2009/97

A Robust Criterion for Determining the Number of Factors in Approximate Factor Models

Lucia ALESSI
Matteo BARIGOZZI
Marco CAPASSO

A Robust Criterion for Determining the Number of Factors in Approximate Factor Models

Lucia Alessi* Matteo Barigozzi† Marco Capasso‡

November 2009

Abstract

We modify the criterion by Bai and Ng (2002) for determining the number of factors in approximate factor models. As in the original criterion, for any given number of factors we estimate the common and idiosyncratic components of the model by applying principal component analysis. We select the true number of factors as the number that minimizes the variance explained by the idiosyncratic component. In order to avoid overparametrization, minimization is subject to penalization. At this step, we modify the original procedure by multiplying the penalty function by a positive real number, which allows us to tune its penalizing power, by analogy with the method used by Hallin and Liška (2007) in the frequency domain. The contribution of this paper is twofold. First, our criterion retains the asymptotic properties of the original criterion, but corrects its tendency to overestimate the true number of factors. Second, we provide a computationally easy way to implement the new method by iteratively applying the original criterion. Monte Carlo simulations show that our criterion is in general more robust than the original one. A better performance is achieved in particular in the case of large idiosyncratic disturbances. These conditions are the most difficult for detecting a factor structure but are not unusual in the empirical context. Two applications on a macroeconomic and a financial dataset are also presented.

Keywords: Approximate factor models, Information criterion, Model selection.
JEL-classification: C52.

---

[*] European Central Bank, Frankfurt am Main, Germany.

[†] European Center for Advanced Research in Economics and Statistics (ECARES), Université Libre de Bruxelles, Bruxelles, Belgium.

[‡] Corresponding author. Urban and Regional research centre Utrecht (URU), Faculty of Geosciences, Utrecht University, P.O. Box 80.115, 3508 TC Utrecht, The Netherlands, and Tjalling C. Koopmans research Institute (TKI), Utrecht School of Economics, Utrecht University, The Netherlands; Tel: +31 (0)30 2533206, Fax: +31 (0)30 253 2037, e-mail: marco.capasso@gmail.com

Matteo Barigozzi gratefully acknowledges financial support from Wallonie-Bruxelles International (WBI).

1 Introduction

Factor analysis is a very popular dimension reduction technique used in many disciplines as e.g. econometrics, statistics, signal processing, psychometrics. It allows to summarize the bulk of the information contained in large datasets by means of few factors which are pervasive and common to all considered variables. Each variable in the dataset can be decomposed into a common component generated by the factors and an idiosyncratic component uncorrelated with the factors.

Determining the number of factors is a crucial step in the identification of the factor structure. Indeed, as explained in Forni et al. (2009), there is a unique number of common factors for which the data satisfy the assumptions of an approximate factor model and the common and idiosyncratic component can be identified and consistently estimated.

Factor models are most useful when datasets are large in both the time (T) and the cross-section (N) dimension. However, this is precisely the case in which determining the number of common factors is particularly difficult as traditional information criteria as BIC or AIC cannot be applied anymore. Relatively few authors have dealt with the model selection problem related to the number of common factors when both N and T diverge. Bai and Ng (2002) pioneered the literature by proposing an information criterion aimed at minimizing the variance of the idiosyncratic components. They modify the usual AIC and BIC taking into account the double asymptotic framework when penalizing the criterion in order to avoid overparametrization. Still, in practice the method proposed by Bai-Ng is known to often deliver non-robust results as the number of factors can be either underestimated or, even worse, overestimated (see e.g. D'Agostino and Giannone, 2007; Forni et al., 2009; Ahn and Horenstein, 2008).

In this paper we propose a generalization of the Bai-Ng criterion. As in the original criterion, for any given number of factors we estimate the common and idiosyncratic components of the model by applying principal component analysis to the sample covariance matrix. The true number of factors is then selected as the number that minimizes the variance explained by the idiosyncratic component. To avoid including too many factors, this minimization is subject to penalization. At this step, we modify the original procedure by multiplying the penalty function times a positive real number which tunes the penalizing power of the function itself. In other words, our criterion is an iterative application of the Bai-Ng criterion, this latter being evaluated for a whole family of penalty functions rather than for just one specific penalty function. By evaluating the criterion for a whole range of values of the tuning parameter, we finally get an estimation of the number of static factors which is empirically more robust than it would be the penalty being fixed. Moreover, the consistency properties of our estimator are exactly the same as those of the original Bai-Ng estimator, the only difference being a multiplicative positive real number.

The method we use is taken from Hallin and Liška (2007), who develop an information criterion with a similar structure of the penalty function in order to estimate the number of factors in the Generalized Dynamic Factor Model by Forni et al. (2000). A distinction has to be made between *static* and *dynamic* factor models. While in the former we consider only the contemporaneous effects of the factors on the variables, in the latter we allow also for lagged dependencies. In the static case we just look at the eigenvalues of the covariance matrix of the data, while in the dynamic case, considered by Hallin and Liška, we consider the eigenvalues of the spectral density matrix. In this paper we examine only the static model as it can be used in many other fields besides time series and, as explained in detail in Section 7, in many econometric applications the number of static factors is a necessary and preliminary step for estimating the dynamic factor model.

Our criterion performs better than the Bai-Ng criterion especially when the variance of the observable data explained by the common factors is relatively low with respect to the variance of the idiosyncratic component. In this sense we say that our criterion is more robust than the Bai-Ng criterion. Our method is also more efficient in the sense that, when considering Monte Carlo simulations, the root mean square deviation from the true number of factors is often smaller when using our criterion.

These results constitute an improvement in the analysis of datasets where comovements among variables are hidden by large idiosyncratic disturbances. For example, in financial applications we often find few common factors explaining a small percentage of total variance. These factors are however of great importance for the structural analysis of financial markets (see e.g. Engle et al., 1990, where the unique factor is interpreted as the "market factor"). Although some authors identify the factor decomposition by requiring the idiosyncratic components to be "small" or "negligible" (e.g. in the case of principal component analysis), such characterization is not reflecting the fundamental nature of factor models: idiosyncratic components indeed can be "large" and strongly autocorrelated, while white noise can be common. The ability of identifying the true number of factors in large but finite and apparently highly heterogeneous datasets is therefore highly desirable.

Other criteria for determining the number of factors have been proposed in the literature. Kapetanios (2005) considers the limit of the empirical distribution of the eigenvalues of the sample covariance matrix. The idea is that the number of eigenvalues diverging as N diverges is equal to the number of factors driving the dataset. Onatski (2007) tests the null hypothesis of r_0 static factors against the alternative of r_1 static factors. The test is based on the few largest eigenvalues of the covariance matrix of a complex-valued sample derived from the original dataset, which asymptotically distribute as a Tracy-Widom. These criteria are rarely used in empirical applications, and therefore, we do not consider them in this paper. Also, we do not consider criteria for the number of dynamic factors as we are interested here only in the purely static factor model.

Among the many applications of factor models, we recall two main streams of literature. In financial econometrics two of the most important models for asset returns, i.e. the Capital Asset Pricing Model of Sharpe (1964) and Lintner (1965) and the Arbitrage Pricing Theory of Ross (1976), assume a factor structure. Factor models have been extensively used also in macroeconomics, generating a huge amount of applications in the fields of forecasting (see Stock and Watson, 1998), nowcasting (see e.g. the EuroCOIN coincident indicator for GDP of Forni et al., 2003; Altissimo et al., 2006), monetary policy (see e.g. the Factor-Augmented Vector AutoRegression of Bernanke and Boivin, 2003), and structural analysis (see Stock and Watson, 2005; Forni et al., 2009).

The paper is structured as follows. In the next Section we briefly recall the approximate factor model, its assumptions, and its estimation. In Section 3 we present our criterion. In Section 4 we provide a practical guide to the choice of the number of factors. In Section 5 we validate our method and compare it to the original criterion on the basis of a Monte Carlo study on four data generating processes. In Section 6 we present two empirical applications of the criterion. In Section 7 we briefly review the link between the static factor model considered in this paper and the more general dynamic model by Forni et al. (2000). In Section 8 we conclude.

