Beta Matrix and Common Factors in Stock Returns

Seung C. Ahn[a]
Arizona State University and Sogang University

Alex R. Horenstein[b]
University of Miami

Na Wang[c]
Hofstra University

This Version: Aug 2013

Many variables have been proposed as common risk factors driving asset returns, which we refer to as "empirical factors" to distinguish them from true latent factors. We examine how many true latent factors are correlated with the empirical factors by estimating the rank of the beta matrix corresponding to the empirical factors. We use a new rank estimation method that can be used for data with a large number of asset returns. Analyzing the U.S. individual and portfolio stock returns in tandem with twenty-six empirical factors we find that the rank of the beta matrix is at most five. Our results have three relevant implications regarding empirical analysis. First, most of the multifactor asset pricing models proposed in the literature lack power to identify risk premiums. Second, our results are consistent with the notion that many of the empirical factors capture the same sources of risk. Third, the Fama-French three factor model is the only multifactor model that consistently generates full rank beta matrices, although it misses one or two additional sources of risk.

Key Words: factor models, beta matrix, rank, eigenvalues.
JEL Classification: C01, C23, C31, G12

---

[*]The title of an earlier version of this paper is "Determining the Rank of the Beta Matrix in a Factor Model with Factor-Candidate Regressors." We are grateful to Raymond Kan, George Korniotis, Alok Kumar, Richard Roll, Aurelio Vazquez, Todd Pronno, the participants at the 2012 International Symposium on Econometric Theory and Applications, the Midwest Finance Association 2010 meeting, the Eastern Finance Association 2010 meeting, the Financial Management Association 2010 annual meeting, the Econometric Society 2009 Far Eastern and South Asia meeting, Columbia University, Instituto Tecnológico Autónomo de México, Universidad Autónoma de Nuevo León, Sogang University, Hofstra University, Arizona State University, University of Miami, and Hitotsubashi University. We also thank Laura Xiaolei Liu, Jonathan Lewellen, and Stefan Nagel for sharing data with us.

[a]Department of Economics, W.P. Carey School of Business, Arizona State University, Tempe, AZ, 85287, USA; School of Economics, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 121 – 742. South Korea. Email: miniahn@asu.edu.

[b]Department of Economics, University of Miami, Coral Gables, FL, 33124. Email: horenstein@bus.miami.edu.

[c]Department of Finance, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY, 11549. Email: na.wang@hofstra.edu.

"We […] thought that the cross-section of expected returns came from the CAPM. Now we have a zoo of new factors" Cochrane (2011).

1. Introduction

Treynor (1962), Sharpe (1964), Lintner (1965), and Mossin (1966) developed the Capital Asset Pricing Model (CAPM), which predicts that the single market factor drives the co-movement in asset returns. Later, the Intertemporal CAPM of Merton (1972) and the Arbitrage Pricing Theory (APT) of Ross (1976) suggest that investors may make their investment decisions considering multiple risk sources, providing the foundations for multifactor asset pricing models. With the advent of these two theories, many variables have been proposed as proxies for the true common risk factors that drive the co-movement in asset returns. We refer to these variables as empirical factors. Some examples are the three factors of Fama and French (1993) and the five macroeconomic factors of Chen, Roll, and Ross (1986). Harvey, Liu and Zhu (2013) categorize 314 empirical factors from 311 different published papers since 1967 in top tier finance journals and current working papers. Many of the proposed multifactor models seem to explain the cross-section of returns better than the CAPM. However, with the richness of empirical factors, some important questions need to be addressed. Three of these questions are: i) are these empirical factors capturing different common risk factors?; ii) how many common risk factors are correlated with the proposed empirical factors? and iii) which empirical factors are really important? This paper attempts to answer these questions.

For this purpose, we estimate the ranks of the beta matrices corresponding to a variety of linear factor models. The rank of the beta matrix corresponding to a set of empirical factors equals the number of true latent factors that are correlated with the empirical factors. A recent study by Lewellen, Nagel, and Shanken (2010) suggests that the relevance of an asset pricing model can be better tested by analyzing a large number of asset returns. Following their suggestion, we analyze large numbers of portfolio and individual stock returns over different time periods. A novelty of our paper is that we develop a new rank estimator that can be applied to a large number of cross-sectional observations. Many methods are available to estimate the rank of a matrix. Examples are the methods proposed by Zhou (1995); Cragg and Donald (1997); Robin and Smith (2000); and Kleibergen and Paap (2006). These methods are designed for data with a relatively small number of risky assets (N) and a large number of time series observations (T). In this paper, we use a new estimator that we refer to as "Modified Bayesian

Information Criterion" (MBIC) estimator. The estimator is a modified version of the Bayesian Information Criterion (BIC) estimator developed by Cragg and Donald (1997). We show that the modified estimator is consistent for any data set with large T whatever the size of N is. Our simulation results also show that the estimator is quite accurate.

The rank of a beta matrix corresponding to a set of empirical factors is not necessarily equal to the *total* number of *true latent* (unobservable) factors. Instead, the beta rank equals the number of linearly independent latent factors that are correlated with the empirical factors. For example, if a beta matrix corresponding to five empirical factors is found to have a rank of two, then only two latent factors are correlated with the five empirical factors. The total number of true latent factors can be greater than two if some latent factors are not correlated with the empirical factors at all. Thus, the total number of true latent factors cannot be directly estimated from an estimated beta matrix generated by a set of empirical factors. However, the estimated rank of a beta matrix can be viewed as a lower bound for the total number of latent factors.

Estimating the rank of a beta matrix is also necessary for the two-pass estimation method of Fama and MacBeth (1973), which has been widely used to estimate the risk premiums of individual empirical factors. The consistency of the two-pass estimator requires that the true (but unobservable) beta matrix, corresponding to the empirical factors used, has full column rank. The estimated beta matrix can have full column rank even if the true beta matrix itself does not. As Kan and Zhang (1999a) and Burnside (2010) have shown, when the true beta matrix fails to have full column rank, the two-pass estimators of risk premiums are not normally distributed (not even asymptotically) and the corresponding *t*-tests are unreliable. Thus, it is important to test whether the true beta matrix has full column rank.

Some special cases have been discussed in the literature in which beta matrices fail to have full column rank. Kan and Zhang (1999a, 1999b) considered a case in which betas corresponding to an empirical factor all equal zeros. They named such an empirical factor as a "useless" factor. This case arises if the empirical factor is not correlated with any of the true latent factors. For the studies using a large number of empirical factors, it is quite possible that some of them may be "useless." Burnside (2010) studied a case in which betas are the same for all individual assets. His empirical study provided evidence that the betas for some asset pricing models are cross-sectionally constant. Ahn, Perez and Gadarowski (APG, 2013) provided further evidence for such betas. Using different sample periods, they estimated the beta matrices of 25 or 10 portfolios using the three factors of Fama and French (1993). They found that the estimated market betas (corresponding to the CRSP value-weighted stock portfolio returns) often have very limited cross-sectional variations. Given the problems created by using the two-pass estimator in the presence of useless factors, multicollinearity among different betas, or cross-sectionally constant betas, APG proposed using two pre-diagnostic statistics to measure levels of multicollinearity and invariance of betas. However, it is important to note that the APG statistics are not designed to estimate the rank of the beta matrix, a necessary condition for the identification of the estimated risk premiums when using the two-pass estimation method.

In our empirical analysis, we apply our rank estimation method to the monthly and quarterly returns of the U.S. stock portfolios and individual stocks over several different time periods during 1952 to 2011. We consider twenty-six empirical factors proposed by previous studies. We analyze both monthly and quarterly returns using the three factors of Fama and French (1993, FF); the five factors of Chen, Roll, and Ross (1986, CRR); the three factor-model of Jagannathan and Wang (1996, JW); the three liquidity-related factors of Pastor and Stambaugh (2003, LIQ); plus the momentum (MOM) factor and the two return reversal (REV) factors (short-term and long-term). For quarterly returns, we also consider the macroeconomic factors used by six additional asset pricing models: the consumption CAPM and the models of Lettau and Ludvigson (2001); Lustig and Van Nieuwerburgh (2004); Li, Vassalou, and Xing (2006); Yogo (2006); and Santos and Veronesi (2006).

