

Eigenvalue Ratio Test for the Number of Factors

By Seung C. Ahn and Alex R. Horenstein¹

This paper proposes two new estimators for determining the number of factors (r) in static approximate factor models. We exploit the well-known fact that the r largest eigenvalues of the variance matrix of N response variables grow unboundedly as N increases, while the other eigenvalues remain bounded. The new estimators are obtained simply by maximizing the ratio of two adjacent eigenvalues. Our simulation results provide promising evidence for the two estimators.

Key words: Approximate factor models, number of factors, eigenvalues.

1. INTRODUCTION

Recently, many estimation methods have been developed for the number of common factors in economic or financial data with both large numbers of cross-section units (N) and time series observations (T). Examples are Bai and Ng (2002), Onatski (2006, 2010), and Alessi, Barigozzi, and Capasso (2010) for static approximate factor models; and Forni, Hallin, Lippi, and Reichlin (2000), Hallin and Liska (2007), Amengual and Watson (2007), Bai and Ng (2007), and Onatski (2009) for dynamic factor models, among others. In this paper, we propose two alternative estimators for static factor models.

Bai and Ng (2002; hereafter BN) proposed to estimate the number of factors (r) by minimizing one of the two model selection criterion functions, named PC and IC. The BN estimators are linked to the eigenvalues of the second-moment matrix of N response variables (see, e.g., Onatski (2006)). Specifically, the PC estimator equals the number of the eigenvalues larger than a threshold value specified by a penalty function. An important contribution, among many, of BN is their finding that the convergence rates of the eigenvalues depend on $\min(N, T)$, and, therefore, the threshold value should be adjusted depending on both N and T .

¹We thank the editor and three anonymous referees for their numerous comments and suggestions that helped us improve the quality of the paper substantially. We also thank Jushan Bai, Alexei Onatski, Marcos Perez, Crocker Liu, Federico Nardari, Manuel Santos, Stephan Dieckmann, Na Wang, and Matteo Barigozzi for their helpful comments and/or sharing codes and data with us. The paper was presented in the econometrics seminars at Tokyo University, Kyoto University, Hitotsubashi University, the Korea Econometric Society Summer Meeting, Korea University, Wilfrid Laurier University, University of Southern California, Texas A&M University, Sam Houston State University, Bar Ilan University, Norwegian School of Economics and Business Administration, University of Alberta, Instituto Tecnológico Autónomo de México, and Seoul National University. We would like to thank the participants in the seminars. Finally, the second author also greatly acknowledges the financial support provided by the Asociación Mexicana de Cultura A.C. All remaining errors are, of course, our own.

There are, however, two issues that need to be addressed to improve the finite-sample properties of the BN estimators. The first issue is that the prespecified threshold functions are not unique (Hallin and Liska (2007)). Any finite multiple of a threshold function is also an asymptotically valid threshold function for consistent estimation of the number of factors. However, the finite-sample properties of the estimators could depend on the threshold function chosen among many alternatives. The second issue is that the BN estimators need to prespecify a maximum possible number of factors ($kmax$) to compute threshold values. Obviously, there are many possible choices for $kmax$. Thus, ideally, the estimators should not be overly sensitive to the choice of $kmax$. However, our simulation results indicate that the BN estimators are quite sensitive to the choice of $kmax$.

The first issue is related to the use of pre-specified threshold functions. Some recent studies have developed data-dependent methods for threshold values. An ideal threshold value would be a value slightly greater than the $(r+1)$ th largest eigenvalue. Onatski (2006) developed a consistent estimator of the $(r+1)$ th eigenvalue under the assumption that the idiosyncratic components of response variables are either autocorrelated or cross-sectionally correlated, but not both. Onatski (2010) also proposed an alternative estimator, named “Edge Distribution” (ED) estimator, which estimates the number of factors using differenced eigenvalues. Instead of estimating an asymptotically valid threshold for consistent estimation of the number of factors, Hallin and Liska (2007) proposed an alternative data-dependent method for general dynamic factor models that consists of two steps: tuning and stability checkup. They suggested estimating the number of factors using different subsamples and different multiples of the BN penalty functions (tuning). The final estimate is the estimate that is invariant to the subsamples used and the changes in the multiplicative constant of the penalty function in a certain range (stability checkup). Alessi, Barigozzi, and Capasso (2010) reported that the BN estimators obtained by this tuning-stability checkup procedure outperform the original estimators in finite samples.

In this paper, we propose two alternative estimators, which we name “Eigenvalue Ratio” (ER) and “Growth Ratio” (GR) estimators, respectively. They are easy to compute. In particular, the ER estimator is obtained simply by maximizing the ratio of two adjacent eigenvalues arranged in descending order. Our simulation results indicate that the finite-sample performances of the two estimators are promising. In most of the cases we consider, they outperform other competing estimators. One exception is the case in which one factor has extremely strong explanatory power for response variables. Even for this case, the GR estimator shows performances

comparable to those of other competing estimators. In addition, the performances of the two estimators are not sensitive to the choice of $kmax$ unless it is too large or too small.

This paper is organized as follows. Section 2 presents the assumptions consistent with approximate static factor models and shows that the proposed estimators are consistent. Section 3 reports our Monte Carlo experiments. Concluding remarks are given in Section 4.

2. ASSUMPTIONS AND ASYMPTOTIC RESULTS

We begin by defining the approximate factor model of Chamberlain and Rothschild (1983). Let x_{it} denote the response variable i ($= 1, \dots, N$) at time t ($= 1, \dots, T$). The variables are generated by an $r \times 1$ vector of factors, $f_t: x_{it} = \Lambda^o f_t + \varepsilon_{it}$, where $x_{it} = (x_{1t}, x_{2t}, \dots, x_{Nt})'$, $\Lambda^o = (\lambda_1^o, \lambda_2^o, \dots, \lambda_N^o)'$, λ_i^o is the $r \times 1$ vector of factor loadings for variable i , and $\varepsilon_{it} = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$ is the vector of the idiosyncratic components of response variables. The factors, factor loadings, and idiosyncratic components are not observed. We can describe the model for the complete panel data by

$$(1) \quad X = F\Lambda^{o'} + E,$$

where $X' = (x_{.1}, \dots, x_{.T})$, $F' = (f_1, \dots, f_T)$, and $E' = (\varepsilon_{.1}, \varepsilon_{.2}, \dots, \varepsilon_{.T})$. Following Bai and Ng (2002), we treat the entries in Λ^o as parameters and those in F as random variables.

We introduce some notation. We denote the norm of a matrix A as $\|A\| = [\text{trace}(A'A)]^{1/2}$. Two scalars, c_1 and c_2 , denote generic positive constants. For any real number z , $[z]$ denotes the integer part of z . We use $\psi_k(A)$ to denote the k^{th} largest eigenvalue of a positive semidefinite matrix A . With this notation, we define

$$\tilde{\mu}_{NT,k} \equiv \psi_k [XX' / (NT)] = \psi_k [X'X / (NT)].$$

Finally, we use $m = \min(N, T)$ and $M = \max(N, T)$.

Our assumptions on the factor model (1) are as follows.

Assumption A: (i) Let $\mu_{NT,k} = \psi_k [(\Lambda^{o'} \Lambda^o / N)(F'F / T)]$ for $k = 1, \dots, r$. Then, for each $k = 1, 2, \dots, r$, $p \lim_{m \rightarrow \infty} \mu_{NT,k} = \mu_k$, and $0 < \mu_k < \infty$. (ii) r is finite.

Assumption B: (i) $E\|f_t\|^4 < c_1$ and $\|\lambda_t^o\| < c_1$ for all i and t . (ii) $E\left(\|N^{-1/2}\sum_i \varepsilon_{it} \lambda_t^o\|^2\right) < c_1$ for all t . (iii) $E\left(N^{-1}\sum_{i=1}^N\|T^{-1/2}\sum_{t=1}^T \varepsilon_{it} f_t\|^2\right) = E\left[(NT)^{-1}\|E'F\|^2\right] < c_1$.

Assumption C: (i) $0 < y \equiv \lim_{m \rightarrow \infty} m/M \leq 1$. (ii) $E = R_T^{1/2} U G_N^{1/2}$, where $U' = [u_{it}]_{N \times T}$, and $R_T^{1/2}$ and $G_N^{1/2}$ are the symmetric square roots of $T \times T$ and $N \times N$ positive semidefinite matrices R_T and G_N , respectively. (iii) The u_{it} are independent and identically distributed (*i.i.d.*) random variables with uniformly bounded moments up to the 4th order. (iv) $\psi_1(R_T) < c_1$ and $\psi_1(G_N) < c_1$, uniformly in T and N , respectively.

Assumption D: (i) $\psi_T(R_T) > c_2$ for all T . (ii) Let $y^* = \lim_{m \rightarrow \infty} m/N = \min(y, 1)$. Then, there exists a real number $d^* \in (1 - y^*, 1]$ such that $\psi_{[d^*N]}(G_N) > c_2$ for all N .

Assumptions A–C are the same assumptions as those in Bai and Ng (2006) and Onatski (2010). Although Assumption C(ii) restricts covariance structure of the errors, it allows both autocorrelation and cross-sectional correlation in the errors.

Assumption D is a new assumption we impose. The matrix G_N governs the cross-section correlations among the errors, while R_T determines the structure of serial correlations. Assumption D(i) states that none of the idiosyncratic components and their linear functions can be perfectly predicted by their past values. Assumption D(ii) states that an asymptotically non-negligible number of the eigenvalues of G_N are bounded below by a positive number. Assumption D(ii) holds with $d^* = 1$ if response variables are not perfectly multicollinear and if none of them have zero idiosyncratic variances.

For macroeconomic or financial data, some variables may be perfectly or almost perfectly correlated with the others or may be factors themselves. An example is the macroeconomic data that contain detailed consumption data such as total consumption expenditure and categorized consumption expenditures for durable and nondurable goods and services. The total expenditure is the sum of the other categorized expenditures. For such data, the smallest eigenvalue of G_N may be close to zero (if logarithms of expenditures are analyzed) or exactly equal to zero (if

level data are used). Another example is the financial data covering both portfolio returns and individual stock returns. If a portfolio is constructed with the individual stocks included in the data, or if a portfolio return itself is a factor, the smallest eigenvalue of G_N should be zero. Assumption D(ii) permits such cases so long as an asymptotically non-negligible portion (d^*) of the eigenvalues of G_N are bounded below by a positive number.