2 The approximate factor model

Let us consider the following model for a large panel $\mathbf{x}_t$, composed of N time series

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \,, \tag{1}$$

where $\mathbf{F}_t$ is the $r \times 1$ vector of factors, r being the number of factors, and $\mathbf{\Lambda}$ is the $N \times r$ matrix of factor loadings. The data are represented as the sum of two components which we assume to be orthogonal, i.e. a common component $\mathbf{\Lambda}\mathbf{F}_t$, and an idiosyncratic component $\boldsymbol{\xi}_t$.

In classical factor models only the case with N fixed and T large is considered (see e.g. Anderson, 1963). Given the large datasets now available, there is the need of factor models that allow both for N and T large. Indeed, the classical models have severe limitations when applied to large datasets. First, while the true covariance can have rank N even if $N > T$, the rank of the usual $\sqrt{T}$-consistent $N \times N$ sample covariance matrix is always less than or equal to $\min(N, T)$. Second, the idiosyncratic covariance matrix must be diagonal whereas assuming some correlation among idiosyncratic components would be desirable. Third, maximum likelihood estimation is not feasible for large cross-sections due to the high number of parameters. Fourth, classical factor models deliver consistent estimates only for the loadings $\mathbf{\Lambda}$ while often the factors are of more interest.

Connor and Korajczyk (1986) consider the case of large N and T fixed. They solve the problem of estimating the covariance matrix by considering a $\sqrt{N}$-consistent $T \times T$ estimator. Moreover, they use principal components to consistently estimate the factors. In fact, principal components were already proposed by Chamberlain and Rothschild (1983) when considering a model with non-diagonal idiosyncratic covariance matrix. Indeed, having N large allows to disentangle the common and idiosyncratic components even when the latter are mildly correlated.

Bai and Ng (2002) and Bai (2003) consider an approximate factor model which allows to consistently estimate both the factors $\mathbf{F}_t$ and the loadings $\mathbf{\Lambda}$ when both N and T diverge. Moreover, they allow for cross-sectional, serial dependence, heteroskedasticity of $\boldsymbol{\xi}_t$, and for weak dependence between factors and idiosyncratic series.

The following assumptions must hold.

Assumption 1 a) The factors $\mathbf{F}_t$ are such that $\mathrm{E}(||\mathbf{F}_t||^4) < \infty$ and $T^{-1}\sum_{t=1}^{T} \mathbf{F}_t\mathbf{F}_t' \to \mathbf{\Gamma}^F$ as $T \to \infty$ for some positive definite matrix $\mathbf{\Gamma}^F$.
b) The loadings $\mathbf{\Lambda}$ are such that $||\boldsymbol{\lambda}|| \leq \bar{\lambda} < \infty$ and $||\mathbf{\Lambda}'\mathbf{\Lambda}/N - \mathbf{D}|| \to 0$ as $N \to \infty$ for some positive definite matrix $\mathbf{D}$.

Assumption 2 There exists a positive constant $M < \infty$ such that for all N and T,
a) $\mathrm{E}[\xi_{it}] = 0$ and $\mathrm{E}[|\xi_{it}|^8] \leq M$;
b) time dependence:
define $\gamma_N(t,s) := \mathrm{E}[\boldsymbol{\xi}_t'\boldsymbol{\xi}_s/N]$, then $|\gamma_N(s,s)| \leq M$ for all s and $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} |\gamma_N(t,s)| \leq M$;
c) cross-sectional dependence.
define $\tau_{ij,t} := \mathrm{E}[\xi_{it}\xi_{jt}]$, then $|\tau_{ij,t}| \leq |\tau_{ij}|$ for all t and $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} |\tau_{ij}| \leq M$;
d) define $\tau_{ij,ts} := \mathrm{E}[\xi_{it}\xi_{js}]$, then $(NT)^{-1} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |\tau_{ij,ts}| \leq M$;
e) for any t, s, $\mathrm{E}[|N^{-1/2}\sum_{i=1}^{N}[\xi_{is}\xi_{it} - \mathrm{E}(\xi_{is}\xi_{it})]|^4] \leq M$.

Assumption 3 The factors and the idiosyncratic components are allowed to be weakly dependent:

$$\mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left|\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{F}_t\xi_{it}\right|\right|^2\right] \leq M.$$

Assumption 1 is standard in approximate factor models. Assumption 2 allows for many different specifications of the idiosyncratic dynamics. Notice that we do not require stationarity. In particular, if we assume ξ_{it} to be stationary then its covariance matrix is bounded and, therefore, Assumption 2.c would imply that the largest eigenvalue of $\mathrm{E}[\boldsymbol{\xi}_t\boldsymbol{\xi}_t']$ is bounded. This is the kind of assumption we would require when considering the static representation of a dynamic factor model (see Forni et al., 2009). Finally, Assumption 3 allows for some mild correlation between common and idiosyncratic components. Both these Assumptions consitute a generalization with respect to the standard approximate factor model in the sense of Chamberlain and Rothschild (1983).

In large cross-sections, the r factors can be consistently estimated by means of asymptotic principal components, the serial and cross-sectional dependence across the idiosyncratic components being not enough to survive aggregation. We have to minimize the variance of the idiosyncratic components

$$(\widehat{\boldsymbol{\Lambda}}, \widehat{\mathbf{F}}_t) = \arg\min_{\boldsymbol{\Lambda},\mathbf{F}_t} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \boldsymbol{\lambda}_i'\mathbf{F}_t)^2, \tag{2}$$

where $\boldsymbol{\lambda}_i$ is the i-th row of $\boldsymbol{\Lambda}$. Define $\boldsymbol{\Gamma}^x := \mathrm{E}[\mathbf{x}_t\mathbf{x}_t']$. There are two possible equivalent solutions to this problem:

1. if $T > N$, we impose $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N = \mathbf{I}_r$, the estimated covariance matrix $\widehat{\boldsymbol{\Gamma}}^x$ is $N \times N$ and $\widehat{\boldsymbol{\Lambda}}$ are $\sqrt{N}$ times the eigenvectors associated to the r largest eigenvalues of $\widehat{\boldsymbol{\Gamma}}^x$, then $\widehat{\mathbf{F}}_t = \widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t/N$;

2. if $T < N$, we impose $\mathbf{F}_t'\mathbf{F}_t/T = \mathbf{I}_r$, the estimated covariance matrix $\widehat{\boldsymbol{\Gamma}}^x$ is $T \times T$ and $\widehat{\mathbf{F}}_t$ are $\sqrt{T}$ times the eigenvectors associated to the r largest eigenvalues of $\widehat{\boldsymbol{\Gamma}}^x$, then $\widehat{\boldsymbol{\Lambda}} = \mathbf{x}_t\widehat{\mathbf{F}}_t'/T$.

In both cases $\mathbf{F}_t$ and $\boldsymbol{\Lambda}$ are identified and consistently estimated only up to an orthogonal transformation. Still, the common component $\boldsymbol{\Lambda}\mathbf{F}_t$ is uniquely identified and consistently estimated. We refer to Theorem 1 in Bai and Ng (2002) for the proof of consistency. Finally, notice also that everything that follows holds for a more general class of factors' estimators (see Corollary 2 in Bai and Ng, 2002), provided that they satisfy Theorem 1 in Bai and Ng (2002). Indeed, all estimators satisfying such Theorem span the same r-dimensional space.