Our main results from actual return data are summarized as follows. First, for both monthly and quarterly portfolio returns, our estimation results provide strong evidence that the beta matrix corresponding to the FF model has full column rank for portfolio returns. That is, the three FF empirical factors appear to be correlated with three linearly independent latent risk factors. In contrast, for monthly and quarterly individual returns, the FF beta matrices have ranks of two or three. Most of the other multi-factor models we consider fail to produce beta matrices of full column rank for both portfolio and individual stock returns. Second, adding a single non-FF factor other than the MOM factor to the FF model does not increase the rank of the beta matrix; the MOM factor often increases the rank of the beta matrix for monthly portfolio returns, but not for quarterly portfolio returns or monthly and quarterly individual stock returns. These results indicate that researchers should be very careful when they test whether an additional non-FF factor is priced or not by the Fama-MacBeth two-pass estimation, because the beta matrix may fail to have full column rank. Third and finally, adding to the FF model all of the non-FF empirical factors increases the rank of the beta matrix at most by two.

Our empirical study is related to some previous studies that test correlations between true latent and empirical factors. For example, Bai and Ng (2006) estimated the latent factors by principal components and then test whether some empirical factors are the same as the latent factors. They found that the FF factors approximate the latent factors estimated from portfolio and individual stock returns much better than the CRR factors do. Kan, Robotti and Shanken (2013) compared performances of different factor models using a test based on the distribution of the cross-sectional *R*-square from the two-pass estimation. However, the goal of our study is different from those of these studies. Our goal is not to identify the empirical factors that are most highly correlated with true factors, nor to find the best performing asset pricing model. For a model, the rank of the beta matrix is determined by the number of the latent factors correlated with the empirical factors used in the model. High correlations and low correlations are not distinguished. Thus, by estimating the ranks of beta matrices from a variety of asset pricing models, we aim to check whether the empirical factors used in different asset pricing models proxy for different latent factors (risk sources) and how many latent factors in return data can be captured by all of the different empirical factors.

The rest of this paper is organized as follows. Section 2 introduces the factor model we investigate and the MBIC estimator. It is shown that the estimator is a consistent estimator when the number of time series observations (T) is large while the number of asset returns (N) could be small or large. The link between the estimator and the BIC estimator of Cragg and Donald (1997) is also discussed. Section 3 reports our Monte Carlo simulation results and Section 4, our estimation results from actual return data. Some concluding remarks follow in Section 5. All of the proofs of our theoretical results are given in the Appendix.

2. Model and Rank Estimation

2.1. Model and MBIC Estimator

We begin with the approximate factor model that was considered by Chamberlain and Rothschild (1983). Let x_{it} be the response variable for the i^{th} cross-section unit at time t, where i = 1, 2, ... , N, and t = 1, 2, ... , T. Explicitly, x_{it} can be the (excess) return on asset i at time t. The response variables x_{it} depend on the J latent factors $g_t = (g_{1t},...,g_{Jt})'$. That is,

$$x_t = \eta + \mathrm{B}_g g_t + u_t , \qquad (1)$$

where $x_t = (x_{1t},...,x_{Nt})'$, η is the N-vector of individual intercepts, B_g is the $N \times J$ matrix of factor loadings, and u_t is the N-vector of idiosyncratic components of individual returns with $E(g_t u_t') = 0_{J\times N}$. The matrix B_g is assumed to have full column rank ($rank(\mathrm{B}_g) = J$) because otherwise, the model (1) can reduce to a $(J-1)$ or smaller factor model. Both g_t and u_t are unobservable.

Observables are K empirical factors, $f_t = (f_{1t},..., f_{Kt})'$, which are correlated with r $(\leq J)$ latent factors in g_t but not with the idiosyncratic errors in u_t. This assumption implies that

$$g_t = \theta + \Xi f_t + v_t, \tag{2}$$

where θ is the J-vector of intercepts and Ξ is a $J \times K$ matrix of coefficients with $rank(\Xi) = r$, $E(f_t u_t') = 0_{K\times N}$, $E(f_t v_t') = 0_{K\times J}$, and $E(v_t u_t') = 0_{J\times N}$. The error vector v_t is the vector of the components of g_t that is not correlated with f_t. If we substitute (2) into (1), we obtain

$$x_t = (\eta + \mathrm{B}_g\theta) + \mathrm{B}_g \Xi f_t + (\mathrm{B}_g v_t + u_t) \equiv \alpha + \mathrm{B} f_t + \varepsilon_t, \tag{3}$$

where we denote the i^{th} row of α and B by α_i and $\beta_i' = (\beta_{i1}, \beta_{i2},..., \beta_{iK})$, respectively. The focus of this paper is to estimate the rank of the beta matrix B, which we denote by r.

Some remarks follow on the linear factor model (3). First, the rank of $\mathrm{B} = \mathrm{B}_g\Xi$ is determined by that of Ξ because B_g is a full column rank matrix. That is, $rank(\mathrm{B}) = r$, which is the number of the latent factors or their linear combinations that are correlated with the empirical factors f_t. Thus, the rank of the beta matrix B equals the maximum number of true latent factors that can be explained by the empirical factors f_t, and the rank can be smaller than the total number of true factors, J. Second, even if individual returns are generated by an exact factor model (in which the idiosyncratic errors in u_t in (1) are mutually independent), the errors in ε_t in (3) could be cross-sectionally correlated through $\mathrm{B}_g v_t$ unless the variables in f_t are perfectly correlated with g_t (so that $v_t = 0_{J\times 1}$). Accordingly, the rank of the beta matrix B needs to be estimated allowing for possible cross-sectional correlations in ε_t. Third, if $r = J$, the beta matrix can perfectly explain expected individual returns. For example, if $\gamma = J = K$, that is if the number of empirical factors f_t equals the number of true latent factors g_t $(K = J)$ and if the former variables are correlated with all of the latter variables ($r = J$), the beta matrix B has the full column rank and can explain expected individual returns perfectly. Specifically, there exists a unique K-vector γ satisfying the pricing restriction, $E(x_t) = \mathrm{B}\gamma$,[1] where x_t contains excess

[1] If x_t contains raw returns, $E(x_t) = [1_N, \mathrm{B}]\gamma$, where 1_N is an N-vector of ones.

returns; see Lewellen, Nagel, and Shanken (2010). However, if $r = J < K$, that is if too many empirical factors are used compared to the number of true latent factors, the beta matrix does not have full column rank. As a consequence, there are an infinite number of K-vectors γ satisfying the pricing restriction. The beta matrix B may still perfectly explain the expected excess returns $E(x_t)$, but the risk prices (γ) are not unique. For this case, Burnside (2010) has shown that the two-pass estimator of γ is not asymptotically normal. This problem arises even if all of the empirical factors f_t are correlated with (linear combinations of) the true latent factors g_t. To explain individual returns, the use of too many empirical factors is not harmful as long as each of them is correlated with true latent factors. However, the resulting two-pass estimates of factor prices would not provide reliable statistical inferences if the number of the empirical factors used is greater than the number of true latent factors.

In order to discuss how to estimate the rank of the beta matrix B in (3), let us introduce some notation. Let

$$\hat{\Sigma}_{xf} = T^{-1}\Sigma_{t=1}^{T}(x_t - \overline{x})(f_t - \overline{f})';\hat{\Sigma}_{ff} = T^{-1}\Sigma_{t=1}^{T}(f_t - \overline{f})(f_t - \overline{f})',$$

where $\overline{f} = T^{-1}\Sigma_{t=1}^{T} f_t$ and $\overline{x} = T^{-1}\Sigma_{t=1}^{T} x_t$. Then, the Ordinary Least Squares (OLS) estimator of B is given by $\hat{\mathrm{B}} = [\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_N]' = \hat{\Sigma}_{xf}\hat{\Sigma}_{ff}^{-1}$. We also define

$$\hat{\sigma}_i^2 = (T-K)^{-1}\Sigma_{t=1}^{T}[(x_{it} - \overline{x}_i) - \hat{\beta}_i'(f_t - \overline{f})]^2;\ \hat{\sigma}_\varepsilon^2 = N^{-1}\Sigma_{i=1}^{N}\hat{\sigma}_i^2,$$

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimator of $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{it})$.

