gives $q = 4$ and Onatski's test gives $q = 3$.

Figure 3 plots the estimated number of dynamic factors as a function of $q_{\max}$; the lag-window and the periodogram-smoothing estimators are shown in the left and the right panel, respectively. The figure shows that all estimators are insensitive to the choice of $q_{\max}$.

Figure 4, left panel, plots the static factor estimators ER(k), GR(k) and DR(k) obtained with $r_{\max} = 20$. The DR criterion reaches its maximum for $k = 8$, in line with Stock and Watson (2005) and Bai and Ng (2007), who find 7 static factors. By contrast, both Ahn and Horenstein criteria (ER and GR) are relatively flat and peak at $k = 2$. A possible interpretation is that ER and GR cannot recognize small factors present in the data, in line with the simulation exercise of the previous section. These results are robust to the choice of $r_{\max}$ (right panel).

8. Conclusion

This paper proposes new criteria to determine the number of dynamic factors in the generalized factor model, as well as a new criterion to determine the number of static factors in a static factor model. Such estimators are based on eigenvalue ratios. Their advantage is that they do not have either penalty functions or nuisance parameters requiring preliminary calibration and/or discretionary choices. We have established consistency of such estimators, as the minimum of n and T approach infinity. We have provide simulation exercises showing that our DDR and DR criteria are valid alternatives to existing estimators.

Appendix A

Proof of Theorem 1. We must show that, for $k = 1, \dots, q-1, q+1, \dots, q_{\max}$,

$$\mathrm{DER}_n(k) < \mathrm{DER}_n(q)$$

for all n larger than some N. By Assumption A5(i), for $k = 1, \dots, q-1, q+1, \dots, q_{\max}$, $\mathrm{DER}_n(k)$ is $O_p(1)$, since, if $n > N_1$, $\lambda_{nk}/\lambda_{n,k+1} < c_k^+/c_{k+1}^-$. On the other hand, for $n > N_1$,

$\mathrm{DER}_n(q) = \lambda_{nq}/\lambda_{n,q+1} > nc_q^-/c_{q+1}^+$. The result follows. Coming to $\mathrm{DGR}_n(k)$, consider the inequalities

$$(a1) \qquad \frac{\alpha}{1+\alpha} \leq \ln(1+\alpha) < \alpha \qquad \alpha \in (0,\infty).$$

Using these inequalities and Assumption A5(i), for $k = 1,\ldots,q-1,q+1,\ldots,q_{\max}$ and $n > N_1$, we have

$$\mathrm{DGR}_n(k) = \frac{\ln(1+\widetilde{\lambda}_{nk})}{\ln(1+\widetilde{\lambda}_{n,k+1})} < \frac{\widetilde{\lambda}_{nk}}{\widetilde{\lambda}_{n,k+1}/(1+\widetilde{\lambda}_{n,k+1})} = \frac{\lambda_{nk}}{\lambda_{n,k+1}} < \frac{c_k^+}{c_{k+1}^-}.$$

On the other hand, we have

$$(a2) \qquad \mathrm{DGR}_n(q) = \frac{\ln(1+\widetilde{\lambda}_{nq})}{\ln(1+\widetilde{\lambda}_{n,q+1})} > \frac{\widetilde{\lambda}_{nq}/(1+\widetilde{\lambda}_{nq})}{\widetilde{\lambda}_{n,q+1}} = \frac{\lambda_{nq}}{\lambda_{n,q+1}}\,\frac{V_n(q+1)}{V_n(q-1)}.$$

But the last ratio on the right-hand side is bounded away from zero by A5(i) and (ii). Precisely, if $n > \max(N_1, N_2)$,

$$\frac{V_n(q+1)}{V_n(q-1)} = \frac{n^{-1}(V_n(q)-\lambda_{n,q+1})}{n^{-1}(V_n(q)+\lambda_{nq})} > \frac{c - n^{-1}c_{q+1}^+}{c_{q+1}^+ + c_q^+} > 0.$$

Moreover, for $n > N_1$, $\lambda_{nq}/\lambda_{n,q+1} > nc_q^-/c_{q+1}^+$. The result follows. Finally, let us consider $\widehat{q}_{n\mathrm{DDR}}$. By Assumption A5(i),

$$\mathrm{DDR}_n(k) = \frac{\lambda_{nk}-\lambda_{n,k+1}}{\lambda_{n,k+1}-\lambda_{n,k+2}} < \frac{c_k^+}{c_{k+1}^- - c_{k+2}^+},$$

for $0 < k < q-1$, $q < k \leq q_{\max}$, $n > N_1$. Moreover, for $n > N_1$,

$$\mathrm{DDR}_n(q-1) = \frac{\lambda_{n,q-1}-\lambda_{nq}}{\lambda_{nq}-\lambda_{n,q+1}} < \frac{c_{q-1}^+}{c_q^- - c_{q+1}^+/n}.$$

On the other hand,

$$\mathrm{DDR}_n(q) = \frac{\lambda_{nq}-\lambda_{n,q+1}}{\lambda_{n,q+1}-\lambda_{n,q+2}} > \frac{nc_q^- - c_{q+1}^+}{c_{q+1}^+}.$$

This concludes the proof. □

Proof of Theorem 2. We use the following result proved in Hallin and Liška (2007), Corollary A.1. Let Assumption A1, A2, B1 and B2 hold. Then for any $\epsilon > 0$ there exist M_ϵ and T_ϵ such that for any fixed $q_{\max}$, n and $T > T_\epsilon$

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

This gives

$$\frac{\lambda_{nk}^T(\theta)}{n} = \frac{\lambda_{nk}(\theta)}{n} + O_p(M_T^*)$$

uniformly in n and $\theta \in [-\pi, \pi]$, where $M_T^* = \max(M_T^{-2}, M_T^{1/2} T^{-1/2})$ goes to zero as $T \to \infty$. Thus we get

$$(a3) \qquad \frac{2\pi}{2M_T+1} \sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}^T(\theta_h)}{n} = \frac{2\pi}{2M_T+1} \sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}(\theta_h)}{n} + O_p(M_T^*).$$

