

# Determining the Number of Factors from Empirical Distribution of Eigenvalues

Alexei Onatski

Economics Department, Columbia University

December 31, 2008

## **Abstract**

We develop a new estimator of the number of factors in the approximate factor models. The estimator is based on the fact, established in the paper, that any finite number of the largest “idiosyncratic” eigenvalues of the sample covariance matrix cluster around a single point. In contrast, all the “systematic” eigenvalues, the number of which equals the number of factors, diverge to infinity. Our estimator consistently separates the diverging eigenvalues from the cluster and counts the number of the separated eigenvalues. The main advantage of our estimator relative to the Bai-Ng estimators is that it works well when the amount of cross-sectional and temporal correlation in the idiosyncratic terms is relatively large. It also improves upon the Bai-Ng criteria when the portion of the observed variance attributed to the factors is small. We apply the new estimation procedure to determine the number of pervasive factors driving returns of stocks traded on NYSE, AMEX, and NASDAQ and the number of pervasive factors influencing the dynamics of a large set of macroeconomic variables.

# 1 Introduction

Factor models with large cross-sectional and temporal dimensions have become a popular tool in finance and macroeconomics. In finance, such models are at the heart of Chamberlain and Rothschild's (1983) extension of the arbitrage pricing theory. In macroeconomics, they are used in an exploding number of applications (for a recent survey see Breitung and Eickmeier (2005)).

Much of the related econometric research, including Connor and Korajczyk (1993), Bai and Ng (2002), Hallin and Liska (2007) and Watson and Amengual (2007), have been exclusively focused on the important question: how many factors there are? A critical assumption made by this research is that factors' cumulative effect on the  $n$  cross-sectional units grows proportionally to  $n$ . Consequently, if there are  $r$  static factors, then  $r$  eigenvalues of the data's covariance matrix grow proportionally to  $n$  while the rest of the eigenvalues stay bounded. Therefore, the number of factors can be estimated by separating relatively large eigenvalues from the rest using threshold functions such as in Bai and Ng's (2002) information criteria.

This paper estimates the number of factors without making any assumptions on the speed of growth of the factors' cumulative effect. One reason to be interested in factors with slowly increasing cumulative effect is provided by the arbitrage pricing theory. As shown by Chamberlain and Rothschild (1983), if asset returns admit the approximate factor structure, an asset's risk premium is approximately equal to a linear combination of its factor loadings. Excluding relatively weakly influential factors from consideration makes the cumulative approximation error arbitrarily large as the number of assets increases.

Another reason to be interested in factors that have cumulative effects growing more slowly than  $n$ , is the empirical observation that the eigenvalues of the sample covariance matrices of large macroeconomic and financial data sets do not obviously separate into

groups of large and small eigenvalues. While this observation does not invalidate the fast-growth assumption, it questions its empirical content and motivates this paper’s search for an alternative approach.

Instead of requiring that the cumulative effect of factors grows as fast as  $n$  this paper imposes a structure on the idiosyncratic components of the data. Precisely, we assume that the  $n \times T$  matrix  $e$  of the idiosyncratic components of the data is such that

$$e = A\varepsilon B, \tag{1}$$

where  $A$  and  $B$  are two largely unrestricted deterministic matrices, and  $\varepsilon$  is an  $n \times T$  matrix with i.i.d. Gaussian entries, so that both the cross-sectional and temporal correlation of the idiosyncratic terms are allowed. We can lift the Gaussianity assumption but then we require that either  $A$  or  $B$  is a diagonal matrix while the other matrix remains relatively unrestricted. As we explain in the next section, structure (1) is more flexible than most of the alternative structures proposed in the similar literature to date. Exploiting the structure, we show that any finite number of the largest of the bounded eigenvalues of the sample covariance matrix cluster around a single point. Our estimator consistently separates the diverging eigenvalues from the cluster and counts the number of the separated eigenvalues. This number is our estimate of the number of factors.

The main difference between our estimator and the Bai-Ng estimator is that, in our case, the implicit threshold separating the bounded and the diverging eigenvalues is sharp, in the sense that it cannot be decreased without hurting the consistency of the estimator, whereas the threshold in the Bai-Ng criterion can be arbitrarily scaled leaving too much freedom for its choice. The Monte Carlo analysis shows that our estimator outperforms estimators of Bai and Ng (2002) in samples of empirically relevant sizes when there is a non-trivial amount of correlation in the idiosyncratic terms. More specifically, the Bai-Ng

criteria tend to severely overestimate the number of factors relative to our criterion.

The rest of the paper is organized as follows. In Section 2 we review the related literature. In Section 3 we describe the approximate factor model, formulate our assumptions, and develop a consistent estimator of the number of factors. In Section 4 we calibrate a parameter that our estimator depends on. Section 5 contains a Monte Carlo analysis of the estimator. Section 6 uses the new method to estimate the number of factors in the arbitrage pricing theory and in a large macroeconomic panel. Section 7 concludes. Technical proofs are in the Appendix, which can be downloaded from the Review of Economics and Statistics' web page.

## 2 Related literature

This paper's general idea to exploit a structure of idiosyncratic terms to find the number of factors is not new. It can be traced back at least to Wachter (1976) who proposes a simple method to decide on the number of principal components in the principal components analysis of large dimensional data. When the data are i.i.d. and thus have no signals, the empirical distribution of eigenvalues of the sample covariance matrix converges to a known non-random distribution as the dimensionality of the data and the number of observations rise (see Marčenko and Pastur's (1967)). Wachter suggests comparing a plot of the actual empirical distribution of the eigenvalues to the cumulative distribution function of such a limiting law and taking the number of principal components equal to the number of the relatively large eigenvalues responsible for a visible mismatch. Wachter's analysis was considerably extended and generalized in Silverstein and Combettes (1992).

Kapetanios (2004) is the first paper that uses the idea of exploiting a structure of idiosyncratic terms in the context of approximate factor models. Kapetanios assumes that the idiosyncratic terms are homoskedastic and unconditionally uncorrelated. He points

out that this assumption implies a closed form expression for a sharp asymptotic upper bound on the “idiosyncratic” eigenvalues of the sample covariance matrix. Counting the eigenvalues above the bound gives an estimate of the number of factors.

This paper allows for the covariance structure of the idiosyncratic terms which is much less restrictive than Kapetanios’. This flexibility costs us the existence of a closed form expression for the upper bound. The upper bound becomes a complicated implicit function of the cross-sectional and temporal correlation structure of the data. We propose a method of consistently estimating the upper bound from the data.

More recently, Kapetanios (2007) develops another estimator of the number of factors which sequentially tests hypotheses that the number of factors is zero, one, two, etc. against the alternative that the number of factors is larger than zero, one, two, etc., respectively.<sup>1</sup> The estimator equals the number of factors specified by the first non-rejected null. To perform the tests, Kapetanios employs a subsampling method to approximate the asymptotic distribution of a test statistic  $\lambda_i - \lambda_j$ , where  $\lambda_i$  is the  $i$ -th largest eigenvalue of the sample covariance matrix. He makes high-level assumptions about the existence of a scaling of  $\lambda_i - \lambda_j$  which converges in distribution to some unknown limit law, about properties of such a law, and about the functional form of the scaling constant.

Unfortunately, the only known sufficient conditions on the correlation structure of the idiosyncratic terms that guarantee Kapetanios’ high level assumptions exclude either cross-sectional or temporal correlation and require the idiosyncratic terms be Gaussian (see El Karoui, 2005). In this paper we explicitly allow for both cross-sectional and temporal dependence of the Gaussian idiosyncratic terms. Alternatively, for non-Gaussian idiosyncratic terms we explicitly allow for time series correlation and cross-sectional heteroskedasticity (general  $B$  and diagonal  $A$  in (1)) or vice versa (general  $A$  and diagonal  $B$  in (1)).

---

<sup>1</sup>See Onatski (2008) for alternative tests in the context of the generalized dynamic factor models.

### 3 Assumptions and basic results

In this paper, we study approximate factor models of the form

$$X^{(n,T)} = \Lambda^{(n,T)} F^{(n,T)} + e^{(n,T)}, \quad (2)$$

where  $X^{(n,T)}$  is an  $n \times T$  matrix of data on  $n$  cross-sectional units observed over  $T$  time periods,  $\Lambda^{(n,T)}$  is an  $n \times r$  matrix whose  $i, j$ -th element is interpreted as the loading of the  $j$ -th factor on the  $i$ -th cross-sectional unit,  $F^{(n,T)}$  is an  $r \times T$  matrix whose  $j, t$ -th element is interpreted as the value of the  $j$ -th factor at time  $t$ , and  $e^{(n,T)}$  is an  $n \times T$  matrix of the idiosyncratic components of the data. Models (2) corresponding to relatively small  $n$  and  $T$  are not necessarily nested into the larger models. The only assumptions on  $\Lambda^{(n,T)}, F^{(n,T)}$  and  $e^{(n,T)}$  made are Assumptions 1 and 2 below. Our goal is to consistently estimate the unknown number of factors  $r$  from the data as both  $n$  and  $T$  go to infinity.

To simplify notation, we will omit the superscript  $(n, T)$  on matrices  $\Lambda, F$  and  $e$  in what follows.

**Assumption 1.** For any constants  $B > 0$  and  $\delta > 0$ , there exist positive integers  $n_0$  and  $T_0$  such that for any  $n > n_0$  and  $T > T_0$ , the probability that the smallest eigenvalue of  $(\Lambda' \Lambda) (\frac{1}{T} F' F)$  is below  $B$ , is smaller than  $\delta$ .

In an important special case when the columns of  $F$  are i.i.d. vectors with an identity covariance matrix, Assumption 1 is equivalent to requiring that the smallest eigenvalue of  $\Lambda' \Lambda$  diverges to infinity as  $n \rightarrow \infty$ . This is the restriction on factor loadings introduced by Chamberlain and Rothschild (1983). Assumption 1 can be loosely interpreted as saying that the cumulative effect of the “least influential factor” diverges in probability to infinity as the sample size grows.

Our assumption is weaker than that in Connor and Korajczyk (1993), Stock and

Watson (1999), Bai and Ng (2002, 2006) and Bai (2003), who assume that the cumulative effect of the “least influential factor” rises proportionally to  $n$ . Such a proportional growth assumption is an essential component of Bai’s (2003) and Bai and Ng’s (2006) proofs of the asymptotic normality of the principal components factor estimators and factor based forecasts. These asymptotic normality results are likely to break down under our Assumption 1.<sup>2</sup>

Assumption 1 does not impose any separate restrictions on the convergence or divergence of  $\Lambda'\Lambda$  and  $\frac{1}{T}F'F$ . Therefore, much room is left for modeling factors. For example, they may be deterministic, or random and stationary, or random and non-stationary. In particular, the factors may follow  $I(1)$  processes as in Bai and Ng (2004). In addition, we do not require factors to be independent from the idiosyncratic terms, which are characterized by our next assumption.