The "Modified Bayesian Information Criterion" (MBIC) estimator we propose is the minimizer of the following criterion function:

$$C_M(p) = T \times \Sigma_{j=1}^{K-p}\psi_j\left(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\mathrm{B}} / \hat{\sigma}_\varepsilon^2\right) - w(T) \times (N-p)(K-p), \quad (4)$$

where the function $w(T)$ should be chosen such that $w(T) \to \infty$ and $w(T)/T \to 0$ as $T \to \infty$. There are an infinite number of possible choices for $w(T)$. However, in unreported simulations we found that the MBIC estimator computed with $w(T) = T^{0.2}$ is more accurate than those with many different $w(T)$ functions. Thus, we use $T^{0.2}$ for our reported simulations and actual data analysis.

Betas corresponding to some empirical factors may have no cross-sectional variations. This possibility is not fictional. For example, Connor and Korajczyk (1989) showed that an intertemporal and competitive equilibrium version of the Arbitrage Pricing Theory (APT) implies the presence of a factor with unitary betas for all returns. Burnside (2010) found evidence that for the 25 Size and Book-to-Market portfolio returns, the betas corresponding to a consumption growth factor (log-growth of real per capita consumption) may be constant. Ahn, Perez and Gadarowski (2013) report that the market betas estimated from many different data sets covering different portfolios and/or different time periods often have very small variations. These results suggest that some empirical factors may have betas with little cross-sectional variations depending on what portfolios and what empirical factors are analyzed. As Burnside (2010) showed, when betas corresponding to a factor are cross-sectionally constant, the two-pass estimation using gross returns (not excess returns) cannot identify risk prices. Thus, it would be important to test whether such betas exist before risk prices are estimated by the two-pass method.

We can test whether some betas are cross-sectionally constant or not by comparing the ranks of two matrices: the beta matrix B and its demeaned version, $Q_N\mathrm{B} = (\dot{\beta}_1, \dot{\beta}_2, ..., \dot{\beta}_N)'$, where $Q_N = I_N - N^{-1}1_N 1_N'$, 1_N is an N-vector of ones, $\dot{\beta}_i = \beta_i - \overline{\beta}$, and $\overline{\beta} = N^{-1}\Sigma_{i=1}^{N}\beta_i$. If a column of B (or a linear combination of the columns of B) is proportional to a vector of ones, the corresponding column of the demeaned beta matrix ($Q_N\mathrm{B}$) becomes a zero vector. Thus, $rank(Q_N\mathrm{B}) = r-1$. For the same reason, if two columns of B are proportional to a vector of ones, $rank(Q_N\mathrm{B}) = r-2$. If no column of B has constant betas, the two matrices B and $Q_N\mathrm{B}$ must have the same ranks. Therefore, comparing the estimated ranks of the beta matrix (B) and the demeaned beta matrix ($Q_N\mathrm{B}$), we can determine whether a constant-beta factor exists in f_t.

The rank of $Q_N\mathrm{B}$ can be also estimated by the MBIC method introduced above with a small modification. The criterion function we can use is

$$D_M(p) = T \times \Sigma_{j=1}^{K-p}\psi_j\left(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}} / \hat{\sigma}_{\varepsilon}^2\right) - w(T)\times(N-1-p)(K-p), \tag{5}$$

where p = 1, ... , K–1, and $D_M(K) = 0$. The MBIC estimator is the minimizer of $D_M(p)$. We refer to this estimator as "MBIC estimator for demeaned betas" (MBICD).

2.2. Consistency of the MBIC and MBICD estimators

In this subsection we show the consistency of the MBIC and MBICD estimators. In what follows, the norm of a matrix A is denoted by $\|A\| = [trace(A'A)]^{1/2}$. We define c as a generic positive constant. With this notation, we make the following assumptions for the factor model (3):

Assumption A (empirical factors): $T^{-1}\Sigma_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})' \to_p \Sigma_f$, and $\bar{f} \to_p \mu_f$, where $\bar{f} = T^{-1}\Sigma_{t=1}^{T} f_t$, Σ_f is a finite and positive definite matrix and μ_f is a finite vector.

Assumption B (betas): (i) $\|\beta_i\| \le c$ for all $i = 1, 2, \cdots, N$ and for any N. (ii) $rank(\mathrm{B}) = r$ and $rank(Q_N \mathrm{B}) = r^d$ $(\le r)$, for all $N > K$, where $0 \le r^d \le r \le K$. (iii) For any $N > K$, $\mathrm{B}'\mathrm{B}/N$ is a finite matrix. If $N \to \infty$, $N^{-1}\Sigma_{i=1}^{N}\beta_i \to \mu_\beta$ and $\mathrm{B}'Q_N\mathrm{B}/N \to \Sigma_{\beta\beta}$, where μ_β is a $K \times 1$ finite vector and $\Sigma_{\beta\beta}$ is $K \times K$ finite matrix with $rank(\Sigma_{\beta\beta}) = r^d$.

Assumption C (idiosyncratic errors): (i) $E(\varepsilon_{it}) = 0$ and $E|\varepsilon_{it}|^4 \le c$ for all i and t, and

$$E\left[\frac{1}{N}\Sigma_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}\varepsilon_{it}\right)^2\right] = \frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}E(\varepsilon_{it}\varepsilon_{is}) \le c .$$

(ii) $p\lim_{T\to\infty} T^{-1}\Sigma_{t=1}^{T}E(\varepsilon_{it}^2) = \sigma_i^2$, and $0 < \sigma_i^2 < c$ for all *i*.

Assumption D (weak dependence between factors and idiosyncratic errors):

$$E\left[\frac{1}{N}\Sigma_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T} f_t\varepsilon_{it}\right\|^2\right] = \frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}E(\varepsilon_{it}\varepsilon_{is} f_t' f_s) \le c .$$

The four assumptions are slightly more general than the assumptions used by Bai and Ng (2002) to estimate the number of true latent factors. Assumption A implies that the empirical factors should be stationary and ergodic. Assumption B(i) simply means that the betas are finite for any individual return. Assumption B(ii) allows the rank of B to be smaller than the number of empirical factors f_t. Assumption B(iii) implies that for the cases where N is large, the $K \times K$ matrix $\mathrm{B}'\mathrm{B}/N$ is asymptotically finite. The MBIC estimator, as well as the MBICD estimator, does not require large *N*. Under Assumption B(iii), the estimator is consistent regardless of the size of *N*. Under Assumption B, we treat the betas as fixed constants, not as random variables. We can relax this assumption, but at the cost of more notation.

Assumption C allows time-series correlation in the errors ε_{it}. It does not impose any restriction on possible cross-sectional correlations among the error terms ε_{it}, either. Our asymptotic results do not depend on the covariance structure of the errors. Assumption C implies that for all *i*, $T^{-1/2}\Sigma_{t=1}^{T}\varepsilon_{it}$ is a random variable with finite variance for each *i*. Similarly, Assumption D implies that the random vectors $T^{-1/2}\Sigma_{t=1}^{T}\varepsilon_{it} f_t$ have finite variance matrices for every *i*. These two assumptions are the general assumptions under which the OLS estimator of each row of the beta matrix B is consistent and asymptotically normal.

As we discussed above, when the empirical factors f_t are proxy variables for true latent factors, the error vector $\varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{Nt})'$ may contain factor components. Assumptions C and D allow such cases.[2] To see why, consider a simple case in which the ε_{it} have a one-factor structure: $\varepsilon_{it} = \xi_i h_t$ where $E(h_t) = 0$, $E(h_t f_t) = 0$, $E(h_t^4) < c$, and $T^{-1}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}E(h_s h_t f_t' f_s) < c$ for all *t* and $|\xi_i| < c$ for all *i*. For this case, the random variables $T^{-1/2}\Sigma_{t=1}^{T}h_t f_t$ have finite variances. Thus, we can easily see that Assumption C holds. In addition,

$$\frac{1}{T}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}E(\varepsilon_{it}\varepsilon_{is} f_t' f_s) = \xi_i^2 \frac{1}{T}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}E(h_t h_s f_t' f_s) < c^3 .$$

Thus, Assumption D holds. Given that ε_{it} can have a factor structure, estimating the rank of B is not equivalent to estimating the number of all of the true latent factors in response variables. The rank of B is the number of true latent factors or their linear combinations that are correlated with the empirical factors f_t. Hence, the rank estimation method works well even if the empirical factors are correlated with only a subset of true latent factors. The uncorrelated latent factors are subsumed in the error terms with a factor structure.

