Eigenvalue Ratio Test for the Number of Factors

Seung C. Ahn
Department of Economics
W.P. Carey School of Business
Arizona State University
Tempe, AZ 85287 USA
miniahn@asu.edu

Alex R. Horenstein
Department of Economics
W.P. Carey School of Business
Arizona State University
Tempe, AZ 85287, USA
alex.horenstein@asu.edu

This Version: November, 2008

Abstract

This paper proposes two new estimators for determining the number of factors (r) in approximate factor models. We exploit the well known fact that the r eigenvalues of the variance matrix of N response variables grow unboundedly as N increases. Bai and Ng (2002) and Onatski (2006) have developed the methods by which the number of factors can be estimated by comparing the eigenvalues with prespecified or estimated threshold values. One limitation of the methods is that they lack a practical guidance on how to choose the appropriate threshold values in finite samples. Asymptotically, any scalar multiple of an appropriate threshold value is also a valid threshold value. However, the finite-sample properties of the estimators critically depend on the choice of the thresholds. As a treatment to this problem, we propose the estimators that do not require use of threshold values. These new estimators are obtained simply by maximizing the ratio of two adjacent eigenvalues. We show that the estimators are consistent under the general conditions of Bai and Ng (2002). Our simulation results show that the estimators have good finite sample properties unless the signal-to-noise-ratio of each factor is too low. They perform better than the Bai-Ng estimators do when either the number of the response variables analyzed or the number of time series observations, T, is small.

Key words: approximate factor models, number of factors, eigenvalues, scree test.

JEL classification**:** C01, C44, C52, G00.

Acknowledgment

We thank Manuel Santos, Marcos Perez, Crocker Liu, Federico Nardari and Stephan Dieckmann for their helpful comments on this paper. We would like to thank participants in the econometrics seminars at Tokyo University, Kyoto University, Hitotsubashi University, the Korea Econometric Society Summer Meeting, Korea University, Wilfrid Laurier University. All remaining errors are of course our own.

1. Introduction

Economic or econometric models often predict that a small number of latent factors drive a large number of response variables. Factor analysis is a natural approach to estimate such models. Since the late 1970's and the early 1980's the factor analysis has been growingly important in the economic realm, especially in the area of finance following the development of the Arbitrage Price Theory (APT) by Ross (1976). The foundation of this theory is the exact factor structure in which returns on risky assets are cross-sectionally correlated only through common latent factors, while their idiosyncratic components are uncorrelated. Chamberlain and Rothschild (1983) have shown that the theory can be extended to approximate factor models in which the idiosyncratic components are cross-correlated. Use of factor analysis has extended gradually to a wider range of economic disciplines. Recent examples are Stock and Watson (1999, 2002a, 2002b) and Bernanke, Boivin, and Eliasz (2004), which propose to forecast major economic variables using the factor estimates from a large number of economic variables.

Identifying the number of common factors that explain the common variations in a set of response variables is one of the major tasks of factor analysis. The literature on this topic can be traced as far as the early 1950's in other disciplines, such as psychology (see Cattell (1966) for more references). The test most widely used to determine the number of factors has been the scree test of Cattell (1966). This is a visual test based on the behaviors of the eigenvalues of the second-moment matrix of the response variables. More recently, Forni, Hallin, Lippi and Reichlin (2000) also proposed a visual test based on the behavior of the eigenvalues for determining the number of factors in the context of dynamic factor models. In this paper, we propose three formal estimators based on the eigenvalues, which can consistently estimate the number of factors in approximate factor models with both large numbers of cross section units (N) and time series observations (T).

While the scree test is intuitively appealing, it is essentially an eyeball test. Recently, more formal statistic methods have been developed for approximate factor models. Bai and Ng (2002, hereafter BN) propose to estimate the number of factors by minimizing certain model section criterion functions, typical form of which is the unconditional variance of the idiosyncratic components plus a penalty function monotonically increasing with the number of factors. Onatski (2006) finds that the BN estimators are linked to scree tests. He shows that a set of the BN estimators are the obtained by comparing the eigenvalues of the covariance matrix and the values of the chosen penalty functions. Specifically, a typical BN estimator is equivalent to the number of the eigenvalues larger than a threshold value which is specified by a chosen penalty function. Onatski develops an alternative estimator based on the fact that for the data with r latent common factors, the largest r eigenvalues of the second-moment matrix of the data grow without limit with N, while the rest of the eigenvalues are bounded (see Chamberlain and Rothschild, 1983). A threshold value is chosen from the empirical distribution of the eigenvalues to differentiate the diverging ones from the bounded ones. Onatski shows that the number of the eigenvalues greater than the threshold value is a consistent estimator of r, under the assumption that the idiosyncratic components of response variables are either serially or cross-sectionally uncorrelated, but not both. His Monte Carlo experiments show that the estimation method has good finite-sample properties even if the idiosyncratic components are both serially and cross-sectionally correlated.[1]

In this paper we propose three alternative estimators that do not require use of the threshold values. One estimator is obtained simply by maximizing the ratio of two adjacent eigenvalues. The other estimators are obtained by a similar method. The new estimators are computed by using eigenvalues only. Our Monte Carlo experiment results indicate that the new estimators outperform the BN and Onatski estimators even in samples with small N and T unless the signal-to-noise ratios are too small.

The paper is organized as follows. Section 2 provides the intuition for our estimators. Section 3 presents the assumptions consistent with approximate factor models. These assumptions are essentially the same as those in Bai and Ng (2002). Under the assumptions, we establish the consistency of the estimators we propose. Section 4 reports our Monte Carlo experiments. Section 5 provides an application of the two estimators to macroeconomic data and stock returns. Concluding remarks are given in section 6.

2. Preliminaries and Motivation

We begin by defining the approximate factor model of Chamberlain and Rothschild (1983). Suppose we have a set of panel data generated from a factor structure. We will try to use the same notation as Bai and Ng (BN, 2002) use as much as we can. Let x_{it} be the observed value of

[1] Kapetanios (2004) also develops an algorithm for testing the number of factors when the idiosyncratic components are serially uncorrelated. Similarly to Onatski (2006), his estimation method is also based on the fact that the (r+1)th largest eigenvalue converges almost surely to a constant.

the response variable i = 1,…,N at time t = 1,…,T. Assume that the response variables are generated by an r×1 vector of common factors, $f_t = (f_{1t}, f_{2t}, ..., f_{rt})'$:

$$x_{i\bullet} = F\lambda_i^o + \varepsilon_{i\bullet}, \tag{1}$$

where $x_{i\bullet} = (x_{i1}, x_{i2}, ..., x_{iT})'$, $F = (f_1, f_2, ..., f_T)'$, $\lambda_i^o = (\lambda_{i1}^o, \lambda_{i2}^o, ..., \lambda_{ir}^o)'$ is the r×1 vector of factor loadings for variable i, , and $\varepsilon_{i\bullet} = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ is the vector of the idiosyncratic components of variable i. The factors, factor loads, and idiosyncratic errors are not observed. Using this notation, we can describe the complete panel data by

$$X = F\Lambda^{o\prime} + \mathrm{E}, \tag{2}$$

where $X = (x_{1\bullet}, ..., x_{N\bullet})$, Λ^o = $(\lambda_1^o, \lambda_2^o, ..., \lambda_N^o)'$, and $\mathrm{E} = (\varepsilon_{1\bullet}, \varepsilon_{2\bullet}, ..., \varepsilon_{N\bullet})$. Both F and Λ_o could be random. But following Bai and Ng (2002), we treat the entries in Λ^o as parameters and those in F as random variables.

For future use, we also define $x_{\bullet t} = (x_{1t}, x_{2t}, ..., x_{Nt})'$. Notice that if the means of x_{it} are all zeros, XX'/N is a sample covariance matrix of $x_{\bullet t}$, while $X'X/T$ is a sample covariance matrix of $x_{i\bullet}$. Let $m = \min(N,T)$ and $M = \max(N,T)$. We use $\lambda_j(A)$ to denote the j^{th} largest eigenvalue of a positive semidefinite matrix A. Define

$$\tilde{\mu}_{NT,k} \equiv \lambda_k\left(\frac{1}{NT}XX'\right) = \lambda_k\left(\frac{1}{NT}X'X\right),$$

$k = 1, ..., m$. Then, $T\tilde{\mu}_{NT,j}$ ($N\tilde{\mu}_{NT,j}$) is the j^{th} largest eigenvalue of the sample covariance matrix of $x_{\bullet t}$ ($x_{i\bullet}$).

One of the most popular methods used to determine the number of common factors is the *scree test* developed by Cattell (1966). The test is a visual way to find out the number of factors from the eigenvalues of XX'. Cattell (1983) describes his method as follows:

> *"To my delight, a very simple finding presented itself, namely, that if I plotted the principal components in their sizes, as a diminishing series, and then joined up the points all through the number of variables concerned, a relatively sharp break appeared where the true number of factors ended and the 'detritus', presumably due to error factors, appeared. From the analogy of the steep descent of a mountain till one comes to the scree of rubble at the foot of it, I decided to call this the scree test"*

In order to visualize the relations of the scree test, BN estimators, and our estimators, we report here some tentative simulation results. The data used for simulation are generated by a five-factor model:

$$x_{it} = \Sigma_{j=1}^{5} \lambda_{ij} f_{jt} + \sqrt{5} e_{it} \tag{3}$$

where e_{it}, λ_{ij} and f_{jt} are independently drawn from $N(0,1)$, and $N = T = 500$. Figure 1 depicts the averages of the k^{th} eigenvalues from 1,000 simulated data sets. When the scree rule is used, the estimated number of factors is the one found just before the "scree of ruble at the foot of a mountain". Thus, Figure 1 suggests presence of five common factors.

One estimator we propose in this paper, which we call *eigenvalue ratio* (ER) estimator, is $\tilde{k}_{ER} = \arg\max_{k \le kmax} (\tilde{\mu}_{NT,k} / \tilde{\mu}_{NT,k+1})$, where *kmax* is the maximum possible number of factors a researcher choose. The ratios of two adjacent eigenvalues are depicted in Figure 2. The ratio is clearly peaked at $k = 5 = r$.

The above exercise is the case in which the scree test can cleanly identify the true number of factors. But the scree test result is not always as obvious as in Figure 1. There are many other cases in which it is hard to identify the point where the *scree* starts. As an example, we generated data from the same model (3), but the errors are now generated to be cross-sectionally and serially correlated.[2] Figures 3.1 and 3.2 show the graphs of the eigenvalues and eigenvalue ratios. In Figure 3.1, it is not obvious which scree point should be chosen to estimate the number of factors. The scree estimator could be 5, 6, 7 or even 11. In contrast, the eigenvalue ratio is clearly maximized at $k = 5 = r$. Once again, the ER estimation successfully picks up the correct number of factors *r*.

Formal consistent estimation methods have been proposed recently by Bai and Ng (BN, 2002) and Onatski (2006). The estimation method by Onatski (2006) is readily related to the scree test. The Onatski estimator, $\tilde{k}_{ON}$, equals the number of the eigenvalues of XX' greater than a threshold value $\hat{c}$: $\tilde{k}_{ON} = \arg\max_{k \le kmax} \{k \mid \tilde{\mu}_{NT,k} > \hat{c}/N\}$. The threshold value is estimated from the empirical distribution of the eigenvalues and the maximum value of *k*, *kmax*. Under the assumption that the idiosyncratic errors are not both cross-sectionally and serially correlated,

---

[2] The data generating process we use for Figures 3.1 and 3.2 are explained in section 3.

$\tilde{k}_{ON}$ is a consistent estimator of the true number of factors, *r*. This method is similar to the scree test in the sense that it uses the eigenvalues of the estimated second moment matrix of $x_{\bullet t}$. Figure 4 depicts the average of the Onatski threshold values from the same generated data from model (3) with *i.i.d.* errors. Again the true number of factors is *r = 5.*

As Onatski has shown, one set of the BN estimators have the form similar to his estimator. The typical form of the criterion functions used by BN is the mean of the squared residuals from the regressions of response variables on estimated principal component factors plus a penalty function. To be specific, let $\tilde{F}^k$ be the *T*×*k* matrix of the eigenvectors corresponding to the largest *k* eigenvalues of XX', normalized so that $\tilde{F}^{k\prime}\tilde{F}^k / T = I_k$. Let $V(k, \tilde{F}^k)$ be the sum of squared residuals from the regressions of $x_{i\bullet}$ on $\tilde{F}^k$. Then, it is straightforward to show:

$$V(k, \tilde{F}^k) = \frac{1}{NT}\Sigma_{i=1}^{N}\left(x_{i\bullet}'x_{i\bullet} - x_{i\bullet}'P(\tilde{F}^k)x_{i\bullet}\right) = \Sigma_{j=k+1}^{T}\tilde{\mu}_{NT,j}, \tag{4}$$

where the last equality is due to Onatski (2006). Let *kmax* be the maximum value of *k* to estimate the true number of factors, *r*; and define $\hat{\sigma}^2 = V(kmax, \tilde{F}^{kmax}) = \Sigma_{j=kmax+1}^{T}\tilde{\mu}_{NT,j}$. Then, the criterion functions used by BN are given:

$$PC(k) = V(k, \tilde{F}^k) + \hat{\sigma}^2 kg(N,T) = \Sigma_{j=k+1}^{T}\tilde{\mu}_{NT,j} + kg(N,T)\Sigma_{j=kmax+1}^{T}\tilde{\mu}_{NT,j}; \tag{5}$$

$$IC(k) = \ln\left(V(k, \tilde{F}^k)\right) + kg(N,T) = \ln\left(\Sigma_{j=k+1}^{T}\tilde{\mu}_{NT,j}\right) + kg(N,T), \tag{6}$$

where $g(N,T)$ is a penalty function such that $g(N,T) \to 0$ and $g(N,T)m \to \infty$ as $N,T \to \infty$. A BN estimator is obtained by minimizing one of the two functions. One of the penalty function suggested by BN is

$$g(N,T) = \left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right). \tag{7}$$

Let $\tilde{k}_{PC} = \min_{k \le kmax} PC(k)$ and $\tilde{k}_{IC} = \min_{k \le kmax} IC(k)$. Then, using (4) and the monotonicity of the eigenvalues, we can show that $\tilde{k}_{PC} = \max_{k \le kmax}\{k \mid \tilde{\mu}_{NT,k} \ge \hat{\sigma}^2 g(N,T)\}$ (see Onatski, 2006). Thus, the PC estimation can be viewed as a scree test using $\hat{\sigma}^2 g(N,T)$ as a threshold value. Conversely, the Onatski estimator can be viewed as a PC estimator using an estimated threshold value as a substitute for a prespecified penalty function.

The IC estimator $\tilde{k}_{IC}$ has a similar property, although Onatski does not consider. Observe that as the number of factors used for estimation increases from k to k+1, the value of $IC_p(k)$ changes by $g(N,T)-\ln(\tilde{\mu}^*_{NT,k+1})$, where $\tilde{\mu}^*_{NT,k}=(\Sigma^T_{j=k}\tilde{\mu}_{NT,j})/\left(\Sigma^T_{j=k+1}\tilde{\mu}_{NT,j}\right)$. For $\tilde{k}_{IC}$ to minimize $IC_p(k)$, the value of $IC_p(k)$ should increase (decrease) as k increases from $\tilde{k}_{IC}$ ($\tilde{k}_{IC}-1$) to $\tilde{k}_{IC}+1$ ($\tilde{k}_{IC}$). Thus, $\tilde{k}_{IC}$ must satisfy the following condition

$$\ln(\tilde{\mu}^*_{\tilde{k}_{IC}+1})<g(N,T)<\ln(\tilde{\mu}^*_{\tilde{k}_{IC}}) \tag{8}$$

However, this condition is a necessary, but not a sufficient condition, because $\tilde{\mu}^*_{NT,k}$ is no longer a monotonic function of k. That is, $\tilde{k}_{IC}\neq\tilde{k}_{IC,M}$ $=\arg\max_{k\leq kmax}\{\tilde{\mu}^*_{NT,k}\mid\ln(\tilde{\mu}^*_{NT,k})>g(N,T)\}$, although the equality held often in our Monte Carlo simulations.[3]

Figures 5.1 and 5.2 show the shapes of $V(k,\tilde{F}^k)$ and $PC(k)$ with the penalty function (7). The figures are obtained using the data generated from model (3) with *i.i.d.* errors. From Figure 5.1, we can observe a kink in the curve of $V(k,\tilde{F}^k)$ at $k=5=r$. The value of $V(k,\tilde{F}^k)$ decreases at slower rate after the kink point. In Figure 5.2, we consider three PC estimators. PC1 represents the PC function with the penalty function (7), while PC2 uses a different penalty function introduced in the next section. As we point out above, the Onatski estimator can be viewed as a PC estimator using his estimated threshold value as a penalty function. ON represents the PC criterion function corresponding to the Onatski estimator. Observe that all of the PC criterion function values rise after the kink point. Thus, the minimum value of each criterion function is achieved at the correct number of factors.

Figure 5.2 suggests a possible pitfall of the $PC(k)$ criterion function, as well as $IC(k)$. That is, the value of the function does not rise after the kink point as sharply as it falls before the kink point. It appears that the function (7) is somewhat generous in penalizing large k. The Onatski threshold value seems to be also generous for large k. In contrast, as we have seen from Figures 2.2 and 3.2, the eigenvalue ratio is sharply peaked at the true number of factors.