The term on the left hand-side is the middle Riemann sum of the function $\lambda_{nk}(\theta)/n$ over a partition of $[-\pi, \pi]$ made up by $2M_T + 1$ sub-intervals of equal length $2\pi/(2M_T + 1)$. Standard results from calculus theory and numerical integration easily imply that

$$(a4) \qquad \left| \frac{\lambda_{nk}}{n} - \frac{2\pi}{2M_T+1} \sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}(\theta_h)}{n} \right| \leq \frac{4\pi^2}{2M_T+1} \frac{L_n}{n}$$

where

$$L_n = \max_{1 \leq k \leq q_{\max}} \sup_{\theta \in [-\pi,\pi]} \left| \frac{d\lambda_{nk}(\theta)}{d\theta} \right|.$$

Assumptions A5(iii) and B2 guarantee the existence and boundedness (uniformly in n) of the above first derivative since $\sup_{n \in \mathbb{N}} L_n/n \leq \sup_{n \in \mathbb{N}} \sup_{\theta \in [-\pi,\pi]} \frac{1}{n} || \frac{d\Sigma(\theta)}{d\theta} || < \infty$ (see Kato, 1982 and Overton and Womersley, 1995). It follows that

$$(a5) \qquad \begin{aligned} \frac{2\pi}{2M_T+1} \sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}^T(\theta_h)}{n} - \frac{\lambda_{nk}}{n} &= \frac{2\pi}{2M_T+1} \left[\sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}^T(\theta_h)}{n} - \sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}(\theta_h)}{n} \right] \\ &+ \frac{2\pi}{2M_T+1} \sum_{h=-M_T}^{M_T} \frac{\lambda_{nk}(\theta_h)}{n} - \frac{\lambda_{nk}}{n} \end{aligned}$$

As $T \to \infty$, the first and second summands of (a5) go to zero by (a3) and (a4), respectively, uniformly in n. So, the left hand-side of (a5) converges to zero as $T \to \infty$, uniformly in n. Hence $\mathrm{DER}_n^T(k) \xrightarrow{p} \mathrm{DER}_n(k)$, $\mathrm{DGR}_n^T(k) \xrightarrow{p} \mathrm{DGR}_n(k)$ and $\mathrm{DDR}_n^T(k) \xrightarrow{p} \mathrm{DDR}_n(k)$, provided that $\lim n M_T^{*-1} = 0$ as n and T go to infinity. Then the result follows from Theorem 1. □

Proof of Theorem 3. Firstly, let us introduce some notation and preliminary results. If A is a symmetric matrix we denote by $\mu_j(A)$ the jth eigenvalue of A in decreasing order. Given a matrix B, $||B||_2$ denotes the spectral norm of B; thus $||B||_2 = \sqrt{\mu_1(BB')}$, which is the Euclidean norm if B is a row matrix. We will make use of the following inequality, which is a consequence of Weyl's inequality (see, e.g., Stewart and Sun, 1990, Cor. 4.10, p. 203): if A and B are two $s \times s$ symmetric matrices, then for $j = 1, \ldots, s$,

$$(a6) \qquad |\mu_j(A+B) - \mu_j(A)| \leq \sqrt{\mu_1(B^2)} = ||B||_2.$$

Now, we prove the following result

$$\frac{1}{n}\,||\Gamma_n^T - \Gamma_n||_2 = O_p(T^{-1/2})$$

uniformly in n. In fact, we have

$$\begin{aligned}\frac{1}{n^2}\,||\Gamma_n^T - \Gamma_n||_2^2 &= \frac{1}{n^2}\,\mu_1[(\Gamma_n^T - \Gamma_n)(\Gamma_n^T - \Gamma_n)^{'}] \\ &\leq \frac{1}{n^2}\,||\Gamma_n^T - \Gamma_n||_F^2 = \frac{1}{n^2}\ \mathrm{tr}[(\Gamma_n^T - \Gamma_n)(\Gamma_n^T - \Gamma_n)^{'}] \\ &= \frac{1}{n^2}\ \sum_{i=1}^{n}\sum_{j=1}^{n}(\gamma_{n,ij}^T - \gamma_{n,ij})^2 \leq \sup_{i,j}\,(\gamma_{n,ij}^T - \gamma_{n,ij})^2\end{aligned}$$

where $\Gamma_n = (\gamma_{n,ij})$, $\Gamma_n^T = (\gamma_{n,ij}^T)$ and $||\cdot||_F$ denotes the Frobenius norm. Using Chebyshev inequality, we get

$$(\gamma_{n,ij}^T - \gamma_{n,ij})^2 \leq \mathrm{var}(\gamma_{n,ij}^T) = E(\gamma_{n,ij}^T - \gamma_{n,ij})^2$$

as $E(\gamma_{n,ij}^T) = \gamma_{n,ij}$ (see, for example, Priestley 1981, p.325). Using results from Priestley (1981), p.326, and Hannan (1970), p.209, we have

$$\mathrm{var}(\gamma_{n,ij}^T) = \frac{1}{T}\sum_{m=-T+1}^{T-1}\left[1 - \frac{|m|}{T}\right]\{\gamma_{n,ii}(m)\gamma_{n,jj}(m) + \gamma_{n,ij}(m)\gamma_{n,ji}(m) + c_{ijij}(0,m,m)\}$$

where $\Gamma_n(m) = E(\mathbf{x}_{nt}\mathbf{x}_{n,t-m}^{'}) = (\gamma_{n,ij}(m))$. Here $c_{ijij}(0,m,m)$ is the cumulant $c_{i_1,\ldots,i_4}(k_1,k_2,k_3)$, where $i_1 = i_3 = i$, $i_2 = i_4 = j$, $k_1 = 0$ and $k_2 = k_3 = m$. Then

$$\begin{aligned}&\frac{1}{n^2}\,||\Gamma_n^T - \Gamma_n||_2^2 \leq \sup_{i,j}\,(\gamma_{n,ij}^T - \gamma_{n,ij})^2 \leq \sup_{i,j}\ \mathrm{var}(\gamma_{n,ij}^T) \\ &\leq \frac{1}{T}\sup_{i,j}\sum_{m=-\infty}^{\infty}\{|\gamma_{n,ii}(m)||\gamma_{n,jj}(m)| + |\gamma_{n,ij}(m)||\gamma_{n,ji}(m)|\} + \frac{1}{T}\sup_{i,j}\sum_{m=-\infty}^{\infty}|c_{ijij}(0,m,m)| \\ &< \frac{1}{T}\rho < \infty\end{aligned}$$