To formulate the next assumption we introduce new notation. Let  $\lambda_1(A) \geq \dots \geq \lambda_n(A)$  be the eigenvalues of a generic  $n \times n$  symmetric matrix  $A$ . We define the eigenvalue distribution function for  $A$  as

$$\mathcal{F}^A(x) = 1 - \frac{1}{n} \# \{i \leq n : \lambda_i(A) > x\}, \quad (3)$$

where  $\# \{\cdot\}$  denotes the number of elements in the indicated set. Note that  $\mathcal{F}^A(x)$  is a valid cumulative probability distribution function (cdf). Further, for a generic probability distribution having a bounded support and cdf  $\mathcal{F}(x)$ , let  $u(\mathcal{F})$  be the upper bound of the support, that is  $u(\mathcal{F}) = \min \{x : \mathcal{F}(x) = 1\}$ .

**Assumption 2.** For any positive integers  $n$  and  $T$ , there exist an  $n \times n$  deterministic matrix  $A$ , a  $T \times T$  deterministic matrix  $B$  and an  $n \times T$  random matrix  $\varepsilon$  with  $i, t$ -th element  $\varepsilon_{it}$  such that (1) holds, where

---

<sup>2</sup>For asymptotic analysis of the principal components estimator of factor models with i.i.d. Gaussian idiosyncratic terms under an assumption similar to Assumption 1 see Onatski (2006).

- i)  $\varepsilon_{it}$ ,  $1 \leq i \leq n$ ,  $1 \leq t \leq T$  are i.i.d. and satisfy moment conditions  $E\varepsilon_{it} = 0$ ,  $E(\varepsilon_{it})^2 = 1$ ,  $E(\varepsilon_{it})^4 < \infty$ ;
- ii)  $\mathcal{F}^{AA^0}$  and  $\mathcal{F}^{BB^0}$  weakly converge to probability distribution functions  $\mathcal{F}_A$  and  $\mathcal{F}_B$ , respectively, as  $n$  and  $T$  go to infinity;
- iii) Distributions  $\mathcal{F}_A$  and  $\mathcal{F}_B$  have bounded support,  $u(\mathcal{F}^{AA^0}) \rightarrow u(\mathcal{F}_A) > 0$  and  $u(\mathcal{F}^{BB^0}) \rightarrow u(\mathcal{F}_B) > 0$  almost surely as  $n$  and  $T$  go to infinity,  $\liminf_{\delta \rightarrow 0} \frac{1}{\delta} \int_{u(\mathcal{F}_A) - \delta}^{u(\mathcal{F}_A)} d\mathcal{F}_A(\lambda) = k_A > 0$  and  $\liminf_{\delta \rightarrow 0} \frac{1}{\delta} \int_{u(\mathcal{F}_B) - \delta}^{u(\mathcal{F}_B)} d\mathcal{F}_B(\lambda) = k_B > 0$ .

Assumption 2 allows the idiosyncratic terms to be non-trivially correlated both cross-sectionally and over time. The cross-sectional correlation is caused by matrix  $A$  linearly combining different rows of  $\varepsilon$ , whereas the correlation over time is caused by matrix  $B$  linearly combining different columns of  $\varepsilon$ . Such a way of creating covariances between different elements of matrix  $e$  is restrictive. Denoting the  $nT \times 1$  vector of stacked columns of  $e$  as  $\text{vec } e$ , we have:  $E(\text{vec } e (\text{vec } e)') = B'B \otimes AA'$ . How well  $B'B \otimes AA'$  can approximate more interesting covariance structures depends on the details of these structures. For a general discussion of the quality of approximations with Kronecker products see Van Loan and Pitsianis (1993).

Part ii of Assumption 2 is satisfied for rich classes of matrices. In particular,  $B'B$  (as well as  $AA'$ ) may be the auto-covariance matrix of any covariance-stationary process with a bounded spectral density. Part iii of the assumption implies clustering of the largest eigenvalues of the population covariance matrix of  $\text{vec } e$ . In particular, it does not allow a few linear combinations of idiosyncratic terms to have “unusually” large variation, which would make these combinations difficult to separate from combinations of the genuine factors in finite samples.

Assumption 2 allows us to use recent results from the large random matrix literature formulated below as Lemmas 1 and 2. Let us, first, introduce some useful concepts and notation. For any probability distribution function  $G$ , we define its Stieltjes transform



$m_G(z)$  as  $m_G(z) \equiv \int \frac{1}{\lambda-z} dG(\lambda)$ , where  $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \text{Im } z > 0\}$ . The Frobenius-Perron inversion formula

$$G\{[a, b]\} = \frac{1}{\pi} \lim_{\eta \rightarrow 0^+} \int_a^b \text{Im } m_G(\xi + i\eta) d\xi, \quad (4)$$

where  $a$  and  $b$  are points of continuity of  $G$ , insures that we can reconstruct a distribution from its Stieltjes transform. We will denote the Stieltjes transforms of  $\mathcal{F}_A$  and  $\mathcal{F}_B$  as  $m_A(z)$  and  $m_B(z)$ , respectively.

**Lemma 1.** (Zhang, 2006) Let  $\{T(n), n \in \mathbb{N}\}$  be a sequence of positive integers such that  $n/T(n) \rightarrow c > 0$  as  $n \rightarrow \infty$ . Then, if Assumption 2 holds, the eigenvalue distribution of  $e'e/T(n)$  converges to a non-random cdf  $\mathcal{F}^{c,A,B}$  almost surely. For each  $z \in \mathbb{C}^+$ , the Stieltjes transform of  $\mathcal{F}^{c,A,B}$ ,  $m(z)$ , together with two other analytic on  $\mathbb{C}^+$  functions  $u(z)$  and  $v(z)$  constitute a solution to the system:

$$\begin{cases} zm(z) + 1 = u(z) m_A(u(z)) + 1 \\ zm(z) + 1 = c^{-1} [v(z) m_B(v(z)) + 1] \\ zm(z) + 1 = -c^{-1} \frac{z}{u(z)v(z)} \end{cases} \quad (5)$$

which is unique in the set  $\{(m(z), u(z), v(z)) : \text{Im } m(z) > 0, \text{Im } (u(z)) > 0, \text{Im } (v(z)) > 0\}$ .

This is a slightly reformulated version of Theorem 1.2.1 in Zhang (2006).<sup>3</sup> Let  $\mathcal{F}^{c_n, A_n, B_n}$  denote the distribution whose Stieltjes transform is given by (5) replacing  $c, m_A$ , and  $m_B$  with  $n/T(n)$ , the Stieltjes transform of  $\mathcal{F}^{AA^0}$ , and the Stieltjes transform of  $\mathcal{F}^{BB^0}$ , respectively. Paul and Silverstein (2008) prove the following Lemma.

**Lemma 2.** (Paul and Silverstein, 2008) Let  $\{T(n), n \in \mathbb{N}\}$  be a sequence of positive integers such that  $n/T(n) \rightarrow c > 0$  as  $n \rightarrow \infty$ . Further, let  $\varepsilon$  have Gaussian entries

---

<sup>3</sup>Zhang (2006) formulates his theorem in terms of the Stieltjes transform  $s(z)$  of  $e'e'/T(n)$  (as opposed to that of  $e'e/T(n)$ ). Further, instead of our  $u(z)$  and  $v(z)$ , he uses  $p(z) = -\frac{1}{u(z)}$  and  $q(z) = -\frac{1}{v(z)}$ . Finally, his theorem is formulated under weaker assumptions than those of Lemma 1.

(alternatively, let either  $A$  or  $B$  in (1) be a diagonal matrix). Then, if Assumption 2 holds,  $\Pr(\text{no eigenvalue of } ee'/T(n) \text{ appear in } [a, b] \text{ for all large } n) = 1$ , where  $[a, b]$  is any interval that lies in an open interval outside the support of  $\mathcal{F}^{c_n, A_n, B_n}$ .

Lemma 2 corresponds to Theorem 1 in Paul and Silverstein (2008). That theorem does not require all parts of Assumption 2 to hold. Precisely, Assumption 2, iii can be substituted by a weaker assumption that functions  $\mathcal{F}_A$  and  $\mathcal{F}_B$  have bounded support and that  $u(\mathcal{F}^{AA^0})$  and  $u(\mathcal{F}^{BB^0})$  are bounded in  $n$ . We use the stronger assumption to prove the next Lemma.

**Lemma 3.** Under conditions of Lemma 2,  $\Pr(\text{no eigenvalue of } ee'/T(n) \text{ appear in } [a, b] \text{ for all large } n) = 1$ , where  $[a, b]$  is any interval that lies in an open interval outside the support of  $\mathcal{F}^{c, A, B}$ .

The proof of Lemma 3 is given in the Appendix, which can be downloaded from the Review of Economics and Statistics' web page. It consists of showing that under Assumption 2, the upper boundary of  $\mathcal{F}^{c_n, A_n, B_n}$  converges to the upper boundary of  $\mathcal{F}^{c, A, B}$ . To show such a convergence, we characterize the upper boundary of support of a distribution whose Stieltjes transform solves system (5). Such a characterization is non-trivial and constitutes a result of independent interest. For similar work in different settings see Marčenko and Pastur (1967), Silverstein and Choi (1995), and Dozier and Silverstein (2004).

For the purpose of this paper, Lemma 3 is important because it implies that the eigenvalues of the sample covariance matrix  $XX'/T$  can be separated into  $r$  “systematic eigenvalues” that diverge to infinity and a group of “idiosyncratic eigenvalues”, any finite number of the largest of which cluster around  $u(\mathcal{F}^{c, A, B})$ . Indeed, we have the following Theorem.

**Theorem 1.** Let model (2) satisfy Assumptions 1 and 2, and let  $\{T(n), n \in \mathbb{N}\}$  be a

sequence of positive integers such that  $n/T(n) \rightarrow c > 0$  as  $n \rightarrow \infty$ . Further, let  $\varepsilon$  have Gaussian entries (alternatively, let either  $A$  or  $B$  in (1) be a diagonal matrix). Then, as  $n \rightarrow \infty$ , we have:

- i) for any sequence of positive integers  $\{i(n), n \in \mathbb{N}\}$  such that  $i(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ , and  $i(n) > r$  for large enough  $n$ , the  $i(n)$ -th eigenvalue of  $XX'/T(n)$  converges almost surely to  $u(\mathcal{F}^{c,A,B})$ ;
- ii) the  $r$ -th eigenvalue of  $XX'/T(n)$  tends to infinity in probability.