[3] In unreported simulation experiments, we found that $\tilde{k}_{IC,M}$ had slightly better finite-sample properties than $\tilde{k}_{IC}$, when they are different.

Another potential problem in using the BN and Onatski estimators is that they require a choice of *kmax*. In asymptotics, the choice does not matter as long as $kmax > r$. But, for small samples, the estimate of *r* could be sensitive to the choice of *kmax*.

To see how the PC estimators are related to the ER (eigenvalue ratio) estimator, observe that

$$ER(k) \equiv \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} = \frac{V(k-1,\tilde{F}^{k-1}) - V(k,\tilde{F}^{k})}{V(k,\tilde{F}^{k}) - V(k+1,\tilde{F}^{k+1})}. \tag{9}$$

Thus, the ER estimator can be viewed as the value of *k* that maximizes the ratio of the changes in the sum of squared residuals at *k*-1 and *k*.

As we formally present in the later section, another way to estimate the number of factors is to maximize the ratio of the growth rates of $V(k,\tilde{F}^{k})$ at *k*-1 and *k*. Specifically, we define the following criterion function:

$$GR(k) = \frac{\ln[V(k-1,\tilde{F}^{k-1})/V(k,\tilde{F}^{k})]}{\ln[V(k,\tilde{F}^{k})/V(k+1,\tilde{F}^{k+1})]} = \frac{\ln(\tilde{\mu}^{*}_{NT,k})}{\ln(\tilde{\mu}^{*}_{NT,k+1})}, \tag{10}$$

where $k \le kmax$. Figure 6, which is obtain from the data generated by model (3) with *i.i.d.* errors, shows that $GR(k)$ achieves a maximum when *k* = 5 = *r*. Similarly to the case of eigenvalue ratios, the $GR(k)$ function is almost symmetric around the true number of factors. We call the value of *k* maximizing $GR(k)$ "growth ratio" estimator, $\tilde{k}_{GR}$.

It is important to note that the growth ratio, as well as the eigenvalue ratio, may not be singly peaked. For example, we found that when factors have different signal-to-noise ratios, the ratios are likely to have multiple peaks. Figure 7 shows the locus of eigenvalue ratios for the case where data are generated from model (3), but with two factors being generated from *N*(0,2) and the other three factors being generated from *N*(0,0.33). The figure also demonstrates that the first peak point is not necessarily the maximum point.

In the next section we formalize the two new criteria presented and we also add a third one that is consistent under more restrictive assumptions.

3. Assumptions and Asymptotic Results

In this section, we denote a norm of a matrix A as $\|A\| = [trace(AA')]^{1/2}$. We also use c_1 and c_2 to denote generic positive constants. For any real number *z*, [*z*] denotes the integer part of *z*.

The followings are the assumptions we made.

Assumption A: Let $\mu_{NT,1} \geq \mu_{NT,2} \geq .. \geq \mu_{NT,r} > 0$ be the eigenvalues of $(\Lambda^{o\prime}\Lambda^o / N)(F'F/T)$. Then, $p\lim_{m\to\infty} \mu_{NT,j} = \mu_j$, where $m = \min(N,T)$, and $0 < \mu_j < \infty$.

Assumption B: $E\left\|F_t^o\right\|^4 < c_1$ and $\left\|\lambda_i^o\right\| < c_1$ for all *i* and *t*.

Assumption C: $E\left(\left\|N^{-1/2}\Sigma_i \varepsilon_{it}\lambda_i^o\right\|^2\right) < c_1$ for all *t*, and $E\left(N^{-1}\Sigma_{i=1}^N \left\|T^{-1/2}\Sigma_{t=1}^T \varepsilon_{it} f_t\right\|^2\right)$ =

$E\left(N^{-1}T^{-1}\left\|\mathrm{E}'F^o\right\|^2\right)$ < c_1.

Assumption D: Let $\mathrm{E} = [\varepsilon_{it}]_{N\times T} = R_T^{1/2} U G_N^{1/2}$, where $U' = [u_{it}]_{N\times T}$, R_T and G_N are $T\times T$ and $N\times N$ matrices, respectively, and $R_T^{1/2}$ and $G_N^{1/2}$ are their symmetric square roots. (i) The u_{it} are independently and identically distributed random variables with uniformly bounded moments up to the 4th order, and (ii) $\lambda_1(R_T) < c_1$ and $\lambda_1(G_N) < c_1$, uniformly in *T* and *N*, respectively.

Assumption E: Let $y = \lim_{m\to\infty} m/M$, where $M = \max(N,T)$. Then, $0 < y \leq 1$.

Assumption F: There exists a real number $d^* \in (0,1]$ such that $p\lim_{m\to\infty} \lambda_{[d^* m]}\left(\mathrm{EE}'/M\right) > 0$.

Assumption A assumes that the common factors are stationary. The methods of BN and Onatski can allow both non-stationary and stationary factors. So, Assumption A is somewhat restrictive. However, when some factors are suspected to be nonstationary, we can use the difference data to estimated number of factors (see Bai and Ng, 2004).

Assumption A could be relaxed. What we need is the assumption that the *r* eigenvalues $\mu_{NT,1} \geq ... \geq \mu_{NT,r}$ have the convergence rates so that $\mu_{NT,j-1}/\mu_{NT,j}$ = $O_p(m^\phi)$ for j = $2,...,r$ where $0 \leq \phi < 1$. This condition would be satisfied if all common factors are I(1), or if $\Lambda^{o\prime}\Lambda^o / N^\phi$ converges to a finite positive definite matrix where $0 < \phi \leq 1$. The latter condition is compatible to the assumption in Onatski that $\Lambda^{o\prime}\Lambda^o \to \infty$ as $N \to \infty$. Because the Onatski method does not require $\Lambda^{o\prime}\Lambda^o / N$ to converge to a finite matrix, it would be able to capture the weak factors whose explanatory power does not proportionally increase with *N*.

As an alternative to Assumption E(i), we may assume that the u_{it} are independently (but not necessarily identically) distributed and their moments are uniformly bounded up to the 7th order. (See Bai and Ng, 2006).

Assumption C allows limited dependence between the factors and the idiosyncratic errors. Assumption D allows the idiosyncratic errors to be weakly correlated over time and individual units. Thus, the assumption is consistent with the approximate factor model of Chamberlain and Rothschild (1983). Onatski (2006, hereafter, ON) assumes that either R_T or G_N are identity matrices. So, Assumption D is more general than his. Assumption E rules out the cases in which N/T or T/N diverge to infinity. However, most of the panel data with both large N and T would satisfy the assumption. Assumption E is also made in Bai and Ng (2006) and Onatski (2006).

Assumption F is the one that neither BN nor ON use. However, the assumption should hold in the general cases in which the factors fail to explain any of the response variables perfectly or near perfectly. To see why, suppose that $R_T = I_T$. Then, $G_N = \text{var}(\varepsilon_{\bullet t})$. If the common factors f_t do not perfectly explain none of the response variables, we can expect that the smallest eigenvalue of G_N would be bounded away from zero. For such cases, we can show that Assumption F holds. Stated formally:

Proposition 1: Suppose that $\lambda_N(G_N) > c_2$, and $\lambda_T(R_T) > c_2$, for any N and T. Then, if $y = 1$, Assumption F holds for any $d^* \in (0,1)$. If $y < 1$, the assumption holds for any $d^* \in (0,1]$

The proof of Proposition 1 is given in the appendix. In the proposition, we assume that both the smallest eigenvalues of G_N and R_T are bounded away from zero. We could obtain the same result even if the smallest eigenvalues converges to zero. The result under a relaxed assumption is available from the authors upon request.

In Onatski, *kmax/N* $\to$ 0 as $N \to \infty$. BN also choose a fixed *kmax*. As we show below, when Assumption F holds, any integer between $r + 1$ and $[d^*m]$ can be used as *kmax*. Note that $[d^*m]$ increases proportionally with $m = \min(T, N)$.

We now state our major finding:

Proposition 2: Let $kmax = [d^* m] - 2r - 1$. Then, under Assumptions A – F and $r \geq 1$,

$$\lim_{N,T\to\infty} \Pr[\tilde{k}_{ER} = r] = \lim_{N,T\to\infty} \Pr[\tilde{k}_{GR} = r] = 1.$$

In general, $[d^* m]$ is unknown. However, $[d^* m]$ increases with the sample size. Any fixed integer greater than r could serve as *kmax*. In addition, because of this property, we could expect that the ER and GR estimators would not be overly sensitive to the choice of *kmax*.

Notice that Proposition 2 holds for $r \geq 1$. For the cases in which the possibility of zero factors cannot be ruled out, one would wish to estimate the number of factors allowing $r = 0$. For such cases, we may define an artificial eigenvalue $\tilde{\mu}_{NT,o} = w(N,T)$ such that $w(N,T) \to 0$ and $w(N,T)m \to \infty$ as $N,T \to \infty$. Then, we obtain the following result:

Proposition 3: Redefine $\tilde{k}_{ER}$ and $\tilde{k}_{GR}$ including $\tilde{\mu}_{NT,o} / \tilde{\mu}_{NT,1}$. Then, under Assumptions A – F and $r \geq 0$, $\lim_{m\to\infty} \Pr[\tilde{k}_{ER} = r] = \lim_{m\to\infty} \Pr[\tilde{k}_{GR} = r] = 1$.

Finally, we consider another estimator that does not require Assumption F. In particular, one may wish to have an estimator that does not require *kmax*. In fact, there are many possible such estimators. The one that we introduce here is what we call a "*log-ratio*" estimator:

$$\tilde{k}_{LR} = aug \max_{k \leq m} \frac{\ln\left(1 + m^{-1} + \tilde{\mu}_{NT,k}\right)}{\ln\left(1 + m^{-1} + \tilde{\mu}_{NT,k+1}\right)}.$$

Proposition 4: Under Assumptions A – E and $r \geq 1$, $\lim_{m\to\infty} \Pr[\tilde{k}_{LR} = r] = 1$.

While $\tilde{k}_{LR}$ is attractive in that it does not require to specify *kmax*, its pitfall is that any estimator replacing m^{-1} by a multiple of m^{-1} (say, am^{-1}) is also consistent. The finite sample property of $\tilde{k}_{LR}$ would crucially depend on the choice of a. We leave the optimal choice of the multiple to future study.

4. Simulation Results

4.1. Simulation Setup

The foundation of our simulation exercises is the following factor model

$$x_{it} = \Sigma_{j=1}^{r} \lambda_{ij} f_{jt} + \sqrt{\theta} u_{it}; u_{it} = \sqrt{\frac{1-\rho^2}{1+2J\beta^2}} e_{it}, \tag{11}$$

where $e_{it} = \rho e_{i,t-1} + v_{it} + \Sigma_{h=\max(i-J,1)}^{i-1} \beta v_{ht} + \Sigma_{h=i+1}^{\min(i+J,N)} \beta v_{ht}$, and the v_{it} ($i \geq 1$) are drawn from $N(0,1)$. This data generating process has been used by Bai and Ng (2002) and Onatski (2005), except that the error terms u_{it} are normalized so that their variances are equal to one for most of the cross-section units (more, specifically, $J+1 \leq i \leq N-J$). We use this normalization to control for the ratios of factor and error variances.

The factors and factor loadings f_{jt} and λ_{ij} are also drawn from $N(0,1)$. The true factors (*F*), the factor loadings (λ) and the idiosyncratic errors (v) are generated as standard-normal and *i.i.d.* random variables. The control parameter θ is the inverse of the signal to noise ratio (SNR) of each factor, that is, $1/\theta = \text{var}(f_{jt}) / \text{var}(\sqrt{\theta} u_{it})$. The magnitude of the time-series correlation is specified by the control parameter ρ. The cross-sectional correlation is governed by two parameters: β specifies the magnitude of the cross sectional correlation and *J*, the number of cross-section units correlated. For example, $J = 8$ means that for $J+1 \leq i \leq N-J$, each cross-section unit is correlated with the 16 (= 2*J*) adjacent cross-section units.

Before we present our Monte Carlo experiment results, we briefly discuss the importance of the signal to noise ratio (SNR) of each factor in the finite-sample properties of the estimators. When *N* and *T* are not sufficiently large, the accuracy of the estimators would be affected by SNR of each factor. As an example, Figure 8 shows the average value of the eigenvalues from 1,000 simulated data sets from model (11) with *i.i.d.* errors. We can clearly see that the "scree" pattern becomes more obscure as SNR decreases. Figure 9 shows the average value of $\tilde{k}_{ER}$ for the data used for Figure 8. It shows that the maximum value of the eigenvalue ratios increases monotonically as SNR increases. Any method for estimating the number of factors would underestimate the true number of factors when the SNRs are low as shown in the weak factor section.

4.2 Basic Results

We compare the ER and GR estimators with the ones recommended by BN ($IC1$, $IC2$, $PC1$, $PC2$), and the one proposed by Onatski (2006). These estimators are given

$$\tilde{k}_{IC1} = \arg\min_{k \le kmax}\left[\ln\left(V(k,\tilde{F}^k)\right) + k\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)\right];$$

$$\tilde{k}_{IC2} = \arg\min_{k \le kmax}\left[\ln\left(V(k,\tilde{F}^k)\right) + k\left(\frac{N+T}{NT}\right)\ln(m)\right];$$

$$\tilde{k}_{PC1} = \arg\min_{k \le kmax}\left[V(k,\tilde{F}^k) + \hat{\sigma}^2 k\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)\right];$$

$$\tilde{k}_{PC2} = \arg\min_{k}\left[V(k,\tilde{F}^k) + \hat{\sigma}^2 k\left(\frac{N+T}{NT}\right)\ln(m)\right];$$

$$\tilde{k}_{ON} = \arg\max_{k \le kmax}\{k \mid \tilde{\mu}_{NT,k} > (1+\delta)\hat{u}/N\},$$

where $\hat{\sigma}^2 = V(kmax, \tilde{F}^{kmax})$. We use $kmax$ = 8. As in Onatski (2006), we choose $\delta = N^{-1/3}$. For the formal definition of $\hat{u}$, see Onatski (2006). For the comparison with the results from BN and Onatski, we first investigate the estimation results obtained ruling out the possibility of $r = 0$. Later, we consider the performances of the tests allowing the possibility of $r = 0$.

Before reporting our simulation results, we want to stress an attractive finite-sample property of the ER and GR estimators, which we do not report here to save space. As we conjecture in section 3, the estimators are robust to the selection of the maximum number (*kmax*) of factors to test. For example, in samples with $N = T = 100$, we obtain the same numerical results whether we use *kmax = 8* or *kmax = 90*, except for a few cases. This is the property that the BN and Onatski estimators do not share when N or T is small. For example, we found from unreported results that these estimators are sensitive to the choice of $kmax \in \{6,7,8,9\}$ when $r = 3$. This sensitivity is not surprising because for the estimators, the threshold values directly depend on the chosen *kmax*, except for the $\tilde{k}_{IC}$'s.

Table 1 shows the case where the idiosyncratic errors are *i.i.d.* ($\beta = \rho = J = 0$). The table is the benchmark table reported in both BN and Onatski (2006). In Table 1, most of the estimators perform well. Notice that for the setup of the simulations used for the table, SNR of each factor decreases with r. The GR estimator slightly performs better than the ER estimator when N is relatively small ($N \leq 20$). The BN estimators overestimate the true number of factors when N = 10. For the cases with $N \geq 40$, the ER, GR and Onatski estimators outperform the BN estimators in a marginal scale. The former three estimators perform equally well, although the GR estimator tends to produce more reliable results when SNR is low.

Tables 2.1-2.3 report the results obtained from the data with autocorrelated errors, but without cross-section correlation ($\beta = J = 0, SNR = 1$). In Table 2.1, the Onatski (ON), ER and GR estimators clearly outperform the IC and PC estimators. For the cases with $\rho = 0.5$, the IC estimators perform well with $N, T \geq 50$, the PC estimators require larger sample sizes for more accurate estimation results. As ρ increases to 0.7 or 0.9, the performances of the IC estimators get worse. In particular, Tables 5.2 shows that the performances of the IC and PC estimators improve as T increases, but not necessarily as N increases. Data with more than 100 time series observations are needed for the IC and PC estimators to obtain accurate estimation results. With $T \leq 100$, both IC and PC estimators perform worse as N increases. In contrast, the ER and GR estimators perform well even if N or T is small and the degree of autocorrelation is high. The performance of the ON estimator is compatible with those of the ER and GR estimators for cases with $\rho \leq 0.7$, but we can observe that the ON estimator does not perform well unless both N and T are sufficiently large.

Table 3.1 reports the finite-sample performances of the estimators when the errors are cross correlated with no autocorrelation ($\rho = 0$). Once again, the ON, ER and GR estimators perform better than the IC and PC estimators for most of the cases considered. When $\beta = 0.2$ and $J = 8$, the ER and GR estimators pick up the correct number of factors quite accurately even for the cases with relatively small N and T ($50 \leq N, T \leq 100$). The performance of the ON estimator becomes compatible to those of the two estimators when $N, T \geq 100$.

The number of cross section units should be substantially large for the IC and PC estimators to produce more reliable statistic inferences. In unreported experiments, we found that the IC and PC estimators perform as well as the ER and GR estimators when $\beta \leq 0.05$. In general, the IC and PC estimators tend to overestimate the correct factor numbers unless the cross correlations are sufficiently weak. In addition, the larger number of time series observations does not necessarily improve the performances of the two estimators: for example, see the cases with $(N,T) = (1000,50)$ and $(N,T) = (1000,500)$.