3 Determining the number of factors

Bai-Ng propose an information criterion to determine the true number of factors in model (1). If we assume to have k common factors, we can estimate $\widehat{\mathbf{F}}_t^{(k)}$ and their loadings $\widehat{\boldsymbol{\Lambda}}^{(k)}$ with asymptotic principal components. Once the factors are estimated, the residual variance of the idiosyncratic components is a function of k

$$V(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \widehat{\boldsymbol{\lambda}}_i^{(k)\prime}\widehat{\mathbf{F}}_t^{(k)})^2. \tag{3}$$

We need to minimize $V(k)$ in order to find the true number of factors. Overparametrizing is then avoided by introducing a penalty function, which counterbalances the fit improvement due to the inclusion of additional common factors. If the factors were known, then a usual BIC criterion would consistently estimate r for T diverging. However, given the double asymptotic framework necessary to have consistency of the estimated common component, we need a crtierion that depends on both N and T. Bai-Ng propose two classes of criteria, while here we consider only the log-version as it is the recommended one:

$$IC_N^T(k) = \log\left[V(k)\right] + kp(N,T). \tag{4}$$

Therefore, the estimated number of factors $\hat{r}_N^T$ is obtained by minimizing (4) and its consistency is proved in Theorem 2 of Bai and Ng (2002), that we restate here.

Theorem [Bai and Ng (2002)] *Suppose that Assumptions 1 to 3 hold, and that we have a consistent estimate of the factors and their loadings as $N, T \to \infty$. Let*

$$\hat{r}_N^T = \arg\min_{0 \le k \le r_{\max}} IC_N^T(k).$$

Assume also that

$$\lim_{\substack{N\to\infty\\T\to\infty}} p(N,T) = 0, \qquad \lim_{\substack{N\to\infty\\T\to\infty}} p(N,T)[\min(\sqrt{N}, \sqrt{T})]^2 = \infty.$$

Then,

$$\lim_{\substack{N\to\infty\\T\to\infty}} \text{Prob}[\hat{r}_N^T = r] = 1.$$

According to what Hallin and Liška (2007) propose in a similar criterion for the number of dynamic factors, a penalty $p(N,T)$ leads to consistent estimation of r if and only if $cp(N,T)$ does, where c is an arbitrary positive real number. Thus, multiplying the penalty by c has no influence on the asymptotic performance of the identification method. However, for given N and T, the value of a penalty function $p(N,T)$ satisfying (4) can be arbitrarily small or arbitrarily large, and this indeterminacy can affect the actual result quite dramatically, as it will be shown in the following Section on the implementation of the criterion.

Bai-Ng propose three choices for the penalty function and indicate the corresponding criteria as IC_1, IC_2, and IC_3. However, only the first two are known to behave well in empirical applications. We therefore propose the two information criteria:

$$\begin{aligned} IC_{1,N}^{T*}(k) &= \log\left[V(k)\right] + c\,k\left(\frac{N+T}{NT}\right)\log\left(\frac{NT}{N+T}\right) \qquad c \in \mathbb{R}^+, \\ IC_{2,N}^{T*}(k) &= \log\left[V(k)\right] + c\,k\left(\frac{N+T}{NT}\right)\log(\min\{\sqrt{N}, \sqrt{T}\})^2 \qquad c \in \mathbb{R}^+. \end{aligned}$$

The estimated number of factors is now a function of c and, depending on the chosen criterion, is given by

$$\hat{r}_{c,N}^T = \underset{0 \le k \le r_{\max}}{\operatorname{argmin}} \; IC_{a,N}^{T*}(k) \quad \text{with} \quad a = 1, 2.$$

4 A guide to the selection of r

The degree of freedom represented by c can be exploited when implementing the criterion in practice. The procedure we propose is the same as the one used by Hallin and Liška (2007) when determining the number of dynamic factors.[1]

The only information we have about the asymptotic behavior of $\hat{r}^T_{c,N}$ comes from considering subsamples of sizes (n_j, τ_j) with $j = 0, \dots, J$ such that $n_0 = 0 < n_1 < n_2 < \dots < n_J = N$ and $\tau_0 = 0 < \tau_1 < \tau_2 < \dots < \tau_J = T$. For any j we can compute $\hat{r}^{\tau_j}_{c,n_j}$, which is a nonincreasing function of c. Assume $r > 0$. According to the value of c we have different behaviors of $\hat{r}^{\tau_j}_{c,n_j}$ as function of j.

No penalty If $c = 0$ then $\hat{r}^{\tau_j}_{0,n_j} = r_{\max}$, indeed no penalization is imposed. Moreover, Theorem 2 does not apply.

Underpenalization If $c > 0$, but small, Theorem 2 applies but, in practice, as j increases $\hat{r}^{\tau_j}_{c,n_j}$ increases to $r_{\max}$ and would converge to r only if N and T would increase without limits. In this case we overestimate r.

Overpenalization When c becomes large, $\hat{r}^{\tau_j}_{c,n_j}$ tends to zero for any j and r is underestimated.

Due to the monotonicity of $\hat{r}^{\tau_j}_{c,n_j}$ as a function of c, there must exist a range of moderate values of c such that $\hat{r}^{\tau_j}_{c,n_j}$ converges from above to r. However, for the stability of the criterion, we require this convergence to be independent of j. The stability with respect to sample size can be measured by the empirical variance of $\hat{r}^{\tau_j}_{c,n_j}$ as a function of j, i.e.

$$S_c = \frac{1}{J}\sum_{j=1}^{J}\left[\hat{r}^{\tau_j}_{c,n_j} - \frac{1}{J}\sum_{j=1}^{J}\hat{r}^{\tau_j}_{c,n_j}\right]^2 .$$

The procedure for selecting the number of static factors basically explores the behavior of the variance S_c of the estimated number of factors for different subpanels of size (n_j, τ_j), for a whole interval of values for c.

In practice we have to follow these steps:

1. Set the maximum number of common factors $r_{\max}$. We will show in the next section that our criterion does not heavily depend on this choice. Thus, in practice, to be sure to find the right number of static factors, we can choose a very high value for $r_{\max}$, up to $\min(N, T)$.

2. Set an upper bound for c, i.e. $c \in (0, c_{\max}]$, where the value $c_{\max}$ must be high enough to have $\hat{r}^T_{c_{\max},N} = 0$.

3. For each considered value of c, and for different subsamples of increasing dimensions n_j and τ_j, where $j = 0, \dots, J$, with j integer, such that $N - n_1$ and $T - \tau_1$ are not too small and $n_0 = \tau_0 = 0$, perform the following:
   (a) compute the number of factors k that minimizes $IC^{\tau_j *}_{a,n_j}(k)$;
   (b) compute the variance S_c of the estimated number of factors as $n_j \to N$ and $\tau_j \to T$.

4. Look for the first interval $[\underline{c}, \overline{c}]$ (starting from small values) of c, such that simultaneously

---

[1] All estimations in this paper were performed using Matlab (R2007a). The code is available at `http://www.barigozzi.eu/ABC_crit.zip`.

(a) no dependence on the sample size is present, i.e. $\exists c \in [\underline{c}, \overline{c}] : S_c = 0$;

(b) when considering the whole $N \times T$ sample (i.e. $j = J$), the number of factors $\hat{r}^T_{c,N}$ is constant for any $c \in [\underline{c}, \overline{c}]$; this number is the estimated number of static factors. The value $r_{\max}$ which corresponds to $c = 0$ cannot be considered as a solution as, in this case, Theorem 2 does not hold.

Figure 1 shows how the $IC^{T*}_{1,N}(k)$ criterion works when applied to a simulated sample with 5 common factors. As c increases, the solid line provides the suggested number of factors. A plateau of the solid line means a region where the suggested number of factors $r^T_{c,N}$ is stable across different values of c. On the other hand, the dashed line provides a measure of the instability of $r^T_{c,N}$ when different subsamples of the dataset are considered. When the dashed line goes to zero, the value provided by the solid line is stable across subsamples of different size, ensuring that the choice of c has not been affected by the size of the sample. Therefore, we have to choose the smallest value of c for which both a plateau of the solid line (not including the extreme left one) and a zero of the dashed line occur. The $IC^T_{1,N}(k)$ criterion by Bai-Ng implicitly considers only the case $c = 1$. In the simulation of Figure 1 when $c = 1$ the suggested number of factors is $r^T_{1,N} = 4$, which is smaller than the true one $r = 5$. Moreover, when $c = 1$ we have $S_c > 0$. Our refinement considers also different values of c, thus finding a range of values for c (i.e. the interval around $c = 0.5$) for which the number of factors suggested by the solid line is the correct one. The estimated number in this case is stable across adjacent values of c (plateau of the solid line) and across different subsample sizes (zero of the dashed line). This way, we avoid an overpenalization of the number of factors and find as result the true number of static factors.