for some positive real number ρ, uniformly in n. The upper bound follows from Assumption A2. Secondly, we prove that

$$\frac{\mu_{nk}^T}{n} = \frac{\mu_{nk}}{n} + O_p(T^{-1/2})$$

uniformly in n. In fact, by using Weyl inequality with $A = \Gamma_n$ and $B = \Gamma_n^T - \Gamma_n$ we obtain

$$\frac{1}{n}|\mu_{nk}^T - \mu_{nk}| \leq \frac{1}{n}||\Gamma_n^T - \Gamma_n||_2$$

Using the first result of the proof, the previous statement follows. Finally, we make use of these results to prove the statement in Theorem 3. By Assumption C2$'$ in Section 6,

$$\mathrm{DR}_n(k) = \frac{\mu_{nk} - \mu_{n,k+1}}{\mu_{n,k+1} - \mu_{n,k+2}} < \frac{d_k^+}{d_{k+1}^- - d_{k+2}^+}$$

for $0 < k < r-1$, $r < k \leq r_{\max}$, $n > N_3$. Moreover, for $n > N_3$,

$$\mathrm{DR}_n(r-1) = \frac{\mu_{n,r-1} - \mu_{nr}}{\mu_{nr} - \mu_{n,r+1}} < \frac{d_{r-1}^+}{d_r^- - d_{r+1}^+/n}$$

and

$$\mathrm{DR}_n(r) = \frac{\mu_{nr} - \mu_{n,r+1}}{\mu_{n,r+1} - \mu_{n,r+2}} > \frac{nd_r^- - d_{r+1}^+}{d_{r+1}^+}$$

Then $\mathrm{DR}_n^T(k)$ converges to $\mathrm{DR}_n(k)$ as T goes to infinity, uniformly in n using the assumption $\lim nT^{-1/2} = 0$. Given the above inequalities, the statement of the theorem follows. $\square$

Appendix B

US Macroeconomic Data - Transformations (T): 1=levels, 2=first differences of levels, 3=logs, 4=first differences of logs.

No.	MNEMONIC	LONG LABEL	T
1	AAA	Moody's Seasoned Aaa Corporate Bond Yield	1
2	AAAFFM	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	1
3	AMBSL	St. Louis Adjusted Monetary Base	2
4	AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	1
5	AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	1
6	BAA	Moody's Seasoned Baa Corporate Bond Yield	1
7	BAAFFM	Moody's Seasoned Baa Corporate Bond Minus Federal Funds Rate	1
8	BUSLOANS	Commercial and Industrial Loans, All Commercial Banks	4
9	CE16OV	Civilian Employment	3
10	CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	1
11	CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing	2
12	CES1021000001	All Employees: Mining and Logging: Mining	4
13	CES2000000008	Average Hourly Earnings of Production and Nonsupervisory Employees: Construction	4
14	CES3000000008	Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing	4
15	CLF16OV	Civilian Labor Force	3
16	CPIAPPSL	Consumer Price Index for All Urban Consumers: Apparel	4
17	CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	4
18	CPIMEDSL	Consumer Price Index for All Urban Consumers: Medical Care	4
19	CPITRNSL	Consumer Price Index for All Urban Consumers: Transportation	4
20	CPIULFSL	Consumer Price Index for All Urban Consumers: All Items Less Food	4
21	CUSR0000SA0L5	Consumer Price Index for All Urban Consumers: All items less medical care	4
22	CUSR0000SAC	Consumer Price Index for All Urban Consumers: Commodities	4
23	CUSR0000SAS	Consumer Price Index for All Urban Consumers: Services	4
24	CUUR0000SA0L2	Consumer Price Index for All Urban Consumers: All items less shelter	4
25	CUUR0000SAD	Consumer Price Index for All Urban Consumers: Durables	4
26	DDURRG3M086SBEA	Personal consumption expenditures: Durable goods (chain-type price index)	4
27	DMANEMP	All Employees: Durable goods	1
28	DNDGRG3M086SBEA	Personal consumption expenditures: Nondurable goods (chain-type price index)	4
29	DPCERA3M086SBEA	Real personal consumption expenditures (chain-type quantity index)	4
30	DSERRG3M086SBEA	Personal consumption expenditures: Services (chain-type price index)	4
31	EXCAUS	Canada / U.S. Foreign Exchange Rate	1
32	EXJPUS	Japan / U.S. Foreign Exchange Rate	1
33	EXSZUS	Switzerland / U.S. Foreign Exchange Rate	1
34	EXUSUK	U.S. / U.K. Foreign Exchange Rate	1
35	FEDFUNDS	Effective Federal Funds Rate	1
36	GS1	1-Year Treasury Constant Maturity Rate	1
37	GS10	10-Year Treasury Constant Maturity Rate	1
38	GS5	5-Year Treasury Constant Maturity Rate	1
39	HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	4
40	HOUSTMW	Housing Starts in Midwest Census Region	4
41	HOUSTNE	Housing Starts in Northeast Census Region	4
42	HOUSTS	Housing Starts in South Census Region	4
43	HOUSTW	Housing Starts in West Census Region	4
44	INDPRO	Industrial Production Index	4
45	IPBUSEQ	Industrial Production: Business Equipment	4
46	IPCONGD	Industrial Production: Consumer Goods	4
47	IPDCONGD	Industrial Production: Durable Consumer Goods	4
48	IPDMAT	Industrial Production: Durable Materials	4
49	IPFINAL	Industrial Production: Final Products (Market Group)	4
50	IPFPNSS	Industrial Production: Final Products and Nonindustrial Supplies	4