For the cases with $J = 20$, the ER and GR estimators pick up the correct number of factors quite accurately for the cases with $N \geq 200$ and $T \geq 100$. For the data with $N \leq 200$ and $T \leq 100$, the ER and GR estimators dominate the ON estimators. The IC and PC estimators tend to overestimate the number of factors even if N is substantially large.

Tables 3.2-3.3 compare the performances of the seven estimators when the degree of cross correlation is higher. The ER and GR estimators still perform well as long as $N \geq 200$ and $T \geq 100$. The ON estimator dominates the IC and PC estimators perform for the data with any size.

Table 4.1 summarizes the results from the data with both cross-correlated and autocorrelated errors. The true number of factors is three, and the auto- and cross-correlation parameters are fixed at $\rho = 0.5$, $\beta = 0.2$ and $J = 8$. High SNR ($\theta = 1/3$, and SNR = 3) is used to generate factors. The ER and GR estimators perform well and clearly dominate the other estimators even if $N,T \leq 50$. The ON estimator also performs reasonably well for the cases with $N,T \geq 100$.

Tables 4.2-4.3 are designed to investigate how sensitive the seven estimators are to SNR of factors. Clearly, all estimators perform worse for the data with low-SNR factors. However, the ER and GR estimators dominate all other estimators for the data with $N,T \leq 100$. For the data with $N,T \geq 100$, the ON estimators perform as well as the ER and GR estimators. This is an interesting result, because the ON estimator is motivated under the assumption of only one-side correlation among errors (auto- or cross-correlations). As Onatski found from his Monte Carlo experiments, his estimator performs reasonably well even if both correlations exist.

Tables 5.1-5.2 consider the cases where factors have different SNRs. We fix the sum of the SNRs of individual factors constant. The data used for Table 5.1 are generated from a three-factor model. Two factors with high SNRs are generated from $N(0,1.5)$ and the other factor from $N(0,0.75)$. The data used for Table 8.2 are generated from a five-factor model with two factors of high SNRs from $N(0,1.2)$ and three factors from $N(0,0.70)$. In Table 5.1, the ER and GR estimators outperform the rest of the estimators. In Table 5.2, for the cases with $N \geq 100$ and $T \geq 50$, the performance of the ON estimator is compatible to those of the ER and GR estimators.

For most of other cases with $N,T \leq 100$, however, the ER and GR estimator dominates the ON estimator.

Table 6 considers the estimation of the number factors when data is generated with no common factors. We only consider the ER estimator obtained following Proposition 3. We denote the estimator by ER* in Table 6. We do not consider the BN estimators because they are designed only for the cases where $r \geq 1$. The ER* and ON criteria performs similarly except in the cases of (N,T) = (100,100) and (N,T) = (1000,60) where the ER* criterion outperforms the ON criterion. In summary, the reported simulation results provide strong evidence that the ER, GR and ER* estimators have better finite-sample properties that the IC, PC and ON estimators. The performances of the ER and GR estimators are especially robust to the degrees of autocorrelation and cross-correlation in the absence of weak factors.

4.3. Weak Factors

The Onatski estimator was designed to capture weak factors. More specifically, his estimator allows factors with $\Sigma_{j=1}^{r}\lambda_j\left(F\Lambda^{o\prime}\Lambda^{o}F'/(NT)\right) < O_p(1)$. In this subsection, we compare the power of the ON, ER and GR estimators to detect weak factors.

ER and GR may fail to capture weak factors in finite samples when the ratios of the SNRs of the weak and strong factors differ substantially. To see this possibility, we first generated 100 different samples, in each of which (N,T) = (1,000,1,000). For each sample, the response variables generated by 4 common factors of 4 different SNRs (1, .5, .25 and .125) and the idiosyncratic errors generated with $\beta = 0.5$, $\rho = 0.5$ and $J = 20$. From each of the samples, we draw subsamples of (N,T) = (110,110), (120,120), and so on, up to (1,000,1,000), subsequently. Then, for each subsample, we calculate the ER, GR, ON estimators and regress the true factors on the first four principal component estimators of factors. The motivation of the latter exercise is to see if the four principal component estimators are consistent estimators of the true factors.

Figure 10 shows the mean results for the ER(k) values over 100 simulations. As we can observe, the only value of ER(k) that grows unbounded with N and T is the one corresponding to the fourth factor (ER4). From the figure, we can observe that when the sample is not large enough, ER4 may not be larger than the rest of the ratios resulting in underestimation.

Figure 11 shows the averages of the estimated number of factors by individual estimators at different sample sizes.[4] This pictures shows that ON overestimates the number of factors in small samples. But in general, ON is able to correctly capture the weak factors in smaller samples than ER and GR. It also shows that the underestimation problem in the presence of weak factors is less severe for GR than for ER.

Figure 12 shows the patterns of the R^2's from the regressions of the actual factors on estimated factors. The first and second factors are highly correlated with actual first and second factors. However, the estimated fourth factor starts to explain 50% or more of the actual factor when $N, T \geq 400$. If we compare figures 10 and 11, we can observe that ER(4) is above the rest of ratios and picks the 4th factor at a sample size of $N = T = 730$, where the estimated 4th factor explains slightly more than 90% of the actual 4th factor. An interesting observation here is that when the ER estimator underestimates the true number of factors, the higher-order principal component factors only have low correlations with the actual factors not captured by the estimator.

To be more specific, we conduct an additional simulation exercise. We generate three factors from N(0,1), N(0,.5), and N(0,SNR), where SNR equals .1, .2, .35 and .45. 1,000 different data are generated. Each simulation uses N = 1,000, T = 60, J = 50 (5% of the idiosyncratic errors are cross-correlated), β = .2 and ρ = .5. The results are reported in Table 7.

When the third factor's SNR is 0.1, the ER and GR estimators fail to capture it while the ON estimator captures the third factor 47.4% of the time. Thus, as we expected, the ON estimator appears to have better power to detect weak factors. However, as shown in the last row of Table 7, only for 26% of the time the third principal component factor is the one most correlated with the real third factor. In fact, for 44% of the times the principal component factors most correlated with the real third factor corresponds to the one corresponding to the 5th eigenvalue, while for 28% of the time the 4th principal component factor is the most highly correlated with the actual third factor. This means that even if the number of factors is correctly estimated, the chances that the 3rd factor is correctly estimated are only 26%. In contrast, the first and second principal component factors always correspond to the first and second strong factors.

As the SNR of the third factor increases, ER and GR start picking the number of factors correctly. In addition, the chance that the third principal component factor is the one most

[4] The results by the BN estimators are not shown because they always choose *kmax* in the simulation setup.

correlated with the actual third factor also increases. When the SNR of the third factor is .45, the GR estimator picks the third factor 95% of the time and the first three principal component factors explain the three actual factors 100% of the time. In contrast, for the same case, the ON estimator criteria picks up three factors only 65% of the time and the other times overestimate the number of factors.

Figures 10 and 11 and Table 7 indicate that the ER and GR estimators would miss weak factors, although the latter estimator has better power to capture them. In our simulations, the ER or GR estimators fail to capture the factors with SNR less than 0.2 ($\mathrm{var}(f)/\mathrm{var}(\varepsilon) < 0.2$). So, we define the factors with SNR less than 0.2 as weak factors. The ON estimator has better power to detect weak factors. However, our simulation results provide some practical guidance. When the estimated number of factors by ER or GR is smaller than the estimate by ON, it would indicate presence of weak factors. But when the ER or GR estimators fail to capture the weak factor, the principal component factors of higher orders may not be correct estimates of actual factors. For example, when the ER and GR estimators fail to capture the third factor, the third principal component factor may not be a consistent estimator of the factor. In contrast, the principal components corresponding to ER or GR appear to be reliable estimators of the actual factors in most of the cases we have considered for simulations.

5. Application

5.1 Factor-Augmented VAR (FAVAR) application

The use of factors extracted from a large panel of data as explanatory variables in a VAR has been proved very useful in several recent works (for a comprehensive analysis and review of the literature see Bernanke, Boivin and Eliaz (2005), from now on BBE). In this section we estimate the number of common factors in the panel data used by BBE (2005) and investigate the possible consequences of using the incorrect number of factors in the FAVAR-based forecasting.

The data set we use here is retrieved from Jean Boivin's webpage. It contains 120 monthly macroeconomic variables with 511 time series observations (February 1959-August 2001). The data set has been transformed to be stationary (for detailed information about the data see appendix 1 of BBE). In order to remove time effects from data, we use time-demeaned variables

Table 8 shows the result of the estimated numbers of factors by different estimators. When we use the maximum number of factors of 15 (*kmax* = 15), the ER and GR estimators find 5 factors while the IC1 estimator finds 15 and the ON estimator finds 13.

To see how robust each estimator is to the choice of *kmax*, we estimate the number of factors with many different values of *kmax*. Among the BN estimators, we consider only the IC1 estimator. Figure 13 shows that ER and GR find 5 factors using any value of *kmax* from 7 to 30. In contrast, the ON estimator seems to be extremely sensitive to the choice of *kmax*. As *kmax* increases, so does the estimated number of factors. The IC1 estimator shows the same pattern.

We now analyze how the number of factors used in VAR influences the forecasting ability of the FAVAR model. BBE analyzed the effect on impulse-response functions of adding factors to a VAR model. They have shown that using either one or five factors lead to similar results. Stock and Watson (2002a, 2002b) also have shown the benefits of using factors calculated through principal components in macroeconomic forecasting.
In this subsection, we do a simple exercise to investigate the effects of using 1, 3, 5, 7, 10 or 13 factors in a FAVAR-based forecasting. We choose to forecast industrial production (IP), CPI (PUNEW in the BBE dataset). We first do 12 month-ahead forecasts for January 1992 using the data up to January 1991. Then, we make 12 month-ahead forecasts for February 1992 using the data up to Feb.1991. We continue this exercise to obtain 120 forecasts for each of IP and CPI. The federal funds rate, IP and CPI are used in every forecast. We add to those variables the factors extracted from the database in the exact same way as BBE did. In each forecast we construct a FAVAR with 7 lags. BBE used 13 lags but they claimed that the results were similar with 7 lags. We choose to use 7 lags because we add up to 13 factors to the VAR.

Figure 14 shows the correlations between the 120 forecasted values and actual data values. We can clearly see that the most correlated series are the ones forecasted using a FAVAR model with 5 factors, which are suggested by the ER and GR estimators. Figure 15 shows the mean square errors (MSE) of the forecasted variables. The forecasted values from the FAVAR models with either 1 or 5 factors have similar MSE. However, MSE increases monotonically after the 5^{th} factor, suggesting that using more factors in a FAVAR model than what is suggested by ER and GR can reduce the accuracy of the forecasts.

5.2 Application to the Stock Market

In the finance literature, factor models have been widely used to estimate the common factors which drive stock returns. In this section, we use estimate the number of latent common factors in the U.S. stock market. The literature of estimating the number of latent factors in the US stock market is extremely rich and the number of factors proposed ranges from one to five or sometimes even more[5].

We used excess returns (including dividends) over the risk free rate on individual stocks downloaded from CRSP. The risk free rate is the one-month Treasury bill rate downloaded from Kenneth French webpage. We exclude REITs (Real Estate Investment Trusts), ADRs (American Depositary Receipts) and the stocks that do not have information for every month during a period. Every dataset contains monthly returns. We also exclude stocks that show more than a 300% excess return in a given month since we are trying to capture common variation and that is most probably due to idiosyncratic risk.

The time span included in the analysis is 1970-2006. We have divided the entire sample into two subsamples (1970-1987 and 1988-2006), three subsamples (1970-78, 1979-92 and 1993-06) and also into seven subsamples (1970-74, 1975-79, 1980-84, 1985-89, 1990-94, 1995-99 and 2000-06). We do so to allow for longer time series (at the expense of cross-sectional observations) and also for a larger number of stocks to be chosen.

Table 9 presents the results obtained from the data. We show the results obtained from the time-demeaned data. Brown (1989) found that when factor loadings have nonzero means, the scree test often predicts one factor in the US stock return data. Demeaning returns eliminates the nonzero means of factor loadings.

An intriguing result from Table 9 is that both the ER and GR estimators capture one common factor in every dataset except for the period 1985-89. The other estimators (ON, PC and IC) capture from one to five factors depending the sample used. However, one factor is the most frequent estimate. Given our simulation results, we can conclude that there is strong evidence for one common factor.

6. Conclusion

In this paper we developed three new estimators, ER, GR and LR, for estimating the number of common factors in approximate factor models. None of them do not require prespecified penalty

[5] See for example Shukla and Trzcinka (1990), Chan et al. (1998) and Ferson and Korajczyk (1995).

functions or estimated threshold values. The ER and GR estimators are designed to be flexible in choosing the maximum number of factors to be tested. The LR estimator does not require the choice of the maximum number. However, we recommend use of ER or GR, because the LR estimator can have infinitely many possible alternatives. In addition, in unreported simulation experiments, the finite-sample performances of the three estimators were compatible, although GR appears to have better power to detect weak factors. We have shown that the three estimators produce consistent estimators under the same assumptions of Bai and Ng (2002). The new estimators only use the eigenvalues of sample covariance matrices of response variables. They can be viewed as a formalization of Cattell's scree test. We also have linked the new estimators to the Bai-Ng and Onatski estimators. Our simulation results indicate that the new estimators outperform the Bai-Ng and Onatski estimators, especially when (i) both N and T are small, or (ii) the signal-to-noise-ratios of individual factors are equally low, or (iii) the idiosyncratic errors are both cross-sectionally and serially correlated. We also have shown how to test for zero factor. While the Onatski estimator sometimes has better power to capture weak factors, the principal component factors corresponding to the eigenvalues suggested by the estimator may not provide accurate factor estimates. While the ER or GR estimator may fail to capture weak factors, the principal component factors suggested by them are the accurate factor estimates in most cases.

Appendix

The following lemmas are useful to prove Proposition 1:

Lemma 1: Under Assumptions D and E,

$$p\lim_{m\to\infty} \lambda_1\left(\frac{1}{M}UU'\right) = \left(1+\sqrt{y}\right)^2; \; p\lim_{m\to\infty} \lambda_{\mathrm{m}}\left(\frac{1}{M}UU'\right) = \left(1-\sqrt{y}\right)^2.$$

Proof: See Bai and Yin (1983) or Bai and Silverstein (1999)

Corollary 1: If $y=1$, $\lambda_{\mathrm{m}}\left(UU'/(NT)\right) = o_p\left(m^{-1}\right)$. If $y<1$, $\lambda_m\left(UU'/(NT)\right) = O_p\left(m^{-1}\right)$.

Lemma 2: A Sturmian Separation Thoerem

Let W_n be an $n\times n$ symmetric matrix and W_{n-1} be an $(n-1)\times(n-1)$ major submatrix of W_n.

Then, $\lambda_{k+1}(W_n) \le \lambda_k(W_{n-1}) \le \lambda_k(W_n)$, for all $k=1,...,n-1$

Proof: See p. 64 of Rao (1973).

Lemma 3: Let W_n be an $n\times n$ symmetric matrix and W_{n-p} be an $(n-p)\times(n-p)$ major submatrix of W_n. Then, $\lambda_{n-p}(W_n) \ge \lambda_{n-p}(W_{n-p})$.

Proof: By Lemma 2, $\lambda_{n-p}(W_{n-p+2}) \ge \lambda_{n-p}(W_{n-p+1}) \ge \lambda_{n-p}(W_{n-p})$.

Lemma 4: Suppose that two matrices A and B are $p\times p$ positive definite and positive semidefinite matrices, respectively. Then, for any $j+k-1\le i$, $\lambda_{p-j+1}(A)\lambda_{p-k+1}(B) \le \lambda_{p-i+1}(AB)$.

Proof: Theorem 2.2 of Anderson and Dasgupta (1963).

Lemma 5: Under Assumption D and E,

$$p\lim_{m\to\infty} \lambda_1\left(\frac{1}{M}\mathrm{EE}'\right) < \infty.$$

Proof: By Lemma 4,

$$\begin{aligned}\lambda_1(\mathrm{EE}'/M) &= \lambda_1(UG_N U'R_T/M) \\ &\le \lambda_1(UG_N U'/M)\lambda_1(R_T) \le \lambda_1(UU'/M)\lambda_1(G_N)\lambda_1(R_T).\end{aligned}$$

Thus, the result is obtained by Lemma 1 and Assumption D(ii).

Corollary 2: Under Assumptions D – E,

$$\lambda_1\left(\frac{1}{TN}\mathrm{EE}'\right) = O_p\left(\frac{1}{m}\right).$$

***Proof of Proposition* 1:** Suppose that $y=1$. For simplicity, assume that $T=N$. Choose any $d^* \in (0,1)$.

$$\lambda_{[d^*m]}\left(\frac{1}{M}R_T^{1/2}UG_NU'R_T^{1/2}\right) = \lambda_{[d^*m]}\left(\frac{1}{M}UG_NU'R_T\right) \geq \lambda_{[d^*m]}\left(\frac{1}{M}UG_NU'\right)\lambda_T(R_T)$$
$$\geq \lambda_{[d^*m]}\left(\frac{1}{M}UU'\right)\lambda_N(G_N)\lambda_T(R_T)$$

where the inequalities are due to Lemma 4. Let $U_{[d^*m]}$ be the $[d^*m]\times N$ main submatrix U. Then, by Lemmas 3 and 1, we have

$$p\lim_{m\to\infty}\lambda_{[d^*m]}\left(\frac{1}{M}UU'\right) \geq p\lim_{m\to\infty}\lambda_{[d^*m]}\left(\frac{1}{M}U_{[d^*m]}U'_{[d^*m]}\right) = \left(1-\sqrt{d^*}\right)^2.$$

Thus, $p\lim_{m\to\infty}\lambda_{[d^*m]}\left(R_T^{1/2}UG_NU'R_T^{1/2}/M\right) > 0$. Similarly, we can show that the limit is positive for any $d^* \in (0,1]$, if $y<1$.