Given Lemma 3, proof of Theorem 1 is simple. We therefore provide it here instead of relegating it to the Appendix. According to a singular value analog of Weyl's eigenvalue inequalities (see Theorem 3.3.16 of Horn and Johnson (1991)), for any  $n \times T$  matrices  $P$  and  $Q$ , we have:

$$\sigma_{i+j-1}(P+Q) \leq \sigma_i(P) + \sigma_j(Q), \quad (6)$$

where  $1 \leq i, j \leq \min(n, T)$  and  $\sigma_i(P)$  denotes the  $i$ -th largest singular value of matrix  $P$ , which is another name for the square root of the  $i$ -th largest eigenvalue of matrix  $PP'$ . Let  $\lambda_i$ ,  $\mu_i$ , and  $\nu_i$  be the  $i$ -th largest eigenvalues of  $XX'/T(n)$ ,  $ee'/T(n)$ , and  $(\Lambda'\Lambda)(F'F/T(n))$ , respectively. Note that since the rank of the latter matrix is  $r$ ,  $\nu_{r+1}$  equals zero. Inequality (6) implies

$$\sqrt{\lambda_r} \geq \sqrt{\nu_r} - \sqrt{\mu_1} \quad (7)$$

$$\lambda_i \leq \mu_{i-r}, \text{ for } i = r+1, \dots, n \quad (8)$$

$$\lambda_i \geq \mu_{i+r}, \text{ for } i = 1, \dots, n-r \quad (9)$$

where the first inequality follows by taking  $P = X/\sqrt{T(n)}$  and  $Q = -e/\sqrt{T}$ , the second inequality follows by taking  $P = e/\sqrt{T}$  and  $Q = \Lambda F'/\sqrt{T}$ , and the last inequality follows by taking  $P = X/\sqrt{T}$  and  $Q = -\Lambda F'/\sqrt{T}$ .

Assumption 1, Lemma 3, and inequality (7) imply that the first  $r$  eigenvalues of the sample covariance matrix diverge to infinity. On the other hand, according to inequalities (8) and (9), the  $i(n)$ -th eigenvalue of the sample covariance matrix is sandwiched between  $\mu_{i(n)+r}$  and  $\mu_{i(n)-r}$ . The latter two converge to  $u(\mathcal{F}^{c,A,B})$  by Lemma 3, by the definition of  $u(\mathcal{F}^{c,A,B})$ , and by the assumption that  $i(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ . Hence  $\lambda_{i(n)}$  also converges to  $u(\mathcal{F}^{c,A,B})$ .  $\square$

Theorem 1 suggests a method of estimation of  $r$ . Indeed, it implies that, for any  $j > r$ , the difference  $\lambda_j - \lambda_{j+1}$  converges to zero, while the difference  $\lambda_r - \lambda_{r+1}$  diverges to infinity. Let  $\{r_{\max}^n, n \in \mathbb{N}\}$  be a slowly increasing sequence of real numbers such that  $r_{\max}^n/n \rightarrow 0$  as  $n \rightarrow \infty$ .  $r_{\max}^n$  will be interpreted as the maximum possible number of factors that a researcher having a sample of size  $n, T(n)$  assumes a priori. We define a family of estimators

$$\hat{r}(\delta) = \max \{i \leq r_{\max}^n : \lambda_i - \lambda_{i+1} \geq \delta\}, \quad (10)$$

parameterized by a positive number  $\delta$ . If the set  $\{i \leq r_{\max}^n : \lambda_i - \lambda_{i+1} \geq \delta\}$  is empty, then  $\hat{r}(\delta) = 0$ . The consistency of  $\hat{r}(\delta)$  for any fixed  $\delta > 0$  and any slowly increasing sequence  $\{r_{\max}^n, n \in \mathbb{N}\}$  follows from Theorem 1 and from the fact that  $r_{\max}^n$  becomes larger than a fixed number  $r$  for large enough  $n$ .

**Corollary 1.** Under assumptions of Theorem 1, for any fixed  $\delta > 0$  and any slowly increasing sequence  $\{r_{\max}^n, n \in \mathbb{N}\}$  such that  $r_{\max}^n/n \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\hat{r}(\delta) \rightarrow r$  in probability as  $n \rightarrow \infty$ .

Although consistent, estimator (10) may perform poorly in finite samples if  $\delta$  is poorly calibrated. In the next section we describe a calibration procedure for  $\delta$ .

## 4 Calibration of $\delta$

Lemma 1 and inequalities (7), (8) and (9) imply that the empirical distribution of  $\lambda_1, \dots, \lambda_n$  converges to  $\mathcal{F}^{c,A,B}$ . Hence, a natural approximation of  $\mathcal{F}^{c,A,B}(\lambda_i)$  is  $\frac{1}{n} \# \{j : \lambda_j \leq \lambda_i\} \equiv \frac{n-i+1}{n}$ , and we may calibrate the difference  $\lambda_i - \lambda_{i+1}$  for  $i > r$  by  $[\mathcal{F}^{c,A,B}]^{-1}(\frac{n-i+1}{n}) - [\mathcal{F}^{c,A,B}]^{-1}(\frac{n-i}{n})$ . It is therefore reasonable to make  $\delta$  in (10) no smaller than  $[\mathcal{F}^{c,A,B}]^{-1}(y + \frac{1}{n}) - [\mathcal{F}^{c,A,B}]^{-1}(y)$ , where  $y$  is close to one.

Unfortunately, as is evident from (5),  $\mathcal{F}^{c,A,B}$  is a very complicated function of  $A$  and  $B$ , and it does not have an explicit form. However, locally to  $u(\mathcal{F}^{c,A,B})$ , the density corresponding to  $\mathcal{F}^{c,A,B}(x)$  is proportional to the square root function  $(u(\mathcal{F}^{c,A,B}) - x)^{1/2}$  for a very general class of matrices  $A$  and  $B$ . At a heuristic level, the idea is as follows. Clearly, the Stieltjes transform of  $\mathcal{F}^{c,A,B}$ ,  $m(z) \equiv \int \frac{d\mathcal{F}^{c,A,B}(\lambda)}{\lambda - z}$ , is a well defined analytic real function on  $z \in (u, +\infty)$ , where  $u \equiv u(\mathcal{F}^{c,A,B})$ . It must, however, have a singularity at  $u$  because otherwise, although not defined on  $z < u$  by integral  $\int \frac{d\mathcal{F}^{c,A,B}(\lambda)}{\lambda - z}$ ,  $m(z)$  can be analytically extended from  $z > u$  to those  $z < u$  which are close enough to  $u$ . The extension will be real for  $z < u$ . On the other hand, according to the inversion formula (4),  $m(z)$  must have a positive imaginary part for  $z \in (-\infty, u)$  arbitrarily close to  $u$ . Since  $m(z)$  has a singularity at  $u$ , the analytic extension of its inverse,  $z(\cdot)$ , should not be invertible at  $m(u)$ , which happens if its second order approximation lacks the first order term:  $z(m) \approx u + k(m - m(u))^2$ . Rearranging, we get  $m(z) \approx m(u) + \frac{1}{k}(z - u)^{1/2}$ . This approximation together with the inversion formula (4) explains the square root phenomenon.<sup>4</sup>

Note that the square root behavior of the density implies that the cdf  $\mathcal{F}^{c,A,B}(x)$  can be approximated by  $1 - a((u - x)^+)^{3/2}$  for some positive  $a$  in the neighborhood of  $u$ , where  $(\cdot)^+$  takes the positive part of the argument. Therefore,  $[\mathcal{F}^{c,A,B}]^{-1}(y)$  can be approximated by  $u - (\frac{1-y}{a})^{2/3}$  for  $y$  close to but smaller than one, and a reasonable

---

<sup>4</sup>Silverstein and Choi (1995) make the above heuristic argument precise when either  $A$  or  $B$  is the identity matrix. We leave a rigorous prove for general  $A$  and  $B$  for future research.

upper bound for  $[\mathcal{F}^{c,A,B}]^{-1}(y + \frac{1}{n}) - [\mathcal{F}^{c,A,B}]^{-1}(y)$  for such  $y$  is given by  $\left(\frac{1-y-1/n}{a}\right)^{2/3} + \left(\frac{1-y}{a}\right)^{2/3} \Big|_{y=1} \equiv \left(\frac{1}{an}\right)^{2/3} \equiv \beta$ . In our Monte Carlo experiments and applications below, we will make  $\delta$  equal  $2 \cdot |\hat{\beta}|$ , where  $\hat{\beta}$  is an estimate of  $\left(\frac{1}{an}\right)^{2/3}$ .

An estimate  $\hat{\beta}$  can be obtained as follows. Define  $y_j \equiv \lambda_j$  and  $x_j \equiv (j-1)^{2/3}$ . Then, an approximate linear relationship  $y_j \approx u - \left(\frac{1}{an}\right)^{2/3} x_j$  holds for small enough  $j > r$ . We obtain  $\hat{\beta}$  by running an ordinary least squares regression of  $y_j, y_{j+1}, \dots, y_{j+4}$  on the constant and  $x_j, x_{j+1}, \dots, x_{j+4}$ , where  $j$  is any integer larger than  $r$ .<sup>5</sup> One possible choice for  $j$  is  $r_{\max} + 1$ . A better choice can be made using an iterative procedure: starting from  $j = r_{\max} + 1$ , calibrate  $\delta$  and estimate  $r$  by  $\hat{r}(\delta)$ ; next, take  $j = \hat{r}(\delta) + 1$ , calibrate  $\delta$  again and estimate  $r$  by  $\hat{r}(\delta)$  using the newly calibrated  $\delta$ . This iterative procedure can be repeated until convergence. In our Monte Carlo experiments the convergence was often achieved at the second iteration.

Let us conclude this section by providing an algorithm for estimating the number of factors based on (10).

1. Compute eigenvalues  $\lambda_1, \dots, \lambda_n$  of the sample covariance matrix  $XX'/T$ . Set  $j = r_{\max} + 1$ .
2. Compute  $\hat{\beta}$ , the slope coefficient in the OLS regression of  $\lambda_j, \dots, \lambda_{j+4}$  on the constant and  $(j-1)^{2/3}, \dots, (j+3)^{2/3}$ . Set  $\delta = 2|\hat{\beta}|$ .
3. Compute  $\hat{r}(\delta) = \max\{i \leq r_{\max}^n : \lambda_i - \lambda_{i+1} \geq \delta\}$ , or if  $\lambda_i - \lambda_{i+1} < \delta$  for all  $i \leq r_{\max}^n$ , set  $\hat{r}(\delta) = 0$ .
4. Set  $j = \hat{r}(\delta) + 1$ . Repeat steps 2 and 3 one more time.

In what follows, we will denote the estimator described by the above algorithm as ED (edge distribution) because it exploits the square-root shape of the edge of the eigenvalue

---

<sup>5</sup>We base the OLS regression on five  $y$ 's and  $x$ 's because this number corresponded to good performance of our estimator in Monte Carlo experiments.

distribution. In the next section, we use Monte Carlo analysis to study the performance of ED and its sensitivity to different choices of  $r_{\max}^n$ .