5 Simulations

In this Section we conduct a set of simulation experiments to evaluate the performance of our refined criterion, relative to that of the original Bai-Ng criterion, in finite samples. We consider four data-generating processes (DGPs) similar to those in Bai and Ng (2002).

The baseline model is:

$$x_{it} = \sum_{j=1}^{r} \lambda_{ij} F_{tj} + \sqrt{\theta}\xi_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \ ,$$

with factors and factor loadings normally distributed with zero mean and unit variance. The different specifications are as follows.

DGP1 Homoskedastic idiosyncratic component

$$\xi_{it} \sim N(0, 1).$$

DGP2 Heteroskedastic idiosyncratic component

$$\xi_{it} = \begin{cases} \xi^1_{it} & \text{if } t \text{ odd} \\ \xi^1_{it} + \xi^2_{it} & \text{if } t \text{ even} \end{cases} , \quad \xi^1_{it}, \xi^2_{it} \sim N(0, 1).$$

DGP3 Cross-sectional correlations across idiosyncratic parts

$$\xi_{it} = v_{it} + \sum_{j \neq 0 \; j=-J}^{J} \beta v_{i-jt} \, , \quad v_{it} \sim N(0, 1).$$

DGP4 Serial correlation across idiosyncratic parts

$$\xi_{it} = \rho \xi_{it-1} + v_{it}\,, \quad \xi_{it} \sim N(0,1)\,, \quad v_{it} \sim N(0,1).$$

For each model we set $r = 1, 3, 5$ and $\theta = \frac{1}{2}r, r, 3r, 5r$, thus assigning to the idiosyncratic part a variance that is respectively one half, one, three or five times the variance of the common part. The values of the other parameters are chosen as in Bai and Ng (2002): $\rho = 0.5$, $\beta = 0.2$, and $J = \max\{N/20, 10\}$. We generate samples having size equal to $(T, N) = (50, 50), (200, 200)$. We thus have four variance-ratio settings. For each variance-ratio setting, we have three "true" numbers of factors, two sample sizes and four data generating processes. We set $r_{\max} = 10$, $n_1 = \frac{3}{4}N$ and $c \in (0, 5]$ with step size of 0.01. For simplicity, we do not consider subsamples in the time dimension. For each case, we simulate 1000 samples.

For each model and each sample we implement the original Bai-Ng criterion IC_1, and our modification IC_1^*. We do not show here results for IC_2 and IC_2^*, but they are available upon request. Moreover, as the original IC_1 and IC_2 Bai-Ng criteria differ only in the weight assigned to the penalty function, it is enough to test only our proposed IC_1^*. Indeed, the main step we add to the original procedure consists exactly in tuning the penalty function, therefore there is no influence exerted by the weight originally assigned to it.

Tables 1, 2, 3 and 4 show, respectively for $\theta = r, \frac{1}{2}r, 3r, 5r$, the frequency with which a given number of factors is selected by the two chosen forms of the original Bai-Ng criteria and our proposed criterion.

In Table 1 we consider the baseline scenario of equal variance for idiosyncratic and common components of the simulated data. When there is only one factor, our criterion performs slightly worse than the original criterion. There is however one exception, which is the case in which DGP3 is simulated with a large sample size: when a small cross-correlation of the idiosyncratic components is allowed, our criterion does not diverge and is able to detect the true number of factors in more than 80% of the Monte Carlo simulations. Instead, the Bai-Ng criterion applied to DGP3 always diverges. As for the other DGPs, the traditional criterion does not seem to outperform our proposed criterion.

Similar results hold for the case of a common variance which is double than the idiosyncratic variance (see Table 2). Our criterion performs similarly to the traditional Bai-Ng criterion for DGP1, DGP2 and DGP4, and performs much better for DGP3 especially in the large sample case.

The advantages of our criterion become clear when we raise the idiosyncratic variance to three times the common variance. As already observed by Bai and Ng (2002) themselves and recently confirmed by Ahn and Horenstein (2008), a high idiosyncratic variance negatively affects the ability of the Bai-Ng criteria in determining the true number of factors on simulated data. Our information criterion is able to work also under such noisy conditions. When considering the large sample case of DGP3 the Bai-Ng criterion diverges and suggest the maximum possible number of factors, i.e. 10 (see Table 3). Also in the small sample case the traditional criterion is often not able to recognize any common variance, especially when the true number of factors is high. Our refinement may underestimate the number of factors, but the underestimation bias is clearly much smaller than the bias obtained with the original Bai-Ng criterion (see the Root Mean Squared Deviations). When there is only one factor in the DGP, our criterion performs slightly worse than the other for DGPs 1, 2 and 3, although this happens in less than 10% of cases. When the true number of factors is higher, our criterion outperforms the Bai-Ng criterion providing in almost all the cases the right number of factors. See for example the case of 5 true factors and long sample ($T = 200$): out of 1000 replications, our criterion was able to retrieve the right number of factors in 999 cases for DGP1 (against 967), in 999 cases for DGP2 (against 32), and in 977 cases for DGP4 (against 418).

If the idiosyncratic variance is raised to 5 times the common variance (see Table 4) and we still consider 5 true factors and long sample, the gain obtained by our proposed criterion is even more evident: out of 1000 replications, the right number of factors is retrieved in 998 cases for DGP1 (against 1), in 969 cases for DGP2 (against 0), and in 482 cases for DGP4 (against 0). For DGP3, again we have the same qualitative results: the traditional criterion diverges, while our criterion suggests 0 factors in 674 cases and a number between 1 and 9 factors for the remaining cases. When we consider the smaller sample, DGP3 is slightly more favourable to the original criterion, but our criterion is outperforming in all the other cases.

Summing up, our refinement makes the Bai-Ng criterion a much more useful tool in practical contexts. Indeed, it enormously reduces the probability of large mistakes and always provides a reliable answer, even when, in the presence of a factor structure in the DGP, the dataset presents some features of high idiosyncratic variance or heterokedasticity that would prevent the traditional criteria from suggesting a finite positive number of factors. Notice that this success is not a technical artifact depending on a tendency to always retrieve a positive number of factors. It is easy to check that when there is no factor structure in the simulated data, the proposed criterion suggests a number of factors equal to zero, exactly as the traditional Bai-Ng criteria would do.

6 Empirical applications

We test the performance of our criterion by means of two empirical applications on macroeconomic and financial datasets. In the first case we take a dataset which has been used in many applications of factor models, (see e.g. Stock and Watson, 2005; Hallin and Liška, 2007). The dataset comprises 132 series of monthly macroeconomic indicators of the U.S. economy from January 1960 to December 2003 for a total of 528 time observations.[2] In a second exercise we consider 89 daily asset returns from the London Stock Exchange, from 1st October 2001 to 31st July 2003 for a total of 456 time observations.

In Figure 2(a) we report results obtained for the macroeconomic application. IC_1^* indicates the presence of 6 factors. The original criteria IC_1 and IC_2 point to 4 factors. However, for $r_{\max} = 100$ the IC criteria lose their power and yield the maximum possible number of factors, while both our criteria are robust, suggesting always 6 factors.

In Figure 2(b) we report the results for the financial dataset. IC_1^* indicates the presence of 6 factors, while the original ICs always indicate 1 or 2 factors. However, this result could underestimate the true number of factors, as shown in Figure 2(b). The plot refers to the IC_1^* criterion with $r_{\max} = 10$. It is clear that the original IC_1 criterion is identifying the largest plateau, while our criterion finds another smaller plateau corresponding to 6 factors and c in the interval around 0.28. Another smaller plateau can be identified when $r = 4$ and c around 0.35.