The following lemmas are useful to prove Proposition 2.

Lemma 6: Suppose that two matrices *A* and *B* are symmetric of order *p*. Then,

$$\lambda_{j+k-1}(A+B) \leq \lambda_j(A) + \lambda_k(B),\ \ j+k \leq p+1.$$

Proof: See Onaski (2006) or Rao (1973, p. 68).

Lemma 7: Suppose that two matrices *A* and *B* are positive semidefinite of order *p*. Then,

$$\lambda_j(A) \leq \lambda_j(A+B),\ \ j=1,\ldots,p.$$

Proof: First, consider the case of $j=1$. Let ξ_A^1 be the eigenvector corresponding to $\lambda_1(A)$. Then,

$$\lambda_1(A) = \frac{\xi_A^{1\prime} A \xi_A^1}{\xi_A^{1\prime}\xi_A^1} \le \frac{\xi_A^{1\prime}(A+B)\xi_A^1}{\xi_A^{1\prime}\xi_A^1} \le \lambda_1(A+B),$$

where the first inequality is due to B being positive semidefinite. We now consider the cases with $j > 1$. Let Ξ^{j-1} be the matrix of the orthonormal eigenvectors corresponding to the first $(j-1)$ largest eigenvalues of $A+B$. Let z be a $p\times1$ nonzero vector. Then,

$$\lambda_j(A) \le \sup_{\Xi^{j-1\prime}z=0} \frac{z'Az}{z'z} \le \sup_{\Xi^{j-1\prime}z=0} \frac{z'(A+B)z}{z'z} = \lambda_j(A+B),$$

where the first inequality comes from Rao (p. 62).

Lemma 8: Let $Q(\Lambda^o) = I_N - \Lambda^o(\Lambda^{o\prime}\Lambda^o)^{-1}\Lambda^{o\prime}$. Then, under Assumptions D – F,

$$0 < p\lim_{m\to\infty} \lambda_j\left(\frac{1}{M}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) < \infty, \text{ for } j = 1,...,[d^*m]-r.$$

Proof: Let $P(\Lambda^o) = I_N - Q(\Lambda^o)$. Then,

$$\frac{1}{NT}\mathrm{EE}' = \frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}' + \frac{1}{NT}\mathrm{E}P(\Lambda^o)\mathrm{E}'.$$

Thus, by Lemma 6,

$$\lambda_{j+r}\left(\frac{1}{NT}\mathrm{EE}'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) + \lambda_{r+1}\left(\frac{1}{NT}\mathrm{E}P(\Lambda^o)\mathrm{E}'\right) = \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right),$$

since $rank\left(\mathrm{E}P(\Lambda^o)\mathrm{E}'\right) = r$. By Lemma 7,

$$\lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}' + \frac{1}{NT}\mathrm{E}P(\Lambda^o)\mathrm{E}'\right) = \lambda_j\left(\frac{1}{NT}\mathrm{EE}'\right).$$

Thus we have

$$\lambda_{j+r}\left(\frac{1}{NT}\mathrm{EE}'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{EE}'\right), \; j = 1,...,T-r. \tag{A.1}$$

By Lemma 5 and Assumption F,

$$0 < p\lim_{m\to\infty} \lambda_j\left(\frac{1}{M}\mathrm{EE}'\right) < \infty, \tag{A.2}$$

for $j = 1,...,[d^*m]$. Thus, by (A.1) and (A.2),

$$0 < p\lim_{m\to\infty} \lambda_{j+r}\left(\frac{1}{M}\mathrm{EE}'\right) \le p\lim_{m\to\infty} \lambda_j\left(\frac{1}{M}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) \le p\lim_{m\to\infty} \lambda_j\left(\frac{1}{M}\mathrm{EE}'\right) < \infty,$$

for all $j = 1, ..., [d^* m] - r$.

Lemma 9: Under Assumptions A – E,

$$\tilde{\mu}_{NT,r+j} = \lambda_{r+j}\left(\frac{1}{NT} XX'\right) = O_p\left(\frac{1}{m}\right), \text{ for } j = 1, ..., m-r.$$

Let *kmax* = $[d^* m] - 2r - 1$. Then,

$$\frac{\tilde{\mu}_{NT,r+j}}{\tilde{\mu}_{NT,r+j+1}} = O_p(1), \text{ for } j = 1, ..., kmax.$$

Proof: Define $F^* = F + \mathrm{E}\Lambda^o(\Lambda^{o\prime}\Lambda^o)^{-1}$. Then, it is straightforward to show

$$\begin{aligned}\frac{1}{NT^2} XX' &= \frac{1}{NT^2} F\Lambda^{o\prime}\Lambda^o F' + \frac{1}{NT^2} F\Lambda^{o\prime}\mathrm{E}' + \frac{1}{NT^2}\mathrm{E}\Lambda^o F' + \frac{1}{NT^2}\mathrm{E}\mathrm{E}' \\ &= \frac{1}{NT^2} F^*\Lambda^{o\prime}\Lambda^o F^{*\prime} + \frac{1}{NT^2}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\end{aligned}$$

Observe

$$\lambda_{j+r}\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) \le \lambda_{j+r}\left(\frac{1}{NT} XX'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right), \text{ for } j = 1, ..., T-r, \qquad \text{(A.3)}$$

where the first inequality is due to Lemma 7. The second inequality is due to Lemma 6 because

$$\lambda_{j+r}\left(\frac{1}{NT} XX'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) + \lambda_{r+1}\left(\frac{1}{NT} F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\right) = \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right).$$

By (A.3) and Lemma 8,

$$\lambda_{r+j}\left(\frac{1}{NT} XX'\right) / \lambda_{r+j+1}\left(\frac{1}{NT} XX'\right) \le \lambda_j\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) / \lambda_{j+r+1}\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right) = O_p(1),$$

for $j = 1, ..., kmax$.

Lemma 10: Under Assumptions D – F,

$$\lambda_j\left(\mathrm{E}Q(\Lambda^o)E'\right) / \lambda_{j+1}\left(\mathrm{E}Q(\Lambda^o)E'\right) = O_p(1), \text{ for } j = 1, ..., [d^* m] - r - 1.$$

Proof: By Lemma 8.

Lemma 11: Under Assumptions B-C, for any $A_{T\times p} = (a_1, ..., a_p)$ such that $A'A = TI_p$,

$$\frac{1}{T^2N}\left|tr(A'F\Lambda^{o\prime}\mathrm{E}'A)\right| = O_p(N^{-1/2});$$

$$tr\left(\frac{1}{NT^2}A'\mathrm{E}P(\Lambda^o)\mathrm{E}'A\right) = O_p(N^{-1}).$$

Proof: Observe that

$$\left|tr(A'F\Lambda^{o\prime}\mathrm{E}'A)\right| \le \left\|AA'F\right\|\left\|\Lambda^{o\prime}\mathrm{E}'\right\| \le \left\|A\right\|^2\left\|F\right\|\left\|\Sigma_i\lambda_i^o\varepsilon_{i\bullet}'\right\|.$$

In addition,

$$\left\|\Sigma_i\lambda_i^o\varepsilon_{i\bullet}'\right\| \le tr\left(\left(\Sigma_i\lambda_i^o\varepsilon_{i\bullet}'\right)\left(\Sigma_i\varepsilon_{i\bullet}\lambda_i^{o\prime}\right)\right)^{1/2} = tr\left(\Sigma_{i,j}\Sigma_t\lambda_i^o\varepsilon_{it}\varepsilon_{jt}\lambda_j^{o\prime}\right)^{1/2}$$
$$= tr\left(\Sigma_t\left(\Sigma_i\lambda_i^o\varepsilon_{it}\right)\left(\Sigma_j\varepsilon_{jt}\lambda_j^{o\prime}\right)\right)^{1/2} = \left(\Sigma_t\left\|\Sigma_i\lambda_i^o\varepsilon_{it}\right\|^2\right)^{1/2}$$

Thus, we have

$$\frac{1}{NT^2}\left|tr(A'F\Lambda^{o\prime}\mathrm{E}'A)\right| \le \frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{T}}A\right\|^2\left\|\frac{1}{\sqrt{T}}F\right\|\left(\frac{1}{T}\Sigma_t\left\|\frac{1}{\sqrt{N}}\Sigma_i\lambda_i^o\varepsilon_{it}\right\|^2\right)^{1/2} = O_p(N^{-1/2}).$$

Similarly,

$$tr\left(\frac{1}{NT^2}A'\mathrm{E}P(\Lambda^o)\mathrm{E}'A\right) = tr\left(\frac{1}{N}\frac{A'}{\sqrt{T}}\frac{\mathrm{E}\Lambda^o}{\sqrt{NT}}\left(\frac{\Lambda^{o\prime}\Lambda^o}{N}\right)^{-1}\frac{\Lambda^{o\prime}\mathrm{E}'}{\sqrt{NT}}\frac{A}{\sqrt{T}}\right)$$
$$\le \frac{1}{N}\left\|\frac{A}{\sqrt{T}}\right\|^2\left\|\frac{\mathrm{E}\Lambda^o}{\sqrt{NT}}\right\|^2 O_p(1) = O_p(N^{-1})$$

Lemma 12: Under Assumptions A-E, for $j = 1,\dots,r$,

$$\lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^oF^{*\prime}\right) = \lambda_j\left(\frac{1}{NT}F\Lambda^{o\prime}\Lambda^oF'\right) + O_p(N^{-1/2}); \tag{A.4}$$

$$\lambda_j\left(\frac{1}{NT}XX'\right) = \lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^oF^{*\prime}\right) + O_p\left(\frac{1}{m}\right); \tag{A.5}$$

$$\lambda_j\left(\frac{1}{NT}XX'\right) = \lambda_j\left(\frac{1}{NT}F\Lambda^{o\prime}\Lambda^oF'\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{m}\right). \tag{A.6}$$

Proof: Observe that

$$F^*\Lambda^{o\prime}\Lambda^oF^{*\prime} = F\Lambda^{o\prime}\Lambda^oF' + \mathrm{E}\Lambda^oF' + F'\Lambda^o\mathrm{E} + \mathrm{E}P(\Lambda^o)\mathrm{E}'.$$

Let Ξ_*^k be the matrix of the eigenvectors corresponding to the first $k\ (\le r)$ largest eigenvalues of $F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}/NT$, normalized such that $\tilde{F}_*^{k\prime}\tilde{F}_*^k = TI_k$. Similarly, define Ξ^k and $\tilde{F}^k$ for the eigenvectors of $F\Lambda^{o\prime}\Lambda^o F'/(NT)$ and $XX'/(NT)$, respectively. By Lemma 11,

$$\begin{aligned}\Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^o F^*\right) &= tr\left(\frac{1}{NT^2}\Xi_*^{k\prime}F\Lambda^{o\prime}\Lambda^o F'\Xi_*^k\right)\\ &\quad +2tr\left(\frac{1}{NT^2}\Xi_*^{k\prime}F\Lambda^{o\prime}\mathrm{E}'\Xi_*^k\right)+tr\left(\frac{1}{NT^2}\Xi_*^{k\prime}\mathrm{E}P(\Lambda^o)\mathrm{E}'\Xi_*^k\right)\\ &\le tr\left(\frac{1}{NT^2}\Xi^{k\prime}F\Lambda^{o\prime}\Lambda^o F'\Xi^k\right)+(N^{-1/2})+O_p(N^{-1})\\ &= \Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}F\Lambda^{o\prime}\Lambda^o F'\right)+O_p(N^{-1/2}),\end{aligned} \tag{A.7}$$

Similarly,

$$\begin{aligned}\Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\right) &\ge tr\left(\frac{1}{NT^2}\Xi^{k\prime}F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\Xi^k\right)\\ &= tr\left(\frac{1}{NT^2}\Xi^{k\prime}F\Lambda^{o\prime}\Lambda^o F'\Xi^k\right)\\ &\quad +2tr\left(\frac{1}{NT^2}\Xi^{k\prime}F\Lambda^{o\prime}\mathrm{E}'\Xi^k\right)+tr\left(\frac{1}{NT^2}\Xi^{k\prime}\mathrm{E}P(\Lambda^o)\mathrm{E}'\Xi^k\right)\\ &= \Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}F\Lambda^{o\prime}\Lambda^o F'\right)+O_p(N^{-1/2})+O_p(N^{-1})\\ &= \Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}F\Lambda^{o\prime}\Lambda^o F'\right)+O_p(N^{-1/2})\end{aligned} \tag{A.8}$$

(A.7) and (A.8) imply (A.4) because they hold for all $k = 1,...,r$.

We now show (A.5). By Lemmas 6 and 8,

$$\begin{aligned}\lambda_j\left(\frac{1}{NT}XX'\right) &\le \lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\right)+\lambda_1\left(\frac{1}{NT}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\right)\\ &= \lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\right)+O_p\left(\frac{1}{m}\right)\end{aligned} \tag{A.9}$$

Also, for any $k = 1,...,r$,

$$\begin{aligned}\Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}XX'\right) &\ge tr\left(\frac{1}{NT^2}\Xi_*^{k\prime}F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\Xi_*^k+\frac{1}{NT^2}\Xi_*^{k\prime}\mathrm{E}Q(\Lambda^o)\mathrm{E}'\Xi_*^k\right)\\ &\ge \Sigma_{j=1}^k\lambda_j\left(\frac{1}{NT}F^*\Lambda^{o\prime}\Lambda^o F^{*\prime}\right)+O_p\left(\frac{1}{m}\right).\end{aligned} \tag{A.10}$$

Thus, by (A.9) and (A.10), we have (A.5). Then, (A.4) and (A.5) imply (A.6).

Proof of Proposition **2**: We first show the consistency of the ER estimator. By Lemmas 9 and 12,

$$\frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} = O_p(1)\,,\ k = 1, 2, ..., r-1, r+1, ..., \mathit{kmax}\text{-}1;$$

$$\frac{\tilde{\mu}_{NT,r}}{\tilde{\mu}_{NT,r+1}} = \frac{O_p(1)}{O_p(m^{-1})} = O_p(m) \to_p \infty .$$

Thus, $\lim_{m\to\infty} \Pr(\tilde{k}_{ER} = r) = 1$. Now consider the GR estimator. Observe that:

$$\frac{\ln(\tilde{\mu}^*_{NT,k})}{\ln(\tilde{\mu}^*_{NT,k+1})} = \frac{\ln\left(1 + \dfrac{\tilde{\mu}_{NT,k}}{V(k,\tilde{F}^k)}\right)}{\ln\left(\dfrac{V(k,\tilde{F}^k)}{V(k,\tilde{F}^k) - \tilde{\mu}_{NT,k+1}}\right)} < \frac{\left(\dfrac{\tilde{\mu}_{NT,k}}{V(k,\tilde{F}^k)}\right)}{\left(\dfrac{\tilde{\mu}_{NT,k+1}}{V(k,\tilde{F}^k)}\right)} = \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} .$$

Thus,

$$\frac{\ln(\mu^*_k)}{\ln(\mu^*_{k+1})} = O_p(1)\text{, for } k = 1, 2, ..., r-1, r+1, ..., \mathit{kmax}\text{-}1.$$

Finally,

$$\frac{\ln(\tilde{\mu}^*_{NT,r})}{\ln(\tilde{\mu}^*_{NT,r+1})} = \frac{\ln\left(1 + \dfrac{\tilde{\mu}_{NT,r}}{V(r,\tilde{F}^r)}\right)}{\ln\left(1 + \dfrac{\tilde{\mu}_{NT,r+1}}{V(r+1,\tilde{F}^{r+1})}\right)} \geq \frac{\dfrac{\dfrac{\tilde{\mu}_{NT,r}}{V(r,\tilde{F}^r)}}{1 + \dfrac{\tilde{\mu}_{NT,r}}{V(r,\tilde{F}^r)}}}{\left(\dfrac{\tilde{\mu}_{NT,r+1}}{V(r+1,\tilde{F}^{r+1})}\right)} = \frac{\tilde{\mu}_{NT,r}}{\tilde{\mu}_{NT,r}} \frac{V(r+1,\tilde{F}^{r+1})}{V(r-1,\tilde{F}^{r-1})} = O_p(m)O_p(1)$$

Thus, we have $\lim_{m\to\infty} \Pr(\tilde{k}_{GR} = r) = 1$.

Proof of Proposition **3**: It is sufficient to show that $\tilde{\mu}_{NT,0} / \tilde{\mu}_{NT,1} \to_p \infty$ if $r = 0$, and $\tilde{\mu}_{NT,0} / \tilde{\mu}_{NT,1}$ $\to_p \infty$, if $r > 0$. Suppose that $r = 0$. Then, $\tilde{\mu}_{NT,j} / \tilde{\mu}_{NT,j+1} = O_p(1)$ for all $j = 1, ..., \mathit{kmax}$. But $\tilde{\mu}_{NT,0} / \tilde{\mu}_{NT,1} = w(N,T) / O_p(m^{-1}) = w(N,T) O_p(m) \to_p \infty$. Now suppose that $r > 0$. Then, $\tilde{\mu}_{NT,0} / \tilde{\mu}_{NT,1}$ $= w(N,T) / O_p(1) = w(N,T) O_p(1) \to_p 0$.