No.	MNEMONIC	LONG LABEL	T
51	IPFUELS	Industrial Production: Fuels	4
52	IPMANSICS	Industrial Production: Manufacturing (SIC)	4
53	IPMAT	Industrial Production: Materials	4
54	IPNCONGD	Industrial Production: Nondurable Consumer Goods	4
55	IPNMAT	Industrial Production: nondurable Materials	4
56	M1SL	M1 Money Stock	4
57	M2REAL	Real M2 Money Stock	4
58	M2SL	M2 Money Stock	4
59	MABMM301USM189S	M3 for the United States	2
60	MANEMP	All Employees: Manufacturing	1
61	NAPM	ISM Manufacturing: PMI Composite Index	1
62	NAPMEI	ISM Manufacturing: Employment Index	1
63	NAPMII	ISM Manufacturing: Inventories Index	1
64	NAPMNOI	ISM Manufacturing: New Orders Index	1
65	NAPMPI	ISM Manufacturing: Production Index	1
66	NAPMPRI	ISM Manufacturing: Prices Index	1
67	NAPMSDI	ISM Manufacturing: Supplier Deliveries Index	1
68	NDMANEMP	All Employees: Nondurable goods	1
69	NONBORRES	Reserves Of Depository Institutions, Nonborrowed	2
70	OILPRICE	Spot Oil Price: West Texas Intermediate	4
71	PAYEMS	All Employees: Total nonfarm	4
72	PCEPI	Personal Consumption Expenditures: Chain-type Price Index	4
73	PERMIT	New Private Housing Units Authorized by Building Permits	4
74	PERMITMW	New Private Housing Units Authorized by Building Permits in the Midwest Census Region	4
75	PERMITNE	New Private Housing Units Authorized by Building Permits in the Northeast Census Region	4
76	PERMITS	New Private Housing Units Authorized by Building Permits in the South Census Region	4
77	PERMITW	New Private Housing Units Authorized by Building Permits in the West Census Region	4
78	PI	Personal Income	4
79	PPICMM	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals	4
80	PPICRM	Producer Price Index: Crude Materials for Further Processing	4
81	PPIFCG	Producer Price Index: Finished Consumer Goods	4
82	PPIFGS	Producer Price Index: Finished Goods	4
83	PPIITM	Producer Price Index: Intermediate Materials: Supplies and Components	4
84	REALLN	Real Estate Loans, All Commercial Banks	4
85	RPI	Real Personal Income	4
86	SRVPRD	All Employees: Service-Providing Industries	4
87	T10YFFM	10-Year Treasury Constant Maturity Minus Federal Funds Rate	1
88	T1YFFM	1-Year Treasury Constant Maturity Minus Federal Funds Rate	1
89	T5YFFM	5-Year Treasury Constant Maturity Minus Federal Funds Rate	1
90	TB3MS	3-Month Treasury Bill: Secondary Market Rate	1
91	TB3SMFFM	3-Month Treasury Bill Minus Federal Funds Rate	1
92	TB6MS	6-Month Treasury Bill: Secondary Market Rate	1
93	TB6SMFFM	6-Month Treasury Bill Minus Federal Funds Rate	1
94	TOTRESNS	Total Reserves of Depository Institutions	2
95	TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies	5
96	UEMP15OV	Number of Civilians Unemployed for 15 Weeks and Over	2
97	UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks	2
98	UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over	2
99	UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks	2
100	UEMPLT5	Number of Civilians Unemployed - Less Than 5 Weeks	2
101	UEMPMEAN	Average (Mean) Duration of Unemployment	2
102	UNEMPLOY	Unemployed	2
103	UNRATE	Civilian Unemployment Rate	1
104	USCONS	All Employees: Construction	4
105	USFIRE	All Employees: Financial Activities	4
106	USGOOD	All Employees: Goods-Producing Industries	4
107	USGOVT	All Employees: Government	4
108	USTPU	All Employees: Trade, Transportation and Utilities	4
109	USTRADE	All Employees: Retail Trade	4
110	USWTRADE	All Employees: Wholesale Trade	4
111	W875RX1	Real personal income excluding current transfer receipts	4

References

[1] Ahn, S.C., and Horenstein, A.R. (2013) Eigenvalue ratio test for the number of factors, *Econometrica* **81**(3), 1203-1227.

[2] Alessi, L., Barigozzi, M., and Capasso, M. (2010) Improved penalization for determining the number of factors in approximate factor models, *Statistics & Probability Letters* **80**, 1806-1813.

[3] Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2010) New Eurocoin: tracking economic growth in real time, *The Review of Economics and Statistics* **92**, 1024-1034.

[4] Amengual, D., and Watson, M. (2007) Consistent estimation of the number of dynamic factors in a large N and T panel, *Journal of Business and Economic Statistics* **25**, 91-96.

[5] Bai, J. (2003) Inferential theory for factor models of large dimensions, *Econometrica* **71**, 135-171.

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

[7] Bai, J., and Ng, S. (2006) Determining the number of factors in approximate factor models, Errata, *Manuscript*, Columbia University.

[8] Bai, J., and Ng, S. (2007) Determining the number of primitive shocks in factor models, *Journal of Business & Economic Statistics* **25**(1), 52-60.

[9] Bernanke, B.S., and Boivin, J. (2003) Monetary policy in a data-rich environment, *Journal of Monetary Economics* **50**, 525-546.

[10] Bernanke, B.S., Boivin, J., and Eliasz, P. (2005) Measuring monetary policy: a factor augmented autoregressive (FAVAR) approach, *The Quarterly Journal of Economics* **120**, 387-422.

[11] Boivin, J., and Ng, S. (2006) Are more data always better for factor analysis?, *Journal of Econometrics* **132**, 169-194.

[12] Brillinger, D.R. (1981) Time series: data analysis and theory, *Holden-Day Series in Time Series Analysis*, San Francisco CA.

[13] Brockwell, P.J., and Davis, R.A. (2009) Time series: theory and methods, *Springer Series in Statistics*, Springer-Verlag, New York Berlin Heidelberg.

[14] Chamberlain, G., and Rothschild, M. (1983) Arbitrage, factor structure and mean-variance analysis in large asset markets, *Econometrica* **51**, 1305-1324.

[15] Corielli, F., and Marcellino, M. (2006) Factor based index tracking, *Journal of Banking & Finance* **30**, 2215-2233.