7 On the static and dynamic representations

In this section we briefly explore the relation between the approximate factor model we consider in this paper and a more general dynamic factor model. The approximate factor model of Bai

---

[2] Data are downloadable at `http://www.princeton.edu/~mwatson`

and Ng (2002) can be seen as the static representation of a purely dynamic model. Indeed, we can add to equation (1) an equation specifying the dynamics of the factors, as follows:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{\Lambda} \mathbf{F}_t + \boldsymbol{\xi}_t, \\ \mathbf{F}_t &= \mathbf{A} \mathbf{F}_{t-1} + \mathbf{B} \mathbf{u}_t, \end{aligned} \tag{5}$$

where $\mathbf{A}$ is $r \times r$, $\mathbf{B}$ is $r \times q$, $\mathbf{u}_t$ is a q-dimensional vector of dynamic factors (white noise in this case), and $q < r$. By plugging the second equation into the first one we get

$$\mathbf{x}_t = \mathbf{\Lambda}(\mathbf{I}_r - \mathbf{A}L)^{-1}\mathbf{B}\mathbf{u}_t + \boldsymbol{\xi}_t, \tag{6}$$

where $\mathbf{I}_r$ is the r-dimensional identity matrix and L is the lag operator. Equation (6) is a particular case of the Generalized Dynamic Factor Model by Forni et al. (2000). It is important to stress that (5) exists with a finite number of factors r if and only if the space spanned by the q dynamic factors is finite dimensional (see Remark R in Forni et al., 2009).

The advantage of representation (5) or (6) versus (1) is that, by taking into account also the dynamic structure, we consider also the non-contemporaneous comovements among the observed variables. In this way the model is particularly useful for forecasting. Morover (5) has an immidiate economic interpretation as it can be considered as the state-space formulation of the steady state of a Dynamic Stochastic General Equilibrium model. The $\mathbf{F}_t$ would be the unobserved state variables (e.g. capital) and the $\mathbf{u}_t$ would be the economic shocks (see e.g. Giannone et al., 2006).

Notice that, even when postulating the existence of a dynamic representation as (5), more general than the static representation we deal with in this paper, a criterion to determine the number of static factors r is still necessary. However, if we consider the static model as a restricted version of the dynamic model, we can use the equivalence between (1) and (5) to confirm ex-post the validity of our criterion. The key point here is that adding an equation which specifies the dynamics of the factors has no impact on the decomposition of each of the observables into a common component $\Lambda \mathbf{F}_t$ and an idiosyncratic component ξ_t. The estimates of the common components should coincide both when assuming a static representation or a dynamic representation. Therefore, we expect the r static factors and the q dynamic factors to explain the same amount of variance.

Take as an example the second empirical application presented in this paper. When applying on this dataset the criterion by Hallin and Liška (2007) for determining q, the suggested number of factors is two. These two dynamic factors explain 44% of the total variance, which approximately corresponds to the percentage of variance explained by 5 static factors, thus supporting the result suggested by our criterion, i.e. the two plateaux at $r = 4$ and $r = 6$.

It is worth stressing, however, that when we compare the variance explained by the static and the dynamic factors, we are implicitly making the assumption of a finite dimensional space spanned by the dynamic factors. If this were not the case, then no static representation with $r < \infty$ would exist and no criteria for determining r would provide a result consistent with reality. Indeed, it is recognized in the literature that in many empirical applications the information criterion proposed by Bai and Ng (2002) is not able to converge to a minimum even when the maximum number of factors allowed for is set to a high value (see e.g. Forni et al., 2009). This lack of convergence may be precisely a hint towards the existence of an infinite dimensional space spanned by the dynamic factors (e.g. one dynamic factor with autoregressive loadings). In this case no criterion for the number of static factors can be expected to converge.

8 Conclusions

In this paper we propose an information criterion for the determination of the number of static factors in approximate factor models. We refine the Bai and Ng (2002) criterion, which is one of the most popular criteria available for addressing this issue. The appeal of our new method is three-fold: (i) it builds on a well known criterion, whose theoretical properties have been proved and are preserved; (ii) it improves the finite sample performance of the original criterion; (iii) it is easy to implement. In particular, our proposed criterion is capable of giving an answer even when the original criterion does not converge. Our criterion and the Bai-Ng benchmark have been compared on the basis of simulations and two empirical applications on macroeconomic and financial data, with encouraging results. In general, our criterion yields more robust results with respect to the Bai-Ng criterion, especially when the variance explained by the common factors is relatively low.

The potential applications of our method go beyond the estimation of the number of static factors. Indeed, as we explain in the previous Section, the static factor model we deal with in this paper can be derived from a restricted version of a more general dynamic factor model. In such a model, the focus is on the dynamic factors, or primitive shocks, which are fewer that the static factors. There exist at least three criteria where the number of dynamic common factors is determined on the basis of a first estimate of the number of static factors. These criteria have been proposed by Amengual and Watson (2007), who study the consistency properties of an estimator proposed in Stock and Watson (2005); Bai and Ng (2007), who focus on the variance explained by static and dynamic factors; and Breitung and Kretschmer (2005), who apply canonical correlation analysis to the estimated static factors. All the properties of these criteria still hold when our refinement of the Bai-Ng criterion is implemented, therefore they could be modified according to what we propose in this paper.

References

Ahn, S. C. and Horenstein, A. (2008). Eigenvalue ratio test for the number of factors. Working paper, Arizona State University. Downloadable at www.wlu.ca/documents/32292/Eigenvalue_Ratio_Test_for_the_Number_of_Factors.pdf.

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2006). New eurocoin: Tracking economic growth in real time. CEPR Discussion Papers 5633, C.E.P.R. Discussion Papers.

Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, 25(1):91–96.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis,. *Annals of Mathematical and Statistics*, 34:122–148.

Bai, J. (2003). Inferential theory for factor models of large dimensions,. *Econometrica*, 71:135–171.

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

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

Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.

Breitung, J. and Kretschmer, U. (2005). Determining the number of dynamic factors in large macroeconomic panels. Working paper, University of Bonn.

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

Connor, G. and Korajczyk, R. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis,. *Journal of Financial Economics*, 15:373–394.

D'Agostino, A. and Giannone, D. (2007). Comparing alternative predictors based on large-panel factor models. Discussion Papers 6564, CEPR.

Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2):213–237.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross-sections. *Econometric Theory*. Online June 2009.

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(4):540–554.

Forni, M., Lippi, M., Reichlin, L., Altissimo, F., and Bassanetti, A. (2003). Eurocoin: A real time coincident indicator of the euro area business cycle. Computing in Economics and Finance 2003 242, Society for Computational Economics.

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.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478).

Kapetanios, G. (2005). A testing procedure for determining the number of factors in approximate factor models with large datasets. Working Papers 551, Queen Mary, University of London, Department of Economics.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budget. *Review of Economics and Statistics*, 47:13–37.

Onatski, A. (2007). A formal statistical test for the number of factors in the approximate factor models. Mimeo, Columbia University, Department of Economics.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Sharpe, W. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Stock, J. H. and Watson, M. W. (1998). Diffusion indexes. NBER Working Papers 6702, National Bureau of Economic Research, Inc.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis. NBER Working Papers 11467, National Bureau of Economic Research, Inc.