***Proof of Proposition* 4**: For $k = 1, 2, ..., r-1$,

$$\frac{\ln(1+m^{-1}+\tilde{\mu}_{NT,k})}{\ln(1+m^{-1}+\tilde{\mu}_{NT,k+1})} \leq \frac{m^{-1}+\tilde{\mu}_{NT,k}}{\dfrac{m^{-1}+\tilde{\mu}_{NT,k+1}}{1+m^{-1}+\tilde{\mu}_{NT,k+1}}} = \frac{m^{-1}+\tilde{\mu}_{NT,k}}{m^{-1}+\tilde{\mu}_{NT,k+1}}(1+m^{-1}+\tilde{\mu}_{NT,k+1}) = O_p(1). \quad \text{(A.11)}$$

For $k = r+1, ..., m$, even if $\tilde{\mu}_{NT,k} < O_p(m^{-1})$, $m^{-1}+\tilde{\mu}_{NT,k} = O_p(m^{-1})$. Thus, (A.11) still holds. Finally,

$$\frac{\ln(1+m^{-1}+\tilde{\mu}_{NT,r})}{\ln(1+m^{-1}+\tilde{\mu}_{NT,r+1})} > \frac{\dfrac{m^{-1}+\tilde{\mu}_{NT,r}}{1+m^{-1}+\tilde{\mu}_{NT,r}}}{m^{-1}+\tilde{\mu}_{NT,r+1}} = \frac{O_p(1)}{O_p(m^{-1})} = O_p(m) \rightarrow_p \infty.$$

References

Anderson, T.W., and S. Dasgupta, 1963, Some inequalities on characteristic roots of matrices, *Biometrika*, 50, 522-524.

Bai, J., 2003, Inferential theory for factor models of large dimensions, *Econometrica,* 71, 135-171.

Bai, J., 2007, Panel data models with interactive fixed effects, mimeo, New York University.

Bai, J., and S. Ng, , 2002, Determining the number of factors in approximate factor models. *Econometrica* , 191-221.

Bai, J, and S. Ng, 2004, A panic attack on unit roots and cointergration, *Econometrica*, 72(4), 1127-1177.

Bai, J., and S. Ng, 2006, Determining the number of factors in approximate factor models, Errata, http://www.columbia.edu/~sn2294/papers/correctionEcta2.pdf

Bai, Z.D., and Y.Q. Yin, 1993, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, *The Annals of Probability*, 21(3), 1275-1294.

Bai, Z.D., and J.W. Silverstein, 1999, Exact separation of eigenvalues of large dimensional sample covariance matrices, *The Annals of Probability*, 27(3), 1536-1555.

Bernanke, B., Boivin J. and Eliasz P., Measuring the Effects of Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach, *The Quarterly Journal of Economics*, February 2005, 387-422.

Brown, S., 1989, The number of factors in security returns, *The Journal of Finance,* 44, 1247-1262.

Cattell, R.B., 1966, The scree test for the number of factors, *Multivariate Behavioral Research,* 1, 245-276.

Cattell, R.B., 1983, This week's citation classic: scree test for the number of factors, *Current Contents*, 5, 16.

Chamberlain, G., and M. Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica,* 51, 1281-1304.

Chan L., Karceski, J. and J. Lakonishok, 1998, The Risk and Return from Factors, *The Journal of Financial and Quantitative Analysis*, 33, 159-188.

Ferson, W. and R. Korajczyk, 1995, Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?, *The Journal of Business*, 68, 309-349.

Forni, M., M. Hallin, M. Lippi and L. Reichlin, 2000, The generalized dynamic factor model: Identification and estimation, *The Review of Economics and Statistics*, 82, 540-554.

Kapetanios, G., 2004, A new method for determining the number of factors in factor models with large datasets, *Queen Mary, University of London,* Working Paper 525.

Onatski, A., 2006, Determining the number of factors from empirical distribution of eigenvalues, Working Paper, Columbia University.

Rao, C.R., 1973, Linear Statistical Inference and Its Applications, 2nd eds., John Wiley & Sons (New York, New York).

Ross, S., 1976, The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory,* 13, 341-360.

Stock, J., and M. Watson, 1999, Forecasting inflation, *Journal of Monetary Economics*, 293-335.

Stock, J., and M. Watson, 2002a, Macroeconomic forecasting using diffusion indexes, *Journal of Business and Economic Statistics*, 20, 147-162.

Stock, J., and M. Watson, 2002b, Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association*, 97, 1167-1179.

Shukla,R. and C. Trzcinka, 1990, On Sequential Test of the Arbitrage Pricing Theory: A Comparison of Principal Components and Maximum Likelihood Factors, *The Journal of Finance,* 45, 1541-1564.

Table 1: Cases with *I.I.D.* Errors

$r = 3, \theta = r,\ \rho = \beta = J = 0$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
				r = 1 (***SNR* = 1**)				
10	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	N/A	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]
20	100	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	4.65 (0.72) [0,0,1000]	3.87 (0.71) [0,0,1000]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]
40	100	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.01 (0.03) [10,990,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]
60	100	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]
100	100	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]	1.00 (0.00) [0,1000,0]
				r = 3 (***SNR* = 1/3**)				
10	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	N/A	1.92 (0.81) [721,277,2]	2.18 (0.79) [576,424,0]
20	100	2.95 (0.21) [47,953,0]	2.92 (0.27) [74,926,0]	5.16 (0.67) [0,2,998]	4.50 (0.67) [0,0,1000]	2.90 (0.30) [97,903,0]	2.66 (0.64) [245,755,0]	2.83 (0.46) [137,863,0]
40	100	3.00 (0.00) [0,1000,0]	2.99 (0.03) [1,999,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	2.99 (0.08) [7,993,0]	2.99 (0.05) [3,997,0]
60	100	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
100	100	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
				r = 5 (***SNR* = 1/5**)				
10	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	N/A	2.14 (1.18) [968,23,9]	2.22 (1.10) [980,20,0]
20	100	4.11 (0.85) [627,373,0]	3.69 (0.99) [781,219,0]	5.88 (0.57) [0,233,767]	5.40 (0.50) [6,592,402]	3.34 (0.66) [977,23,0]	3.21 (1.47) [731,269,0]	3.68 (1.40) [602,398,0]
40	100	4.88 (0.33) [117,883,0]	4.68 (0.53) [291,709,0]	4.99 (0.04) [2,998,0]	4.98 (0.11) [13,987,0]	4.74 (0.45) [251,749,0]	4.78 (0.71) [115,885,0]	4.90 (0.46) [64,936,0]
60	100	4.99 (0.07) [9,994,0]	4.93 (0.25) [65,935,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	4.99 (0.09) [8,992,0]	4.98 (0.14) [14,986,0]	4.99 (0.05) [4,996,0]
100	100	5.00 (0.00) [0,1000,0]	4.99 (0.03) [1,999,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \le kmax]$.

Table 2.1: Cases with Autocorrelated Errors

$r = 3, \theta = 1\ (SNR = 1),\ \beta = 0$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$\rho = 0.5$								
30	50	3.39 (0.75) [0,720,280]	3.03 (0.16) [0,971,29]	7.01 (0.62) [0,0,1000]	5.94 (0.67) [0,0,1000]	3.01 (0.09) [0,992,8]	2.98 (0.18) [20,980,0]	2.98 (0.12) [12,988,0]
50	50	3.21 (0.52) [0,830,170]	3.001 (0.03) [0,999,1]	6.42 (0.65) [0,0,1000]	4.92 (0.63) [0,2,998]	3.00 (0.04) [0,998,2]	2.99 (0.05) [3,997,0]	2.99 (0.04) [2,998,0]
100	50	3.11 (0.35) [0,894,106]	3.004 (0.06) [0,996,4]	6.09 (0.64) [0,0,1000]	5.08 (0.63) [0,1,999]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	30	5.65 (1.72) [0,153,847]	3.27 (0.60) [0,788,212]	7.71 (0.48) [0,0,1000]	6.86 (0.62) [0,0,1000]	3.06 (0.26) [0,936,64]	2.96 (0.22) [40,960,0]	2.98 (0.23) [17,983,0]
50	100	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	4.00 (0.60) [0,172.828]	3.28 (0.45) [0,726,274]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
100	100	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.35 (0.50) [0,656,344]	3.002 (0.04) [0,998,2]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \leq kmax]$.

Table 2.2: Cases with Autocorrelated Errors

$r = 3, \theta = 1\ (SNR = 1),\ \beta = 0$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$\rho = 0.7$								
30	50	7.78 (0.60) [0,0,1000]	5.96 (1.61) [0,83,917]	7.98 (0.14) [0,0,1000]	7.69 (0.48) [0,0,1000]	3.32 (0.51) [0,700,300]	2.92 (0.31) [68,932,0]	2.98 (0.18) [24,973,3]
50	50	7.94 (0.26) [0,0,1000]	5.84 (1.52) [0,70,930]	7.99 (0.06) [0,0,1000]	7.57 (0.51) [0,0,1000]	3.17 (0.39) [0,830,170]	2.98 (0.15) [17,983,0]	2.99 (0.06) [4,996,0]
100	50	8.00 (0.00) [0,0,1000]	7.94 (0.28) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.08) [0,0,1000]	3.06 (0.24) [0,943,57]	2.99 (0.03) [1,999,0]	3.00 (0.00) [0,1000,0]
50	30	7.99 (0.05) [0,0,1000]	7.82 (0.47) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.97 (0.16) [0,0,1000]	4.23 (0.69) [0,129,871]	2.90 (0.40) [95,900,5]	2.98 (0.35) [45,933,22]
50	100	4.83 (1.47) [0,211,789]	3.30 (0.57) [0,757,243]	7.34 (0.61) [0,0,1000]	6.41 (0.64) [0,0,1000]	3.01 (0.05) [0,997,3]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
100	100	7.22 (1.14) [0,4,996]	3.37 (0.66) [0,720,280]	7.86 (0.34) [0,0,1000]	6.43 (067) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.03 (0.17) [0,970,30]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \leq kmax]$.

Table 2.3: Cases with Autocorrelated Errors

$r = 3, \theta = 1\ (SNR = 1),\ \beta = 0$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$\rho = 0.9$								
30	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.51 (0.57) [0,0,1000]	3.06 (0.85) [131,720,149]	3.67 (1.30) [52,604,344]
50	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.63 (0.54) [0,0,1000]	3.04 (0.56) [50,872,78]	3.50 (1.10) [26,724,250]
100	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.75 (0.47) [0,0,1000]	3.01 (0.37) [25,947,28]	3.30 (0.90) [8,857,135]
50	30	8.00 (0.00) [0,0,1000]	7.99 (0.05) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.96 (0.46) [0,0,1000]	3.21 (0.84) [92,687,221]	4.09 (1.40) [34,438,528]
50	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	4.26 (0.68) [0,114,886]	2.99 (0.12) [12,987,1]	3.00 (0.19) [8,984,8]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	4.26 (0.68) [0,110,890]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.58 (0.59) [0,465,535]	3.41 (0.57) [0,542,458]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.98 (0.13) [0,0,1000]	2.99 (0.08) [6,994,0]	3.02 (0.23) [1,989,10]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \leq kmax]$.

Table 3.1: Cases with Cross-Correlated Errors

$$r = 3, \theta = 1\ (SNR = 1),\ \rho = 0,\ \beta = .2$$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
J=8								
50	100	6.93 (0.26) [0,0,1000]	6.69 (0.00) [0,0,1000]	7.01 (0.37) [0,0,1000]	6.95 (0.00) [0,0,1000]	6.58 (0.22) [0,0,1000]	2.99 (0.18) [14,984,2]	3.12 (0.62) [4,954,42]
100	50	7.98 (0.12) [0,0,1000]	7.81 (0.44) [0,0,1000]	7.99 (0.08) [0,0,1000]	7.94 (0.24) [0,0,1000]	3.43 (0.57) [0,610,390]	2.99 (0.05) [3,997,0]	3.00 (0.00) [0,1000,0]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.05 (0.22) [0,951,49]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
200	100	8.00 (0.00) [0,0,1000]	7.98 (0.14) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	7.00 (0.05) [0,0,1000]	2.99 (0.04) [2,998,0]	3.10 (0.62) [0,975,25]
1000	50	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	4.21 (0.91) [0,220,280]	3.02 (0.13) [0,983,17]	6.22 (0.67) [0,0,1000]	4.09 (0.58) [0,120,880]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
J=20								
50	100	5.04 (0.21) [0,0,1000]	4.99 (0.07) [0,0,1000]	5.82 (0.53) [0,0,1000]	5.29 (0.46) [0,0,1000]	5.00 (0.04) [0,610,390]	3.22 (0.42) [1,769,230]	3.75 (0.55) [0,311,689]
100	50	6.35 (0.48) [0,0,1000]	6.10 (0.32) [0,0,1000]	6.76 (0.48) [0,0,1000]	6.44 (0.50) [0,0,1000]	6.72 (0.45) [0,0,1000]	3.55 (0.99) [8,725,267]	4.69 (1.21) [0,300,700]
100	100	7.16 (0.36) [0,0,1000]	7.00 (0.08) [0,0,1000]	7.28 (0.44) [0,0,1000]	7.02 (0.14) [0,0,1000]	6.97 (0.16) [0,0,1000]	3.36 (0.86) [0,846,154]	4.87 (1.24) [0,278,722]
200	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	4.34 (0.76) [0,118,882]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	5.80 (0.40) [0,0,1000]	5.74 (0.44) [0,0,1000]	6.06 (0.25) [0,0,1000]	6.04 (0.20) [0,0,1000]	5.01 (0.09) [0,0,1000]	3.10 (0.30) [0,899,101]	3.83 (0.47) [0,207,793]
1000	50	3.92 (1.08) [0,439,561]	3.71 (0.92) [0,56,484]	6.91 (0.64) [0,0,1000]	6.76 (0.64) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \le kmax]$.

Table 3.2: Cases with Cross-Correlated Errors

$r = 3, \theta = 1\ (SNR = 1),\ \rho = 0,\ \beta = .5$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$J = 8$								
50	100	7.99 (0.07) [0,0,1000]	7.90 (0.29) [0,0,1000]	7.99 (0.03) [0,0,1000]	7.98 (0.11) [0,0,1000]	7.41 (0.49) [0,0,1000]	5.49 (1.90) [6,354,540]	6.89 (0.58) [0,20,980]
100	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.63 (0.64) [0,451,549]	2.98 (0.17) [18,981,1]	3.01 (0.21) [6,989,5]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.10 (0.30) [0,896,104]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
200	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.97 (0.17) [0,0,1000]	6.40 (1.43) [0,150,850]	7.00 (0.00) [0,0,1000]
1000	50	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.70 (0.60) [0,369,631]	3.59 (0.58) [0,458,542]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
$J = 20$								
50	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.01 (0.07) [0,0,1000]	3.79 (0.73) [0,388,612]	4.82 (0.38) [0,0,1000]
100	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	6.94 (0.22) [0,0,1000]	4.78 (1.16) [3,251,746]	5.63 (0.72) [0,37,963]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	6.99 (0.03) [0,0,1000]	4.79 (1.21) [0,286,714]	5.77 (0.55) [0,18,982]
200	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	4.73 (0.74) [0,39,961]	3.00 (0.00) [0,1000,0]	3.01 (0.13) [0,998,2]
50	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.00 (0.00) [0,0,1000]	3.85 (0.85) [0,442,558]	4.99 (0.06) [0,0,1000]
1000	50	7.99 (0.03) [0,0,1000]	7.99 (0.05) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \le kmax]$.

Table 3.3: Cases with Cross-Correlated Errors

$r = 3, \theta = 1\ (SNR = 1),\ \rho = 0,\ \beta = .9$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
				$J = 8$				
50	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.01 (0.11) [0,0,1000]	4.60 (1.79) [9,520,471]	6.39 (1.24) [1,104,895]
100	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.83 (0.69) [0,327,673]	2.99 (0.30) [18,978,4]	3.06 (0.30) [10,983,7]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.18 (0.41) [0,823,177]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
200	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.00 (0.00) [0,0,1000]	4.32 (1.86) [0,661,339]	6.88 (0.67) [0,28,972]
1000	50	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	4.04 (0.63) [0,173,827]	3.90 (0.62) [0,241,759]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
				$J = 20$				
50	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.17 (0.39) [0,0,1000]	3.53 (0.52) [0,481,519]	4.38 (0.71) [0,40,960]
100	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	6.83 (0.37) [0,0,1000]	4.65 (1.11) [3,266,731]	5.50 (0.75) [0,44,956]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	6.91 (0.27) [0,0,1000]	4.73 (1.13) [0,270,730]	5.68 (0.57) [0,17,983]
200	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	4.91 (0.73) [0,23,977]	3.00 (0.00) [0,1000,0]	3.01 (0.13) [0,998,2]
50	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.01 (0.06) [0,0,1000]	3.46 (0.50) [0,538,462]	4.27 (0.50) [0,5,995]
1000	50	8.00 (0.00) [0,0,1000]	7.99 (0.03) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \le kmax]$.