[16] Cristadoro, R., Forni, M., Reichlin, L., and Veronese, G. (2005) A Core Inflation Indicator for the Euro Area, *Journal of Money, Credit and Banking* **37**(3), 539-560.

[17] D'Agostino, A., and Giannone, D. (2012) Comparing Alternative Predictors Based on Large Panel Factor Models, *Oxford bulletin of economics and statistics* **74**(2), 306-326.

[18] Eickmeier, S. (2007) Business cycle transmission from the US to Germany - A structural factor approach, *European Economic Review* **51**, 521-551.

[19] Favero, C., Marcellino, M., and Neglia, F. (2005) Principal components at work: empirical analysis of monetary policy with large data sets, *Journal of Applied Econometrics* **20**, 603-620.

[20] Forni, M., and Gambetti, L. (2010) The dynamic effects of monetary policy: A structural factor model approach, *Journal of Monetary Economics* **57**, 203-216.

[21] Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009) Opening the black box: structural factor models with large cross-sections, *Econometric Theory* **25**, 1319-1347.

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

[23] Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005) The generalized dynamic factor model: one-sided estimation and forecasting, *Journal of the American Statistical Association* **100**(471), 830-840.

[24] Forni, M., and Lippi, M. (2001) The generalized dynamic factor model: representation theory, *Econometric theory* **17**(06), 1113-1141.

[25] Forni, M., and Reichlin, L. (1998) Let's get real: a factor analytical approach to disaggregated business cycle dynamics, *The Review of Economic Studies* **65**(3), 453-473.

[26] Geweke, J. (1977) The dynamic factor analysis of economic time series, in D.J. Aigner and A.S. Goldberger, Eds., *Latent Variables in Socio-Economic Models*, Amsterdam: North Holland.

[27] Giannone, D., Reichlin, L., and Sala, L. (2005) Monetary Policy in Real Time, in NBER Macroeconomic Annual 2004, eds. M. Gertler and K. Rogoff, Cambridge, MA, MIT Press, 161-200.

[28] Giannone, D., Reichlin, L., and Sala, L. (2006) VARs, common factors and the empirical validation of equilibrium business cycle models, *Journal of Econometrics* **132**(1), 257-279.

[29] Giannone, D., Reichlin, L., and Small, D. (2008) Nowcasting: The real-time informational content of macroeconomic data, *Journal of Monetary Economics* **55**, 665-676.

[30] Hallin, M., and Lippi, M. (2013) Factor models in high-dimensional time series - A time domain approach, *Stochastic processes and their applications* **123**, 2678-2695.

[31] Hallin, M., and Liška, R. (2007) Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**, 603-617.

[32] Hallin, M., Mathias, C., Pirotte, H., and Veredas, D. (2011) Market liquidity as dynamic factors, *Journal of Econometrics* **163**, 42-50.

[33] Hannan, E.J. (1970) *Multiple Time Series*, J. Wiley series in probability and mathematical statistics, New York- London-Sydney-Toronto.

[34] Kato, T. (1982) *A Short Introduction to Perturbation Theory for Linear Operators*, Springer-Verlag, New York.

[35] Lancaster, P. (1964) On eigenvalues of matrices dependent on a paramter, *Numerische Mathematik* **6**, 377-387.

[36] Ludvigson, S.C., and Ng, S. (2007) The empirical risk-return relation: a factor analysis approach, *Journal of Financial Economics* **83**, 171-222.

[37] Ludvigson, S.C., and Ng, S. (2009) Macro factors in bond risk premia, *The Review of Financial Studies* **22**, 5027-5067.

[38] Marcellino, M., Stock, J.H., and Watson, M.W. (2003) Macroeconomic forecasting in the euro area: Country specific versus area-wide information, *European Economic Review* **47**(1), 1-18.

[39] Onatski, A. (2009) Testing hypotheses about the number of factors in large factor models, *Econometrica* **77**(5), 1447-1479.

[40] Onatski, A. (2010) Determining the number of factors from empirical distribution of eigenvalues, *The Review of Economics and Statistics* **92**(4), 1004-1016.

[41] Overton, M.L., and Womersley, R.S. (1995) Second derivatives for optimizing eigenvalues of symmetric matrices, *SIAM Journal on Matrix Analysis and Applications* **16**(3), 697-718.

[42] Priestley, M.B. (1981) *Spectral Analysis and Time Series*, Vol.1, Univariate Series, Academic Press, London-New York-San Francisco.

[43] Sargent, T.J., and Sims, C.A. (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.

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

[45] Stock, J.H., and Watson, M.W. (2002b) Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**, 1167-1179.

[46] Stock, J.H., and Watson, M.W. (2005) Implications of dynamic factor models for VAR analysis, *National Bureau of Economic Research* No. w11467.

Tables

q	n	T	HL	O	DER	DGR	DDR	DER*	DGR*	DDR*
						MA loadings				
2	60	100	**99**	62	80	92	98	80	92	97
	100	100	**99**	79	86	97	**99**	89	97	**99**
	70	120	**99**	67	85	96	**99**	86	97	**99**
	120	120	99	87	96	99	**100**	96	99	**100**
	150	120	**100**	88	97	**100**	**100**	97	**100**	**100**
3	60	100	60	30	24	46	69	27	45	**71**
	100	100	**92**	40	25	57	87	32	56	83
	70	120	**94**	32	36	59	85	37	59	84
	120	120	**99**	38	43	68	94	39	68	96
	150	120	**99**	45	44	72	96	44	71	95
						AR loadings				
2	60	100	94	82	83	93	95	83	92	**97**
	100	100	**99**	94	87	94	98	87	94	**99**
	70	120	**100**	91	89	95	98	88	94	98
	120	120	**100**	97	96	98	**100**	94	98	**100**
	150	120	**100**	97	96	99	**100**	95	98	**100**
3	60	100	34	44	40	51	63	44	53	**68**
	100	100	62	60	51	64	80	51	65	**82**
	70	120	71	56	51	63	**79**	46	61	78
	120	120	**92**	76	64	77	91	60	76	91
	150	120	**94**	79	62	78	**94**	58	76	**94**