The following theorem presents the asymptotic properties of the eigenvalues of the two matrices $\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\mathrm{B}}/N$ and $\hat{\Sigma}_{ff}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}/N$.

Theorem 1: Under Assumptions A – D,

(i) $p\lim_{T\to\infty}\psi_j(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\mathrm{B}}/N) > 0$ for $K - r + 1 \le j \le K$;

(ii) $\psi_j(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\mathrm{B}}/N) = O_p(T^{-1})$, for $1 \le j \le K - r$;

(iii) $p\lim_{T\to\infty}\psi_j(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}/N) > 0$, for $K - r^d + 1 \le j \le K$;

(iv) $\psi_j(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}/N) = O_p(T^{-1})$, for $1 \le j \le K - r^d$.

[2] In the model of Bai and Ng (2002), the error terms are not allowed to have a factor structure.

Theorem 1 shows that the first K- r (K- r^d) small eigenvalues of $\hat{\Sigma}_{ff}\hat{B}'\hat{B}/N$ ($\hat{\Sigma}_{ff}\hat{B}'Q_N\hat{B}/N$) have the same convergence rates, which are different from those of the other eigenvalues. This difference in convergence rate is used to identify the rank of the beta matrix B ($Q_N B$). Notice that the asymptotic properties of the eigenvalues do not require any restriction on *N*. Theorem 1 holds for any fixed number *N*.

Theorem 1 implies our main theoretical results. Stated formally:

Theorem 2: Under Assumptions A – D, the MBIC estimator is a consistent estimator of the rank of beta matrix B, r. Similarly, the MBICD estimator is a consistent estimator of the rank of demeaned beta matrix $Q_N B$, r^d.

A technical point is worth noting here. Theorem 2 is proven under the assumption that data are balanced; that is, all individual returns are observed for all *T* time periods. However, in practice the data need not be balanced. The MBIC and MBICD estimators can be computed even if each asset in the data has a different number of time series observations (T_i). The betas can be estimated by asset-by-asset time series regressions, as long as the betas of each asset are estimated with a sufficiently large number of time series observations. The average time series observations, $\bar{T} = N^{-1}\Sigma_{i=1}^{N}T_i$, can be used for *T* in the MBIC and MBICD estimators.

While Theorem 2 shows the consistency of the MBIC and MBICD estimators, it does not provide any prediction about the estimators' finite-sample performances. Later in Section 3, we investigate its finite-sample performance through Monte Carlo simulation exercises.

2.3. Related Estimators

This subsection discusses some related estimators and the link between the MBIC estimator and the BIC estimator of Cragg and Donald (1997). We discuss the link between the two estimators under Assumptions A – D, and the additional assumption that *N* is fixed. We do so because the BIC estimator is designed for the cases with fixed *N*.

Under Assumptions A – D and the assumption of fixed *N*, it can be shown that as $T \to \infty$,

$$\sqrt{T}vec(\hat{B}' - B') \to_d N(0,\Omega),$$

where $vec(\bullet)$ is a matrix operator stacking all the columns in a matrix into a column vector, and "$\to_d$" means "converges in distribution." Let $\hat{\Omega}$ be a consistent estimator of Ω as $T \to \infty$ with a fixed number (N) of individual returns; and use the notation to denote an $N \times K$ matrix $G_{K,p}$ ($0 \le p \le K$) that minimizes an objective function

$$\Pi_T(G_{K,p}, p) = Tvec(\hat{\mathrm{B}} - G_{K,p})'\hat{\Omega}^{-1}vec(\hat{\mathrm{B}} - G_{K,p}) . \tag{6}$$

Cragg and Donald (1997) showed (i) that $\Pi_T(\hat{G}_{K,p}, p) \to_p \infty$ if $p < r$ and (ii) that $\Pi_T(\hat{G}_{K,r}, r) \to_d \chi^2_{(N-r)(K-r)}$, where r is the true rank of B and "$\to_p$" means "converges in probability." Based upon these findings, they develop two different rank estimation methods. One estimator, which they refer to as the TC (Testing Criterion) method, is obtained by repeatedly testing the null hypotheses of $r = p$ (p = 0, 1 , 2, ..., K – 1; where K is the number of empirical factors used) against the alternative hypothesis of full-column rank. Each hypothesis is tested by using $\Pi_T(\hat{G}_{K,p}, p)$ as a $\chi^2_{(N-p)(K-p)}$ statistic. The TC estimate is the minimum value of p that does not reject the hypothesis of $r = p$. If all of the null hypotheses are rejected, the TC estimate equals K.

Recently, Burnside (2010) proposed to use the $\Pi_T(\hat{G}_{K,K-1}, K-1)$ statistic to test the null hypothesis of $r = K-1$ against the alternative of $r = K$. His simulation results show that the test performs well in small samples, especially when it is applied to covariance matrices instead of beta matrices. His method is designed to determine whether the beta matrix (or the covariance matrix of empirical factors and individual returns) has full column rank or not. In contrast, using the TC estimator, we can estimate the true rank of r itself.

The other estimator, which Cragg and Donald (1997) refer to as BIC (Bayesian Information Criterion) estimator, is obtained by finding a value of p, which minimizes the criterion function

$$C(p) = \Pi_T(\hat{G}_{K,p}, p) - \ln(T) \times (N-p)(K-p), \tag{7}$$

where p = 0, 1, ... , K. For (7), $\ln(T)$ can be replaced by any $w(T)$ function such that $w(T) \to \infty$ and $w(T)/T \to 0$ as $T \to \infty$. Clearly, $\ln(T)$ is a possible $w(T)$ function to use. The BIC estimator computed with any $w(T)$ is a consistent estimator as $T \to \infty$ with fixed N. Replacing $\ln(T)$ by $w(T)$ in (7), we can see that the MBIC criterion function $C_M(p)$ in (4) has a form similar to that of the criterion function $C(p)$. In fact, as we show below, the MBIC estimator numerically equals a BIC estimator computed with a different weighting matrix for $\hat{\Omega}$.

While the BIC estimator is consistent as $T \to \infty$, its finite-sample performance may depend on the choice of the $w(T)$ function. In some unreported simulations we found that the BIC estimator performs better in finite samples when $w(T) = \ln(T)$ is used, while the MBIC estimator performs better with $w(T) = T^{0.2}$.

While both the TC and BIC estimators have desirable large-sample properties, they are computationally burdensome to use in practice, especially for the cases with large *N*. This is so because the matrix $G_{K,p}$ contains a large number of unknown parameters to be estimated especially for the cases with large *N* and/or *p*. In unreported experiments, we attempted to compute the TC and BIC estimators using the same simulated data that are used for the results reported in the next section. We observed that standard minimization algorithms failed to find $\hat{G}_{K,p}$ too often.

This computational problem can be resolved if some restrictions are imposed on the covariance structure of the error terms. For example, suppose that the idiosyncratic error vectors ε_t are independently and identically distributed (*i.i.d.*) conditionally on the empirical factors f_t with the conditional variance-covariance matrix, $Var(\varepsilon_t \mid f_t) = \Sigma_{\varepsilon\varepsilon}$. The individual errors ε_{it} are still allowed to be cross-sectionally correlated; that is, the off-diagonal elements of $\Sigma_{\varepsilon\varepsilon}$ need not be zero. For this case, the computation procedures for the TC and BIC estimators are considerably simplified. When the error vectors are *i.i.d.* over time, $\hat{\Omega}_1 = \hat{\Sigma}_{\varepsilon\varepsilon} \otimes \hat{\Sigma}_{ff}^{-1}$ is a consistent estimator of Ω, where "$\otimes$" means the Kronecker product and $\hat{\Sigma}_{\varepsilon\varepsilon}$ is a consistent estimator of $\Sigma_{\varepsilon\varepsilon}$; *i.e.*,

$$\hat{\Sigma}_{\varepsilon\varepsilon} = \frac{1}{T-k}\Sigma_{t=1}^{T}[(x_t - \overline{x}) - \hat{\mathrm{B}}(f_t - \overline{f})][(x_t - \overline{x}) - \hat{\mathrm{B}}(f_t - \overline{f})]'.$$

Cragg and Donald (1997) show that when $\hat{\Omega}_1$ is used for $\hat{\Omega}$,

$$\Pi_T(\hat{G}_{K,K}, K) = 0;\ \Pi_T(\hat{G}_{K,p}, p) = T \times \Sigma_{j=1}^{K-p}\psi_j\left(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\mathrm{B}}\right), \qquad (8)$$

for $p = 0, 1, \ldots, K-1$. We denote by the "BIC1" estimator the minimizer of the criterion function (7) with (8):

$$C_1(p) = T \times \Sigma_{j=1}^{K-p}\psi_j\left(\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\mathrm{B}}\right) - \ln(T) \times (N-p)(K-p). \qquad (9)$$

We also refer to the TC estimator computed with (8) as the TC1 estimator. Both the TC1 and BIC1 estimators can be easily computed with any software that can compute eigenvalues.