r	Sample	DGP		0	1	2	3	4	5	6	7	8	9	10	RMSD
1	1	1	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	1	1	IC_1^*	0	991	9	0	0	0	0	0	0	0	0	0.09
1	1	2	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	1	2	IC_1^*	0	983	17	0	0	0	0	0	0	0	0	0.13
1	1	3	IC_1	0	0	0	0	659	268	62	9	2	0	0	3.49
1	1	3	IC_1^*	0	0	0	0	764	163	55	13	5	0	0	3.40
1	1	4	IC_1	0	904	81	10	4	1	0	0	0	0	0	0.42
1	1	4	IC_1^*	0	876	91	22	5	2	2	1	1	0	0	0.63
1	2	1	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	1	IC_1^*	0	999	1	0	0	0	0	0	0	0	0	0.03
1	2	2	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	2	IC_1^*	0	998	2	0	0	0	0	0	0	0	0	0.04
1	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	9
1	2	3	IC_1^*	0	825	27	6	7	5	0	12	38	80	0	2.76
1	2	4	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	4	IC_1^*	0	966	32	2	0	0	0	0	0	0	0	0.20
3	1	1	IC_1	0	0	5	995	0	0	0	0	0	0	0	0.07
3	1	1	IC_1^*	0	0	0	974	24	2	0	0	0	0	0	0.18
3	1	2	IC_1	0	8	116	876	0	0	0	0	0	0	0	0.38
3	1	2	IC_1^*	1	2	32	940	24	1	0	0	0	0	0	0.28
3	1	3	IC_1	0	0	0	0	0	1	589	313	80	14	3	3.60
3	1	3	IC_1^*	2	0	0	1	3	13	733	185	49	14	0	3.37
3	1	4	IC_1	0	0	24	838	117	18	2	1	0	0	0	0.50
3	1	4	IC_1^*	11	16	72	779	94	19	5	2	0	2	0	0.74
3	2	1	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	1	IC_1^*	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	2	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	2	IC_1^*	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	7
3	2	3	IC_1^*	3	1	12	909	36	7	5	3	11	13	0	0.97
3	2	4	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	4	IC_1^*	0	0	0	968	30	2	0	0	0	0	0	0.19
5	1	1	IC_1	0	1	7	66	306	620	0	0	0	0	0	0.81
5	1	1	IC_1^*	11	10	15	20	88	838	17	1	0	0	0	0.87
5	1	2	IC_1	21	93	209	316	279	82	0	0	0	0	0	2.33
5	1	2	IC_1^*	133	79	69	118	221	363	12	4	0	1	0	2.44
5	1	3	IC_1	0	0	0	0	0	0	1	34	532	304	129	3.61
5	1	3	IC_1^*	52	29	16	12	22	31	48	142	512	136	0	3.06
5	1	4	IC_1	1	10	50	150	319	379	76	9	4	2	0	1.32
5	1	4	IC_1^*	237	125	129	109	173	168	34	13	4	8	0	3.15
5	2	1	IC_1	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	1	IC_1^*	0	0	0	0	0	998	2	0	0	0	0	0.04
5	2	2	IC_1	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	2	IC_1^*	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	5
5	2	3	IC_1^*	734	95	42	29	36	49	9	3	1	2	0	4.52
5	2	4	IC_1	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	4	IC_1^*	0	0	0	0	0	957	41	2	0	0	0	0.22

Table 1: Number of times the true number of factors is retrieved from simulated data. **Ratio between idiosyncratic and common variance = 1**. IC_1: Bai-Ng criterion; IC_1^*: our refinement; r: true number of factors; Sample = 1: small sample ($N = 50, T = 50$); Sample = 2: big sample ($N = 200, T = 200$); DGP: data generating process as defined in the text; RMSD: root mean square deviation from the true number of factors.

r	Sample	DGP		0	1	2	3	4	5	6	7	8	9	10	RMSD
1	1	1	IC_1	0	999	1	0	0	0	0	0	0	0	0	0.03
1	1	1	IC_1^*	0	985	14	1	0	0	0	0	0	0	0	0.13
1	1	2	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	1	2	IC_1^*	0	972	27	1	0	0	0	0	0	0	0	0.18
1	1	3	IC_1	0	0	0	0	501	341	124	28	4	1	1	3.80
1	1	3	IC_1^*	0	0	0	0	776	165	46	12	1	0	0	3.36
1	1	4	IC_1	0	767	176	42	11	4	0	0	0	0	0	0.71
1	1	4	IC_1^*	0	866	106	18	8	2	0	0	0	0	0	0.53
1	2	1	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	1	IC_1^*	0	999	1	0	0	0	0	0	0	0	0	0.03
1	2	2	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	2	IC_1^*	0	998	2	0	0	0	0	0	0	0	0	0.04
1	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	9
1	2	3	IC_1^*	0	805	37	11	3	3	9	12	32	88	0	2.83
1	2	4	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	4	IC_1^*	0	959	35	5	0	1	0	0	0	0	0	0.27
3	1	1	IC_1	0	0	0	999	1	0	0	0	0	0	0	0.03
3	1	1	IC_1^*	0	0	0	963	36	1	0	0	0	0	0	0.20
3	1	2	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	1	2	IC_1^*	0	0	0	958	39	3	0	0	0	0	0	0.23
3	1	3	IC_1	0	0	0	0	0	0	426	384	141	43	6	3.91
3	1	3	IC_1^*	0	0	0	0	0	1	727	198	52	22	0	3.44
3	1	4	IC_1	0	0	0	657	244	79	13	5	2	0	0	0.90
3	1	4	IC_1^*	0	0	1	836	117	29	12	2	2	1	0	0.68
3	2	1	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	1	IC_1^*	0	0	0	998	2	0	0	0	0	0	0	0.04
3	2	2	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	2	IC_1^*	0	0	0	995	5	0	0	0	0	0	0	0.07
3	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	7
3	2	3	IC_1^*	0	0	0	896	47	17	7	3	4	26	0	1.12
3	2	4	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	4	IC_1^*	0	0	0	961	35	3	1	0	0	0	0	0.24
5	1	1	IC_1	0	0	0	0	6	994	0	0	0	0	0	0.08
5	1	1	IC_1^*	0	0	0	0	3	955	37	4	0	1	0	0.27
5	1	2	IC_1	0	0	0	1	86	912	1	0	0	0	0	0.30
5	1	2	IC_1^*	7	3	4	10	54	877	41	3	1	0	0	0.64
5	1	3	IC_1	0	0	0	0	0	0	0	0	381	396	223	3.92
5	1	3	IC_1^*	10	1	0	0	0	2	2	23	789	173	0	3.20
5	1	4	IC_1	0	0	0	0	7	656	220	79	26	8	4	1.00
5	1	4	IC_1^*	19	8	3	16	78	746	90	26	8	6	0	1.06
5	2	1	IC_1	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	1	IC_1^*	0	0	0	0	0	997	3	0	0	0	0	0.05
5	2	2	IC_1	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	2	IC_1^*	0	0	0	0	0	995	5	0	0	0	0	0.07
5	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	5
5	2	3	IC_1^*	0	0	1	0	2	927	39	17	8	6	0	0.53
5	2	4	IC_1	0	0	0	0	0	1000	0	0	0	0	0	0
5	2	4	IC_1^*	0	0	0	0	0	954	41	4	1	0	0	0.26

Table 2: Number of times the true number of factors is retrieved from simulated data. **Ratio between idiosyncratic and common variance = 0.5**. IC_1: Bai-Ng criterion; IC_1^*: our refinement; r: true number of factors; Sample = 1: small sample ($N = 50, T = 50$); Sample = 2: big sample ($N = 200, T = 200$); DGP: data generating process as defined in the text; RMSD: root mean square deviation from the true number of factors.