For the data with $N \leq T$ (so that $m = N$ and $y^* = 1$), Assumption D(ii) only requires that $d^* > 0$. However, for the data with $T < N$ (so that $m = T$ and $y^* < 1$), d^* needs to be sufficiently large so that $d^* + y^* > 1$. This condition is likely to hold unless the ratio T/N is extremely small or a majority of variables are almost perfectly correlated (or their idiosyncratic components have near zero variances). For example, Assumption D(ii) holds if the number of time series observations (T) is more than a half of the number of cross-section units ($y^* > 0.5$) and if more than 50% of the cross-section variables are linearly independent and have non-negligible idiosyncratic components ($d^* > 0.5$).

We note that Assumptions C and D are sufficient, but not necessary, conditions for our main results. Weaker conditions sufficient for our results are

$$(2) \quad \psi_1(\mathbb{E}\mathbb{E}'/M) = O_p(1),$$

$$(3) \quad \psi_{[d^c m]}(\mathbb{E}\mathbb{E}'/M) \geq c + o_p(1),$$

for some positive and finite real number c and some $d^c \in (0, 1]$. The condition (2) rules out the possibility that the error matrix \mathbb{E} contains common factors. Bai and Ng (2006) have shown that Assumption C implies (2). The condition (3) indicates that the first largest $[d^c m]$ eigenvalues of $\mathbb{E}\mathbb{E}'/M$ are bounded away from zero. In the Appendix (Lemma A.9), it is shown that Assumptions C and D are sufficient for both (2) and (3).

We now turn to our estimators. A criterion function we use to estimate the number of factors (r) is simply the ratio of two adjacent eigenvalues of $XX'/(TN)$:

$$ER(k) \equiv \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}}, \quad k = 1, 2, \dots, kmax,$$

where ‘‘ER’’ refers to ‘‘eigenvalue ratio’’. Another criterion function we consider is given by

$$GR(k) \equiv \frac{\ln[V(k-1)/V(k)]}{\ln[V(k)/V(k+1)]} = \frac{\ln(1 + \tilde{\mu}_{NT,k}^*)}{\ln(1 + \tilde{\mu}_{NT,k+1}^*)}, \quad k = 1, 2, \dots, kmax,$$

where $V(k) = \sum_{j=k+1}^m \tilde{\mu}_{NT,j}$ and $\tilde{\mu}_{NT,k}^* = \tilde{\mu}_{NT,k} / V(k)$. Here, $V(k)$ equals the sample mean of the squared residuals from the time series regressions of individual response variables on the first k principal components of $XX' / (TN)$ (see Onatski (2006)). The term GR refers to ‘‘Growth Ratio’’ because both the numerator and denominator of $GR(k)$ are the growth rates of residual variances as one fewer principal component is used in the time series regressions. The estimators of r we propose are simply the maximizers of $ER(k)$ and $GR(k)$, which we call ‘‘ER’’ and ‘‘GR’’ estimators, respectively:

$$\tilde{k}_{ER} = \max_{1 \leq k \leq kmax} ER(k); \quad \tilde{k}_{GR} = \max_{1 \leq k \leq kmax} GR(k).$$

Our main result follows.

Theorem 1: Suppose that Assumptions A–D hold with $r \geq 1$. Then, there exists $d^c \in (0,1]$ such that $\lim_{m \rightarrow \infty} \Pr(\tilde{k}_{ER} = r) = \lim_{m \rightarrow \infty} \Pr(\tilde{k}_{GR} = r) = 1$, for any $kmax \in (r, [d^c m] - r - 1]$.²

While a formal proof of the theorem is given in the Appendix, a brief sketch of proof provides some explanation. As discussed above, Assumptions C and D are sufficient for the two conditions (2) and (3) to hold. That is, the first $[d^c m]$ largest eigenvalues of $EE' / (NT)$ are $O_p(m^{-1})$, and the ratios of two adjacent eigenvalues are $O_p(1)$. The first r eigenvalues of $XX' / (NT)$ are asymptotically determined by the eigenvalues of $F\Lambda^o\Lambda^oF' / (NT)$ and other eigenvalues by the eigenvalues of $EE' / (NT)$. Accordingly, $\tilde{\mu}_{NT,j} / \tilde{\mu}_{NT,j+1} = O_p(1)$ for $j \neq r$, and $\tilde{\mu}_{NT,r} / \tilde{\mu}_{NT,r+1} = O_p(m)$. That is, while the ratio of the r^{th} and $(r+1)^{\text{th}}$ eigenvalues of $XX' / (TN)$ diverges to infinity, all other ratios of two adjacent eigenvalues are asymptotically bounded.

The possibility of zero factor ($r = 0$) can be allowed by using slightly modified $ER(k)$ and $GR(k)$ criterion functions. Let us define a mock eigenvalue $\tilde{\mu}_{NT,0} = w(N, T)$ such that $w(N, T) \rightarrow 0$ and $w(N, T)m \rightarrow \infty$ as $m \rightarrow \infty$. Then, we obtain the following result:

²The ER estimator can be viewed as a BN estimator using an estimated threshold value, $\tilde{\mu}_{NT,k+1}$ with $k = \tilde{k}_{ER}$. We thank an anonymous referee for providing this interpretation.

Corollary 1: Redefine \tilde{k}_{ER} and \tilde{k}_{GR} using $\tilde{\mu}_{NT,0}$ for $k = 0$. Then, under Assumptions A – D with $r \geq 0$, $\lim_{m \rightarrow \infty} \Pr[\tilde{k}_{ER} = r] = \lim_{m \rightarrow \infty} \Pr[\tilde{k}_{GR} = r] = 1$.

This corollary holds for any multiple of $w(N, T)$. Accordingly, the finite-sample properties of the modified ER and GR estimators depend on the choice of the multiple and the functional form of $w(N, T)$. Fortunately, our simulation experiments show that estimation results are not excessively sensitive to the choice of the mock eigenvalue. The mock eigenvalue used for our simulations is

$$(4) \quad \tilde{\mu}_{NT,0} = V(0) / \ln(m) = \sum_{k=1}^m \tilde{\mu}_{NT,k} / \ln(m).$$

We have found that while the ER and GR estimators perform better with some other choices of the mock value, the improvement is not substantial.

Theorem 1 and Corollary 1 indicate that $kmax$ can be chosen to increase with $m = \min(T, N)$. This requirement is less restrictive than the condition, $kmax/m \rightarrow 0$ as $m \rightarrow \infty$, that is required for the ED estimator of Onatski (2010). In practice, however, we do not recommend that researchers use an excessively large value for $kmax$ so as to avoid the danger of choosing a value smaller than r . We suggest two possible choices for $kmax$. First, Theorem 1, as well as our finding from simulations, suggests that it should not be a problem to choose much a larger $kmax$ than r . Thus, if one has *a priori* information about a possible maximum (fixed) number of factors, say r_{max} , she could use $kmax_1 = 2r_{max}$ for $kmax$. So long as r_{max} is fixed, the ER and GR estimators computed with $kmax_1$ must be consistent. Second, when such information is not available, one may consider using a sequence, $kmax_2 = \min(kmax^*, 0.1m)$, where $kmax^* = \#\{k \mid \tilde{\mu}_{NT,k} \geq V(0)/m, k \geq 1\}$. As shown in the Appendix, $V(0) = O_p(1)$ and $m\tilde{\mu}_{NT,k} = O_p(m)$ for $k = 1, \dots, r$. Thus, $\Pr(kmax^* \leq r) \rightarrow 0$ as $m \rightarrow \infty$. Accordingly, if $d^c > 0.1$, $kmax_2$ satisfies all of the conditions that warrant the consistency of the ER and GR estimators.

Our results apply to a factor model with time and/or individual effects:

$$(5) \quad x_{it} = \alpha_i + \delta_t + f_t' \lambda_i^o + \varepsilon_{it},$$

where α_i is an individual-specific effect and δ_t is a time-specific effect. The two effects can be controlled by subtracting from the x_{it} their time and individual means and adding their overall mean. The ER and GR estimators applied to these demeaned data are still consistent with a small adjustment for the possible range of $kmax$.

Even for the data without time or individual effects, we suggest that practitioners estimate the number of factors using demeaned data. Brown (1989) has found that for the data (with small N and large T) generated by four factors of the same explanatory power, the tests based on eigenvalues tend to predict only one factor. To obtain an intuition for his result, consider a simple case in which $F'F/T = \tilde{\Lambda}'\tilde{\Lambda}'/N = I_r$ for all T and N , where $\bar{\lambda}' = N^{-1}\Sigma_i\lambda_i'$, $\tilde{\Lambda}' = \Lambda' - 1_N\bar{\lambda}'$, and 1_N is an N -vector of ones. Observe that $F\Lambda'\Lambda'F' = NF\bar{\lambda}'\bar{\lambda}'F' + F\tilde{\Lambda}'\tilde{\Lambda}'F'$. For this case, we can easily show (using Lemmas A.5 and A.6 in the Appendix) that

$$(6) \quad \bar{\lambda}'\bar{\lambda}' \leq \psi_1 \left[F\Lambda'\Lambda'F'/(NT) \right] \leq \bar{\lambda}'\bar{\lambda}' + 1,$$

$$(7) \quad \psi_k \left[F\Lambda'\Lambda'F'/(NT) \right] = \psi_k \left[F\tilde{\Lambda}'\tilde{\Lambda}'F'/(NT) \right] = 1, \quad k = 2, \dots, r.$$

The first r eigenvalues of XX' mainly depend on the eigenvalues of $F\Lambda'\Lambda'F'$. Thus, (6) implies that the first eigenvalue of $XX'/(NT)$ must be asymptotically bounded below by $\bar{\lambda}'\bar{\lambda}'$, while the probability limits of the next $(r-1)$ eigenvalues are all ones. Thus, we can expect that the ER and GR estimators are likely to predict one factor in small samples when the means of factor loadings deviate from zeros substantially. This problem is alleviated if demeaned data are used. To see why, suppose we use demeaned data $(x_{it} - N^{-1}\Sigma_i x_{it})$ for X instead of raw data x_{it} . Then, the ER and GR estimators are obtained from the eigenvalues of $XQ_N X'/(NT)$, where $Q_N = I_N - N^{-1}1_N 1_N'$. The first r eigenvalues of $XQ_N X'/(NT)$ now depend on the eigenvalues of $F\Lambda'Q_N\Lambda'F'/(NT) = F\tilde{\Lambda}'\tilde{\Lambda}'F'/(NT)$, which are all ones.