It is important to note that the TC1 estimator is inconsistent if the error vectors ε_t are autocorrelated or heteroskedastic (conditionally on f_t) over time. This is because, when $\hat{\Omega}_1$ is used, the statistic $\Pi_T(\hat{G}_{K,r}, r)$ is no longer a $\chi^2_{(N-r)(K-r)}$ random variable asymptotically. Thus, it is inappropriate to use the sequential χ^2-tests to estimate the true rank, *r*.

In contrast, the BIC1 estimator is still consistent. This is a fact that is not well known in the literature. As Ahn, Lee and Schmidt (2013, p. 6) point out, the consistency of the BIC estimator requires the statistic $\Pi_T(\hat{G}_{K,r}, r)$ to be a random variable that is bounded in probability. The statistic needs not be a χ^2 random variable. When the error vectors are not *i.i.d.* over time, $\hat{\Omega}_1$ is not a consistent estimator of Ω. However, following Jagannathan and Wang (1996), we can easily show that the statistic $\Pi_T(\hat{G}_{K,r}, r)$ computed with $\hat{\Omega}_1$ is asymptotically a weighted sum of independent χ^2_1 random variables, which is bounded in probability. Thus, the BIC estimator computed with $\hat{\Omega}_1$, which is the BIC1 estimator, remains consistent even if the error vectors are autocorrelated and/or heteroskedastic over time.

The BIC1 estimator can be easily modified to estimate the rank of the demeaned beta matrix. Define the following criterion function:

$$D_1(p) = T \times \Sigma_{j=1}^{K-p} \psi_j \left(\hat{\Sigma}_{ff} \hat{\mathrm{B}}' Q_N (Q_N \hat{\Sigma}_{\varepsilon\varepsilon} Q_N)^{+} Q_N \hat{\mathrm{B}} \right) - \ln(T) \times (N-1-p)(K-p) \quad (10)$$

where p = 1, ... , K–1, and $D_1(K) = 0$. Then, the BIC1 estimator of the demeaned beta matrix ($Q_N \mathrm{B}$) equals the minimizer of $D_1(p)$. We refer to this estimator as "BICD1" estimators. We note that even for the cases in which $\hat{\Sigma}_{\varepsilon\varepsilon}$ has full column rank, $Q_N \hat{\Sigma}_{\varepsilon\varepsilon} Q_N$ does not. That is why we use the Moore-Penrose generalized inverse of $Q_N \hat{\Sigma}_{\varepsilon\varepsilon} Q_N$ in (10).

Because the BIC1 estimator is consistent as $T \to \infty$ with fixed *N*, we can expect that the estimator would have good finite-sample properties for the data with large *T* and relatively small *N*. However, it is unknown whether the BIC1 estimator would remain consistent as both N and *T* grow infinitely. One immediate problem in using the BIC1 estimator for the data with large *N* is that $\hat{\Sigma}_{\varepsilon\varepsilon}$ is not invertible if $N > T$. This numerical problem can be resolved if we use the

Moore-Penrose generalized inverse matrix of $\hat{\Sigma}_{\varepsilon\varepsilon}$ ($\hat{\Sigma}_{\varepsilon\varepsilon}^{+}$) instead of $\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$. However, it is still difficult to determine whether the BIC1 estimator computed with $\hat{\Sigma}_{\varepsilon\varepsilon}^{+}$ would be consistent for the data with both large *N* and *T*. In this paper we do not attempt to investigate the asymptotic distribution of the BIC1 estimator when both *N* and *T* are large. Instead, we will consider in Section 3 the estimator's finite-sample properties and compare them to those of the MBIC estimator.

Similarly to the BIC1 estimator, our MBIC estimator is a BIC estimator computed with a different weighting matrix for $\hat{\Omega}^{-1}$. Specifically, if we compute the BIC estimator using $\hat{\Omega}_2 = \hat{\sigma}_\varepsilon^2 (I_N \otimes \hat{\Sigma}_{ff}^{-1})$ for $\hat{\Omega}$ and using $w(T) = T^{0.2}$ instead of $\ln(T)$, we obtain the MBIC estimator. Interestingly, $\hat{\Omega}_2$ is a consistent estimator of Ω under the assumption that the errors ε_{it} are *i.i.d.* over both different time and individual returns. This assumption is stronger than the assumption under which $\hat{\Omega}_1$ is a consistent estimator of Ω (the errors are *i.i.d.* only over time). Thus, from the perspective of BIC estimation, the MBIC estimator is an estimator motivated under quite restrictive assumptions. However, as we have shown already in subsection 2.2, the MBIC estimator is consistent as $T \to \infty$ regardless of the size of *N*. The errors therefore can be both serially and cross-sectionally correlated.

3. Simulation Results

3.1. Basic Simulation Setup

The foundation of our simulation exercises is the following data generating process:

$$x_{it} = \alpha_i + \Sigma_{j=1}^{K} \beta_{ij} f_{jt} + \varepsilon_{it} ; \varepsilon_{it} = \phi \frac{(\xi_{i1} h_{1t} + \xi_{i2} h_{2t})}{\sqrt{\xi_{1i}^2 + \xi_{2i}^2}} + \sqrt{1-\phi^2} v_{it} ,$$

where the empirical factors f_{jt} and the ξ_{i1}, ξ_{i2}, h_{1t}, h_{2t}, and v_{it} in ε_{it} are all randomly drawn from $N(0,1)$. For simplicity, we set $\alpha_i = 0$ for all *i*. Under this setup, the variance of error ε_{it} is equal to one for all *i* and *t*. The factor components h_{1t} and h_{2t} can be viewed as common latent factors that are not correlated with the empirical factors f_{jt}. The errors ε_{it} are cross-sectionally correlated through h_{1t} and h_{2t} if $\phi \neq 0$. We have also considered the cases in which the errors are serially correlated. We do not report the results because they are not materially different from the results reported below. For the reported simulations, we set ϕ = 0.2. The use of greater values for ϕ does not change estimation results substantially.

We generate the beta matrix B by the following three steps. First, we draw an $N \times r$ random matrix B_g such that its first column equals the vector of ones, and the entries in the other columns are drawn from $N(0,1)$. Second, we draw a random $K \times K$ positive definite matrix, compute the first r orthonormalized eigenvectors of the matrix, and set a $K \times r$ matrix C using the eigenvectors.[3] Finally, we set $\mathrm{B} = \mathrm{B}_g \Lambda^{1/2} C'$, where $\Lambda = diag(\lambda_1, ..., \lambda_r)$. This setup is equivalent to the case in which individual returns are generated by r true factors $g_t = (g_{1t}, ..., g_{rt})' = \Lambda^{1/2} C' f_t$ with $Var(g_t) = \Lambda$. The factor loading matrix corresponding to g_t is B_g. By construction, the factors in g_t are mutually independent, and the betas corresponding to g_{1t} are constant over different individual response variables. Under this setup, $rank(\mathrm{B}) = r$ and $rank(Q_N \mathrm{B}) = r - 1$. We use this setup to investigate the finite-sample performances of the BIC1, MBIC, BICD1 and MBICD estimators, as well as that of the TC1 estimator.