Table 4.1: Cases with Both Autocorrelated and Cross-Correlated Errors

$r = 3,\ \rho = .5,\ \beta = .2,\ J = 8$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$\theta = 1/3$ (***SNR* = 3**)								
30	50	7.72 (0.58) [0,0,1000]	6.64 (0.82) [0,0,1000]	7.95 (0.22) [0,0,1000]	7.54 (0.54) [0,0,1000]	5.14 (0.41) [0,0,1000]	2.99 (0.10) [6,989,5]	3.29 (0.74) [3,843,154]
50	50	7.94 (0.23) [0,0,1000]	7.39 (0.61) [0,0,1000]	7.99 (0.09) [0,0,1000]	7.68 (0.46) [0,0,1000]	5.57 (0.58) [0,0,1000]	2.99 (0.03) [1,999,0]	3.00 (0.07) [1,998,1]
100	50	8.00 (0.00) [0,0,1000]	7.99 (0.09) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.03) [0,0,1000]	3.89 (0.68) [0,284,716]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	30	8.00 (0.12) [0,0,1000]	7.71 (0.55) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.04) [0,0,1000]	5.22 (0.62) [0,1,999]	2.99 (0.09) [6,992,2]	3.02 (0.18) [0,986,14]
50	100	7.92 (0.26) [0,0,1000]	7.70 (0.46) [0,0,1000]	7.97 (0.16) [0,0,1000]	7.83 (0.37) [0,0,1000]	6.12 (0.49) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.33 (0.52) [0,696,304]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.00 (0.03) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	50	5.15 (1.66) [0,190,810]	4.67 (1.65) [0,280,720]	7.52 (0.55) [0,0,1000]	7.38 (0.59) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \le kmax]$.

Table 4.2: Cases with Both Autocorrelated and Cross-Correlated Errors

$r = 3,\ \rho = .5,\ \beta = .2,\ J = 8$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$\theta = 1$ (***SNR* = 1**)								
30	50	7.72 (0.59) [0,0,1000]	6.62 (0.83) [0,0,1000]	7.95 (0.21) [0,0,1000]	7.54 (0.54) [0,0,1000]	5.13 (0.41) [0,0,1000]	3.16 (0.67) [65,746,189]	4.50 (1.17) [23,249,728]
50	50	7.94 (0.24) [0,0,1000]	7.38 (0.61) [0,0,1000]	7.99 (0.10) [0,0,1000]	7.68 (0.47) [0,0,1000]	5.55 (0.59) [0,0,1000]	3.02 (0.44) [38,911,51]	3.59 (1.32) [35,727,238]
100	50	8.00 (0.00) [0,0,1000]	7.99 (0.10) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.03) [0,0,1000]	3.88 (0.68) [0,289,711]	2.99 (0.11) [8,990,2]	3.00 (0.08) [1,996,3]
50	30	7.99 (0.12) [0,0,1000]	7.71 (0.55) [0,0,1000]	7.99 (0.04) [0,0,1000]	7.93 (0.25) [0,0,1000]	5.19 (0.62) [0,1,999]	2.99 (0.04) [2,998,0]	3.52 (0.87) [36,738,226]
50	100	7.92 (0.27) [0,0,1000]	7.70 (0.46) [0,0,1000]	7.97 (0.16) [0,0,1000]	7.83 (0.37) [0,0,1000]	6.10 (0.49) [0,0,1000]	3.02 (0.68) [108,788,104]	3.15 (0.65) [5,927,68]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.32 (0.51) [0,700,300]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,999,1]
50	1000	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.00 (0.03) [0,0,1000]	3.01 (0.29) [16,968,16]	3.06 (0.50) [0,984,16]
1000	50	5.14 (1.65) [0,195,805]	4.66 (1.53) [0,279,721]	7.51 (0.55) [0,0,1000]	7.39 (0.58) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \le kmax]$.

Table 4.3: Cases with Both Autocorrelated and Cross-Correlated Errors

$r = 3,\ \rho = .5,\ \beta = .2,\ J = 8$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
θ=3 (***SNR* = 1/3**)								
30	50	7.71 (0.60) [0,0,1000]	6.57 (0.83) [0,0,1000]	7.95 (0.21) [0,0,1000]	7.53 (0.55) [0,0,1000]	5.04 (0.44) [0,1,999]	3.65 (1.26) [188,188,624]	4.97 (1.18) [56,52,892]
50	50	7.93 (0.24) [0,0,1000]	7.34 (0.63) [0,0,1000]	7.98 (0.11) [0,0,1000]	7.67 (0.47) [0,0,1000]	5.04 (0.43) [0,0,1000]	3.89 (1.62) [194,246,560]	5.31 (1.77) [95,87,818]
100	50	8.00 (0.00) [0,0,1000]	7.99 (0.10) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.80 (0.66) [1,333,666]	2.89 (1.03) [257,605,138]	3.20 (1.28) [214,527,259]
50	30	7.98 (0.13) [0,0,1000]	7.67 (0.59) [0,0,1000]	7.99 (0.04) [0,0,1000]	7.92 (0.26) [0,0,1000]	5.03 (0.65) [0,7,993]	3.42 (1.65) [316,206,478]	4.03 (1.65) [198,165,637]
50	100	7.92 (0.27) [0,0,1000]	7.69 (0.46) [0,0,1000]	7.97 (0.16) [0,0,1000]	7.82 (0.38) [0,0,1000]	6.05 (0.50) [0,0,1000]	4.49 (1.60) [103,222,675]	5.33 (1.39) [31,114,955]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.29 (0.50) [2,727,271]	2.93 (0.58) [96,864,40]	2.99 (0.68) [95,827,78]
200	200	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
50	1000	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.00 (0.04) [0,0,1000]	6.21 (1.47) [21,115,864]	6.86 (0.51) [1,7,992]
1000	50	5.07 (1.64) [0,205,795]	4.62 (1.53) [0,291,709]	7.49 (0.55) [0,0,1000]	7.37 (0.59) [0,0,1000]	3.00 (0.00) [0,1000,0]	2.99 (0.06) [4,996,0]	2.99 (0.05) [3,997,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \leq kmax]$.

Table 5.1: Cases with One Strong and Two Weak Factors

$\theta = 1$, $\rho = 0.5$, $\beta = 0.2$ and $J = 8$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$r = 3$, $f_1, f_2 \sim N(0,.75)$ and $f_3 \sim N(0,1.5)$								
30	50	7.71 (0.57) [0,0,1000]	6.62 (0.82) [0,0,1000]	7.95 (0.21) [0,0,1000]	7.53 (0.53) [0,0,1000]	5.14 (0.42) [0,0,1000]	2.95 (0.98) [207,581,212]	3.60 (0.96) [57,484,459]
50	50	7.94 (0.24) [0,0,1000]	7.38 (0.60) [0,0,1000]	7.99 (0.09) [0,0,1000]	7.67 (0.47) [0,0,1000]	5.56 (0.60) [0,0,1000]	2.83 (0.80) [158,782,60]	3.32 (1.04) [56,746,198]
100	50	8.00 (0.00) [0,0,1000]	7.99 (0.07) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.04) [0,0,1000]	3.85 (0.66) [0,304,696]	2.91 (0.39) [53,945,2]	2.99 (0.18) [10,983,7]
50	30	7.99 (0.09) [0,0,1000]	7.69 (0.56) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.95 (0.22) [0,0,1000]	5.17 (0.64) [0,3,997]	2.77 (0.89) [228,686,86]	3.33 (1.08) [96,644,260]
50	100	7.93 (0.26) [0,0,1000]	7.71 (0.46) [0,0,1000]	7.97 (0.17) [0,0,1000]	7.83 (0.37) [0,0,1000]	6.11 (0.51) [0,0,1000]	2.92 (0.59) [80,894,26]	3.30 (0.94) [22,844,134]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.31 (0.51) [0,707,293]	2.99 (0.10) [3,997,0]	3.00 (0.03) [0,999,1]
200	200	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.33 (1.10) [2,915,83]
50	1000	7.99 (0.03) [0,0,1000]	7.99 (0.03) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.00 (0.03) [0,0,1000]	2.97 (0.22) [21,979,0]	3.00 (0.00) [0,1000,0]
1000	50	5.03 (1.60) [0,191,809]	4.64 (1.51) [0,272,728]	7.51 (0.54) [0,0,1000]	7.39 (0.56) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]	3.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \leq kmax]$.

Table 5.2: Cases with Three Strong and Two Weak Factors
$\theta=1$, $\rho=0.5$, $\beta=0.2$ and $J=8$

N	T	$IC1$	$IC2$	$PC1$	$PC2$	ON	ER	GR
$r=5$, $f_1,f_2,f_3 \sim N(0,1.2)$ and $f_4,f_5 \sim N(0,.7)$								
30	50	7.99 (0.03) [0,0,1000]	7.96 (0.20) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.07) [0,0,1000]	6.47 (0.54) [0,17,983]	5.20 (1.08) [140,528,332]	5.74 (1.00) [51,394,555]
50	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.71 (0.66) [5,387,608]	4.95 (0.88) [134,741,125]	5.28 (0.84) [64,693,243]
100	50	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.03 (0.19) [3,961,36]	4.96 (0.32) [36,958,6]	5.01 (0.22) [16,965,19]
50	30	8.00 (0.00) [0,0,1000]	7.99 (0.09) [0,0,1000]	8.00 (0.00) [0,0,1000]	7.99 (0.03) [0,0,1000]	5.70 (0.64) [6,379,615]	4.79 (1.22) [229,590,181]	5.24 (1.13) [114,569,317]
50	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.64 (0.62) [2,429,569]	5.01 (0.49) [45,904,51]	5.24 (0.66) [17,829,154]
100	100	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.00 (0.00) [0,1000,0]	4.99 (0.03) [0,999,1]	5.00 (0.03) [0,999,1]
200	200	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]
50	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.31 (0.48) [0,700,300]	4.99 (0.14) [10,988,2]	5.03 (0.25) [3,981,16]
1000	50	7.01 (0.1.06) [0,122,878]	6.76 (1.11) [0,176,824]	7.82 (0.38) [0,0,1000]	7.74 (0.45) [0,0,1000]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]
500	1000	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]
1000	500	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	8.00 (0.00) [0,0,1000]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]	5.00 (0.00) [0,1000,0]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0<\tilde{k}<r, \tilde{k}=r, r<\tilde{k}\le kmax]$.

Table 6: No Common Factor

$r = 0, \theta = 1, J = 8$

N	*T*	*ON*	*ER**	*ON*	*ER**	*ON*	*ER**
		$\rho = 0.25, \beta = 0.1$		$\rho = 0.5, \beta = 0.2$		$\rho = 0.5, \beta = 0.5$	
100	100	2.11 (0.73) [10,178,812]	0.73 (1.47) [730,95,175]	4.29 (0.81) [0,0,1000]	2.55 (2.39) [364,90,546]	5.78 (0.72) [0,0,1000]	1.70 (2.50) [659,20,321]
1000	60	0.00 (0.00) [1000,0,0]	0.001 (0.03) [999,1,0]	0.04 (0.20) [955,45,0]	0.02 (0.17) [982,14,4]	0.34 (0.53) [687,287,26]	0.07 (0.43) [956,26,15]
1000	250	0.00 (0.00) [1000,0,0]	0.00 (0.00) [1000,0,0]	0.00 (0.00) [1000,0,0]	0.00 (0.00) [1000,0,0]	0.001 (0.03) [999,1,0]	0.00 (0.00) [1000,0,0]
150	500	0.00 (0.00) [1000,0,0]	0.00 (0.00) [1000,0,0]	0.001 (0.03) [999,1,0]	0.00 (0.00) [1000,0,0]	0.005 (0.07) [995,5,0]	0.00 (0.00) [1000,0,0]
40	150	2.71 (0.45) [0,0,1000]	2.64 (0.59) [0,3,997]	3.64 (0.48) [0,0,1000]	3.28 (0.68) [0,0,1000]	4.01 (0.7) [0,0,1000]	4.32 (1.13) [0,0,1000]
500	150	0.001 (0.03) [999,1,0]	0.006 (0.08) [994,6,0]	0.02 (0.12) [984,16,0]	0.04 (0.21) [962,35,3]	0.06 (0.25) [942,55,3]	0.05 (0.26) [955,40,5]

The maximum number of factors tested in simulations is eight (*kmax*=8). The values in the parenthesis are the standard deviations of the estimates. The values in brackets provide information on the frequencies of the estimated numbers of factors. The first value indicates the frequency of underestimation out of 1,000 times. The second value is the frequency of correct estimation, while the third is that of overestimation: $[0 < \tilde{k} < r, \tilde{k} = r, r < \tilde{k} \leq kmax]$.

Table 7: Weak Factors

$f_1 \sim N(0,1)$, $f_2 \sim N(0,.5)$ and $f_3 \sim N(0, SNR)$

	***SNR*=.10**	***SNR*=.20**	***SNR*=.35**	***SNR*=.45**
ER	1.87 [127,872,**1**,0]	1.71 [345,597,**58**,0]	2.44 [246,67,**686**,1,0]	2.78 [98,16,**886**,0]
GR	1.97 [30,967,**3**,0]	2.03 [148,676,**172**,4]	2.83 [75,35,**877**,12]	2.95 [24,8,**958**,10]
Onatski	2.73 [0,397,**474**,129]	3.31 [0,18,**669**,313]	3.36 [0,0,**660**,340]	4.20 [0,0,**658**,342]
PC estimator of 1st Factor	1.006 [**994**,6,0,0]	1.006 [**994**,6,0,0]	1.006 [**944**,6,0,0]	1.007 [**993**,7,0,0]
PC estimator of 2nd Factor	1.99 [11,**989**,0,0]	1.99 [10,**988**,2,0]	2.05 [13,**922**,65,0]	2.28 [15,**688**,297,0]
PC estimator of 3rd Factor	4.58 [1,0,**267**,732]	3.06 [1,3,**943**,53]	2.88 [0,116,**884**,0]	2.60 [3,393,**604**,0]

The maximum number of factors tested in every simulation is eight (*kmax*=8). The values in brackets provide information on the frequencies of the estimated numbers of factors in the case of the criteria. In the case of the "PC estimator of the Factor", the values in brackets provide the frequency of the factor extracted through principal components most correlated with the real factor. The first value indicates the frequency of estimation of 1 factor in the case of the criteria and in the case of the "PC estimator of the Factor" indicates the number of times the first PC is most correlated with the real factor analyzed. The frequencies analyzed are $[0,1,2,3,\geq 4]$.

Table 8: Estimated Number of Factors in Monthly Macroeconomic Variables

	N	*T*	*IC1*	*IC2*	*PC1*	*PC2*	*ON*	*ER*	*GR*
BBE dataset	120	511	**15**	**14**	**15**	**15**	**13**	**5**	**5**

The estimates are obtained using *kmax*=15 and time-demeaned data.

Table 9: Estimated Number of Latent Factors in the U.S. Stock Market

	N	*T*	*IC1*	*IC2*	*PC1*	*PC2*	*ON*	*ER*	*GR*
January 1970- December 2006	313	444	**4**	**3**	**4**	**4**	**2**	**1**	**1**
January 1970- December 1987	816	216	**3**	**3**	**3**	**3**	**3**	**1**	**1**
January 1988- December 2006	1288	228	**2**	**1**	**2**	**2**	**3**	**1**	**1**
January 1970- December 1978	1384	108	**1**	**1**	**3**	**3**	**3**	**1**	**1**
January 1979- December 1992	1640	168	**1**	**1**	**2**	**2**	**3**	**1**	**1**
January 1993- December 2006	1855	168	**1**	**1**	**2**	**2**	**3**	**1**	**1**
January 1970- December 1974	1859	60	**1**	**1**	**3**	**3**	**3**	**1**	**1**
January 1975- December 1979	3087	60	**1**	**1**	**2**	**1**	**1**	**1**	**1**
January 1980- December 1984	3091	60	**1**	**1**	**1**	**1**	**2**	**1**	**1**
January 1985- December 1989	3401	60	**1**	**1**	**1**	**1**	**1**	**3**	**3**
January 1990- December 1994	3858	60	**1**	**1**	**1**	**1**	**2**	**1**	**1**
January 1995- December 1999	4142	60	**1**	**1**	**1**	**1**	**1**	**1**	**1**
January 2000- December 2006	3436	84	**2**	**2**	**3**	**3**	**3**	**1**	**1**

The estimates are obtained using *kmax*=15 and time-demeaned data.