The one-factor bias problem identified by Brown (1989) also arises when the factor means deviate from zeros by a large margin. Thus, it is recommended to use doubly demeaned data, that is, $x_{it} - T^{-1}\Sigma_t x_{it} - N^{-1}\Sigma_i x_{it} + (NT)^{-1}\Sigma_{i,t} x_{it}$, for better results from our estimation methods. By some unreported simulations, we have found that the ER and GR estimators often predict one factor in small samples when the means of factor loading and/or the means of factors are large in absolute value. This problem disappears if demeaned data are used.³

³The time effect δ_t itself can be viewed as a factor with constant loadings. The time effect can be estimated by the time mean of response variables, $\bar{x}_t = N^{-1}\Sigma_i x_{it}$. If the mean has significant explanatory power for individual response variables, it should be used as an estimated factor.

Finally, we note two cases in which use of the ER and GR estimators may be inappropriate. The first is the case in which some factors are $I(1)$ while the others are $I(0)$, and the second is the case in which some factors have dynamic factor loadings of infinite order (generalized factor model). The first case is a case violating Assumption A(i). For this case, the ER or GR estimators may pick up only the $I(1)$ factors. Thus, when some factors are suspected to be $I(1)$, the number of factors can be estimated with first differenced data as suggested by Bai and Ng (2004). The second case violates Assumption A(ii). Hallin and Liska (2007) estimated the number of dynamic factors applying the BN estimation methods (with a tuning-stability checkup procedure) to the spectral density matrix of response variables. Although not pursued here, it might be interesting to investigate whether the ER and GR methods can be generalized to estimation of the number of dynamic factors.

3. SIMULATIONS AND RESULTS

The foundation of our simulation exercises is the following model:

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

where $e_{it} = \rho e_{i,t-1} + v_{it} + \sum_{h=\max(i-J,1)}^{i-1} \beta v_{ht} + \sum_{h=i+1}^{\min(i+J,N)} \beta v_{ht}$, and the v_{ht} and λ_{ij} are all drawn from $N(0,1)$. The factors f_{jt} are drawn from normal distributions with zero means. Bai and Ng (2002) and Onatski (2010) have used the same data generating process. The only exception is that we normalize the idiosyncratic components (errors) u_{it} so that their variances are equal to 1 for most of the cross-section units (more specifically, $J+1 \leq i \leq N-J$).

The control parameter θ is the inverse of the signal to noise ratio (SNR) of each factor when $\text{var}(f_{jt}) = 1$ because $1/\theta = \text{var}(f_{jt}) / \text{var}(\sqrt{\theta} u_{it})$. When it is necessary to change SNRs of all factors, we adjust the value of θ while fixing variances of factors at 1. To change SNR of a single factor, we adjust the variance of the factor with θ fixed at 1. The magnitude of the time series correlation is specified by the control parameter ρ . Cross-sectional correlation is governed by two parameters: β specifies the magnitude of cross sectional correlation and J specifies the number of cross-section units correlated.

Our simulations are categorized into four parts. The first part is designed to investigate how error covariance structure influences the finite-sample performances of the ER and GR estimators. Data are generated with errors of four different covariance structures: (a) *i.i.d.* errors

($\rho = \beta = J = 0$); (b) serially correlated errors ($\rho = 0.7$ and $\beta = J = 0$); (c) cross-sectionally correlated errors ($\rho = 0$, $\beta = 0.5$, and $J = \max(10, N/20)$); and (d) both serially and cross-sectionally correlated errors ($\rho = 0.5$, $\beta = 0.2$, and $J = \max(10, N/20)$).

In the second part, we examine the effects of weak factors on the estimators. We consider two cases. The first is the case in which all three factors have weak explanatory power ($\text{SNR} = 0.17$). The second is the case in which two factors are strong ($\text{SNR} = 1$) and one factor is weak ($\text{SNR} < 1$).

In the third part, we investigate how the use of large k_{\max} may influence estimation results when the eigenvalues $\tilde{\mu}_{NT,k}$ are close to zero for some large k ($< m$). As discussed in Section 2, this could happen if many response variables are highly multicollinear or if many response variables have very small idiosyncratic variations. These cases are related to the case in which d^* in Assumption D is smaller than 1. If too large a value of k_{\max} is used for such data, the ER and GR estimators may over estimate the true number of factors because the ratios $ER(k)$ and $GR(k)$ may explode for some $k > r$. We examine this possibility using the data generated with heteroskedastic errors.

The fourth and final part of our simulations considers the case in which one factor has a dominantly strong explanatory power. For such a case, the value of $ER(k)$ and $GR(k)$ may peak at $k = 1$. We examine how large a difference in the explanatory power of two factors is needed to make the ER and GR estimators underestimate the true number of factors. To do so, we generate data using two factors with different SNRs.

For each case we consider, we compute root mean squared errors (RMSEs) or frequencies of incorrect estimation by estimators, from 1,000 simulated data sets. The modified ER and GR estimators introduced in Corollary 1 are used for our simulations. Although the means of factors and factor loadings are all zero in our data generating process, we use doubly demeaned data to compute ER and GR estimators, to be consistent with our suggestions in Section 2. The performances of the two estimators are compared with those of the BIC3 estimator of BN and the ED estimator of Onatski (2010).⁴ We also consider the estimator by Alessi, Barigozzi, and Capasso (2010; hereafter, ABC), which is the IC1 estimator of BN with the tuning-stability

⁴Professor Jushan Bai kindly suggested that we consider the BIC3 estimator in simulations. We only report the performances of the BIC3 estimator, because the estimator outperforms the other BN estimators in our simulations.

checkup procedure of Hallin and Liska (2007). The BIC3, ED, and ABC estimators are computed with raw data (not with demeaned data).

Figure 1 reports the results from the first part of our simulations. Three factors ($r = 3$) are drawn from $N(0,1)$ and θ is fixed at 1. Thus, all factors have SNRs equal to 1. Sample size (N,T) increases from $(25,25)$ to $(200,200)$. Panel A shows the results from the data generated with *i.i.d.* errors. The results from the BN and ED estimators are essentially the same as the benchmark results reported in both Bai and Ng (2002) and Onatski (2010). The BIC3 estimator outperforms other estimators for the data with $N = T \leq 50$, and shows perfect accuracy for the data with $N = T \geq 50$. For the data with $N = T = 25$, the ED and ABC estimators outperform the ER and GR estimators, but the latter two estimators perform better for the data with $N = T \geq 50$.

Panels B, C and D report the estimation results from the data with serially or/and cross-sectionally correlated errors. For the cases with $N = T \geq 75$, the ER and GR estimators perform equally to or better than the other estimators. When the errors are cross-sectionally correlated, the BIC3 estimator overestimates the correct number of factors even if large samples are used. It appears that the performance of the BIC3 estimator is much more sensitive to cross-sectional correlation than autocorrelation in the errors. The ABC estimator clearly outperforms the BIC3 estimator when errors are cross-sectionally correlated.

Figure 2 reports the results from the second part of our simulations. The figure shows the finite-sample performance of each estimator when all or some factors have weak explanatory power (low SNRs). Comparing Panel A of Figure 2 and Panel D of Figure 1, we can see that all of the estimators have lower power to detect weak factors. The ER and GR estimators no longer show perfect accuracy for the sample sizes reported, but they still outperform the other estimators. Panel B of Figure 2 reports the estimation results from data with $N = T = 100$ and with two strong factors and one weak factor. The first two factors are drawn from $N(0,1)$, and the other, from $N(0, SNR3)$, where $0 < SNR3 < 1$. The value of θ is set at 1. Thus, the SNRs of the first two factors are equal to 1, while that of the third factor equals $SNR3$. We try many different values for $SNR3$ (0.45 - 0.10). As in Panel A, the ER and GR estimators outperform other estimators for any value of $SNR3$.

So far, we have reported the estimation results obtained using $kmax = 8$. Figure 3 shows how the choice of $kmax$ may influence the performances of the estimators. Six different values are used for $kmax$. The data generating process is the same as the one used for Panel D of Figure 1 with $N = T = 150$. Figure 3 shows that the performances of the ER and GR estimators are not

sensitive to $kmax$. In contrast, the RMSE of the BIC3 estimator increases with $kmax$. This is because the bias in the BIC3 estimator increases with $kmax$ (although not shown in the figure). The RMSE of the ABC (ED) estimator also increases until $kmax = 12$ (16). The ED and ABC estimators are less sensitive to $kmax$ than the BIC3 estimator.

The third part of our simulations examines how the use of a large $kmax$ may influence the finite-sample properties of the ER and GR estimators when many response variables have small idiosyncratic variations. As before, the data are generated from a three-factor model ($r = 3$) with both serially and cross-sectionally correlated errors and with $N = T = 150$. For the first half of the cross-section units, we fix their error variances at 1; $\text{var}(u_{it}) = 1$, for $i \leq 75$. However, for the second half, error terms are generated with variances equal to $V2$ ($\text{var}(u_{it}) = V2$, for $i \geq 76$), where $V2$ varies from 0.5 to 0.001. In our setup, $V2 = 0.001$ means that the idiosyncratic variances of the first half of response variables are 1,000 times greater than those of the second half. We also vary the explanatory power of three factors by using six different values for θ , from 1 to 6. We choose $kmax = 100$ to make sure that the heteroskedasticity structure we use for simulations can influence the performances of the ER and GR estimators.

For each possible combination of $V2$ and θ , we compute the frequency of incorrect estimation by each estimator. The results are reported in Figure 4. Panels A and B show that the accuracies of the ER and GR estimators remain fairly stable when $V2$ changes. Panels C and D show that the ED and ABC estimators miss the correct number of factors in every case when $V2 < 0.5$. In contrast, as θ increases, the accuracies of the ER and GR estimators fall for any level of $V2$. These results indicate that using too large a value for $kmax$ can hurt the performances of the ER and GR estimators when some response variables have very small idiosyncratic variations. The seriousness of this problem, however, depends on the explanatory power of factors. When factors are reasonably strong (*e.g.*, $\theta \leq 3$), use of large $kmax$ would have only limited effects on the ER and GR estimators, unless too many response variables (or linear combinations of them) have extremely small idiosyncratic variations. Intuitively, however, if a larger value of $kmax$ is used for actual data analysis, it is increasingly more likely that the value of an eigenvalue $\tilde{\mu}_{NT,k}$ drops substantially at some value of k greater than r . To mitigate this possibility, it is important to avoid choosing an excessively large value for $kmax$.