Under our data generating setup, each of the empirical factors f_t can have non-zero explanatory power for individual response variables x_{it}, even if the beta matrix B does not have full column rank. The parameter λ_j equals the variance of the j^{th} true factor, g_{jt}. Given that B_g is drawn from $N(0,1)$, the λ_j equals the signal to noise ratio (SNR) of g_{jt} (e.g., ratio of the return variations caused by the true factor g_{jt} and by idiosyncratic errors ε_{it}). The population average *R*-square (average explanatory power of the empirical factors f_t for individual response variables x_{it}) equals $(\Sigma_{j=1}^{r} \lambda_j) / (1 + \Sigma_{j=1}^{r} \lambda_j)$.

We try 3 different values of *T*: *T* = 60, 120, and 240. For each *T*, we generate seven different numbers of response variables:

$N \in \{25, 30, 36, 40, 50, 100, 200\}$ for $T = 60$;

$N \in \{25, 50, 60, 75, 80, 100, 200\}$ for $T = 120$;

$N \in \{25, 50, 100, 120, 145, 160, 200\}$ for $T = 240$.

We have chosen different numbers of response variables for each *T* because the finite-sample performances of the BIC1 and BICD1 estimators critically depend on the relative sizes of *T* and

[3]The random matrix is of the form $M'M$ where the entries of the $K \times K$ matrix *M* are drawn from *N*(0,1).

N. What we find from the reported and unreported simulations is that the two estimators tend to begin over-estimating the ranks of the beta and demeaned beta matrices as N increases further from one-half of T (when $T \leq 120$) or two-thirds of T (when $T \leq 240$). We use different values of N for each T to highlight this pattern.

For each combination of N and T, we also consider two cases: one with five empirical factors ($K = 5$) with two different beta ranks, r = 1 and 3; and the other with ten empirical factors ($K = 10$) with two different beta ranks, r = 1 and 3. For each simulation, we set the SNRs of the true factors (λ_j) at values not greater than 0.05. For each combination of N, T, K, and r, 1,000 samples are drawn.

Our simulation setup may not represent the true data generating processes of asset returns. However, we choose parameter values such that the simulated data have properties similar to those of actual data. First, empirical studies of asset pricing models routinely use monthly data over five, ten, or twenty years. The values of T are chosen to be consistent with this practice. Second, the empirical factors proposed in the literature generally have low explanatory power for individual stock returns although they have higher explanatory power for portfolio returns. To investigate the cases in which empirical factors have limited explanatory power for response variables, we have generated data with latent factors with low SNRs (λ_j). Third, the idiosyncratic error components of actual returns are likely to be cross-sectionally correlated. Under our simulation setup, the error terms are cross-sectionally correlated through the unobserved factor components h_{1t} and h_{2t}. We could have generated cross-sectionally correlated errors using the estimated variance-covariance matrix of the errors from actual data, but from the actual data with N close to or greater than T, we could not consistently estimate the variance matrix of the error vector $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, ..., \varepsilon_{Nt})'$. For example, the estimated variance matrix is not invertible if $N > T$ although the true variance matrix would be invertible. Thus, the errors generated based on an estimated variance matrix from actual data are likely to have a different cross-sectional covariance structure from that of the idiosyncratic error components of actual returns.

Finally, the empirical factors proposed in the literature (f_t in our notation) are unlikely to be perfectly correlated with true latent factors (g_t in our notation). When the empirical factors are not perfectly correlated with true latent factors, the finite-sample performances of the rank estimators could depend on the degrees of correlation between f_t and g_t. Our simulations, however, can provide useful information for such general cases. When empirical factors are imperfect proxy variables, the errors ε_{it} should be cross-sectionally correlated as we discussed in section 3. In addition, for the cases of imperfect correlation, the SNR of a latent factor (λ_j) in our simulations can be interpreted as the SNR of a linear combination of empirical factors that is maximally correlated with the factor. For example, if a latent factor has SNR of 0.01, it can be interpreted as the linear combination of the empirical factors maximally correlated with the latent factor having SNR of 0.01 and explaining 1% (=0.01/1.01) of total variation in response variables.

3.2. Simulation Results

Our simulation is designed to answer three questions. First, given that the TC1 and BIC1 estimators are designed for data with relatively large *T* and small *N*, we wish to know what data size is required to obtain reliable inferences from the estimators. Second, the main difference between the BIC1 and MBIC estimators is that the former is computed controlling for cross-sectional correlation in the errors. While both estimators are consistent as $T \to \infty$ with fixed *N*, controlling for cross-sectional correlation might improve the accuracy of the BIC estimator when a relatively small number of response variables are analyzed. We investigate this possibility. Third, and most importantly, we wish to (i) know the data size with which the BIC1 and MBIC estimators can accurately estimate the rank of the beta matrix and (ii) assess the relative performance between the two.

We begin by considering the finite-sample performance of the TC1 estimator. Table 1 reports the TC1 estimation results from our simulations with five empirical factors (K = 5). We consider two cases: $r = 1$ and $r = 3$. Data are generated such that the true latent factors (or linear combinations of the five empirical factors) have low SNRs: 0.02, 0.03, and 0.05. With these small values of SNR, each of the true latent factors can explain smaller than 5% of total variation in response variables (R^2 in the table). We have chosen these low values of SNRs because some empirical factors proposed in the literature have only limited explanatory power for portfolio or individual stock returns. Using the low values of SNRs, our simulation results provide better guidance for the data of actual returns. Table 1 reports the percentages (%) of correct estimation by the TC1 estimator. The percentages of underestimation and overestimation are reported below in parentheses.

Table 1 shows that the TC1 estimator performs rather poorly even if $N < T$. For the cases with $N = 25$, the accuracy of the TC1 estimator is not greater than 78.9% (when $N = 240$) if r = 1 and not greater than 87.8% if r = 3. The estimator's performance improves with T but deteriorates with N. The accuracy of the estimator is not greater than 8.7% for the cases with $N \geq 50$ and $T \leq 120$, not greater than 55.0% for the cases with $N \geq 50$ and $T = 240$. Furthermore, in all of the cases considered in Table 1, the performance of the TC1 estimator is dominated by that of the BIC1 estimator, as shown in Table 2.

The performance of the BIC1 and MBIC estimators are reported in Table 2. The data generating process is the same as the one described in the beginning of this section. We report the results from the simulated data with SNRs of 0.02, 0.03 and 0.05 if $T \leq 120$ and with SNRs of 0.01, 0.02 and 0.03 if $T = 240$. We do so because the BIC1 and MBIC estimators can capture much weaker latent factors when $T = 240$.

The accuracy of the BIC1 estimator appears to have a non-monotonic relationship with the number of response variables (N). For the cases with $r = 1$, the accuracy of the estimator increases with N when $N \leq T / 2$ and decreases with N when $T / 2 < N \leq T$. The estimator overestimates the rank of the beta matrix when $T / 2 < N \leq T$. When $N > T$, the accuracy of the estimator increases with N up to some points (e.g., (T, N) = (60, 100) and (120, 200)). However, as N increases further, the estimator begins to underestimate the beta rank, and its accuracy drops sharply. In order to investigate this irregular behavior of the BIC1 estimator further, we conducted some additional experiments using data with $N > T / 2$. To save space, we just summarize the results here without reporting them. For a given T, the degree of overestimation by the BIC1 estimator increases as N increases from $N = T / 2$ to $N = T$. However, the accuracy of the estimator improves as N increases from $N = T$ to $N = 2T$ (or near $2T$). Then, as N increases further from $N = 2T$ (or near $2T$), the estimator starts to under-estimate the beta rank. In general, the BIC1 estimator tends to over-estimate the beta rank if $T / 2 < N < 2T$, while it severely under-estimates the rank when $N > 2T$. The tendency of over-estimation reverses to the tendency of under-estimation at some point in $T < N < 2T$. Thus, the BIC1 estimator occasionally performs well at some points when $T < N < 2T$. However, even for such cases, the BIC1 estimator is outperformed by the MBIC estimator.

We can find a similar pattern for the cases with $r = 3$. The accuracy of the BIC1 estimator increases with N up to some points where $T / 2 < N \leq 2T / 3$. However, the accuracy drops quickly and monotonically with N after such points. The results reported in Table 2 suggest that the BIC1 estimator should be used with caution for data with $N > T/2$. The estimator may not be appropriate to use for data with $N > 2T/3$.