## 5 Monte Carlo Analysis

In the Monte Carlo experiments below we compare our estimator ED with estimators  $PC_{p1}$ ,  $IC_{p1}$ , proposed by Bai and Ng (2002), Bayesian Information Criterion estimator  $BIC_3$  studied in the same paper, and estimators ME and MED, proposed by Kapetanios (2004) and Kapetanios (2007), respectively.

Estimators  $PC_{p1}$ ,  $IC_{p1}$ , and  $BIC_3$  are based on the minimization of the following criterion functions:

$$PC_{p1}(k) = V(k) + k\hat{\sigma}^2 \left( \frac{n+T}{nT} \right) \ln \left( \frac{nT}{n+T} \right), \quad (11)$$

$$IC_{p1}(k) = \ln(V(k)) + k \left( \frac{n+T}{nT} \right) \ln \left( \frac{nT}{n+T} \right), \quad (12)$$

$$BIC_3(k) = V(k) + k\hat{\sigma}^2 \frac{(n+T-k) \ln(nT)}{nT}, \quad (13)$$

where  $V(k)$  is  $\frac{1}{nT}$  times the minimal possible sum of squared errors of a rank  $k$  approximation of the data matrix, and  $\hat{\sigma}^2$  is an estimator of the unconditional variance of the idiosyncratic terms.<sup>6</sup> Note that  $V(k)$  and  $\hat{\sigma}^2$  in (11),(12), and (13) are equal to  $\frac{1}{nT} \sum_{j=k+1}^n \lambda_j$  and  $\frac{1}{nT} \sum_{j=r_{\max}^n+1}^n \lambda_j$  respectively, where  $\lambda_i$  is the  $i$ -th largest eigenvalue of the sample covariance matrix. Therefore, similar to ED, estimators  $PC_{p1}$ ,  $IC_{p1}$  and  $BIC_3$  are based entirely on the empirical distribution of eigenvalues  $\lambda_i$ .

Kapetanios' (2004) ME (maximum eigenvalue) estimator compares the largest eigen-

---

<sup>6</sup>Bai and Ng (2002) propose four more estimators that differ from  $PC_{p1}$  and  $IC_{p1}$  by the penalty term in the corresponding criterion functions. Performance of these other estimators in the experiments reported by Bai and Ng are similar to that of  $PC_{p1}$  and  $IC_{p1}$ , so we limit our attention to estimators  $PC_{p1}$  and  $IC_{p1}$ .

value of the sample covariance matrix of the standardized data with the theoretical large sample limit  $(1 + \sqrt{c})^2$  of the largest eigenvalue of the sample covariance matrix of i.i.d. data. If the largest eigenvalue is smaller than  $(1 + \sqrt{c})^2 + 1$ , ME equals zero. Otherwise, the first principal component is subtracted from the data, and the largest eigenvalue of the standardized residuals is compared with  $(1 + \sqrt{c})^2$ . If the largest eigenvalue is smaller than  $(1 + \sqrt{c})^2 + 1$ , ME equals one. Otherwise, the two principal components are subtracted from the data etc., until the largest eigenvalue is smaller than  $(1 + \sqrt{c})^2 + 1$  or until  $r_{\max}$  is reached. Kapetanios' (2007) MED (maximum eigenvalues distribution) estimator sequentially tests hypotheses that the number of factors is zero, one, two, etc. against the alternative that the number of factors is larger than zero, one, two, etc., respectively. MED<sup>7</sup> equals the number of factors specified by the first non-rejected null, or to  $r_{\max}$  if all the nulls are rejected.

Our first experiment consists of generating 1000 replications of data produced by the data-generating process used in Bai and Ng (2002):

$$\begin{aligned} X_{it} &= \sum_{k=1}^r \Lambda_{ik} F_{kt} + \sqrt{\theta} e_{it}; \\ e_{it} &= \rho e_{i,t-1} + v_{it} + \sum_{1 \leq |j| \leq 8} \beta v_{i-j,t}; \quad F_{kt}, \Lambda_{ik}, v_{it} \sim \text{i.i.d. } N(0, 1). \end{aligned} \tag{14}$$

We consider three pairs of parameters  $\rho$  and  $\beta$  :  $(\rho, \beta) = (0, 0)$ ,  $(0.3, 0.1)$ , and  $(0.85, 0)$ . The latter two pairs are tailored<sup>8</sup> to fit our financial and macroeconomic applications described in the next section, respectively. The true number of factors equals  $r = 1, 3, 5$ , and 15. For  $r = 1, 3, 5$ , we set  $r_{\max} = 8$ . For  $r = 15$ , we set  $r_{\max} = 20$ . Further, we set

---

<sup>7</sup>We compute MED based on the subsampling size  $b(n) = 0.6n$  and 5% significance level tests.

<sup>8</sup>First, we removed 10 principal components from our financial data to obtain an estimate  $e_F$  of the matrix of its idiosyncratic components. Then, we chose  $\rho$  and  $\beta$  so as to minimize  $\sum_{i=1}^n (\lambda_i(ee') - \lambda_i(e_F e_F'))^2$ , where  $e$  was generated according to (14) and normalized so that  $\text{tr}(ee') = \text{tr}(e_F e_F')$ . The minimizing  $\rho$  and  $\beta$  were 0.3 and 0.1, respectively. For our macroeconomic data, a similar calibration procedure gave us  $\rho = 0.85$  and  $\beta = 0$ .



$\theta = r(1 - \rho^2) / (1 + 16\beta^2)$ . Such a choice implies that the factors explain exactly 50% variation in the data. We consider sample sizes  $(n, T) \in \{(40, 100), (200, 60), (1000, 60), (100, 100), (150, 150), (150, 500), (1000, 250)\}$ .

Tables 1, 2 and 3 report the percentage of the replications that result in overestimation and underestimation of the number of factors in the cases  $(\rho, \beta) = (0, 0)$ ,  $(\rho, \beta) = (0.3, 0.1)$  and  $(\rho, \beta) = (0.85, 0)$ , respectively. The tables' entries have form  $x/y$  which means that  $x\%$  of the replications result in the overestimation,  $y\%$  of the replications result in the underestimation, and  $(100 - x - y)\%$  of the replications result in the correct estimation of the number of factors.

The tables reveal that there are at least three sources of problems for the considered estimators: correlation in the idiosyncratic terms, a relatively large number of factors, and a relatively small sample size. Correlation in the idiosyncratic terms ( $\rho \neq 0$  or/and  $\beta \neq 0$ ) leads to the situation when the estimators very often overestimate the true number of factors. The exception from this rule is our estimator and, for the case  $\rho = 0.3, \beta = 0.1$ ,  $\text{BIC}_3$ . Our estimator remains very good for the cases  $\rho = 0.3, \beta = 0.1, r < 15$ , and  $\rho = 0.85, \beta = 0, r = 1$ . For the other cases, it very often underestimates. Overall, our estimator correctly estimates the number of factors far more often than any other of the studied estimators.

A relatively large number of factors ( $r = 15$ ) leads to the situation when all the estimators are almost always wrong about the true number of factors. Depending on the values of  $\rho$  and  $\beta$  and on the sample size, the estimators may either overestimate or underestimate the true number of factors. When our estimator works poorly, it underestimates.

Note that depending on the application, underestimation of the number of factors may be more problematic than overestimation. For example, the principal components based estimators of the factor space remain consistent only when the assumed number of factors is greater than or equal to the true number. However, note also that when  $\lambda_r$  is

only marginally larger than  $\lambda_{r+1}$ , as is typically the case when  $r$  is relatively large, the principal components based estimators estimate the factor space very poorly even if the true number of factors is known (see Onatski (2006)).

Returning to Tables 1, 2 and 3, we see that relatively small sample sizes in general worsen the quality of the estimators. For some of the estimators, even the largest of the considered sample sizes ( $n = 1000, T = 250$ ) is not large enough to guarantee good performance in the face of the non-trivial correlation in the idiosyncratic terms or/and the large number of factors. Our estimator benefits the most from the relatively large sample sizes. For the two largest sample sizes ( $n = 1000, T = 250$  and  $n = 150, T = 500$ ), it practically always correctly estimates the true number of factors for 10 out of 12 considered combinations of  $(\rho, \beta)$  and  $r$ . The second best is  $\text{BIC}_3$  with 6 out of 12 combinations with correctly estimated number of factors.

In our second Monte Carlo experiment, we explore the sensitivity of the estimators to the amount of the noise in the data. Figure 1 shows performance of  $\text{PC}_{p1}, \text{IC}_{p1}, \text{BIC}_3, \text{ME}, \text{MED}$  and  $\text{ED}$  when  $\rho = \beta = 0, r = 3$ , and  $\theta/r$ , interpreted as the relative variance of the idiosyncratic terms, varies from 1 to 25. We see that the Bai-Ng estimators start to underestimate the true number of factors as  $\theta/r$  increases. Note that our estimator remains correct for a considerably larger range of  $\theta/r$ . Eventually, however, it is also bound to underestimate the true number of factors because the systematic information becomes significantly dissolved in the noise which makes correct identification of the true number of factors virtually impossible. Kapetanios' ME estimator outperforms the Bai-Ng estimators but works worse than our estimator. Kapetanios' MED estimator differs from the other estimators in that it overestimate the true number of factors for most of the considered range of  $\theta/r$ . In addition, its mean is not monotone with respect to  $\theta/r$ .

In our next experiment, we explore in more detail the relationship between the quality of the estimators and the amount of the dependence in the idiosyncratic terms. We set

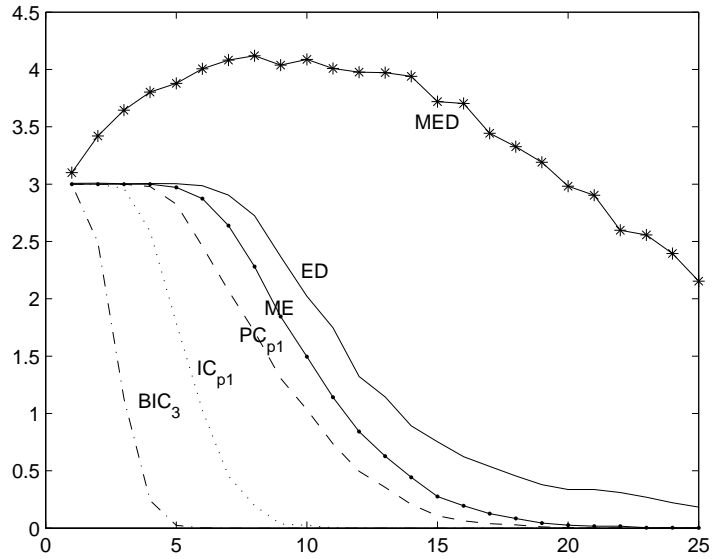


Figure 1: Average estimated number of factors according to  $BIC_3$ ,  $PC_{p1}$ ,  $IC_{p1}$ , ME, MED, and ED. The true number of factors  $r = 3$ . Horizontal axis:  $\theta/r$ .

the true number of factors at 3 and vary  $\rho$  on a grid  $0:0.1:0.9$  and  $\beta$  on a grid  $0:0.05:0.3$ . Figure 2 reports the average estimates of the number of factors produced by  $PC_{p1}$ ,  $IC_{p1}$ ,  $BIC_3$ , ME, MED, and our estimator ED for  $\rho$  and  $\beta$  on the grid across 1000 Monte Carlo replications, for sample size  $n = T = 100$ .<sup>9</sup> Table 4 reports the numerical height of the surfaces shown in Figure 2.