We now turn to the fourth and final part of our simulations. We consider a two-factor model ($r = 2$) in which both factors have strong explanatory power, but one factor's power is

increasingly dominant. The two factors are drawn from $N(0,1)$ and $N(0,SNR2)$, respectively, where $SNR2$ is an integer between 1 and 20. The simulation results are reported in Figure 5.

The GR estimator performs better than the ER estimator, especially when $SNR2$ is large. For example, although not shown clearly in Figure 5, when $N = T = 150$, the ER estimator captures the true number of factors more than 90% until $SNR2 \leq 5$, while the GR estimator does until $SNR2 = 20$. In our simulation setup, when $SNR2 = 20$, the average R -squared from regressions of individual response variables on the second factor alone is about 0.90. This is an extreme case that is unlikely to happen in actual data analysis. For less extreme cases ($SNR2 < 20$), the GR estimator performs quite well when data are sufficiently large ($N = T \geq 150$).

Figure 5 shows that the accuracies of the ED and ABC estimators are not affected by difference in explanatory power between the two factors. This is an expected result because both the estimators determine the number of factors comparing the eigenvalues $\tilde{\mu}_{NT,k}$ with given threshold values. Large differences among the first r eigenvalues have little impact on these estimators. In addition, the ED and ABC estimators outperform the ER estimator when $SNR2$ is very large. Indeed, the cases with large differences in the explanatory power of factors are the only cases we found from all of our reported and unreported simulations in which the ED and ABC estimators outperform the ER estimator. However, the performance of the GR estimator is comparable to, if not better, those of the ED and ABC estimators, unless one factor is unrealistically dominant. The GR estimator uses logarithmic functions of eigenvalues, not eigenvalues directly. It appears that use of logarithmic functions mitigates the effect of the dominant factor.

Figures 1–5 show that the ER and GR estimators are generally better estimators when the same $kmax$ is used for all estimators. The last but important question is what $kmax$ should be used for the ER and GR estimators if the information about a possible maximum number of factors (r_{max}) is not available. We have suggested using $kmax_2$ in the previous section. When we repeat the simulations reported in Figure 1 with $kmax_2$ (not reported here), the performances of the two estimators remain the same. When the simulations reported in Figure 3 are repeated, the two estimators are perfectly accurate. However, we found from some unreported simulations that with $kmax_2$, the estimators tend to overestimate the number of factors when factors' SNRs are low ($\theta > r$) and (not or) the degree of cross-sectional correlation is high ($\beta \geq 0.2$). For such cases, the estimation results are sensitive to the choice of $kmax$. Fortunately, applying the ER

and GR estimators to the macro data of Bernanke, Boivin, and Eliasz (2005) and other stock return data, we found that the estimation results were insensitive to $kmax$. This result is consistent with the notion that idiosyncratic components in the data we analyzed are not too highly cross-sectionally correlated or factors are relatively strong. Overall, the results from our simulations and actual data analysis provide positive evidence for the use of $kmax_2$.

4. CONCLUDING REMARKS

In this paper, we have introduced two new estimators, ER and GR, for the number of common factors in approximate factor models. The estimators are easy to compute. Some simulation experiments are conducted to compare the performances of the estimators with those of the estimators by Bai and Ng (2002), Onatski (2010), and Alessi, Barigozzi, and Capasso (2010). The simulation results indicate that the ER and GR estimators generally outperform these competing estimators, especially when the idiosyncratic components of response variables are both cross-sectionally and serially correlated. When a dominant factor (in terms of explanatory power) exists, the ER estimator might not perform well. However, the GR estimator performs well unless a dominant factor has unrealistically high explanatory power.

Dep. of Economics, Arizona State University, Tempe, AZ 85287, U.S.A.; and Sogang University, South Korea; miniahn@asu.edu

and

Dept. of Economics, University of Miami, Coral Gables, FL 33124, USA and, Dept. of Business, Instituto Tecnológico Autónomo de México, México, 1080; horenstein@bus.miami.edu

APPENDIX

The following lemmas are useful to prove Theorem 1.

Lemma A.1: Under Assumption C,

$$p \lim_{m \rightarrow \infty} \psi_1(UU' / M) = (1 + \sqrt{y})^2; \quad p \lim_{m \rightarrow \infty} \psi_m(UU' / M) = (1 - \sqrt{y})^2.$$

Proof: See Bai and Yin (1993).

Lemma A.2: For a given $b \in (0,1]$, let $U_{[bm]}$ be the $[bm] \times N$ major submatrix (upper block) of U . Then, under Assumption C,

$$p \lim_{m \rightarrow \infty} \psi_{[bm]}(U_{[bm]}U'_{[bm]} / N) = \left(1 - \sqrt{by^*}\right)^2.$$

Proof: The result follows by Lemma A.1 and the fact that $\lim_{m \rightarrow \infty} [bm] / N = by^*$.

Lemma A.3: Let W_n be an $n \times n$ symmetric matrix; and W_{n-k} be an $(n-k) \times (n-k)$ major submatrix of W_n , where $k \leq p$. Then, $\psi_{n-p}(W_{n-p}) \leq \psi_{n-p}(W_n)$.

Proof: $\psi_{n-p}(W_{n-p}) \leq \psi_{n-p}(W_{n-p+1}) \leq \dots \leq \psi_{n-p}(W_{n-1}) \leq \psi_{n-p}(W_n)$, where each inequality is due to Sturmian Separation Theorem (Rao (1973), p. 64).

Lemma A.4: Suppose that A and B are $p \times p$ positive definite and positive semi-definite matrices, respectively. Then, for any $j+k-1 \leq i$,

$$\psi_i(AB) \leq \psi_j(A)\psi_k(B); \quad \psi_{p-j+1}(A)\psi_{p-k+1}(B) \leq \psi_{p-i+1}(AB).$$

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

Lemma A.5: If A and B are $p \times p$ symmetric matrices,

$$\psi_{j+k-1}(A+B) \leq \psi_j(A) + \psi_k(B), \quad j+k \leq p+1.$$

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

Lemma A.6: If A and B are $p \times p$ positive semi definite matrices,

$$\psi_j(A) \leq \psi_j(A+B), \quad j=1, \dots, p.$$

Proof: First, consider the case of $j=1$. Let ξ_A^1 be the eigenvector corresponding to $\psi_1(A)$. Then, $\psi_1(A) = \xi_A^{1'} A \xi_A^1 / \xi_A^{1'} \xi_A^1 \leq \xi_A^{1'} (A+B) \xi_A^1 / \xi_A^{1'} \xi_A^1 \leq \psi_1(A+B)$, where the first inequality is due to B being positive semi definite. 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 $p \times 1$ nonzero vector. Then,

$$\psi_j(A) \leq \sup_{\Xi^{j-1} z = 0} z' A z / z' z \leq \sup_{\Xi^{j-1} z = 0} z' (A+B) z / z' z = \psi_j(A+B),$$

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

Lemma A.7: Under Assumptions C and D, choose real numbers b and v such that $b, v \in (0,1)$ and $d^c \equiv d^* + b(y^* - v) - 1 > 0$. Then, for sufficiently large m ,

$$c_2^2(N/M)\psi_{[bm]}(U_{[bm]}U'_{[bm]}/N) \leq \psi_{[d^c m]}(EE'/M) \leq \psi_1(EE'/M) \leq c_1^2\psi_1(UU'/M).$$

Proof: Lemma A.4 and Assumption C imply

$$\psi_1(EE'/M) \leq \psi_1(UU'/M)\psi_1(G_N)\psi_1(R_T) \leq c_1^2\psi_1(UU'/M).$$

For a moment, assume that for sufficiently large m ,

$$(8) \quad [d^c m] \leq [d^* N] + [bm] - N \leq m.$$

Under this assumption, using Lemmas A.4 and A.3, we can show that

$$\begin{aligned} \psi_{[d^c m]}(EE'/M) &\geq \psi_{[d^* N] + [bm] - N}(UG_N U' R_T / M) \geq \psi_{[d^* N] + [bm] - N}(UG_N U' / M)\psi_T(R_T) \\ &\geq \psi_{[bm]}(UU'/M)\psi_{[d^* N]}(G_N)\psi_T(R_T) \\ &\geq \psi_{[bm]}(U_{[bm]}U'_{[bm]})\psi_{[d^* N]}(G_N)\psi_T(R_T) \geq c_2^2(N/M)\psi_{[bm]}(U_{[bm]}U'_{[bm]}/N). \end{aligned}$$

Thus, we can complete the proof by showing (8). We replace $[\bullet]$ by its inside argument (e.g., $[d^* N]$ by $d^* N$) without loss of generality. If $m = N \leq T$ ($y^* = 1$), (8) immediately follows.