Figure 1: Scree Test (Locus of $NT\tilde{\mu}_{NT,k}$)

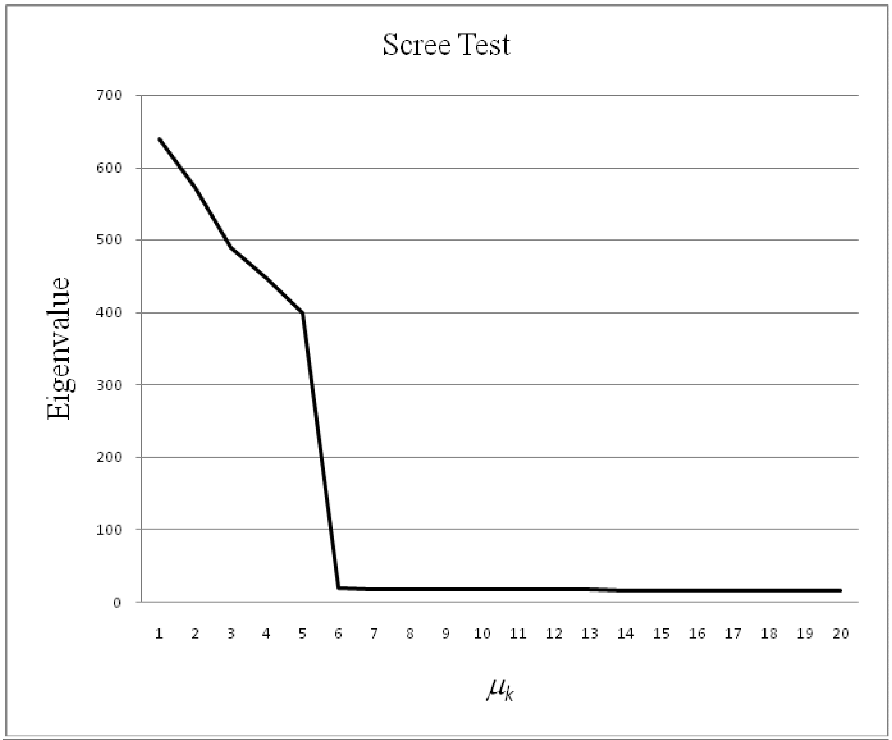

Figure 2

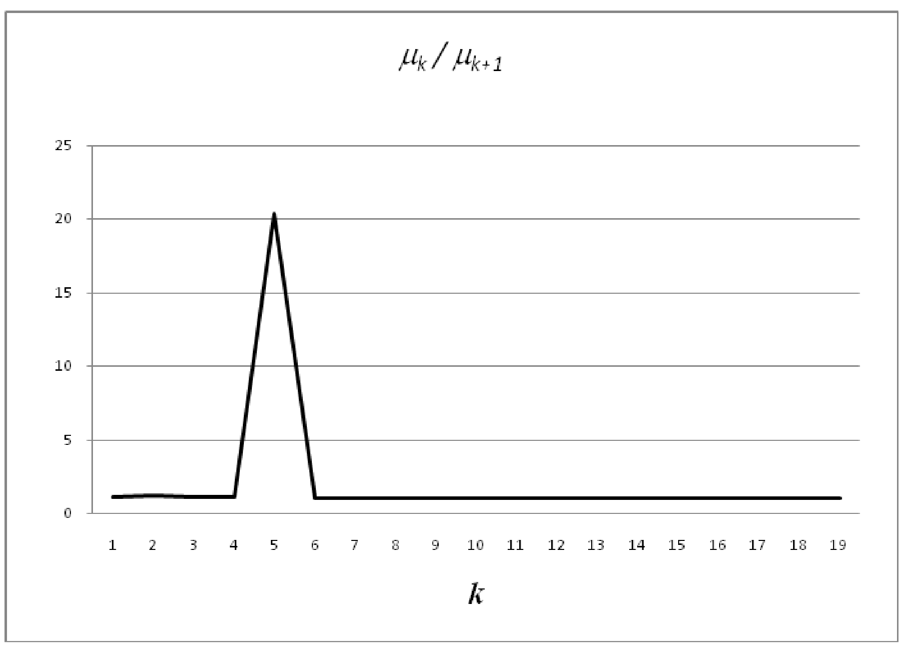

Figure 3.1

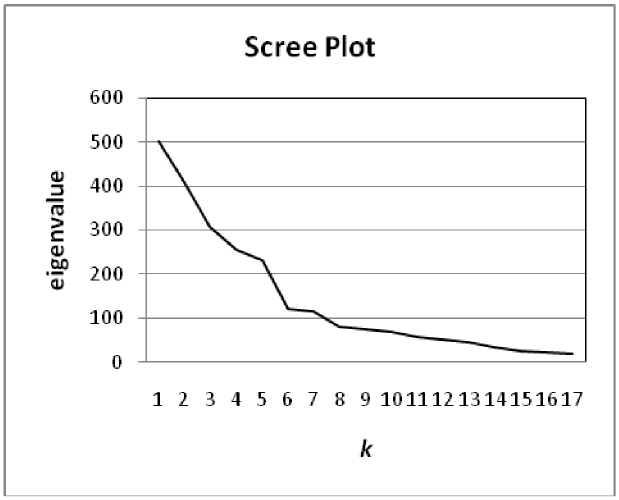

Figure 3.2

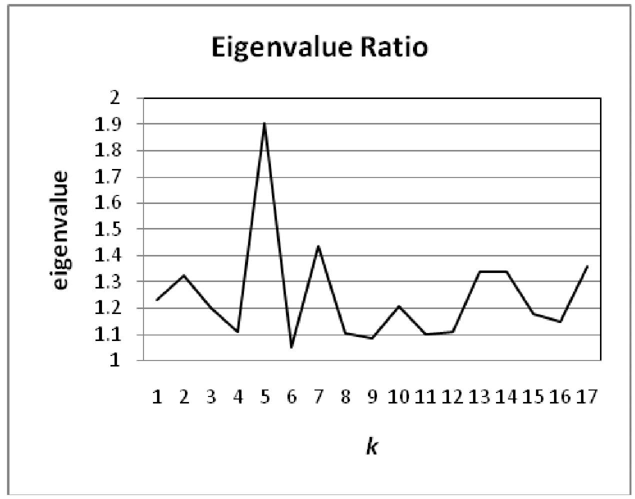

Figure 4

Onatski's Threshold Value

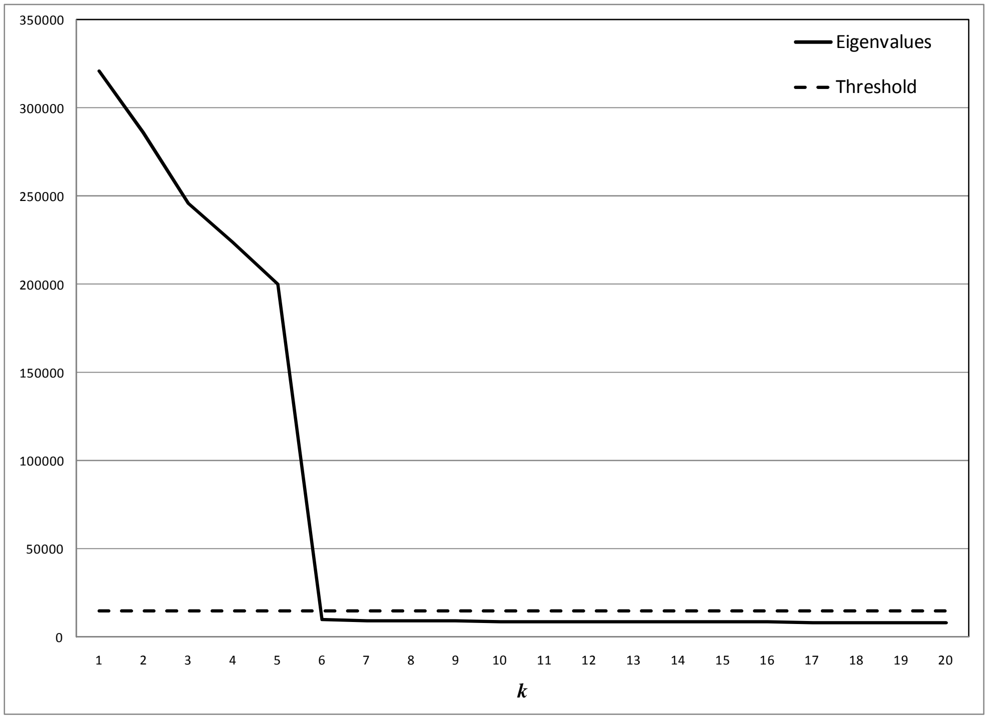

Figure 5.1

$V(k, \tilde{F}^k)$

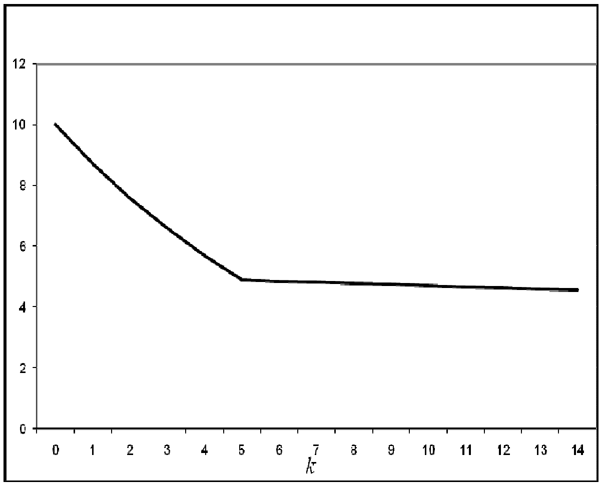

Figure 5.2

PC Functions of BN and ON

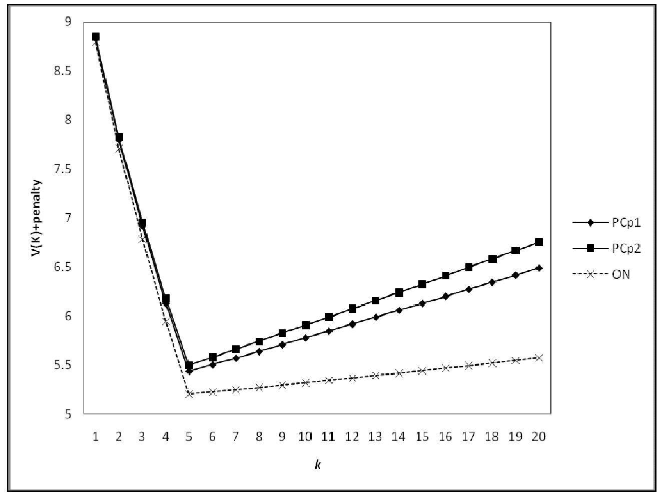

Figure 6

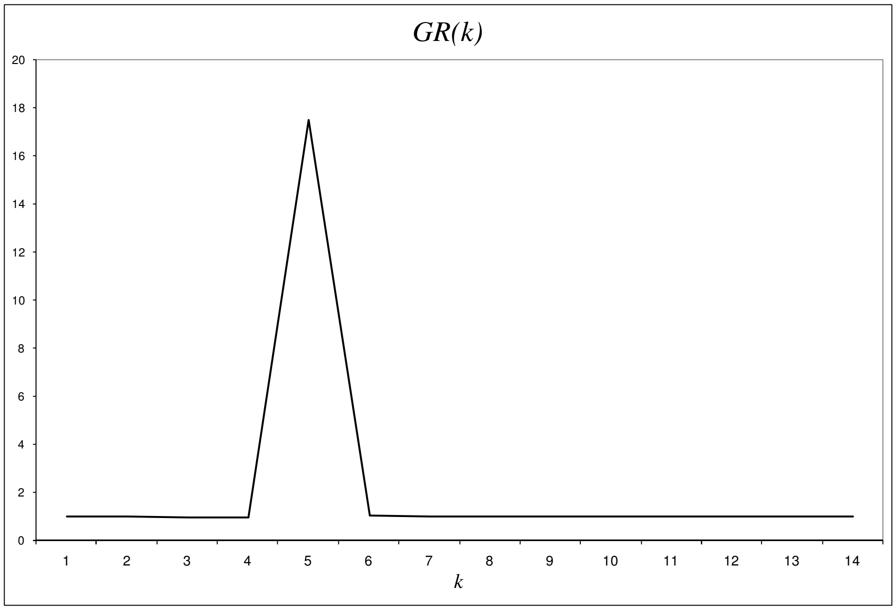

Figure 7

$\tilde{\mu}_{NT,k} \,/\, \tilde{\mu}_{NT,k+1}$

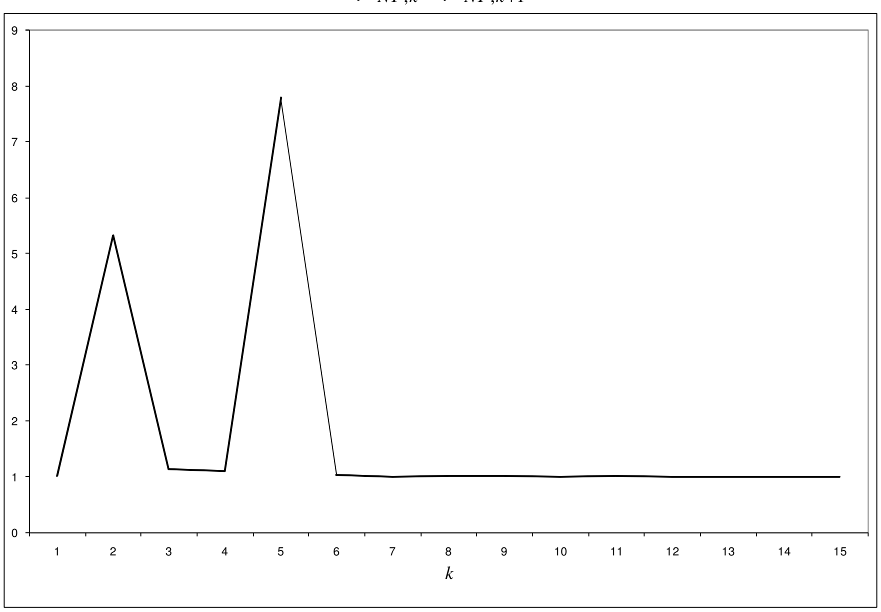

Figure 8

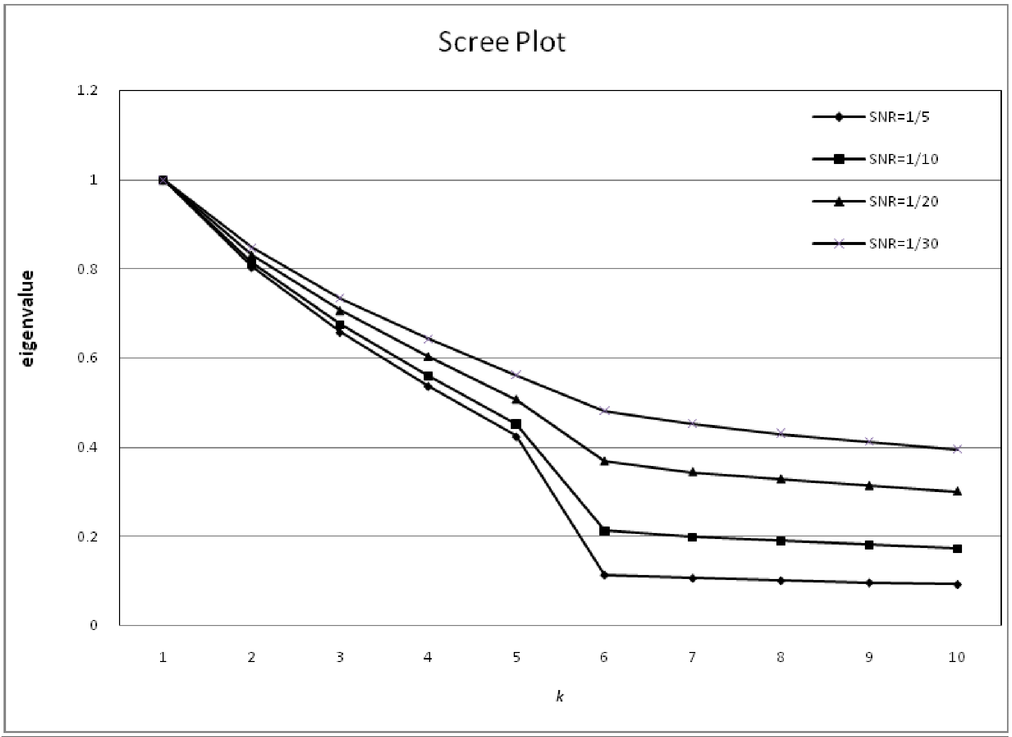