We see that the mean of our estimator remains very close to the true number of factors 3 for the widest area in the  $(\rho, \beta)$ -plane. Indeed, the percentage of the grid nodes where the mean remains in the  $\pm 10\%$ -of-the-true-number-of-factors segment equals 13% for  $PC_{p1}$ , 19% for  $IC_{p1}$ , 33% for  $BIC_3$ , 6% for ME, 4% for MED, and 51% for our estimator. As  $\rho$  and  $\beta$  rise, our estimator starts to underestimate the true number of factors. The mean of our estimator is below 50% of the true number of factors for 20% of the grid nodes.

<sup>9</sup>We also considered combinations  $n = 200, T = 60$ ,  $n = 40, T = 100$ , and  $n = T = 150$ . The results for these sample sizes were similar to those reported.

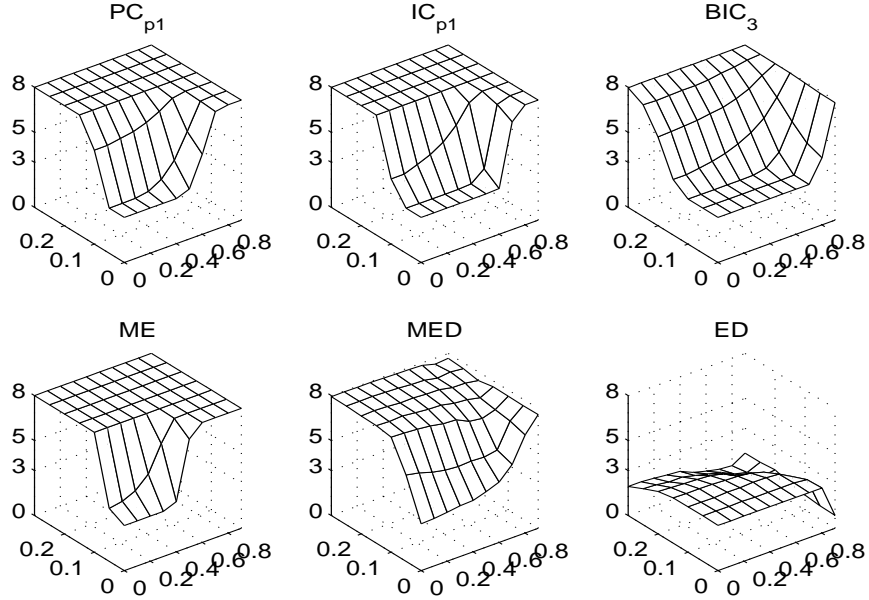


Figure 2: Monte Carlo average of  $PC_{p1}$ ,  $IC_{p1}$ ,  $BIC_3$ ,  $ME$ ,  $MED$ , and  $ED$ . Right axis:  $\rho$ . Left axis:  $\beta$ . The true number of factors is 3.

The worst underestimation happens at the node  $\rho = 0.8, \beta = 0.3$  where the mean of our estimator is 0.8. In contrast, as  $\rho$  and  $\beta$  rise,  $PC_{p1}$ ,  $IC_{p1}$ ,  $BIC_3$ ,  $ME$ , and  $MED$  start to substantially overestimate the true number of factors quickly reaching the maximum  $r_{\max} = 8$ .

To explore the sensitivity of different estimators to the choice of  $r_{\max}$ , we computed the corresponding average estimates in 1000 MC replications on the grid  $r_{\max} = 6 : 1 : 20$ . Figure 3 reports the  $r_{\max}$ -sensitivity results for  $\rho = 0.3, \beta = 0.1$  and for the sample size  $n = T = 80, 120, 160$  and 200.

Our estimator (solid line) is the least sensitive to the choice of  $r_{\max}$ . Estimator  $BIC_3$  (dash-dotted line) behaves very similar to ours for  $n = T \geq 120$ , but it is more sensitive to  $r_{\max}$  for  $n = T = 80$ . Similar to our estimator,  $IC_{p1}$  (dotted line) is not sensitive to the choice of  $r_{\max}$ . However, it substantially overestimates the number of factors for the two smallest sample sizes studied. Estimators  $PC_{p1}, ME$ , and  $MED$  (dashed line, solid

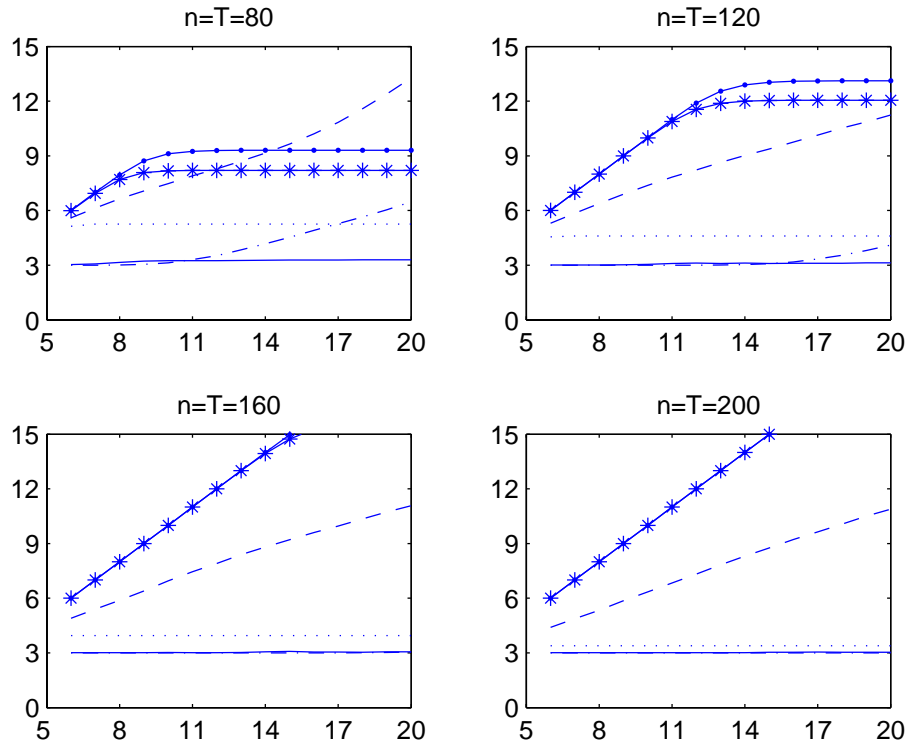


Figure 3: Average estimated number of factors according to BIC<sub>3</sub> (dash-dotted line), PC<sub>p1</sub> (dashed line), IC<sub>p1</sub> (dotted line), ME (solid line with dot markers), MED (solid line with star markers), and ED (solid line). The true number of factors  $r = 3$ . Horizontal axis:  $r_{\max}$ .

line with dot markers, and solid line with star markers, respectively) are very sensitive to  $r_{\max}$ .

The setting of our last Monte Carlo experiment is the same as that of experiment B in Kapetanios (2007). The data-generating process is:

$$\begin{aligned} X_{it} &= \sum_{k=1}^{r_0} \Lambda_{1,ik} F_{kt} + \sum_{k=1}^{r_0} \Lambda_{2,ik} F_{kt-1} + e_{it}; \quad F_{kt}, \Lambda_{1,ik}, \Lambda_{2,ik} \sim \text{i.i.d. } N(0, 1); \\ e_t &= \Sigma^{1/2} v_t; \quad v_t = 0.5v_{t-1} + \xi_t; \quad \xi_t \sim \text{i.i.d. } N(0, 2\theta I_n), \end{aligned} \quad (15)$$

where  $e_t$  is the  $n \times 1$  vector with components  $e_{it}$ ;  $v_t$  is an  $n$ -dimensional AR(1) process with innovations  $\xi_t$ , which are independent from  $F_{ks}, \Lambda_{1,ik}, \Lambda_{2,ik}$ ; and matrix  $\Sigma = [\sigma_{ij}]$  is such that  $\sigma_{ii} = 1, \sigma_{ij} = \sigma_{ji} \sim U(-0.1, 0.1)$  for  $0 < |i - j| \leq 5$ , and  $\sigma_{ij} = \sigma_{ji} = 0$  for  $|i - j| > 5$ . We consider  $r_0 = 0$  and  $r_0 = 2$  so that the true number of static factors equals 0 and 4, respectively. Finally, we consider all combinations of sample sizes  $n = 50, 100, 200; T = 50, 100, 200$ , and  $\theta = 1, 3, 5, 10$ . Note that the ratio of  $\sum_{i,t} E \left( \sum_{k=1}^{r_0} \Lambda_{1,ik} F_{kt} + \sum_{k=1}^{r_0} \Lambda_{2,ik} F_{kt-1} \right)^2$  to  $\sum_{i,t} EX_{it}^2$  equals  $\frac{r_0}{r_0 + \theta/(1-\rho^2)}$ . Hence, for  $r_0 = 2$ , the systematic components explain 60%, 33%, 23%, and 13% of the variation in the data when  $\theta = 1, 3, 5$ , and 10, respectively.

Tables 5 and 6 report, respectively, the average estimated number of factors and the percent of the overestimation and underestimation in 1000 Monte Carlo replications of the data-generating process (15). As can be seen from the tables, estimators  $PC_{p1}$  and ME work considerably worse than the other estimators. They poorly estimate the number of factors even in the easiest of the cases when  $\theta = 1$ . Estimator  $BIC_3$  works similar to  $IC_{p1}$  and ED when  $r = 0$ , and when  $\theta = 1$  and  $r = 4$ . However, it more substantially underestimates the number of factors that  $IC_{p1}$  and ED when  $\theta > 1$  and  $r = 4$ . Our estimator performs similarly to  $IC_{p1}$ . It is more difficult to compare the performances of MED and of  $IC_{p1}$  and ED. Estimator MED almost always overestimates, and it very substantially overestimates when  $r = 0$ . On the other hand,  $PC_{p1}$  and ED underestimate

(except for the largest sample sizes) for relatively large  $\theta$  when  $r = 4$ .<sup>10</sup>

## 6 Applications

We apply the new estimation procedure to determine the number of pervasive factors driving stock returns and the number of pervasive factors influencing the dynamics of a large set of macroeconomic variables. First, we estimate the number of factors in the approximate factor model of the stock returns. Our data consist of monthly returns on 1148 stocks ( $n = 1148$ ) traded on the NYSE, AMEX, and NASDAQ during the period from January 1983 to December 2003 ( $T = 252$ ), obtained from CRSP data base. We included a stock in the data set if it was traded during the whole sample period.