Table 1: First experiment described in Section 5. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER, DGR and DDR: lag window version of our estimators described in Section 4.1, DER*, DGR* and DDR*: periodogram smoothing version of our estimators described in Section 4.2. Boldface numbers denote the estimator(s) which perform best in each row.

n	T	σ^2	HL	O	DER	DGR	DDR	DER*	DGR*	DDR*
					MA loadings					
70	70	1	**100**	**100**	**100**	**100**	**100**	99	93	**100**
70	70	2	96	**100**	90	97	**100**	91	96	99
70	70	4	1	**90**	50	59	80	50	63	79
100	120	1	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
100	120	2	**100**	**100**	**100**	**100**	**100**	99	**100**	**100**
100	120	6	2	**96**	53	63	88	55	64	88
150	500	1	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
150	500	8	**100**	**100**	98	99	**100**	99	99	**100**
150	500	16	35	**97**	38	46	91	47	55	95
					AR loadings					
70	70	1	98	98	96	99	**100**	85	92	99
70	70	2	86	91	84	90	**98**	67	75	95
70	70	4	13	66	61	71	**85**	45	54	79
100	120	1	**100**	98	**100**	**100**	**100**	99	**100**	**100**
100	120	2	**100**	98	98	99	**100**	96	98	**100**
100	120	6	45	83	78	83	**97**	77	81	96
150	500	1	**100**	99	**100**	**100**	**100**	**100**	**100**	**100**
150	500	8	**100**	99	99	**100**	**100**	99	99	**100**
150	500	16	99	92	91	93	**100**	82	85	99

Table 2: Second experiment described in Section 5. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER, DGR and DDR: lag window version of our estimators described in Section 4.1, DER*, DGR* and DDR*: periodogram smoothing version of our estimators described in Section 4.2. Boldface numbers denote the estimator(s) which perform best in each row.

q	n	T	HL	O	DER	DGR	DDR	DER*	DGR*	DDR*
					Large idiosyncratic components					
2	50	80	74	82	83	91	**99**	78	89	97
	120	80	**100**	94	96	99	**100**	93	98	**100**
	50	240	**100**	99	99	**100**	**100**	99	**100**	**100**
	120	240	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	98	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	50	80	0	16	3	10	**27**	2	6	22
	120	80	1	27	12	30	**65**	6	17	47
	50	240	8	40	54	77	**86**	44	66	82
	120	240	**100**	93	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	97	**100**	**100**	**100**	**100**	**100**	**100**	**100**
6	50	80	0	**10**	0	1	6	0	0	6
	120	80	0	**10**	0	1	**10**	0	0	6
	50	240	0	15	2	11	**24**	1	4	16
	120	240	**98**	35	45	73	90	26	57	85
	240	480	92	99	**100**	**100**	**100**	**100**	**100**	**100**
					Small idiosyncratic components					
2	50	80	**100**	89	98	**100**	**100**	97	**100**	99
	120	80	96	89	**100**	**100**	**100**	**100**	**100**	99
	50	240	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	120	240	58	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	11	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	50	80	8	35	37	68	**74**	35	56	58
	120	80	83	45	73	89	**93**	60	78	82
	50	240	**100**	88	**100**	**100**	**100**	99	**100**	**100**
	120	240	97	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	84	**100**	**100**	**100**	**100**	**100**	**100**	**100**
6	50	80	0	16	1	10	**17**	2	4	14
	120	80	0	14	4	18	**33**	3	6	26
	50	240	9	38	80	93	**94**	69	89	87
	120	240	99	82	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	94	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table 3: Third experiment described in Section 5, case (a), i.e. common shocks with equal variance. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER, DGR and DDR: lag window version of our estimators described in Section 4.1, DER*, DGR* and DDR*: periodogram smoothing version of our estimators described in Section 4.2. Boldface numbers denote the estimator(s) which perform best in each row.

q	n	T	HL	O	DER	DGR	DDR	DER*	DGR*	DDR*
					Large idiosyncratic components					
2	50	80	71	76	73	85	**96**	67	81	93
	120	80	**100**	92	92	97	**100**	87	96	**100**
	50	240	**100**	97	99	**100**	**100**	98	99	**100**
	120	240	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	98	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	50	80	0	14	3	9	**20**	1	3	17
	120	80	0	27	8	24	**58**	4	13	43
	50	240	5	36	40	64	**82**	33	56	78
	120	240	**100**	87	96	99	**100**	92	97	**100**
	240	480	97	**100**	**100**	**100**	**100**	**100**	**100**	**100**
6	50	80	0	7	0	0	**8**	0	0	5
	120	80	0	**10**	0	0	7	0	0	7
	50	240	0	15	0	4	**16**	0	2	**16**
	120	240	**95**	33	30	61	83	17	43	75
	240	480	94	99	**100**	**100**	**100**	**100**	**100**	**100**
					Small idiosyncratic components					
2	50	80	**99**	88	94	**99**	**99**	94	**99**	**99**
	120	80	92	86	97	**100**	99	98	**100**	99
	50	240	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	120	240	45	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	5	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	50	80	5	29	29	54	**65**	23	44	51
	120	80	76	36	53	78	**85**	46	67	68
	50	240	98	84	99	**100**	**100**	99	99	**100**
	120	240	96	99	**100**	**100**	**100**	**100**	**100**	**100**
	240	480	73	**100**	**100**	**100**	**100**	**100**	**100**	**100**
6	50	80	0	12	1	6	**14**	1	2	12
	120	80	0	20	3	17	**31**	2	3	20
	50	240	6	33	64	**87**	**87**	52	78	80
	120	240	99	76	**100**	**100**	**100**	98	99	**100**
	240	480	94	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table 4: Third experiment described in Section 5, case (b), i.e. common shocks with different variances. Percentage of correct outcomes over 500 replications. HL: Hallin and Liška (2007) estimator, O: Onatski (2009) estimator, DER, DGR and DDR: lag window version of our estimators described in Section 4.1, DER*, DGR* and DDR*: periodogram smoothing version of our estimators described in Section 4.2. Boldface numbers denote the estimator(s) which perform best in each row.