Different from the BIC1 estimator, the accuracy of the MBIC estimator generally increases with *N*. In particular, when $N \geq 100$, the MBIC estimator outperforms the BIC1 estimator in all of the cases with $\lambda \geq 0.03$ (in which each latent factor has SNR of 0.03 or greater and can explain at least 2.9% of total variation in response variables). The power of the estimator to identify weaker latent factors increases with *T*. For example, when $T = 60$ and $N \geq 100$, the MBIC estimator can predict the correct beta rank with at least 93.5% accuracy if the latent factors correlated with empirical factors have SNRs of 0.05 or greater ($\lambda \geq 0.05$). When T = 120 and $N \geq 100$, the accuracy of the estimator is greater than 97.8% if $\lambda \geq 0.03$. When T = 240 and $N \geq 100$, the estimator can predict the correct rank of beta matrix with at least 99.8% accuracy when all of the latent factors (correlated with empirical factors) have SNRs of 0.02 or greater. From unreported experiments with data with $T = 360$ and $N \geq 100$, we also find that even if latent factors' SNRs equal 0.01, the MBIC estimator can predict the correct beta rank with at least 99.8% accuracy if $r = 1$ and with at least 82.9% of accuracy if $r = 3$. In contrast, for the cases with $T = 360$, $N = 100$, and $\lambda = 0.01$, the BIC1 estimator predicts the correct beta rank with 69.4% accuracy if $r = 1$, with as low as 10.2% accuracy if $r = 3$.

When N = 25 and $T > 60$, the MBIC estimator often has greater power than the BIC1 estimator to identify the weak factors with SNRs of 0.02 or 0.01, although its accuracy is not greater than 77.3% (see the case with T = 120, N = 25, and r = 1). For the cases with $N < T/2$, the MBIC estimator has greater power to detect such weak factors. The BIC1 estimator's accuracy increases with *N* up to some point where $T/2 < N < 2T/3$. For the cases in which $T/2 \leq N \leq 2T/3$, the BIC1 estimator often has greater power for the factors with SNRs of 0.02 or 0.01 (0.02 for the cases with $T \leq 120$ and 0.01 for the cases with $T \geq 240$), especially when r = 3. However, for the cases with either $N < T/2$ or $N > 2T/3$, the MBIC estimator outperforms even for weak factors with SNRs of 0.02 or 0.01. Overall, the results reported in Table 2 indicate that the MBIC estimator is generally the better estimator to use.

The accuracy of the BIC1 and MBIC estimators may depend on the number of empirical factors (K). Table 3 reports the estimation results from the cases with K = 10. Comparing the results from Table 2, we can see that the accuracy of the two estimators generally falls as more empirical factors are used (while their explanatory power remains the same). However, the general performance patterns of the two estimators remain the same. When $N \geq 80$ and $T \leq 120$, the MBIC estimator outperforms the BIC1 estimator if the latent factors correlated with the empirical factors have SNRs of 0.02 or greater. When $T = 240$, the MBIC estimator outperforms the BIC1 estimator as long as all of the latent factors have SNRs of 0.02 or greater. Overall, the MBIC estimator remains a better estimator.

Finally, we consider the performances of the BICD1 and MBICD estimators. Under our simulation setup, the rank of the demeaned beta matrix equals r – 1. We use the simulated data with r = 3. The estimation results are reported in Table 4. The results are similar to those reported in Tables 2 and 3. As N gets close to or larger than T, the MBICD estimator outperforms the BICD1 estimator for most of the cases and the difference between the relative performances of the two estimators becomes wider as the SNRs of the empirical factors increase. Overall, the MBICD estimator performs better than the BICD1 estimator.

Our simulation results are summarized as follows. First, the accuracy of the BIC1 estimator has a non-monotonic relationship with the number of response variables (N). The power of the estimator initially increases with N but falls as N increases further from some points (smaller than T). Given this non-monotonic relationship, it is difficult to determine the size of data for which the estimator would be appropriate. The BICD1 estimator shows the same pattern. Second, the power of the MBIC estimator generally increases with N. When $N \geq 100$, the estimator has high power to identify the latent factors with SNRs of 0.05 or higher if T = 60, those with SNRs of 0.03 or higher if T = 120, and those with SNRs of 0.02 or higher if T = 240. The MBICD estimator performs equally well. Given these findings, the MBIC and MBICD estimators appear to be the better estimators to use.

4. Application

In this section, we estimate the ranks of different beta matrices using a variety of combinations of empirical factors. Our estimation is conducted with monthly and quarterly data from 1952 to 2011. For the estimation with monthly data, we consider fifteen non-repetitive empirical factors from the three factors of Fama and French (1993, FF); the five factors of Chen, Roll, and Ross (1986, CRR); the three factors of Jagannathan and Wang (1996, JW); the three liquidity-related factors of Pastor and Stambaugh (2003, LIQ); the momentum factor (MOM, selling losers and buying winners 6 – 12 months ago); and the two reversal factors (REV, one by selling winners and buying losers 1 month ago and the other by selling winners and buying losers 13 – 60 months ago). The FF factors are the CRSP value weighted portfolio return minus the return on the one-month Treasury bill (VW), SMB, and HML factors.[4] The CRR factors are industrial production (MP), unexpected inflation (UI), change in expected inflation (DEI), the term premium (UTS), and the default premium (UPR),[5] while the JW factors are the VW, LAB (growth rate of labor income), and UPR factor.[6] The three LIQ factors are aggregate liquidity level, traded liquidity, and innovation in aggregate liquidity.[7]

Five sets of portfolios are used for regressions. Four of them are the 25 Size and Book-to-Market (B/M) portfolios, 30 Industrial portfolios, the 25 Size and Momentum portfolios, and the 100 Size and B/M portfolios. Following the suggestion of Lewellen, Nagel, and Shanken (2010), we also consider the combined set of the 25 Size and B/M and 30 Industrial portfolios.[8] Excess returns on each portfolio are computed using the one-month Treasury bill rate as the risk-free rate.

The data on the 100 Size and B/M portfolios are unbalanced because some portfolios have missing observations. Specifically, twelve portfolios have some missing observations, with the maximum (average) number of missing observations being equal to 48 (21) out of 720 monthly observations from January 1952 to December 2011. As we discussed in Section 3, the MBIC and MBICD estimators can be computed with the portfolio-by-portfolio time series regressions using all of the observations available for each portfolio. The MBIC and MBICD estimators defined in Section 2 are for balanced data in which T is the same for all cross-section units. For unbalanced data, we use the average number of time series observations on individual portfolios for T.

We also analyze individual stock returns (which include dividends). Excess returns are used for regression. The data are downloaded from CRSP. Excluded from our data are REITs

---

[4]All of the FF factors are available from Kenneth French's website.

[5]The CRR factors are available from Laura Xiaolei Liu's webpage (http://www.bm.ust.hk/~fnliu/research.html). For detailed information on how these factors have been constructed, see Liu and Zhang (2008). The UPR factor (default premium) equals the yield spread between BAA- and AAA-rated bonds.

[6]The LAB factor is constructed using the NIPA 2.1 and NIPA 2.6 tables for quarterly and monthly data, respectively. The tables are available at the Bureau of Economic Analysis' webpage: http://www.bea.gov/iTable. Specifically, the factor is the growth rate of total personal income minus personal dividend income divided by total population.

[7]The LIQ factors are available from Lubos Pastor's webpage, http://faculty.chicagobooth.edu/lubos.pastor/research.

[8]According to Lewellen, Nagel, and Shanken (2010), the 25 Size and B/M portfolios have a strong factor structure favoring the FF model, and, thus, model specification tests can produce more reliable inferences when the tests are done with additional portfolios that are not strongly correlated with the SMB and HML factors.

(Real Estate Investment Trusts) and ADRs (American Depositary Receipts). We have also excluded the stock-month observations in which the stocks show more than 300% excess returns in a given month because such huge variations are unlikely due to changes in common factors. Excessively high or low returns are most likely to be driven by idiosyncratic shocks. Expectedly, the data on individual stock returns are heavily unbalanced. Thus, to make sure the number of time series observations is sufficiently large for each stock, we have chosen the stocks whose numbers of time series observations are greater than or equal to $2T/3$ for a given time span, T. Then, the average number of time series observations on individual stocks in the data is used for T in the MBIC and MBICD estimators.