Suppose now that $m = T < N$. By Assumption D, there exists $m_v \in \mathbb{N}$, such that $|y^* - (T/N)| < v$ for all $m \geq m_v$. Thus, for $m \geq m_v$,

$$d^c m \leq d^c N = d^* N + b(y^* - v)N - N \leq d^* N + bT - N \leq [d^* N] + [bm] - N < m.$$

Lemma A.8: Under Assumptions A, C, and D, for sufficiently large m and $j \leq [d^c m] - 2r$,

$$c_2^2(N/(mM))\psi_{[bm]}(U_{[bm]}U'_{[bm]}/N) \leq \tilde{\mu}_{NT,r+j} \leq c_1^2 m^{-1} \psi_1(UU'/M).$$

Proof: Let $P(\Lambda^o) = \Lambda^o(\Lambda^{o'}\Lambda^o)^{-1}\Lambda^{o'}$ and $Q(\Lambda^o) = I_N - P(\Lambda^o)$. Let $F^* = F + E\Lambda^o(\Lambda^{o'}\Lambda^o)^{-1}$ so that $XX' = F^*\Lambda^{o'}\Lambda^o F^{*'} + EQ(\Lambda^o)E'$. Since $\text{rank}(F^*\Lambda^{o'}\Lambda^o F^{*'}) \leq r$, $\psi_{r+1}(F^*\Lambda^{o'}\Lambda^o F^{*'}) = 0$. Thus, using Lemmas A.6 and A.5, we can show that

$$(9) \quad \psi_{r+j}(EQ(\Lambda^o)E') \leq \psi_{r+j}(XX') \leq \psi_j(EQ(\Lambda^o)E') + \psi_{r+1}(F^*\Lambda^{o'}\Lambda^o F^{*'}) = \psi_j(EQ(\Lambda^o)E').$$

Using the same lemmas, we can also show that

$$(10) \quad \psi_j(EQ(\Lambda^o)E') \leq \psi_j(EQ(\Lambda^o)E' + EP(\Lambda^o)E') = \psi_j(EE'),$$

$$(11) \quad \psi_{2r+j}(\mathbf{E}\mathbf{E}') \leq \psi_{r+j}(\mathbf{E}Q(\Lambda^\circ)\mathbf{E}') + \psi_{r+1}(\mathbf{E}P(\Lambda^\circ)\mathbf{E}') = \psi_{r+j}(\mathbf{E}Q(\Lambda^\circ)\mathbf{E}'),$$

because $\text{rank}(\mathbf{E}P(\Lambda^\circ)\mathbf{E}') \leq r$. Equations (9) – (11) imply that

$$(12) \quad \psi_{2r+j}(\mathbf{E}\mathbf{E}' / (NT)) \leq \tilde{\mu}_{NT,r+j} \leq \psi_j(\mathbf{E}\mathbf{E}' / (NT)), \text{ for } j = 1, \dots, m - 2r.$$

Lemma A.7 and (12) imply the result.

Lemma A.9: Under Assumptions A, C, and D, for $j = 1, \dots, [d^c m] - 2r$,

$$\underline{c} + o_p(1) \leq m\tilde{\mu}_{NT,r+j} \leq \bar{c} + o_p(1),$$

where $y^{**} = \lim_{m \rightarrow \infty} (N/M)$, $\underline{c} = c_2^2 y^{**} (1 - \sqrt{by^*})^2$, and $\bar{c} = c_1^2 (1 + \sqrt{y})^2$.

Proof: The result immediately follows from Lemmas A.8, A.1, and A.2.

Lemma A.10: Under Assumptions A and B, for any $A_{T \times p} = (a_1, \dots, a_p)$ such that $A'A = TI_p$,

$$\frac{1}{T^2 N} \left| \text{trace}(A' F \Lambda^\circ \mathbf{E}' A) \right| = O_p(N^{-1/2}), \quad \text{trace} \left(\frac{1}{T^2 N} A' \mathbf{E} P(\Lambda^\circ) \mathbf{E}' A \right) = O_p(N^{-1}).$$

Proof: Observe that

$$\begin{aligned} \left| \text{trace}(A' F \Lambda^\circ \mathbf{E}' A) \right| &\leq \|AA' F\| \|\Lambda^\circ \mathbf{E}'\| \leq \|A\|^2 \|F\| \|\Sigma_i \lambda_i^\circ \varepsilon_i'\|, \\ \|\Sigma_i \lambda_i^\circ \varepsilon_i'\| &\leq \sqrt{\text{trace} \left((\Sigma_i \lambda_i^\circ \varepsilon_i') (\Sigma_i \varepsilon_i \lambda_i^{\circ'}) \right)} = \sqrt{\text{trace} \left(\Sigma_{i,j} \Sigma_i \lambda_i^\circ \varepsilon_{it} \varepsilon_{jt} \lambda_j^{\circ'} \right)} \\ &= \sqrt{\text{trace} \left(\Sigma_t (\Sigma_i \lambda_i^\circ \varepsilon_{it}) (\Sigma_j \varepsilon_{jt} \lambda_j^{\circ'}) \right)} = \sqrt{\Sigma_t \|\Sigma_i \lambda_i^\circ \varepsilon_{it}\|^2}. \end{aligned}$$

Thus, we have

$$\frac{1}{T^2 N} \left| \text{trace}(A' F \Lambda^\circ \mathbf{E}' A) \right| \leq \frac{1}{\sqrt{N}} \left\| \frac{1}{\sqrt{T}} A \right\|^2 \left\| \frac{1}{\sqrt{T}} F \right\| \sqrt{\frac{1}{T} \Sigma_t \left\| \frac{1}{\sqrt{N}} \Sigma_i \lambda_i^\circ \varepsilon_{it} \right\|^2} = O_p(N^{-1/2}).$$

Similarly,

$$\begin{aligned} \text{trace} \left(\frac{1}{T^2 N} A' \mathbf{E} P(\Lambda^\circ) \mathbf{E}' A \right) &= \text{trace} \left(\frac{1}{N} \frac{A'}{\sqrt{T}} \frac{\mathbf{E} \Lambda^\circ}{\sqrt{NT}} \left(\frac{\Lambda^\circ \Lambda^\circ}{N} \right)^{-1} \frac{\Lambda^\circ \mathbf{E}'}{\sqrt{NT}} \frac{A}{\sqrt{T}} \right) \\ &\leq \frac{1}{N} \left\| \frac{A}{\sqrt{T}} \right\|^2 \left\| \frac{\mathbf{E} \Lambda^\circ}{\sqrt{NT}} \right\|^2 O_p(1) = O_p(N^{-1}). \end{aligned}$$

Lemma A.11: Under Assumptions A – D, for $j = 1, \dots, r$,

$$\tilde{\mu}_{NT,j} = \mu_{NT,j} + O_p(N^{-1/2}) + O_p(m^{-1}).$$

Proof: We can complete the proof by showing that, for $j = 1, \dots, r$,

$$(13) \quad \psi_j \left(F^* \Lambda^{o'} \Lambda^o F^{*'} / (NT) \right) = \mu_{NT,j} + O_p(N^{-1/2});$$

$$(14) \quad \tilde{\mu}_{NT,j} = \psi_j \left(F^* \Lambda^{o'} \Lambda^o F^{*'} / (NT) \right) + O_p(m^{-1}).$$

Observe that $F^* \Lambda^{o'} \Lambda^o F^{*'} = F \Lambda^{o'} \Lambda^o F' + E \Lambda^o F' + F' \Lambda^o E + EP(\Lambda^o) E'$. Let Ξ_*^k be the matrix of the eigenvectors corresponding to the first k ($\leq r$) largest eigenvalues of $F^* \Lambda^{o'} \Lambda^o F^{*'} / (NT)$, normalized such that $\Xi_*^{k'} \Xi_*^k = TI_k$. Similarly, define Ξ^k and \tilde{F}^k for the eigenvectors of $F \Lambda^{o'} \Lambda^o F' / (NT)$ and $XX' / (NT)$, respectively. Then, by Lemma A.10,

$$(15) \quad \begin{aligned} & \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F^* \Lambda^{o'} \Lambda^o F^{*'} \right) \\ &= \text{trace} \left(\frac{1}{NT^2} \Xi_*^{k'} F \Lambda^{o'} \Lambda^o F' \Xi_*^k + 2 \frac{1}{NT^2} \Xi_*^{k'} F \Lambda^{o'} E' \Xi_*^k + \frac{1}{NT^2} \Xi_*^{k'} EP(\Lambda^o) E' \Xi_*^k \right) \\ &\leq \text{trace} \left(\frac{1}{NT^2} \Xi_*^{k'} F \Lambda^{o'} \Lambda^o F' \Xi_*^k \right) + O_p(N^{-1/2}) + O_p(N^{-1}) \\ &= \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F \Lambda^{o'} \Lambda^o F' \right) + O_p(N^{-1/2}). \end{aligned}$$

Similarly,

$$(16) \quad \begin{aligned} & \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F^* \Lambda^{o'} \Lambda^o F^{*'} \right) \\ &\geq \text{trace} \left(\frac{1}{NT^2} \Xi^{k'} F^* \Lambda^{o'} \Lambda^o F^{*'} \Xi^k + 2 \frac{1}{NT^2} \Xi^{k'} F \Lambda^{o'} E' \Xi^k + \frac{1}{NT^2} \Xi^{k'} EP(\Lambda^o) E' \Xi^k \right) \\ &= \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F \Lambda^{o'} \Lambda^o F' \right) + O_p(N^{-1/2}) + O_p(N^{-1}) \\ &= \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F \Lambda^{o'} \Lambda^o F' \right) + O_p(N^{-1/2}). \end{aligned}$$

The fact that (15) and (16) hold for all $k = 1, \dots, r$ implies (13). We now show (14). By (10), Lemmas A.7 and A.1,

$$\psi_1[EQ(\Lambda^o)E' / (NT)] \leq \psi_1[EE' / (NT)] \leq c_1^2 m^{-1} \psi_1(UU' / M) = O_p(m^{-1}).$$

Thus, by Lemma A.5,

$$\begin{aligned}
(17) \quad \sum_{j=1}^k \psi_j \left(\frac{1}{NT} XX' \right) &\leq \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F^* \Lambda^{o'} \Lambda^o F^{*'} \right) + k \times \psi_1 \left(\frac{1}{NT} EQ(\Lambda^o)E' \right) \\
&= \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F^* \Lambda^{o'} \Lambda^o F^{*'} \right) + O_p(m^{-1}).
\end{aligned}$$

Also, for any $k = 1, \dots, r$,

$$\begin{aligned}
(18) \quad \sum_{j=1}^k \psi_j \left(\frac{1}{NT} XX' \right) &\geq \text{trace} \left(\frac{1}{NT^2} \Xi^{k'} F^* \Lambda^{o'} \Lambda^o F^{*'} \Xi^k + \frac{1}{NT^2} \Xi^{k'} EQ(\Lambda^o)E' \Xi^k \right) \\
&\geq \sum_{j=1}^k \psi_j \left(\frac{1}{NT} F^* \Lambda^{o'} \Lambda^o F^{*'} \right) + O_p(m^{-1}).
\end{aligned}$$

Then, (14) follows from (17) and (18).

Lemma A.12: Under Assumptions A – D, $V(r+1) = O_p(1)$.