r	n	T	DGP1 ER	DGP1 DR	DGP2 ER	DGP2 DR	DGP3 ER	DGP3 DR	DGP4 ER	DGP4 DR
2	50	80	**81**	43	62	**83**	**95**	86	65	**71**
	120	80	**100**	91	73	**93**	**100**	95	77	**88**
	50	240	**92**	40	71	**95**	**100**	95	77	**87**
	120	240	**100**	98	78	**99**	**100**	99	89	**98**
	240	480	**100**	99	87	**100**	**100**	**100**	98	**100**
4	50	80	**61**	28	40	**70**	**94**	76	44	**56**
	120	80	**100**	88	55	**88**	**100**	94	66	**81**
	50	240	**77**	28	54	**87**	**100**	92	68	**78**
	120	240	**100**	97	68	**98**	**100**	99	90	**96**
	240	480	**100**	99	85	**100**	**100**	**100**	99	**100**
6	50	80	**47**	24	35	**63**	**90**	71	38	**45**
	120	80	**100**	84	49	**85**	**100**	90	70	**71**
	50	240	**54**	24	51	**86**	**100**	91	66	**74**
	120	240	**100**	97	70	**98**	**100**	99	**94**	**94**
	240	480	**100**	99	91	**100**	**100**	**100**	**100**	99

Table 5: Simulation experiment described in Section 6. Percentage of correct outcomes over 500 replications. ER: Ahn and Horestein (2013) Eigenvalue Ratio estimator; DR: the eigenvalue Difference Ratio estimator described in Section 6.

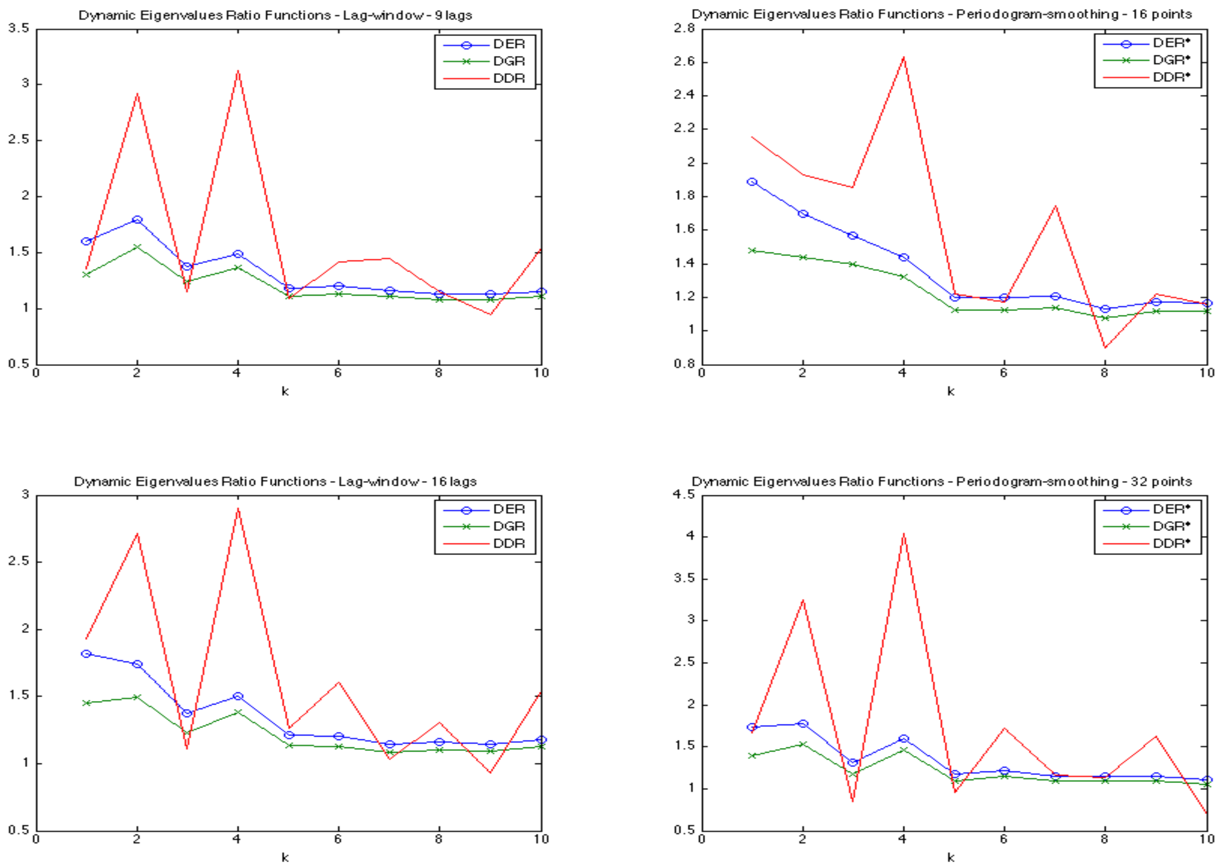

Figure 2: Dynamic eigenvalue ratio estimators. Left column: lag-window criteria $\mathrm{DER}_n^T(k)$, $\mathrm{DGR}_n^T(k)$ and $\mathrm{DDR}_n^T(k)$ as functions of k, window sizes 9 (upper panel) and 16 (lower panel). Right column: periodogram smoothing criteria $\mathrm{DER}_n^{*T}(k)$, $\mathrm{DGR}_n^{*T}(k)$ and $\mathrm{DDR}_n^{*T}(k)$ as functions of k, bandwidth 16 points (upper panel) and 32 points (lower panel). The data set includes 111 US macroeconomic time series from January 1973 to December 2011. Source: FRED database.