r	Sample	DGP		0	1	2	3	4	5	6	7	8	9	10	RMSD
1	1	1	IC_1	1	999	0	0	0	0	0	0	0	0	0	0.03
1	1	1	IC_1^*	0	992	8	0	0	0	0	0	0	0	0	0.09
1	1	2	IC_1	21	979	0	0	0	0	0	0	0	0	0	0.14
1	1	2	IC_1^*	2	981	16	1	0	0	0	0	0	0	0	0.15
1	1	3	IC_1	0	0	0	0	756	212	31	1	0	0	0	3.31
1	1	3	IC_1^*	0	0	0	2	740	184	54	15	5	0	0	3.42
1	1	4	IC_1	4	959	35	2	0	0	0	0	0	0	0	0.22
1	1	4	IC_1^*	6	892	77	11	6	6	0	0	2	0	0	0.61
1	2	1	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	1	IC_1^*	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	2	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	2	IC_1^*	0	998	2	0	0	0	0	0	0	0	0	0.04
1	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	9
1	2	3	IC_1^*	2	834	30	5	1	2	4	14	37	71	0	2.66
1	2	4	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	4	IC_1^*	0	961	36	2	1	0	0	0	0	0	0	0.23
3	1	1	IC_1	198	377	351	74	0	0	0	0	0	0	0	1.90
3	1	1	IC_1^*	50	83	246	617	4	0	0	0	0	0	0	1.01
3	1	2	IC_1	736	228	35	1	0	0	0	0	0	0	0	2.75
3	1	2	IC_1^*	322	268	247	160	3	0	0	0	0	0	0	2.05
3	1	3	IC_1	0	0	0	0	29	206	556	176	29	3	1	3.09
3	1	3	IC_1^*	1	0	4	139	156	208	356	97	31	8	0	2.61
3	1	4	IC_1	322	403	221	54	0	0	0	0	0	0	0	2.18
3	1	4	IC_1^*	504	256	153	60	18	7	2	0	0	0	0	2.40
3	2	1	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	1	IC_1^*	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	2	IC_1	0	0	1	999	0	0	0	0	0	0	0	0.03
3	2	2	IC_1^*	0	0	0	999	1	0	0	0	0	0	0	0.03
3	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	7
3	2	3	IC_1^*	956	32	5	1	0	0	0	1	0	5	0	2.99
3	2	4	IC_1	0	0	0	1000	0	0	0	0	0	0	0	0
3	2	4	IC_1^*	0	0	0	971	27	2	0	0	0	0	0	0.19
5	1	1	IC_1	780	201	18	1	0	0	0	0	0	0	0	4.78
5	1	1	IC_1^*	429	227	135	132	61	14	2	0	0	0	0	4.02
5	1	2	IC_1	981	19	0	0	0	0	0	0	0	0	0	4.98
5	1	2	IC_1^*	688	225	71	16	0	0	0	0	0	0	0	4.64
5	1	3	IC_1	0	0	0	62	172	281	276	135	48	22	4	1.46
5	1	3	IC_1^*	0	1	6	500	216	138	61	43	18	17	0	1.72
5	1	4	IC_1	829	152	18	1	0	0	0	0	0	0	0	4.83
5	1	4	IC_1^*	785	153	44	12	5	0	1	0	0	0	0	4.75
5	2	1	IC_1	0	0	0	0	33	967	0	0	0	0	0	0.18
5	2	1	IC_1^*	0	0	0	0	0	999	1	0	0	0	0	0.03
5	2	2	IC_1	1	33	214	460	260	32	0	0	0	0	0	2.14
5	2	2	IC_1^*	0	0	0	0	0	999	1	0	0	0	0	0.03
5	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	5
5	2	3	IC_1^*	910	42	9	5	5	3	5	6	5	10	0	4.87
5	2	4	IC_1	0	0	7	146	429	418	0	0	0	0	0	1.04
5	2	4	IC_1^*	0	0	0	0	0	977	21	2	0	0	0	0.17

Table 3: Number of times the true number of factors is retrieved from simulated data. **Ratio between idiosyncratic and common variance = 3**. IC_1: Bai-Ng criterion; IC_1^*: our refinement; r: true number of factors; Sample = 1: small sample ($N = 50, T = 50$); Sample = 2: big sample ($N = 200, T = 200$); DGP: data generating process as defined in the text; RMSD: root mean square deviation from the true number of factors.

r	Sample	DGP		0	1	2	3	4	5	6	7	8	9	10	RMSD
1	1	1	IC_1	44	956	0	0	0	0	0	0	0	0	0	0.21
1	1	1	IC_1^*	0	995	5	0	0	0	0	0	0	0	0	0.07
1	1	2	IC_1	284	716	0	0	0	0	0	0	0	0	0	0.53
1	1	2	IC_1^*	30	948	21	1	0	0	0	0	0	0	0	0.23
1	1	3	IC_1	0	0	0	0	726	232	41	1	0	0	0	3.36
1	1	3	IC_1^*	0	0	0	11	734	191	53	9	2	0	0	3.38
1	1	4	IC_1	122	857	20	1	0	0	0	0	0	0	0	0.38
1	1	4	IC_1^*	117	767	83	25	5	1	1	1	0	0	0	0.65
1	2	1	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	1	IC_1^*	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	2	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	2	IC_1^*	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	9
1	2	3	IC_1^*	350	504	24	5	0	2	3	4	36	72	0	2.64
1	2	4	IC_1	0	1000	0	0	0	0	0	0	0	0	0	0
1	2	4	IC_1^*	0	969	29	1	1	0	0	0	0	0	0	0.20
3	1	1	IC_1	848	145	7	0	0	0	0	0	0	0	0	2.87
3	1	1	IC_1^*	336	280	274	109	1	0	0	0	0	0	0	2.10
3	1	2	IC_1	991	9	0	0	0	0	0	0	0	0	0	2.99
3	1	2	IC_1^*	678	254	58	10	0	0	0	0	0	0	0	2.68
3	1	3	IC_1	0	0	0	105	318	342	188	42	5	0	0	2.04
3	1	3	IC_1^*	0	0	0	425	258	154	107	37	13	6	0	1.72
3	1	4	IC_1	836	145	17	2	0	0	0	0	0	0	0	2.85
3	1	4	IC_1^*	756	175	48	13	4	2	1	1	0	0	0	2.75
3	2	1	IC_1	0	0	12	988	0	0	0	0	0	0	0	0.11
3	2	1	IC_1^*	0	0	0	999	1	0	0	0	0	0	0	0.03
3	2	2	IC_1	8	174	538	280	0	0	0	0	0	0	0	1.14
3	2	2	IC_1^*	0	0	0	999	1	0	0	0	0	0	0	0.03
3	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	7
3	2	3	IC_1^*	886	33	5	5	3	5	6	13	12	32	0	3.13
3	2	4	IC_1	1	8	249	742	0	0	0	0	0	0	0	0.54
3	2	4	IC_1^*	0	0	0	975	22	3	0	0	0	0	0	0.18
5	1	1	IC_1	994	6	0	0	0	0	0	0	0	0	0	4.99
5	1	1	IC_1^*	726	187	66	18	3	0	0	0	0	0	0	4.67
5	1	2	IC_1	1000	0	0	0	0	0	0	0	0	0	0	5
5	1	2	IC_1^*	889	96	12	2	1	0	0	0	0	0	0	4.89
5	1	3	IC_1	0	0	0	489	369	104	29	8	1	0	0	1.54
5	1	3	IC_1^*	0	0	0	680	209	65	28	9	5	4	0	1.76
5	1	4	IC_1	975	25	0	0	0	0	0	0	0	0	0	4.98
5	1	4	IC_1^*	874	103	16	5	1	0	1	0	0	0	0	4.86
5	2	1	IC_1	28	230	436	259	46	1	0	0	0	0	0	3.06
5	2	1	IC_1^*	0	0	0	0	0	998	2	0	0	0	0	0.04
5	2	2	IC_1	947	52	1	0	0	0	0	0	0	0	0	4.95
5	2	2	IC_1^*	3	1	1	4	21	969	1	0	0	0	0	0.37
5	2	3	IC_1	0	0	0	0	0	0	0	0	0	0	1000	5
5	2	3	IC_1^*	674	37	6	4	2	5	15	39	68	150	0	4.55
5	2	4	IC_1	637	319	42	2	0	0	0	0	0	0	0	4.63
5	2	4	IC_1^*	141	71	63	82	143	482	17	1	0	0	0	2.39

Table 4: Number of times the true number of factors is retrieved from simulated data. **Ratio between idiosyncratic and common variance = 5**. IC_1: Bai-Ng criterion; IC_1^*: our refinement; r: true number of factors; Sample = 1: small sample ($N = 50, T = 50$); Sample = 2: big sample ($N = 200, T = 200$); DGP: data generating process as defined in the text; RMSD: root mean square deviation from the true number of factors.