Figure 9

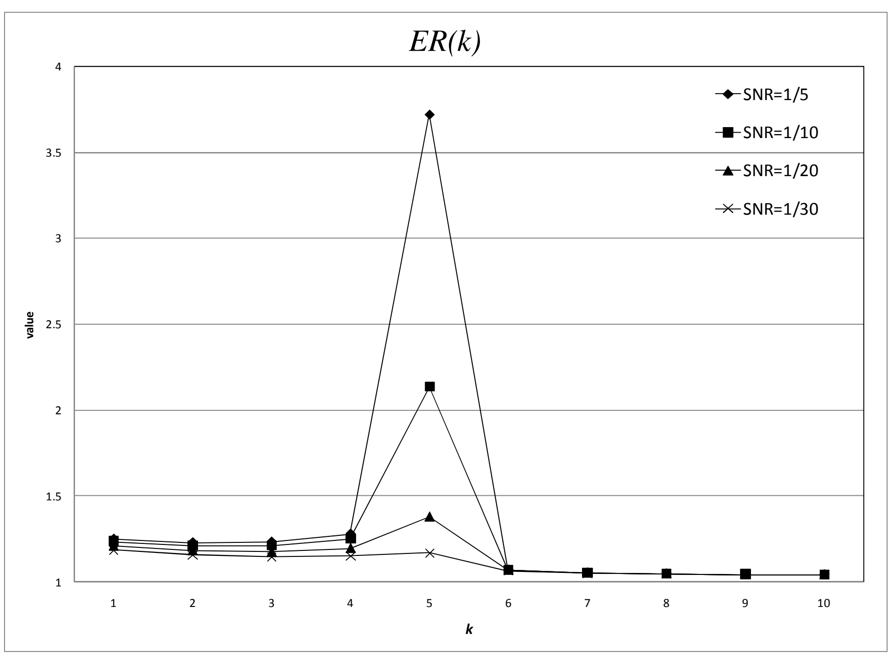

Figure 10
Evolution of *ER*(*k*) in the Presence of Weak Factors

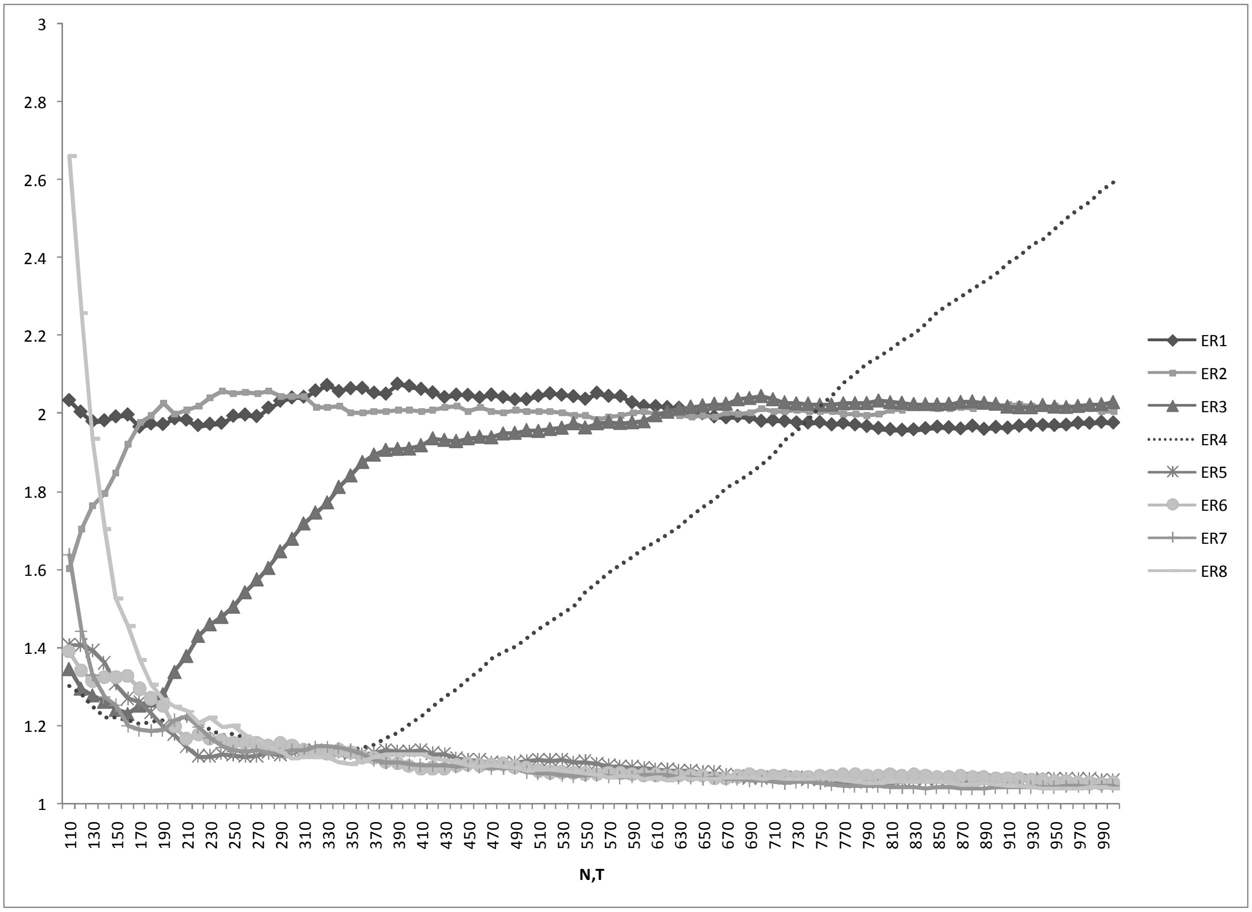

Figure 11
Evolution of GR, ER and ON Criterion Functions in the Presence of Weak Factors

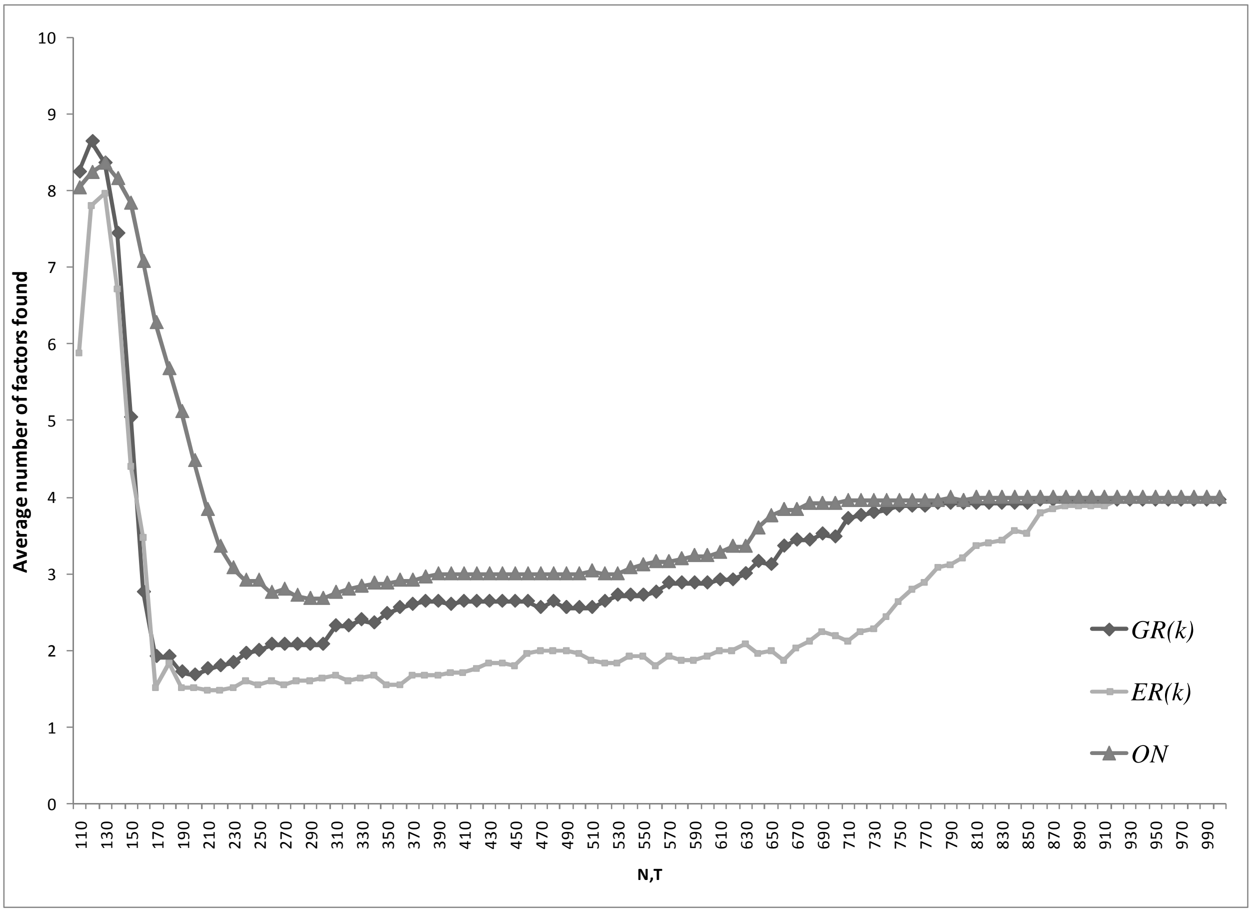

Figure 12
Evolution of R^2 in the Presence of Weak Factors

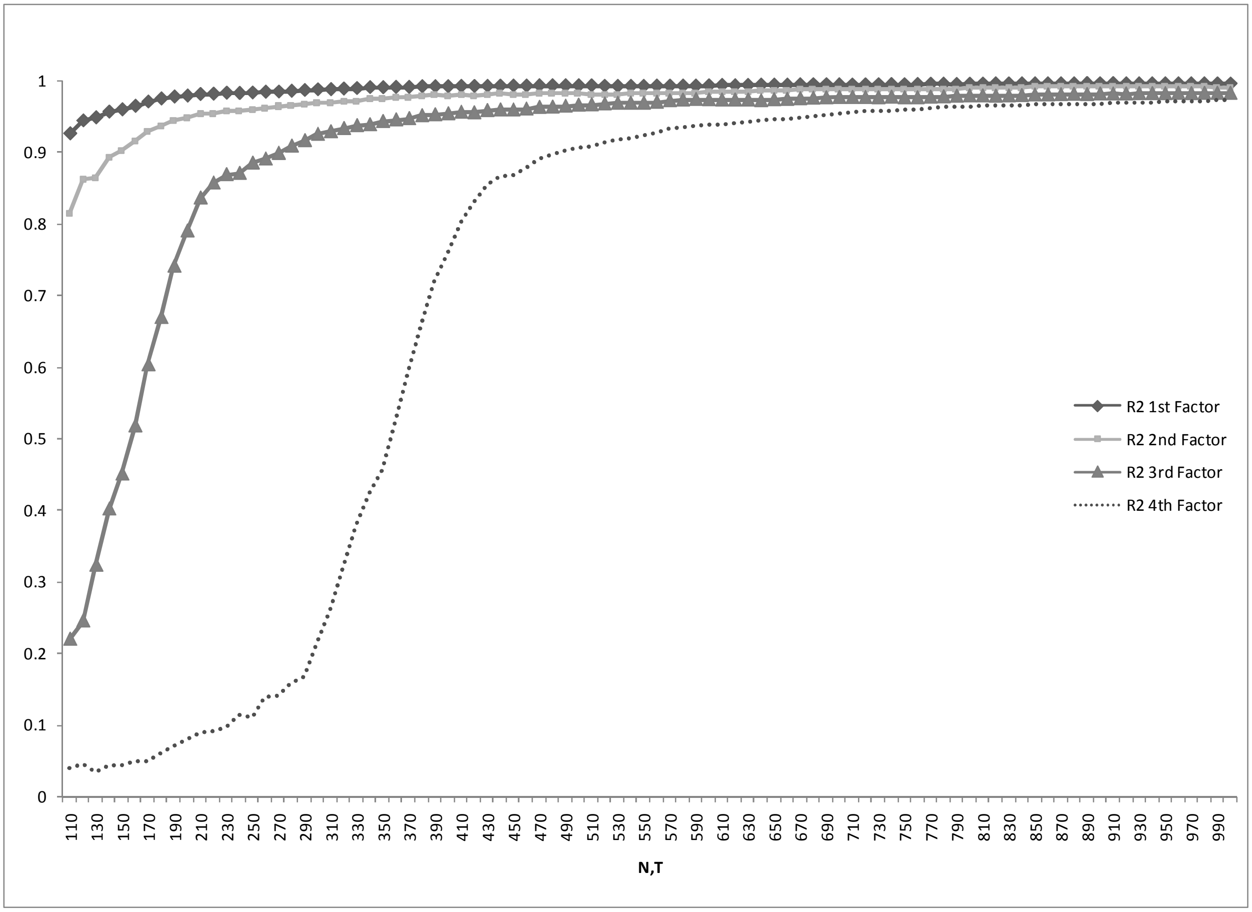

Figure 13

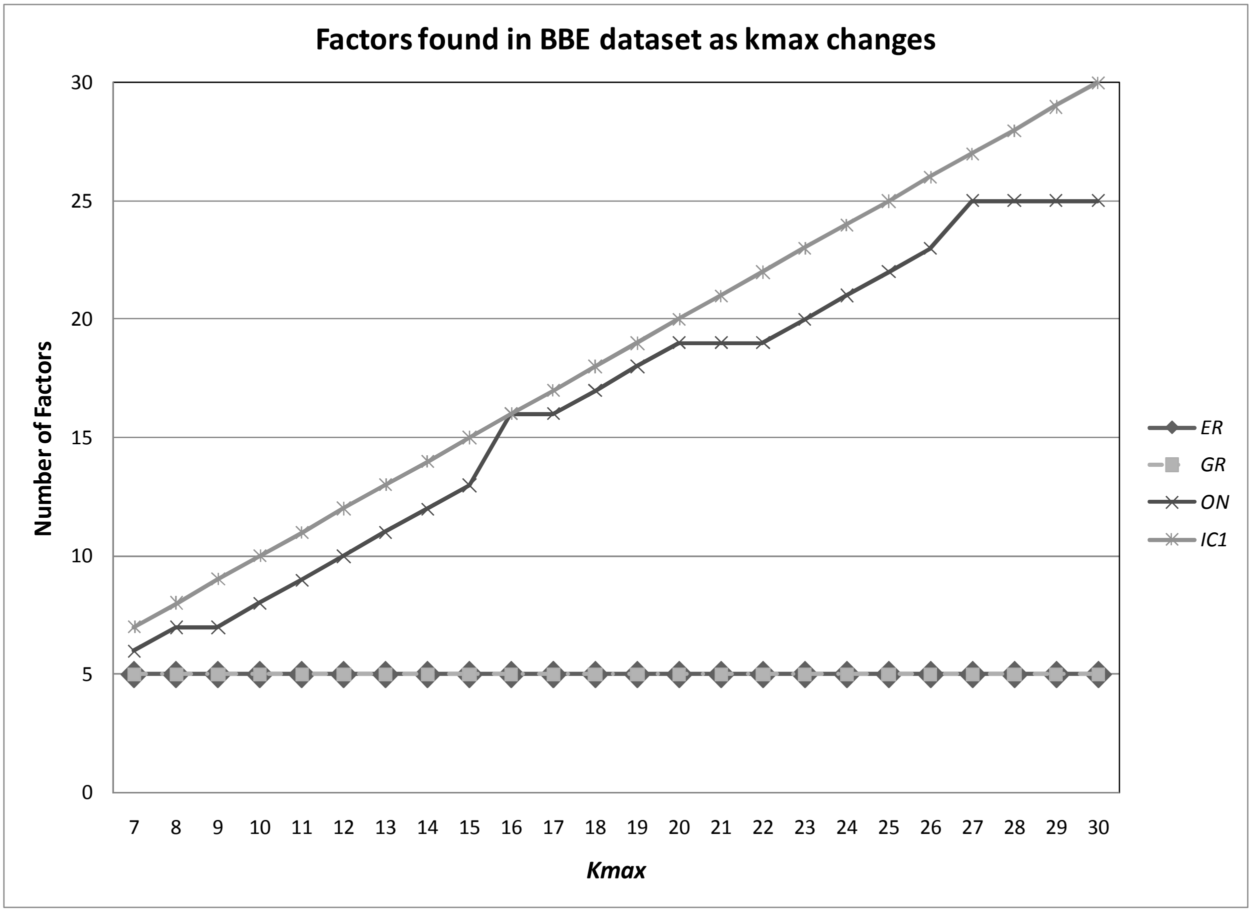
Factors found in BBE dataset as kmax changes
Number of Factors
0
5
10
15
20
25
30
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
Kmax
ER
GR
ON
IC1

Figure 14

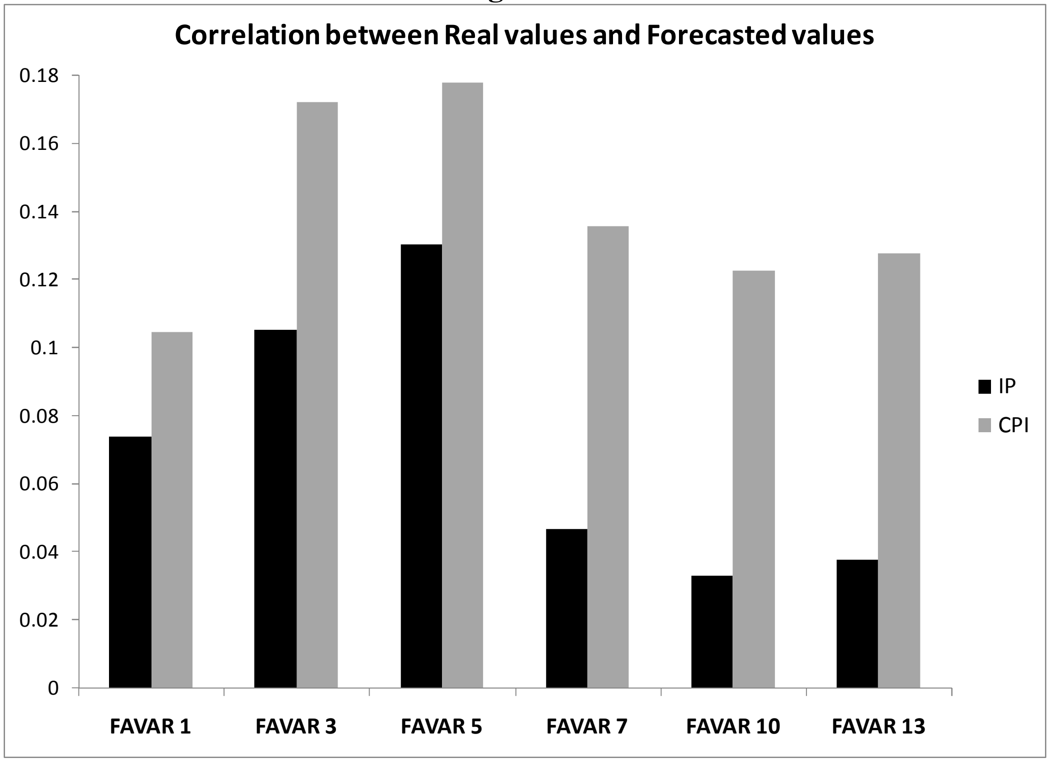

Figure 15