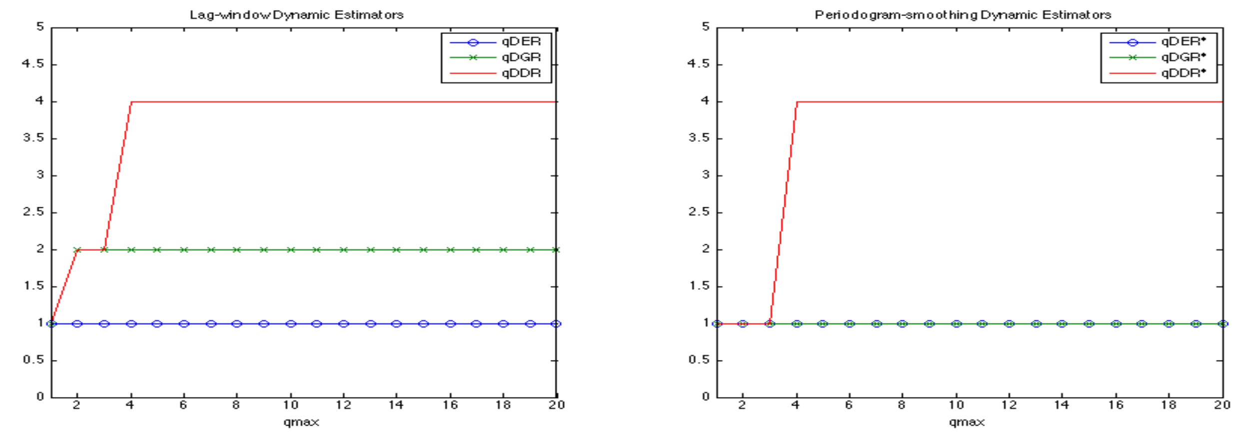

Figure 3: : Left panel: estimated number of factors as functions of $q_{\max}$, lag-window estimators (16 lags). Right panel: estimated number of dynamic factors as functions of $q_{\max}$, periodogram-smoothing estimators (16 points). The data set includes 111 US macroeconomic time series from January 1973 to December 2011. Source: FRED database.

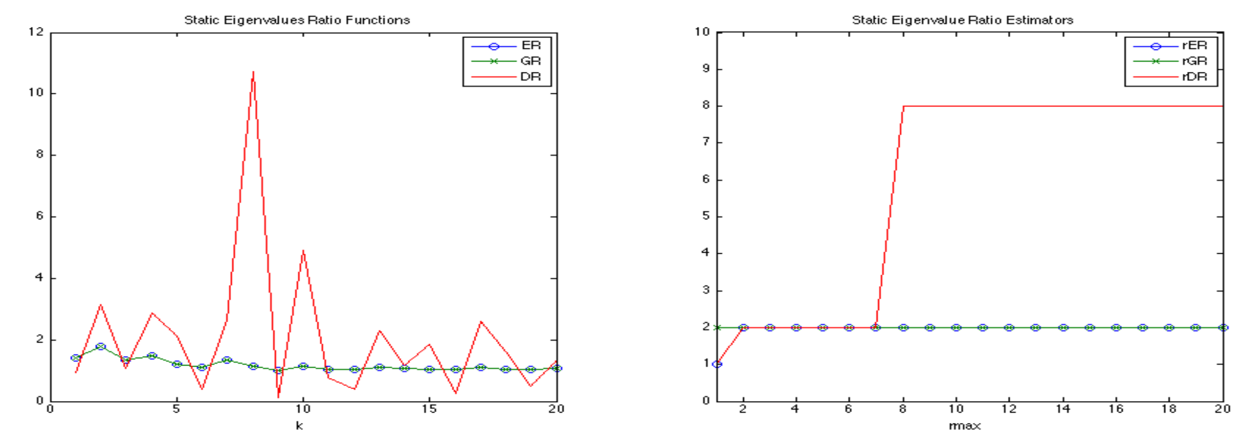

Figure 4: Static eigenvalue ratio estimators. Left panel: eigenvalue ratio criteria $\mathrm{ER}_n^T(k)$, $\mathrm{GR}_n^T(k)$ and $\mathrm{DR}_n^T(k)$ as functions of k. Right panel: estimated number of static factors as functions of $r_{\max}$. The data set includes 111 US macroeconomic time series is from January 1973 to December 2011. Source: FRED database.

RECent Working Papers Series

The 10 most RECent releases are:

No. 123 EIGENVALUE RATIO ESTIMATORS FOR THE NUMBER OF DYNAMIC FACTORS (2016)
M. Cavicchioli, M. Forni, M. Lippi and P. Zaffaroni

No. 122 DETERMINANTS OF CENTRAL BANK INDEPENDENCE: A RANDOM FOREST APPROACH (2016)
M. Cavicchioli, A. Papana, A. Papana Dagiasis and B. Pistoresi

No. 121 THE INTERPLAY OF CULTURALAVERSION AND ASSORTATIVITY FOR THE EMERGENCE OF COOPERATION (2016)
E. Bilancini, L. Boncinelli and J. Wu

No. 120 DYNAMIC FACTOR MODEL WITH INFINITE DIMENSIONAL FACTOR SPACE: FORECASTING (2016)
M. Forni, A. Giovannelli, M. Lippi and S. Soccorsi

No. 119 VAR INFORMATION AND THE EMPIRICAL VALIDATION OF DSGE MODELS (2016)
M. Forni, L. Gambetti and L. Sala

No. 118 DYNAMIC FACTOR MODEL WITH INFINITE DIMENSIONAL FACTOR SPACE: FORECASTING (2016)
M. Forni, A. Giovannelli, M. Lippi and S. Soccorsi

No. 117 INEFFICIENCY IN CHILDCARE PRODUCTION. EVIDENCE FROM ITALIAN MICRODATA (2016)
L. Brighi and P. Silvestri

No. 116 TESTING THRESHOLD COINTEGRATION IN WAGNER'S LAW: THE ROLE OF MILITARY SPENDING (2016)
M. Cavicchioli and B. Pistoresi

No. 115 DYNAMIC FACTOR MODELS WITH INFINITE-DIMENSIONAL FACTOR SPACE: ASYMPTOTIC ANALYSIS (2015)
M. Forni, M. Hallin, M. Lippi and P. Zaffaroni

No. 114 ECONOMIC SHOCKS AND THEIR EFFECTS ON UNEMPLOYMENT IN THE EURO AREA PERIPHERY UNDER THE EMU (2015)
P. Dallari and A. Ribba

The full list of available working papers, together with their electronic versions, can be found on the RECent website: http://www.recent.unimore.it/workingpapers.asp