Proof: Note that $V(r+1) = \sum_{j=r+2}^{[d^c m]-2r} \tilde{\mu}_{NT,j} + \sum_{j=[d^c m]-2r+1}^m \tilde{\mu}_{NT,j}$. By Lemma A.8,

$$\begin{aligned}
\frac{[d^c m]-3r-1}{m} c_2^2 \frac{N}{M} \psi_{[bm]} \left(\frac{1}{N} U_{[bm]} U'_{[bm]} \right) &\leq \sum_{j=r+2}^{[d^c m]-2r} \tilde{\mu}_{NT,j} \leq \frac{[d^c m]-3r-1}{m} c_1^2 \psi_1 \left(\frac{1}{M} UU' \right), \\
0 \leq \sum_{j=[d^c m]-2r+1}^m \tilde{\mu}_{NT,j} &\leq \frac{m-[d^c m]+2r}{m} c_1^2 \psi_1 \left(\frac{1}{M} UU' \right).
\end{aligned}$$

Then, $A_1 \leq V(r+1) \leq A_2$, where

$$\begin{aligned}
A_1 &= \frac{[d^c m]-3r-1}{m} c_2^2 \frac{N}{M} \psi_{[bm]} \left(\frac{1}{N} U_{[bm]} U'_{[bm]} \right) \rightarrow_p d^c c_2^2 y^{**} \left(1 - \sqrt{by^*} \right)^2, \\
A_2 &= p \lim_{m \rightarrow \infty} \frac{m-r-1}{m} c_1^2 \psi_1 \left(\frac{1}{M} UU' \right) \rightarrow_p c_1^2 \left(1 + \sqrt{y} \right)^2,
\end{aligned}$$

as $m \rightarrow \infty$, by Lemmas A.1 and A.2, and “ \rightarrow_p ” means “converges in probability.”

Proof of Theorem 1: By Lemma A.11, $\tilde{\mu}_{NT,j} / \tilde{\mu}_{NT,j+1} = \mu_{NT,j} / \mu_{NT,j+1} + o_p(1) = O_p(1)$ for $j = 1, 2, \dots, r-1$. By Lemmas A.11 and A.9,

$$\frac{\tilde{\mu}_{NT,r}}{\tilde{\mu}_{NT,r+1}} \geq \frac{\mu_{NT,r} + O_p(N^{-1/2}) + O_p(m^{-1})}{[\bar{c} + o_p(1)]/m} \rightarrow_p \infty.$$

By Lemma A.9, for $j = 1, \dots, [d^c m] - 2r - 1$, $\tilde{\mu}_{NT,r+j} / \tilde{\mu}_{NT,r+j+1} \leq (\bar{c} + o_p(1)) / (\underline{c} + o_p(1))$. These results indicate that the ER estimator is consistent.

We now show the consistency of the GR estimator. Consider the inequalities

$$(19) \quad c / (1 + c) < \ln(1 + c) < c, \text{ for } c \in (0, \infty).$$

Using these inequalities, we have that

$$\frac{\ln(1 + \tilde{\mu}_{NT,k}^*)}{\ln(1 + \tilde{\mu}_{NT,k+1}^*)} < \frac{\tilde{\mu}_{NT,k}^*}{\tilde{\mu}_{NT,k+1}^* / (1 + \tilde{\mu}_{NT,k+1}^*)} = \frac{\tilde{\mu}_{NT,k}}{\tilde{\mu}_{NT,k+1}} = O_p(1),$$

for $k = 1, 2, \dots, r-1, r+1, \dots, kmax$. Lemma A.12 implies that

$$\frac{V(r+1)}{V(r-1)} = \frac{V(r+1)}{\tilde{\mu}_{NT,r} + \tilde{\mu}_{NT,r+1} + V(r+1)} = O_p(1).$$

Using this and the inequalities (19), we have that

$$\frac{\ln(1 + \tilde{\mu}_{NT,r}^*)}{\ln(1 + \tilde{\mu}_{NT,r+1}^*)} > \frac{\tilde{\mu}_{NT,r}^* / (1 + \tilde{\mu}_{NT,r}^*)}{\tilde{\mu}_{NT,r+1}^*} = \frac{\tilde{\mu}_{NT,r}}{\tilde{\mu}_{NT,r+1}} \frac{V(r+1)}{V(r-1)} = O_p(m) O_p(1) = O_p(m).$$

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

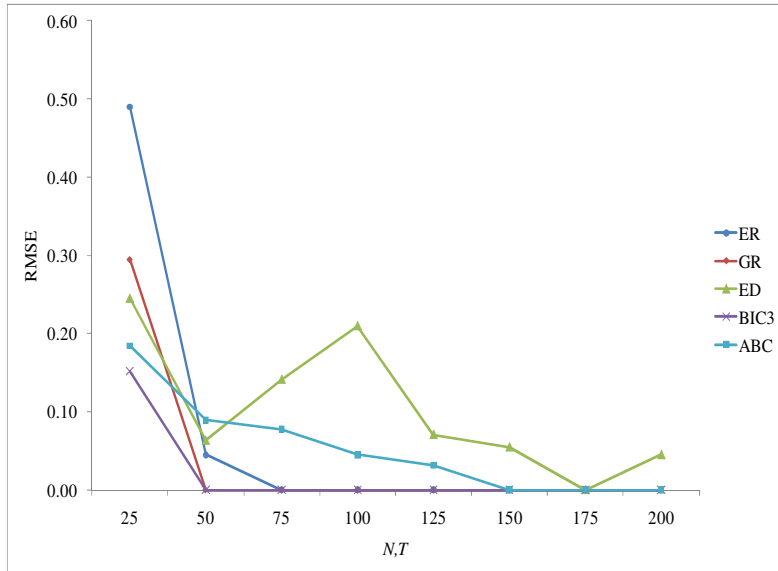
References

- Alessi, L., M. Barigozzi, and M. Capasso, 2010, Improved Penalization for Determining the Number of Factors in Approximate Factor Models, *Statistics and Probability Letters*, 80, 1806 – 1813.
- Amengual, D, and M.W. Watson, 2007, Consistent estimation of the number of dynamic factors in a large N and T panel, *Journal of Business & Economic Statistics*, 25, 91 – 96.
- 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., 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 Cointegration, *Econometrica*, 72, 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, J., and S. Ng, 2007, Determining the Number of Primitive Shocks in Factor Models, *Journal of Business & Economic Statistics*, 25, 52 – 60.
- 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, 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, 1536 – 1555.
- Brown, S., 1989, The number of Factors in Security Returns, *The Journal of Finance*, 44, 1247 – 1262.
- Bernanke, B., J. Boivin, and P. Elias, 2005, Measuring the Effects of Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach, *The Quarterly Journal of Economics*, 120, 387-422.
- Chamberlain, G., and M. Rothschild, 1983, Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, 51, 1281 – 1304.
- 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.
- Hallin, M., and R. Liska, 2007, Determining the Number of Factors in the Generalized Dynamic Factor Model, *Journal of the American Statistical Association*, 102, 603 – 617.
- Onatski, A., 2006, Determining the Number of Factors from Empirical Distribution of Eigenvalues, Working Paper, Columbia University.
- Onatski, A., 2010, Determining the Number of Factors from Empirical Distribution of Eigenvalues, *Review of Economic and Statistics*, 92, 1004 – 1016.
- Onatski, A., 2009, Testing hypotheses about the number of factors in large factor models, *Econometrica*, 77, 1447 – 1479.
- Rao, C.R., 1973, *Linear Statistical Inference and Its Applications*, 2nd eds., John Wiley & Sons (New York, New York).

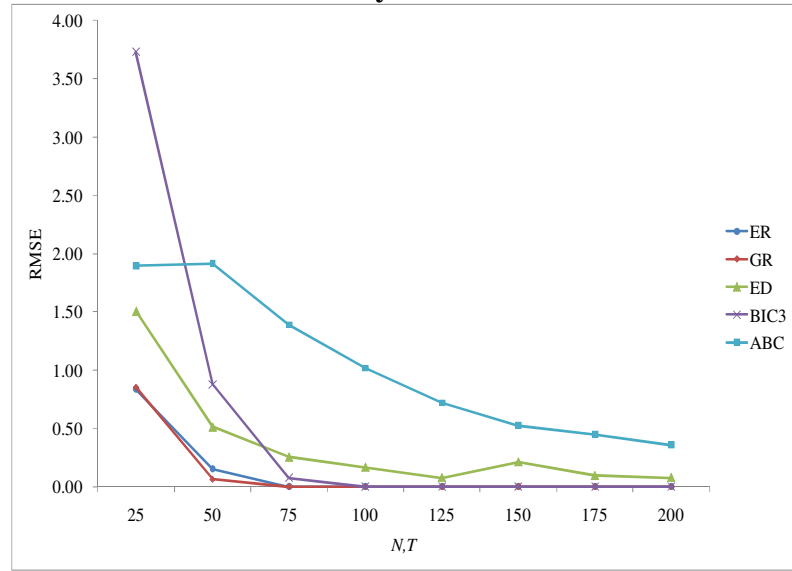
Figure 1: Effects of Error Covariance Structure (Three-Factor Model)

Panel A: *I.I.D.* Errors



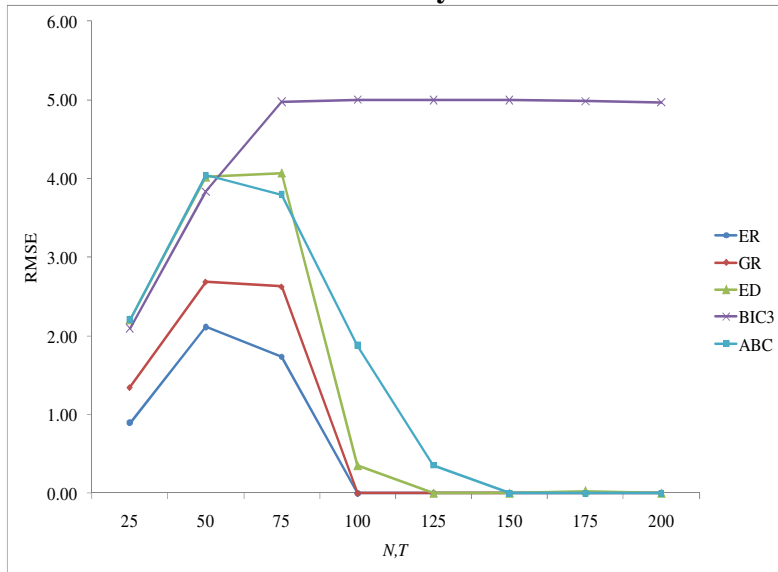
$r = 3, \theta = 1, kmax = 8, \text{ and } \rho = \beta = J = 0.$