We computed our estimator ED on the grid  $r_{\max} = 6 : 1 : 20$ . For all  $r_{\max}$  on the grid, ED equals two. For comparison, Bai and Ng (2002) detect two pervasive factors in the stock returns data. Connor and Korajczyk (1993) find evidence for between one and six pervasive factors in the stock returns. Trzcinka (1986) finds some support for the existence of five pervasive factors. Five seems also to be a preferred number for Roll and Ross (1980) and Reinganum (1981). A study by Brown and Weinstein (1983) also suggests that the number of factors is unlikely to be greater than five.<sup>11</sup> Huang and Jo (1995) identify only two common factors.

To check the robustness of our estimator to the choice of the sample, and as an empirical approach to decide on the number of factors, we repeat our analysis over different subgroups of the variables. Precisely, we estimate the number of factors in the returns of randomly selected 50% of the original stocks. We repeat the random selection procedure

---

<sup>10</sup>We also explored the quality of the estimators when the noise is much more persistent than in (15). Precisely,  $v_t = 0.95v_{t-1} + \xi_t$  instead of  $v_t = 0.5v_{t-1} + \xi_t$ . Such a setting corresponds to experiment C in Kapetanios (2007). We have found that all the estimators work poorly in this case.

<sup>11</sup>Dhrymes, Friend, and Gultekin (1984) find that the estimated number of factors grows with the sample size. However, their setting was the classical factor model as opposed to the approximate factor model.

1000 times. For  $r_{\max} = 20$ , 460 selected samples give the estimate of one factor, 447 selected samples give the estimate of two factors, and the other 93 selected samples give the estimate of three or more factors. For  $r_{\max} < 20$ , the results are similar. It is, therefore, likely that there are only one or two common factors in the stock returns.

Our second application is about determining the number of pervasive factors influencing the dynamics of a large set of macroeconomic variables. The pervasive factors can be viewed as corresponding to the basic macroeconomic shocks driving the economy. The existence of such shocks is in the spirit of modern dynamic stochastic general equilibrium macroeconomic models.

The dataset we use is the same as in Watson (2003). It includes 215 monthly time series for the United States from 1959:1 to 1998:12 ( $T = 480$ ). The series represent 14 main categories of macroeconomic time series: real output and income; employment and hours; real retail, manufacturing, and trade sales; real inventories and inventory-sales ratios; orders and unfilled orders; stock prices; exchange rates; interest rates; money and credit quantity aggregates; price indexes; average hourly earnings, and miscellaneous. The variables in the dataset were transformed, standardized and screened for outliers as described in Stock and Watson (2002). The determination of the number of factors was based on the data subset of the transformed and screened 148 variables available for the full sample period ( $n = 148$ ).

As in the case of the financial data, for all  $r_{\max}$  on the grid  $6 : 1 : 20$ , our estimate of the number of pervasive factors equals two. For comparison, Stock and Watson (2005) use PC and IC criteria to estimate seven to nine pervasive factors in the data set, which is almost identical to ours but has a larger time dimension. When we apply  $PC_{p1}$  and  $IC_{p1}$  to our data set, we find that they estimate the number of factors as equal to  $r_{\max}^n$ , for  $r_{\max}^n$  in the range  $6 : 18$ .  $BIC_3$  finds 2 factors when  $r_{\max}^n$  is 6; 3 factors when  $r_{\max}^n$  is 7 and 8; 4 factors when  $r_{\max}^n$  is 9, 10, and 11; and more factors when  $r_{\max}^n$  increases further.



There is a growing literature which estimates the number of dynamic factors in the generalizations of the approximate factor model proposed in Forni et al. (2000) and Stock and Watson (2002). This literature finds at least two dynamic factors in macroeconomic datasets. If the true data generating process is indeed a dynamic factor model, then our estimate should in principle capture the number of factors and their lags. Hence, the number of static factors should be larger than two.

As we know from our Monte Carlo experiment calibrated to fit the macroeconomic data ( $\rho = 0.85, \beta = 0, n = 150, T = 500$ ), our estimator could underestimate the true number of factors if this number is really large, say 15. Such a large number of static factors is not unrealistic a priori. Another, related reason for possible underestimation may be that the dynamic factors are strongly serially correlated so that the effect of their lags can be easily confused with the contemporaneous effect. Therefore some of the static factors capturing lagged effects may be weak in the sense that the smallest “systematic” eigenvalues would be close to the largest “idiosyncratic” eigenvalues. Such a situation will lead to the underestimation by ED. Note that if some static factors are indeed weak, they can only be estimated with huge error by the principal components technique (see Onatski, 2006).

To check the robustness of our estimator to the choice of the sample, we repeat our analysis over different subgroups of the variables. We set  $r_{\max} = 20$ . Our estimate of the number of factors remains ‘two’ when we remove from the data the 20 variables that constitute “real output and income” category, or the 14 variables that constitute “price indexes” category, or the 18 variables that constitute “interest rates” category. However, when we remove 25 variables that constitute “employment and hours” category, we get an estimate of only one factor in the data. When we estimate the number of factors in the data that consist of randomly selected 50% of the original series, we get 31 out of 1000 cases of zero-factors-estimate, 267 cases of one-factor-estimate, 647 cases of two-factors-

estimate, and 55 cases of three-or-more-factors-estimate. For  $r_{\max} < 20$ , the results are similar.

## 7 Conclusion

In this paper we have developed a new consistent estimator for the number of factors in the approximate factor models. The estimator is based on the fact, established in the paper, that any finite number of the largest “idiosyncratic” eigenvalues of the sample covariance matrix cluster around a single point. In contrast, all the “systematic” eigenvalues, the number of which equals the number of factors, diverge to infinity. Our estimator consistently separates the diverging eigenvalues from the cluster and counts the number of the separated eigenvalues.

The clustering result is obtained under assumption that the matrix of idiosyncratic terms,  $e$ , has form  $A\varepsilon B$ , where  $\varepsilon$  is a matrix with i.i.d. Gaussian entries, and matrices  $A$  and  $B$  are relatively unconstrained matrices that encode cross-sectional and temporal correlation of the entries of  $e$ . We can lift the requirement of the Gaussianity of  $\varepsilon$  at the expense of making one of the matrices  $A, B$  diagonal.

To prove the clustering, we have established a novel result in the literature on large random matrices (see Lemma A4 in the Appendix). Precisely, we have characterized the upper boundary of support of the limiting spectral distribution of  $ee'/T$  in terms of the limiting spectral distributions of matrices  $A$  and  $B$ . The most advanced results of this type up to date considered the special case when either  $A$  or  $B$  is proportional to the identity matrix (see, for example, Silverstein and Choi, 1995).

The main advantage of our estimator relative to the Bai-Ng estimators is that it works well in realistically small samples when the amount of cross-sectional and temporal correlation in the idiosyncratic terms is relatively large. It also improves upon the Bai-

Ng criteria when the portion of the observed variance attributed to the factors is small relative to the variance due to the idiosyncratic term.

We applied our methodology to estimate the number of factors driving stock returns and the number of factors influencing a large set of macroeconomic variables. In both cases our estimate of the number of factors equals 2. If factors not captured by our estimator exist, their influence on the cross-sectional terms is likely to be weak, so that it is easily confused with the idiosyncratic influences. As shown by Onatski (2006), such weak factors are poorly estimated by the principal components methods.

## References

- [1] Bai, J. (2003). "Inferential theory for factor models of large dimensions", *Econometrica*, 71, pp.135-171
- [2] Bai, J. and Ng, S (2002). "Determining the number of factors in approximate factor models", *Econometrica*, 70, pp 191-221
- [3] Bai, J. and Ng, S (2004). "A PANIC Attack on Unit Roots and Cointegration," *Econometrica* 72, 1127-1177.
- [4] Bai, J. and Ng, S (2006). "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions", *Econometrica* 74, pp.1133-1150
- [5] Bai, Z. and Silverstein, J. (1998) "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices", *The Annals of Probability*, 26, pp 316-345.
- [6] Bai, Z. and Silverstein, J. (2004) "CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices", *The Annals of Probability*, Vol. 32, pp. 553-605

- [7] Breitung, J. and Eickmeier, S. (2005) “Dynamic Factor Models”, Deutsche Bundesbank Discussion Paper 38/2005.
- [8] Brown, S. and Weinstein, M., (1983) “A new approach to testing asset pricing models: the bilinear paradigm”, *Journal of Finance*, 38, pp 711-743.
- [9] Chamberlain, G. and Rothschild, M. (1983) “Arbitrage, factor structure, and mean-variance analysis on large asset markets”, *Econometrica*, 51, pp.1281-1304.
- [10] Connor, G. and Korajczyk, R. (1993) “A test for the number of factors in an approximate factor model”, *The Journal of Finance*, 58, pp. 1263-1291
- [11] Dhrymes, P., Friend, I., and Gultekin, N. (1984) “A critical reexamination of the empirical evidence on the arbitrage pricing theory”, *The Journal of Finance*, 39, pp 323-346.
- [12] Dozier, R.B. and J.W. Silverstein (2004) “Analysis of the Limiting Spectral Distribution of Large Dimensional Information-Plus-Noise Type Matrices”, *Journal of Multivariate Analysis* 98, pp. 1099-1122.
- [13] El Karoui, N. (2005) ”Recent results about the largest eigenvalue of random covariance matrices and statistical application”, *Acta Physica Polonica B*, Vol.36 No9
- [14] 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, pp 540-554,
- [15] Hallin, M., and Liska R. (2007). The generalized dynamic factor model: determining the number of factors. *Journal of the American Statistical Association* 102, 603-617.
- [16] Harding, M.C. (2008) “Explaining the single factor bias of arbitrage pricing models in finite samples”, *Economics Letters* 99, 85-88.