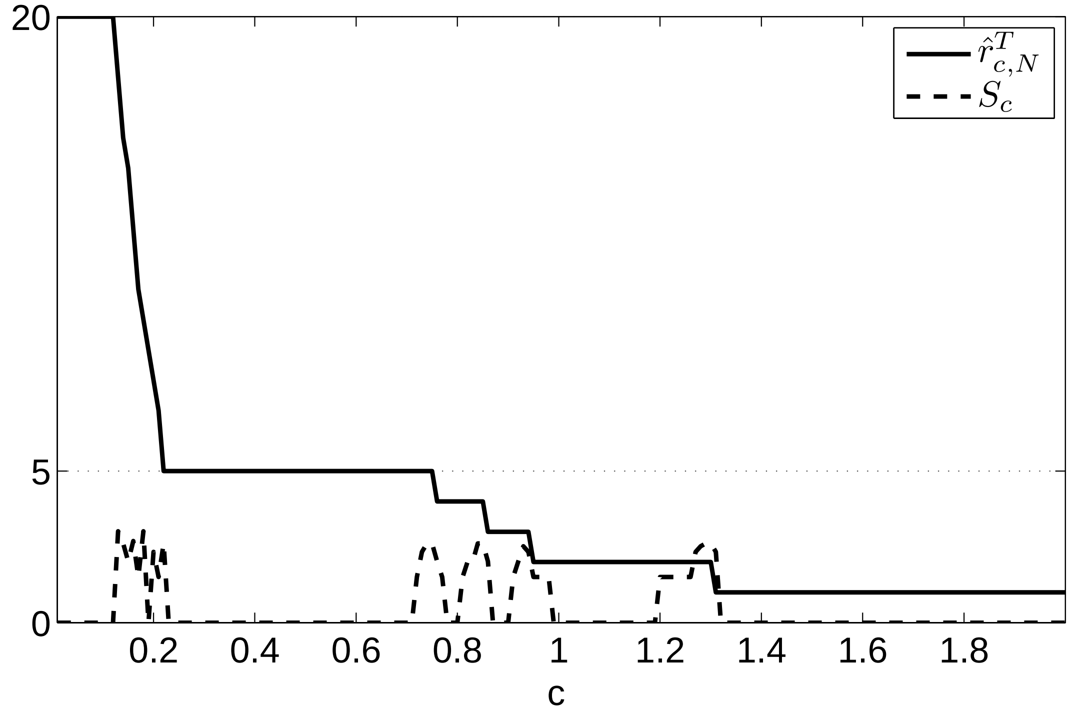
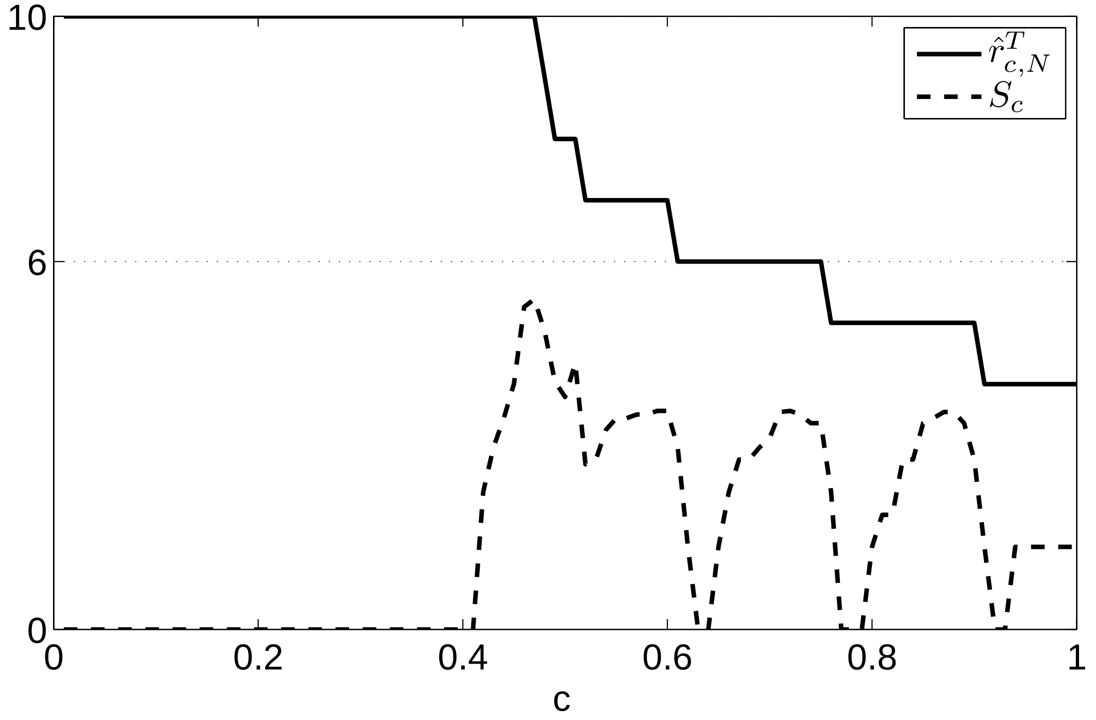
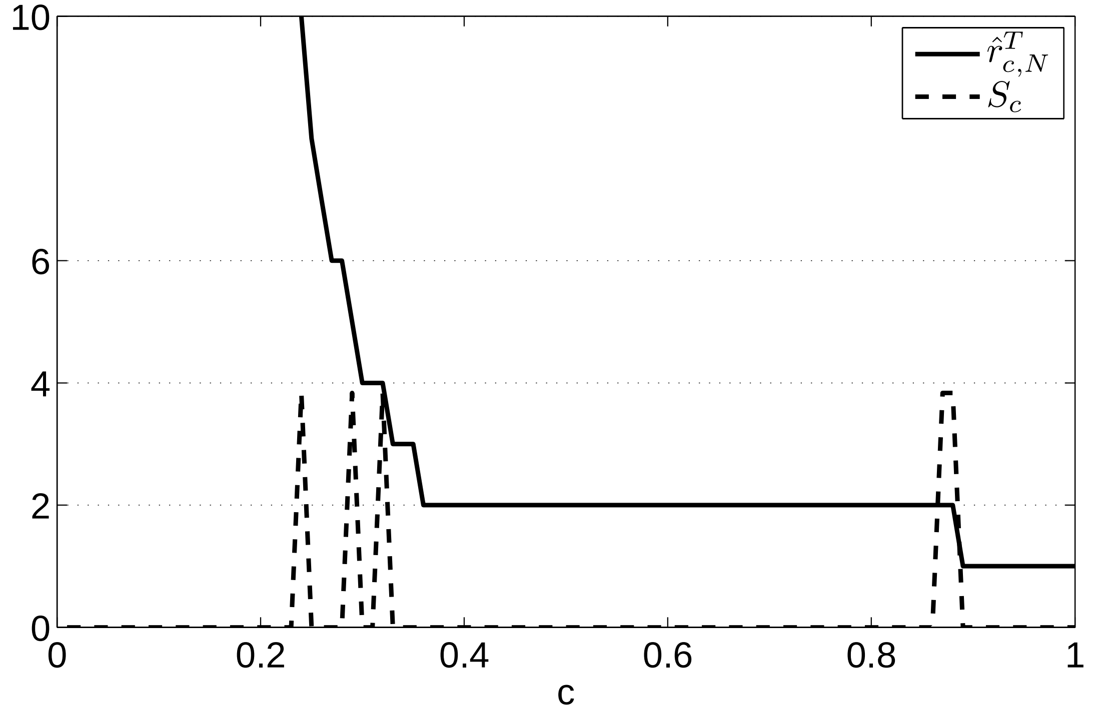


Figure 1: An example of the application of IC_1^* criterion (true number of factors: $r = 5$). $\hat{r}^{T}_{c,N}$: estimated number of factors as a function of c; S_c variance of $\hat{r}^{T}_{c,n_j}$ as $n_j \to N$.



(a) Macroeconomic case



(b) Financial case

Figure 2: IC_1^* criterion in the two empirical cases. $\hat{r}^T_{c,N}$: estimated number of factors as a function of c; S_c variance of $\hat{r}^T_{c,n_j}$ as $n_j \to N$.