Panel B: Serially Correlated Errors



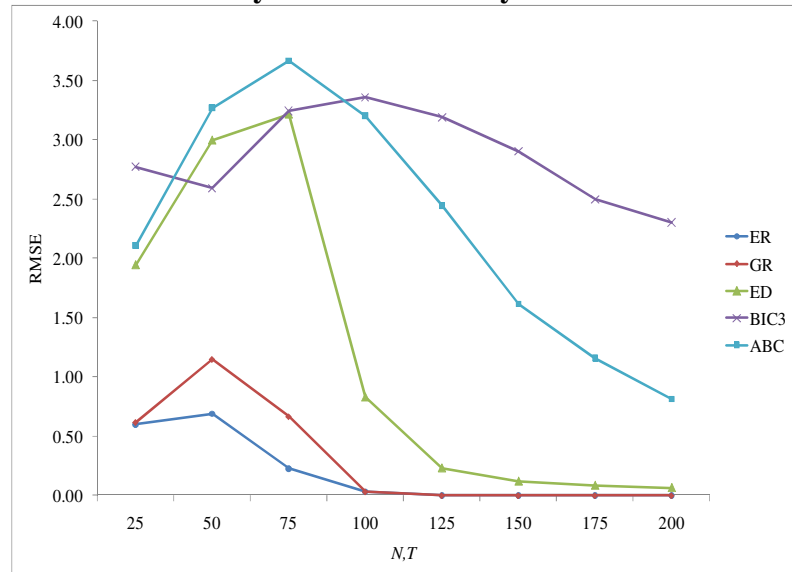
$r = 3, \theta = 1, kmax = 8, \rho = 0.7, \text{ and } \beta = J = 0.$

Panel C: Cross-Sectionally Correlated Errors



$r = 3, \theta = 1, kmax = 8, \rho = 0, \beta = 0.5, \text{ and } J = \max\{10, N/20\}.$

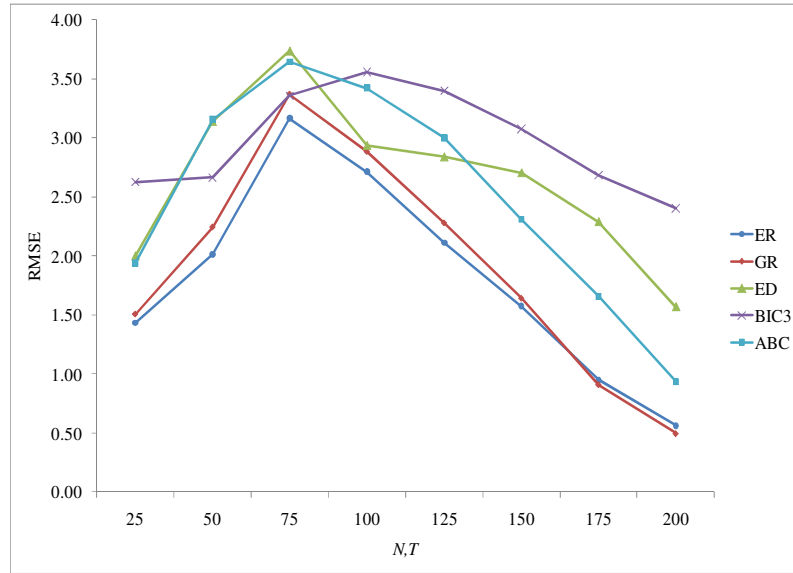
Panel D: Serially/Cross-Sectionally Correlated Errors



$r = 3, \theta = 1, kmax = 8, \rho = 0.5, \beta = 0.2, \text{ and } J = \max\{10, N/20\}.$

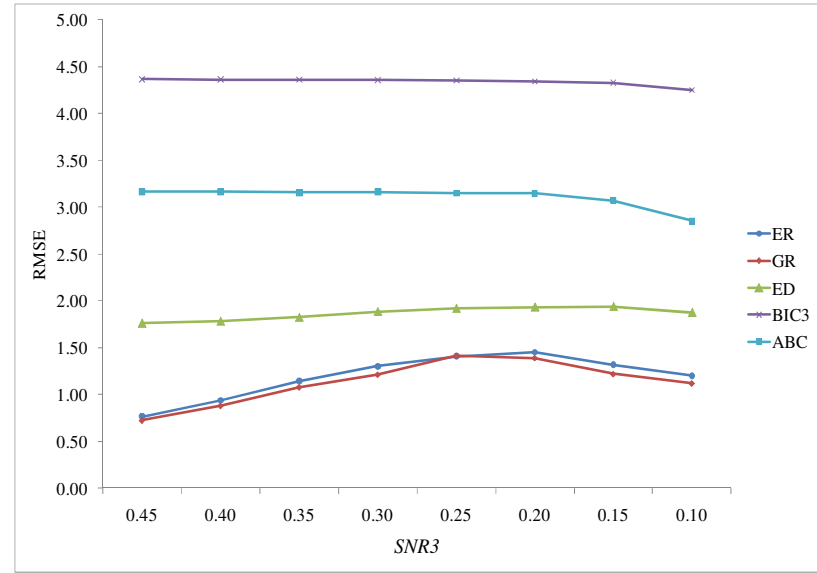
Figure 2: Effects of Weak Factors (Three-Factor Model)

Panel A: When All Factors Are Weak



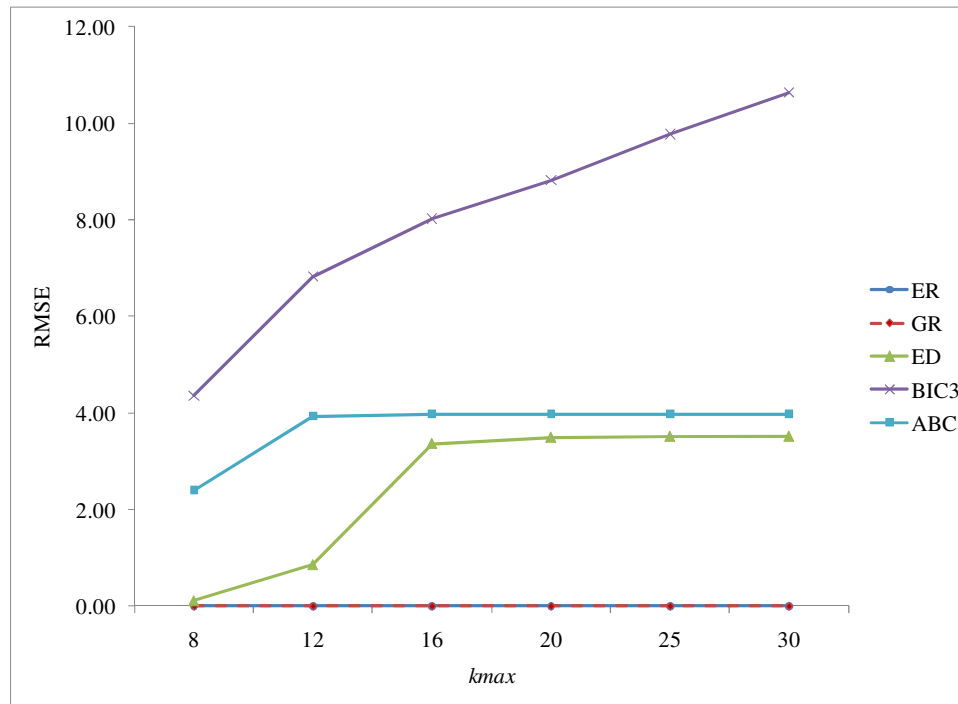
$r = 3, \theta = 6, \rho = 0.5, \beta = 0.2, J = \max(10, N / 20), kmax = 8,$ and $f_1, f_2, f_3 \sim N(0,1)$.

Panel B: When One Factor Is Weak



$N = T = 100, r = 3, \theta = 1, \rho = 0.5, \beta = 0.2, J = \max(10, N / 20), kmax = 8,$ $f_1, f_2 \sim N(0,1),$ and $f_3 \sim N(0, SNR3)$.

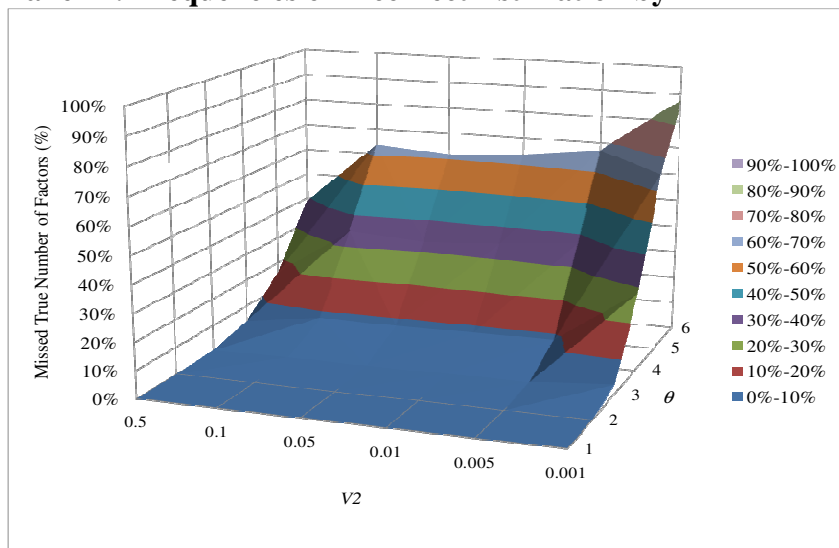
Figure 3: Estimation with Different Values of $kmax$ (Three Factor Model)



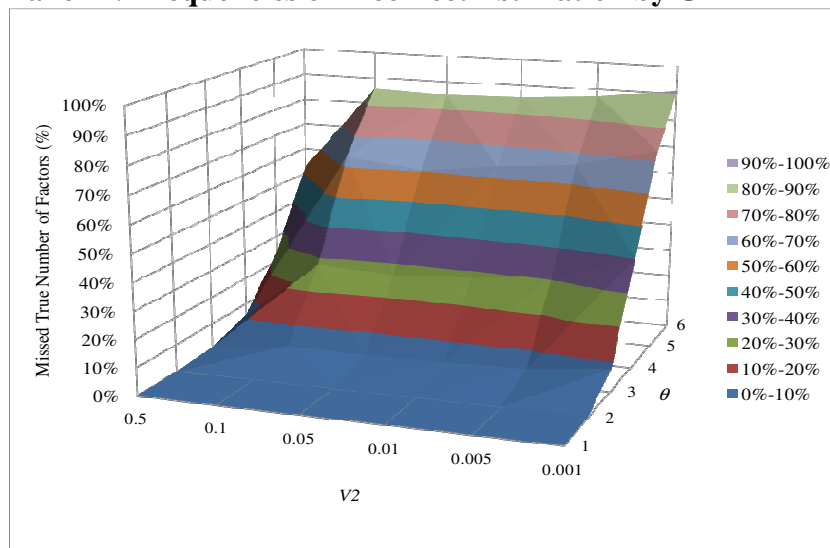
$N = T = 150, r = 3, \theta = 1, \rho = 0.5, \beta = 0.2, J = \max(10, N / 2),$ and $f_1, f_2, f_3 \sim N(0,1).$