- [17] Horn, R.A. and Johnson C.R. (1991) Topics in Matrix Analysis, Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney.
- [18] Huang, R. and Jo, H. (1995). “Data frequency and the number of factors in stock returns”, Journal of Banking and Finance, 19, pp 987-1003.
- [19] Kapetanios, G. (2004) “A new method for determining the number of factors in factor models with large datasets”, Department of Economics, Queen Mary University of London, working paper 525.
- [20] Kapetanios, G. (2007) “A Testing Procedure for Determining the Number of Factors in Approximate Factor Models with Large Datasets”, Department of Economics, Queen Mary University of London, working paper 551.
- [21] Marčenko, V.A., and L.A. Pastur (1967) “Distribution of eigenvalues for some sets of random matrices”, Math. USSR-Sbornik, vol. 1, no. 4, 457-483
- [22] Onatski, A. (2006) “Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak”, manuscript, Columbia University.
- [23] Onatski, A. (2008) “Testing hypotheses about the number of factors in large factor models”, manuscript, Columbia University.
- [24] Paul, D., and Silverstein, J. (2008) “No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix”, working paper, North Carolina State University.
- [25] Reinganum, M., (1981) “The arbitrage pricing theory: some empirical results”, Journal of Finance, 36, pp 313-321.

- [26] Roll, R., and Ross S. (1980) “An empirical investigation of the arbitrage pricing theory”, *Journal of Finance*, 5, pp 1073-1103
- [27] Silverstein, J. and Choi, S. (1995) “Analysis of the limiting spectral distribution of large dimensional random matrices”, *Journal of Multivariate Analysis*, 54, pp 295-309
- [28] Silverstein, J. and Combettes P. (1992) “Signal Detection via Spectral Theory of Large Dimensional Random Matrices”, *IEEE Transactions on Signal Processing*, Vol. 40, No. 8, pp 2100-2105
- [29] Stock, J. and Watson, M. (2002) “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business and Economic Statistics*, 20, pp. 147-162.
- [30] Stock, J. and Watson, M. (2005) “Implications of Dynamic Factor Models for VAR Analysis”, manuscript, Harvard University.
- [31] Trzcinka, C. (1986) “On the number of factors in the arbitrage pricing model”, *The Journal of Finance*, XLI, pp 347-368.
- [32] Van Loan, C. F. and Pitsianis, N. P. (1993) “Approximation with Kronecker products”, in M. S. Moonen and G. H. Golub, editors, *Linear Algebra for Large Scale and Real Time Applications*, Kluwer Publications, pp. 293–314.
- [33] Wachter, K.W. (1976) “Probability plotting points for principal components”, in Hoaglin and Welsch, eds, “Ninth Interface Symposium on Computer Science and Statistics”, Prindle, Weber and Schmidt, Boston, pp. 299-308
- [34] Watson, M.W. (2003) “Macroeconomic Forecasting Using Many Predictors”, in M. Dewatripont, L. Hansen and S. Turnovsky (eds), *Advances in Economics and Econometrics, Theory and Applications, Eight World Congress of the Econometric Society*,” Vol. III, page 87-115.

- [35] Watson, M. and D. Amengual (2007) “Consistent Estimation of the Number of Dynamic Factors in a Large N And T Panel”, *Journal of Business and Economic Statistics*
- [36] Zhang, L. (2006). *Spectral Analysis of Large Dimensional Random Matrices*. Ph. D. Thesis. National University of Singapore.

Table 1: (% of overestimation)/(% of underestimation) in 1000 Monte Carlo replications of data-generating process (14). Uncorrelated idiosyncratic terms:  $\rho = 0, \beta = 0$ .

$n$	$T$	$PC_{p1}$	$IC_{p1}$	$BIC_3$	ME	MED	ED
$r = 1$							
40	100	1/0	0/0	0/0	0/0	2/0	2/0
200	60	0/0	0/0	0/0	0/0	4/0	1/0
1000	60	0/0	0/0	0/0	0/0	0/0	1/0
100	100	0/0	0/0	0/0	0/0	8/0	1/0
150	150	0/0	0/0	0/0	0/0	7/0	1/0
150	500	0/0	0/0	0/0	0/0	14/0	1/0
1000	250	0/0	0/0	0/0	0/0	3/0	1/0
$r = 3$							
40	100	0/0	0/0	0/7	0/0	10/0	1/0
200	60	0/0	0/0	0/0	0/0	4/0	0/0
1000	60	0/0	0/0	0/0	7/0	2/0	0/0
100	100	0/0	0/0	0/0	0/0	17/0	1/0
150	150	0/0	0/0	0/0	0/0	10/0	0/0
150	500	0/0	0/0	0/0	0/0	23/0	0/0
1000	250	0/0	0/0	0/0	0/0	4/0	1/0
$r = 5$							
40	100	0/0	0/12	0/99	0/6	17/0	1/1
200	60	0/0	0/0	0/85	0/0	14/0	0/0
1000	60	0/0	0/0	0/90	55/0	3/0	0/0
100	100	0/0	0/0	0/60	0/0	19/0	0/0
150	150	0/0	0/0	0/2	0/0	15/0	0/0
150	500	0/0	0/0	0/0	0/0	34/0	0/0
1000	250	0/0	0/0	0/0	0/0	9/0	0/0
$r = 15$							
40	100	98/0	0/100	0/100	0/100	1/99	0/100
200	60	0/99	0/100	0/100	78/5	3/56	0/92
1000	60	0/99	0/100	0/100	100/0	0/3	0/0
100	100	0/100	0/100	0/100	0/80	34/10	0/93
150	150	0/85	0/100	0/100	0/1	41/0	0/2
150	500	0/1	0/18	0/100	0/0	44/0	0/0
1000	250	0/0	0/0	0/100	0/0	26/0	0/0



Table 2: (% of overestimation)/(% of underestimation) in 1000 Monte Carlo replications of data-generating process (14). Correlated idiosyncratic terms:  $\rho = 0.3, \beta = 0.1$ .

$n$	$T$	$PC_{p1}$	$IC_{p1}$	$BIC_3$	ME	MED	ED
$r = 1$							
40	100	100/0	100/0	91/0	100/0	100/0	69/0
200	60	100/0	3/0	0/0	100/0	94/0	5/0
1000	60	0/0	0/0	0/0	100/0	13/0	1/0
100	100	100/0	93/0	0/0	100/0	100/0	7/0
150	150	100/0	70/0	0/0	100/0	100/0	3/0
150	500	100/0	100/0	0/0	100/0	100/0	0/0
1000	250	0/0	0/0	0/0	100/0	70/0	2/0
$r = 3$							
40	100	100/0	100/0	57/0	100/0	100/0	52/27
200	60	99/0	4/0	0/0	100/0	95/0	2/0
1000	60	0/0	0/0	0/0	100/0	17/0	1/0
100	100	100/0	92/0	0/0	100/0	100/0	2/0
150	150	100/0	70/0	0/0	100/0	99/0	1/0
150	500	100/0	100/0	0/0	100/0	91/0	0/0
1000	250	0/0	0/0	0/0	100/0	64/0	0/0
$r = 5$							
40	100	100/0	100/0	2/53	100/0	100/0	24/74
200	60	71/0	4/0	0/63	100/0	90/0	1/2
1000	60	0/0	0/0	0/79	100/0	26/0	0/0
100	100	100/0	91/0	0/21	100/0	99/0	1/22
150	150	100/0	72/0	0/0	100/0	98/0	0/0
150	500	100/0	100/0	0/0	100/0	72/0	0/0
1000	250	0/0	0/0	0/0	100/0	60/0	0/0
$r = 15$							
40	100	100/0	0/100	0/100	8/90	53/39	0/100
200	60	93/0	0/100	0/100	100/0	94/3	0/100
1000	60	0/58	0/100	0/100	100/0	73/2	0/98
100	100	99/0	0/100	0/100	100/0	100/0	0/100
150	150	100/0	3/88	0/100	100/0	100/0	0/100
150	500	100/0	99/0	0/100	100/0	100/0	0/100
1000	250	0/0	0/0	0/100	100/0	99/0	0/0

Table 3: (% of overestimation)/(% of underestimation) in 1000 Monte Carlo replications of data-generating process (14). Correlated idiosyncratic terms:  $\rho = 0.85, \beta = 0$ .

$n$	$T$	$PC_{p1}$	$IC_{p1}$	$BIC_3$	ME	MED	ED
$r = 1$							
40	100	100/0	100/0	100/0	100/0	100/0	14/0
200	60	100/0	100/0	100/0	100/0	100/0	4/0
1000	60	100/0	100/0	100/0	100/0	100/0	0/0
100	100	100/0	100/0	100/0	100/0	100/0	4/0
150	150	100/0	100/0	100/0	100/0	93/0	1/0
150	500	100/0	97/0	0/0	100/0	25/0	2/0
1000	250	100/0	100/0	100/0	100/0	0/0	0/0
$r = 3$							
40	100	100/0	100/0	100/0	100/0	99/0	8/50
200	60	100/0	100/0	100/0	100/0	100/0	3/85
1000	60	100/0	100/0	100/0	100/0	100/0	0/86
100	100	100/0	100/0	100/0	100/0	100/0	2/18
150	150	100/0	100/0	100/0	100/0	92/0	1/0
150	500	100/0	97/0	0/0	100/0	31/0	1/0
1000	250	100/0	100/0	100/0	100/0	0/0	0/0
$r = 5$							
40	100	100/0	100/0	51/5	100/0	94/0	1/98
200	60	100/0	100/0	100/0	100/0	100/0	0/100
1000	60	100/0	100/0	100/0	100/0	100/0	0/100
100	100	100/0	100/0	99/0	100/0	97/0	0/97
150	150	100/0	100/0	76/0	100/0	88/0	0/44
150	500	100/0	97/0	0/0	100/0	36/0	0/0
1000	250	100/0	100/0	87/0	100/0	0/0	0/0
$r = 15$							
40	100	100/0	100/0	90/0	100/0	99/0	0/100
200	60	100/0	100/0	98/0	100/0	100/0	0/100
1000	60	100/0	100/0	77/1	100/0	100/0	0/100
100	100	100/0	100/0	59/5	100/0	100/0	0/100
150	150	100/0	100/0	7/51	100/0	100/0	0/100
150	500	100/0	97/0	0/100	100/0	99/0	0/99
1000	250	100/0	100/0	100/0	100/0	100/0	0/100

Table 4: Numerical height of the surfaces shown in Figure 2.