Figure 4: Effects of Small Error Variances When Large $kmax$ is Used (Three-Factor Model)

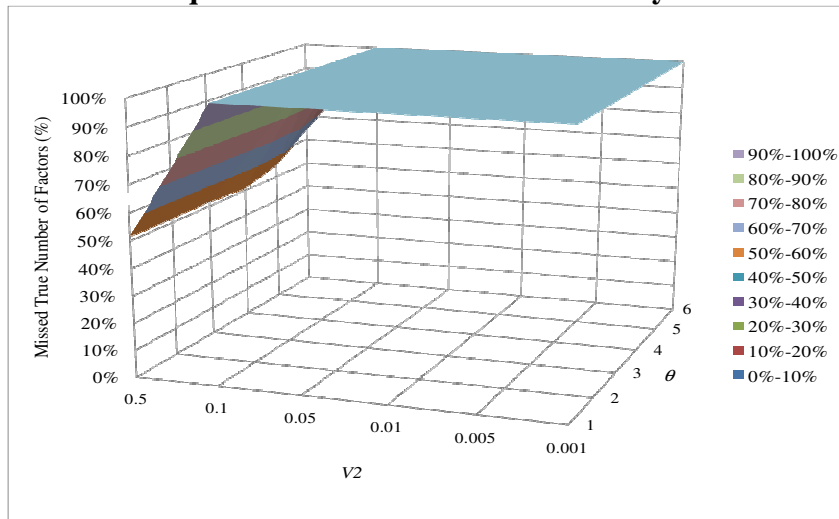
Panel A: Frequencies of Incorrect Estimation by ER



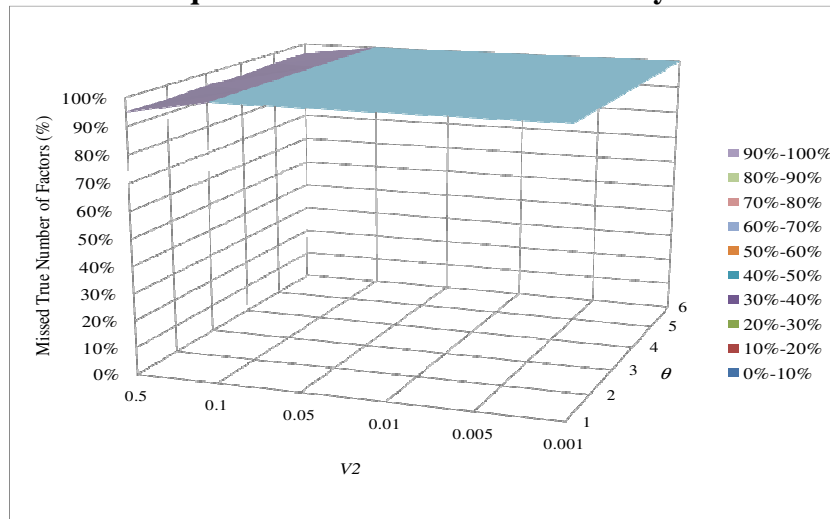
Panel B: Frequencies of Incorrect Estimation by GR



Panel C: Frequencies of Incorrect Estimation by ED



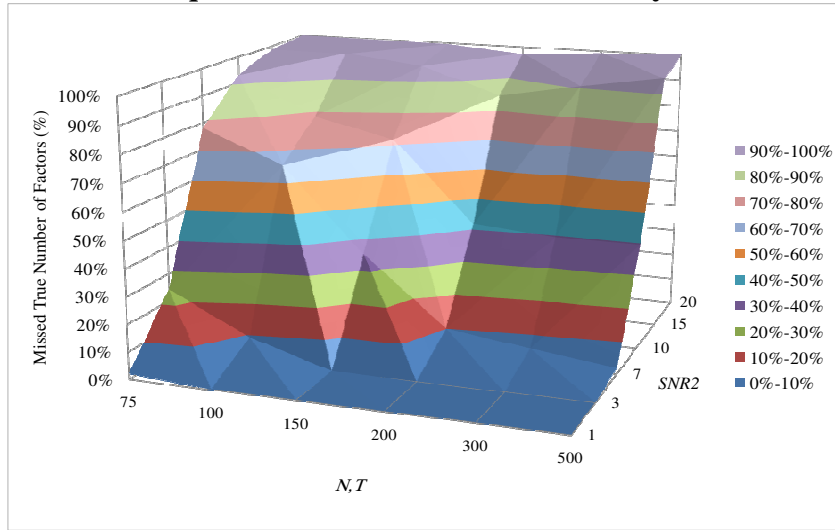
Panel D: Frequencies of Incorrect Estimation by ABC



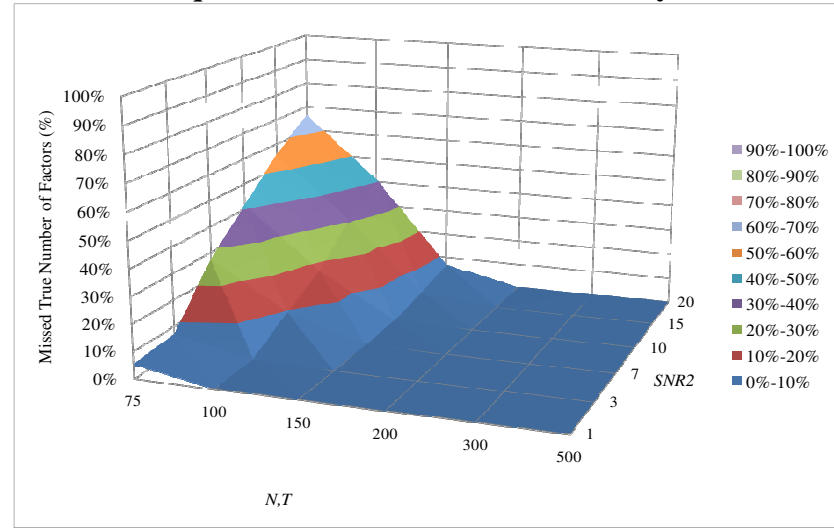
$N = T = 150$, $kmax = 100$, $r = 3$, $\theta = 1$, $\rho = 0.5$, $\beta = 0.2$, $J = \max(10, N/2)$, and $f_1, f_2, f_3 \sim N(0,1)$.

Figure 5: Effects of Dominant Factor (Two-Factor Model)

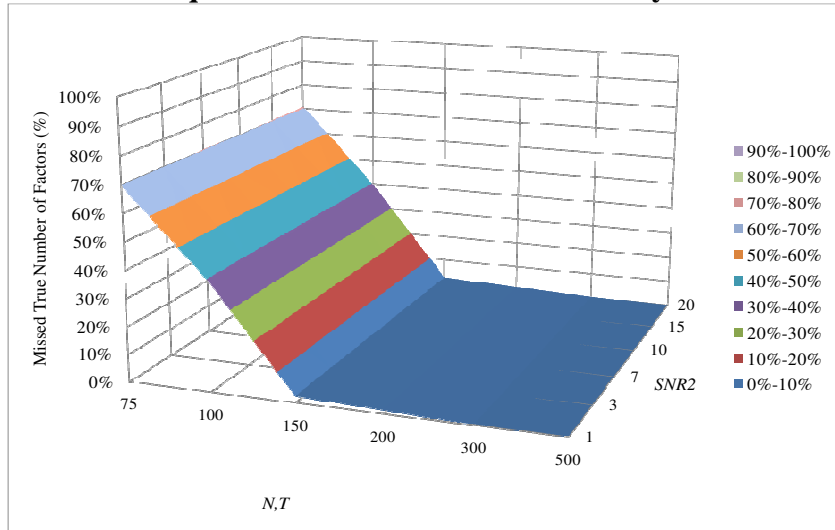
Panel A: Frequencies of Incorrect Estimation by ER



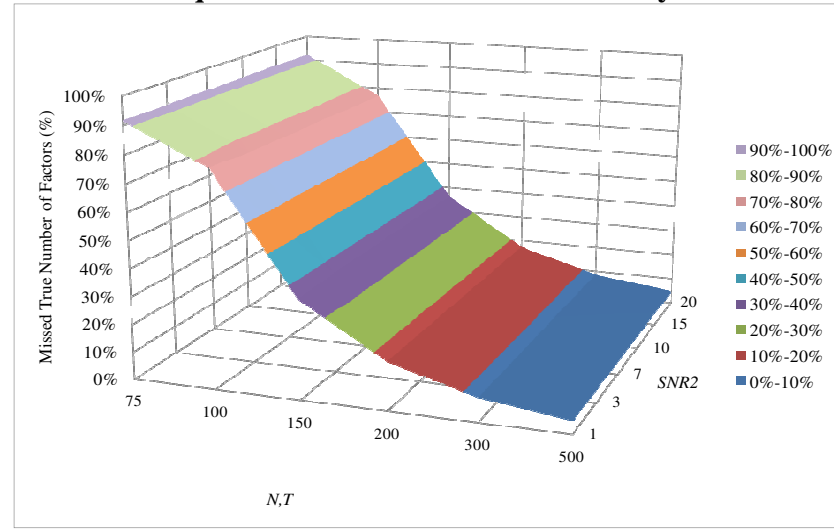
Panel B: Frequencies of Incorrect Estimation by GR



Panel C: Frequencies of Incorrect Estimation by ED



Panel D: Frequencies of Incorrect Estimation by ABC



$r = 2$, $\theta = 1$, $k_{max} = 8$, $\rho = 0.5$, $\beta = 0.2$, $J = \max\{10, N/20\}$, $f_1 \sim N(0,1)$ and $f_2 \sim N(0, SNR2)$