$\rho \setminus \beta$	0	.05	0.1	.15	0.2	.25	0.3	0	.05	0.1	.15	2	.25	0.3
	PC <sub>p1</sub>							IC <sub>p1</sub>						
0	3	3	6.3	8	8	8	8	3	3	4.3	8	8	8	8
0.1	3	3	6.3	8	8	8	8	3	3	4.4	8	8	8	8
0.2	3	3	6.4	8	8	8	8	3	3	4.7	8	8	8	8
0.3	3	3.2	6.6	8	8	8	8	3	3	5	8	8	8	8
0.4	3	3.7	6.9	8	8	8	8	3	3	5.5	8	8	8	8
0.5	3.3	4.6	7.3	8	8	8	8	3	3.1	6.4	8	8	8	8
0.6	5.3	6.2	7.8	8	8	8	8	3.1	3.8	7.5	8	8	8	8
0.7	7.9	8	8	8	8	8	8	7.2	7.7	8	8	8	8	8
0.8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
0.9	8	8	8	8	8	8	8	8	8	8	8	8	8	8
	BIC <sub>3</sub>							ME						
0	3	3	3	3.7	5.9	7.5	8	3	3.5	8	8	8	8	8
0.1	3	3	3	3.7	6	7.4	8	3	3.7	8	8	8	8	8
0.2	3	3	3	3.8	6	7.4	8	3	4.1	8	8	8	8	8
0.3	3	3	3	4	6.1	7.5	8	3	4.9	8	8	8	8	8
0.4	3	3	3	4.2	6.2	7.5	8	3.3	6.6	8	8	8	8	8
0.5	3	3	3	4.5	6.3	7.5	8	7.2	8	8	8	8	8	8
0.6	3	3	3.2	4.9	6.6	7.7	8	8	8	8	8	8	8	8
0.7	3	3	3.9	5.6	6.9	7.8	8	8	8	8	8	8	8	8
0.8	4.5	4.9	5.7	6.7	7.5	7.9	8	8	8	8	8	8	8	8
0.9	7.8	7.9	7.9	8	8	8	8	8	8	8	8	8	8	8
	MED							ED						
0	3.1	5.8	7.7	7.8	7.9	7.9	7.9	3	3	3	3	2.8	2.4	1.9
0.1	3.2	5.6	7.6	7.8	7.9	7.9	7.9	3	3.1	3	3	2.8	2.4	1.9
0.2	3.3	5.3	7.6	7.8	7.9	7.9	7.9	3	3	3	3	2.7	2.2	1.9
0.3	3.4	5.2	7.5	7.8	7.8	7.9	7.9	3	3.1	3.1	3	2.7	2.2	1.8
0.4	3.4	5.1	7.3	7.7	7.8	7.9	7.9	3	3	3.1	2.9	2.6	2	1.7
0.5	3.7	5.1	7.3	7.6	7.8	7.8	7.9	3	3	3.1	2.9	2.4	1.8	1.3
0.6	4	5.1	6.9	7.4	7.7	7.8	7.9	3	3	3.1	2.8	2	1.6	1.1
0.7	4.9	5.2	6.7	7.4	7.6	7.6	7.8	3	3	3	2.5	1.7	1.2	0.9
0.8	6.5	6.9	7	7.2	7.6	7.7	7.6	3	3	2.6	1.8	1.4	1	0.8
0.9	7.6	7.6	7.6	7.6	7.4	7.6	7.7	0.9	0.9	1	1.2	1.5	1.4	1.3

Table 5: Mean of different estimators in 1000 Monte Carlo replications of the data generating process (15).

n	T	PC	IC	BIC	ME	MED	ED	PC	IC	BIC	ME	MED	ED
		$\theta = 1, r = 0$						$\theta = 1, r = 4$					
50	50	5.8	0.1	0	0.7	5.1	0.1	6.8	4.4	4	5.8	4.9	4
100	50	2.9	0	0	0.1	3.6	0.1	4.7	4	4	4.3	4.6	4
200	50	0.2	0	0	0	1.9	0.1	4	4	4	4	4.4	4
50	100	5.4	0.1	0	7.9	6	0.1	6.5	4.2	4	8	4.9	4
100	100	1.8	0	0	4.3	4	0.1	4.2	4	4	7.9	4.4	4
200	100	0	0	0	0.2	2.8	0	4	4	4	4.9	4.2	4
50	200	5.5	0	0	8	6.2	0	6.5	4.1	4	8	4.5	4
100	200	1.2	0	0	8	3	0	4	4	4	8	4.2	4
200	200	0	0	0	8	2.3	0	4	4	4	8	4.1	4
		$\theta = 3, r = 0$						$\theta = 3, r = 4$					
50	50	5.8	0.1	0	0.9	5.2	0.1	6.7	3.4	2.1	4.9	5.9	0.9
100	50	2.9	0	0	0.1	3.2	0.1	4.7	3.6	1.8	4.2	5.6	2.7
200	50	0.1	0	0	0	2.1	0	4	3.8	1.6	4	5.2	4
50	100	5.5	0.1	0	7.8	5.8	0.1	6.4	3.8	2.1	8	6.3	1.5
100	100	1.8	0	0	3.9	3.7	0.1	4.2	4	2.4	7.8	5.7	3.9
200	100	0	0	0	0.4	2.7	0.1	4	4	2.8	4.7	5.2	4
50	200	5.5	0	0	8	6	0	6.4	3.9	2	8	6	2.5
100	200	1.1	0	0	8	3.3	0	4	4	3	8	5.1	4
200	200	0	0	0	8	2.3	0	4	4	3.9	8	5	4
		$\theta = 5, r = 0$						$\theta = 5, r = 4$					
50	50	5.8	0.1	0	0.7	5.1	0.1	6.4	1.4	0.9	3.1	5.9	0.4
100	50	2.9	0	0	0.1	3.3	0.1	4.5	1.7	0.3	3.3	5.7	0.8
200	50	0.2	0	0	0	2	0	3.8	2.1	0.1	3.7	5.4	2.7
50	100	5.4	0.1	0	7.9	6.2	0.1	6.2	2	0.6	8	6.2	0.5
100	100	1.8	0	0	4	3.8	0	4.1	3	0.4	7.7	6	2
200	100	0	0	0	0.3	2.6	0	4	3.8	0.5	4.6	5.7	4
50	200	5.4	0	0	8	5.8	0	6.1	2.5	0.4	8	6.2	0.6
100	200	1.1	0	0	8	3.4	0	4	3.8	0.7	8	5.5	3.6
200	200	0	0	0	8	2.6	0	4	4	1.6	8	5.3	4
		$\theta = 10, r = 0$						$\theta = 10, r = 4$					
50	50	5.8	0.1	0	0.9	5.2	0.1	5.9	0.3	0.1	1.1	5	0.2
100	50	2.9	0	0	0.1	3.2	0.1	3.7	0.1	0	0.9	4.6	0.3
200	50	0.1	0	0	0	2.1	0	2.2	0	0	1.2	5.4	0.4
50	100	5.5	0.1	0	7.8	5.8	0.1	5.5	0.2	0	8	5.5	0.2
100	100	1.8	0	0	3.9	3.7	0.1	3.1	0.2	0	7.1	5.8	0.3
200	100	0	0	0	0.4	2.7	0.1	2.5	0.4	0	4.3	5.9	1.2
50	200	5.5	0	0	8	6	0	5.3	0.3	0	8	5.2	0.2
100	200	1.1	0	0	8	3.3	0	3.1	0.7	0	8	5.3	0.5
200	200	0	0	0	8	2.3	0	3.4	2.1	0	8	5.8	3.4

Table 6. Performance of different estimators in 1000 Monte Carlo replications of the data generating process (15):  
 % of overestimation for  $r = 0$ , and (% of overestimation)/(% of underestimation) for  $r = 4$ .

n	T	PC	IC	BIC	ME	MED	ED	PC	IC	BIC	ME	MED	ED
		$\theta = 1, r = 0$						$\theta = 1, r = 4$					
50	50	100	10	3	51	89	10	100/0	29/0	0/3	74/0	49/0	2/4
100	50	100	0	0	11	82	6	67/0	0/0	0/2	25/0	40/0	1/0
200	50	16	0	0	0	57	5	0/0	0/0	0/1	1/0	29/0	1/0
50	100	100	5	0	100	93	5	100/0	18/0	0/1	100/0	52/0	1/0
100	100	99	0	0	100	85	5	23/0	0/0	0/0	100/0	30/0	1/0
200	100	0	0	0	22	72	3	0/0	0/0	0/0	50/0	18/0	1/0
50	200	100	1	0	100	96	2	100/0	9/0	0/1	100/0	34/0	0/0
100	200	89	0	0	100	73	3	5/0	0/0	0/0	100/0	13/0	0/0
200	200	0	0	0	100	65	2	0/0	0/0	0/0	100/0	11/0	0/0
		$\theta = 3, r = 0$						$\theta = 3, r = 4$					
50	50	100	11	3	60	94	9	100/0	11/49	0/98	54/13	80/3	1/91
100	50	100	0	0	11	75	7	63/0	0/36	0/100	20/5	76/0	1/42
200	50	14	0	0	0	62	4	0/0	0/15	0/100	0/1	67/0	1/2
50	100	100	5	0	100	96	5	100/0	9/28	0/98	100/0	88/2	1/77
100	100	99	0	0	97	79	5	21/0	0/2	0/96	100/0	81/0	1/4
200	100	0	0	0	33	73	5	0/0	0/0	0/86	47/0	63/0	1/0
50	200	100	2	0	100	98	2	100/0	6/12	0/98	100/0	76/1	0/49
100	200	88	0	0	100	75	3	4/0	0/0	0/77	100/0	62/0	0/0
200	200	0	0	0	100	61	2	0/0	0/0	0/8	100/0	59/0	0/0
		$\theta = 5, r = 0$						$\theta = 5, r = 4$					
50	50	100	11	3	47	93	9	100/0	1/97	0/100	16/66	79/9	0/99
100	50	100	0	0	5	79	6	46/3	0/98	0/100	8/57	75/5	1/92
200	50	16	0	0	0	63	4	0/18	0/94	0/100	0/25	76/0	1/41
50	100	100	6	0	100	97	5	100/0	1/92	0/100	100/0	86/5	0/99
100	100	99	0	0	99	82	4	16/2	0/67	0/100	99/0	87/0	1/59
200	100	0	0	0	27	71	3	0/1	0/22	0/100	48/0	80/0	1/1
50	200	100	1	0	100	95	3	100/0	1/86	0/100	100/0	81/7	0/98
100	200	88	0	0	100	79	2	3/0	0/21	0/100	100/0	76/0	1/14
200	200	0	0	0	100	68	3	0/0	0/0	0/100	100/0	69/0	0/0
		$\theta = 10, r = 0$						$\theta = 10, r = 4$					
50	50	100	11	2	54	93	10	99/0	0/100	0/100	1/99	61/26	0/100
100	50	100	0	0	11	83	7	9/40	0/100	0/100	0/99	54/33	0/100
200	50	13	0	0	0	58	4	0/99	0/100	0/100	0/99	66/14	0/98
50	100	100	5	0	100	93	4	96/0	0/100	0/100	100/0	72/19	0/100
100	100	99	0	0	98	83	5	1/74	0/100	0/100	93/2	78/8	0/100
200	100	0	0	0	26	70	4	0/97	0/100	0/100	32/6	82/2	0/80
50	200	100	1	0	100	94	2	92/0	0/100	0/100	100/0	65/26	0/100
100	200	88	0	0	100	76	3	0/80	0/100	0/100	100/0	66/14	0/99
200	200	0	0	0	100	57	3	0/55	0/96	0/100	100/0	81/0	1/19