For sensitivity analysis, we also estimate the above factor models using quarterly observations. Analyzing quarterly portfolio and individual stock returns, we can examine seven additional factor models that are discussed in Lewellen, Nagel, and Shanken (2010): the CAPM; the consumption CAPM (CCAPM); the two conditional CCAPMs of Lettau and Ludvigson (2001, LL) and Lustig and Van Nieuwerburgh (2004, LVN); the durable-consumption CAPM of Yogo (2006, Y); the conditional CAPM of Santos and Veronesi (2006, SV); and the investment-based CAPM of Li, Vassalou, and Xing (2006, LVX). Lewellen, Nagel, and Shanken (2010) examined how well the seven models can explain expected returns of the 25 Size and B/M plus the 30 industrial portfolios. Our goal here is not to replicate their analysis but to estimate how many true latent factors are correlated with the empirical factors proposed in their models.

The empirical factors used by the seven models are VW for the CAPM model; CG (aggregate consumption growth rate) for the CCAPM; CG, CAY (aggregate consumption-to-wealth ratio), and CG×CAY for the LL model; CG, MYMO (housing collateral ratio), and CG×MYMO for the LVN model; VW, DCG (durable-consumption growth rate), and NDCG (nondurable-consumption growth rate) for the Y model; VW and VW×LC (labor income-to-consumption ratio) for the SV model; and DHH (change in the gross private investment for households), DCORP (change in the gross private investment for non-financial corporate firms), and DNCORP (change in the gross private investment for non-financial non-corporate firms) for the LVX model.[9]

[9] We are grateful to Jonathan Lewellen and Stefan Nagel for sharing their data with us. The CG, CAY, and LC factors can be directly downloaded or constructed using the data available from Sydney Ludvigson's website, http://www.econ.nyu.edu/user/ludvigsons. The DCG and NDCG factors are constructed using data from the *NIPA* 2.3.3 and *NIPA* 2.3.5 tables. We also use the *Consumer-Durables Goods: Chain-Type Quantity Indexes for Net Stock* table for constructing DCG. All these tables are available at the Bureau of Economic Analysis webpage: http://www.bea.gov/iTable. For the DHH, DCORP, and DNCORP factors we use the *Flow of Funds Accounts* tables available at the Federal Reserve Board's webpage: http://www.federalreserve.gov. Specifically, we

4.1. Results from Monthly Stock Portfolio Returns

In this subsection we report the estimation results obtained using the five sets of monthly portfolio returns. The sample period is from January 1952 to December 2011 (T = 720). The monthly observations on the LAB factor of Jagannathan and Wang (1996, JW) are available only from March 1959. Thus, whenever we estimate a factor model with the JW factors, we use the data from March 1959 to December 2011 (T = 634). As discussed above, the data on the 100 Size and B/M portfolios are unbalanced. Thus, we use the average of the time series observations on individual portfolios to compute the MBIC and MBICD estimates.

The cross-sectional dimension N equals the number of portfolios used to estimate a beta matrix. The results from the entire sample period and two subsample periods are reported in Table 5. For each combination of portfolio sets and empirical factors, we report the adjusted R-square ($\bar{R}^2$, explanatory power of empirical factors) and the estimated rank of the beta matrix by the MBIC estimator. The MBICD estimation results are reported in parentheses. Our simulation results reported in Section 4 indicate that the BIC1 estimator produces reliable inferences when using data with $T \geq 240$ and $N \leq T/2$. The data used for Table 5 satisfy all these conditions. Thus, we also estimated the ranks of the beta matrices using the BIC1 and BICD1 estimators. The estimation results are not materially different from those from the MBIC and MBICD estimators, which are reported in Table 5.

The results from the entire sample (T = 720, or T = 634 if the LAB factor of JW is used) and two sub-samples (T = 360, or T = 274 in the first sub-sample when JW is used) are in Panels A, B, and C of Table 5, respectively.[10] The main observations from Panel A are as follows. First, for all of the five portfolio sets, the MBIC estimator predicts that the beta matrix corresponding to the FF factors has the rank of three. This result is consistent with the notion that the three FF factors are correlated with three linearly independent latent risk factors.

---

used the table FA155019005 for the DHH factor, the tables FA105019005 and FA105020005 for the DCORP factor, and the tables FA115019005 and FA115020005 for the DNCORP factor.

[10] We do not report the estimation results using the LIQ factors. The data on the LIQ factors are only available from December 1969 to December 2008. Our unreported estimation results (with the data from December 1969 to December 2008) show that the three LIQ factors generate a beta matrix with a rank of one for all of the five portfolio sets we analyze. When the LIQ factors are added to the FF model, the rank of the beta matrix does not change for four out of the five portfolio sets we test. The rank increases to four for only one case (the 25 Size and Momentum portfolios). We do not find any evidence that the LIQ factors are correlated with additional latent factors that are not explained by the FF and CRR factors or by the FF and MOM+REV factors.

Second, the beta matrices corresponding to the MOM+REV, CRR or JW factors all fail to have full column rank. This implies that the two-pass estimation would not be able to identify each of the risk premiums related to the MOM+REV, CRR, and JW factors. The explanatory power of the FF factors is much stronger than that of other factors.[11] The explanatory power of the CRR factors is particularly low: the factors can explain no more than 2% of the average total variation in the portfolio returns analyzed.

Third, the four factor model of Carhart (1997), which uses the three FF factors and the MOM factor, produces beta matrices with a rank of four for the four portfolio sets other than the 25 Size and B/M portfolios. For the four portfolio sets, the MOM factor appears to be correlated with one latent factor that cannot be identified by the FF factors alone. Adding the two REV factors to the FF model increases the rank of the beta matrix by one for the 25 Size and Momentum factors but by none for the four other portfolio sets. Adding the REV factors to the Carhart model does not change the rank of the beta matrix for any portfolio set. The REV factors appear to have little information about the latent factors that cannot be explained by the FF and MOM factors alone.

Fourth, while the CRR factors fail to produce full rank beta matrices, they appear to be correlated with an additional latent factor that is not explained by the FF factors alone. When the CRR factors are used in tandem with the FF factors, the rank of the beta matrix increases by one for four sets of portfolio returns but by none for the set of the 100 Size and B/M portfolios. When we add the JW factors (LAB and UPR) to the FF model, the rank of the beta matrix increases by one for the portfolios sorted by industry (the 30 Industrial portfolios and the 25 Size and B/M plus 30 Industrial portfolios). However, when both the CRR and JW factors (CRR+LAB) are added to the FF model, the rank of the beta matrix increases at most by one. If the CRR and JW factors are respectively correlated with two different latent factors, we should expect that the beta matrix corresponding to the FF, CRR, and LAB factors has a rank of five. Given that the beta matrix has a rank of at most four for all of the portfolios we consider, the CRR and LAB factors appear to be correlated with the same single latent factor that cannot be identified by the FF factors alone.

We have also run some unreported tests to detect which of the five CRR (MP, UI, DEI, UTS, and UPR) and LAB factors, or which linear combinations amongst them, can increase the rank of the beta matrix. We find the following. First, adding any single factor to the FF model

[11] Among the FF three factors, the VW factor has the strongest explanatory power.

generally does not increase the rank of the beta matrix. One exception is the case in which the UI factor is added to the FF model to analyze the 30 Industry portfolios. Adding the UI factor and one other single factor to the FF model often increases the rank of the beta matrix by one for the portfolios sorted by industry (the 30 Industry portfolio and the 25 Size and B/M plus 30 Industrial portfolios). For the 25 Size and B/M portfolios and the 25 Size and Momentum portfolios, adding the UI factor and two other factors to the FF model occasionally increases the rank of the beta matrix by one.[12] These results seem to indicate that the UI factor is an important determinant of the rank of the beta matrix, suggesting that the FF model might be missing an inflation-related risk factor. However, the UI factor alone does not have sufficient power to increase the rank of the beta matrix. It does only when some other factors are also added.

Finally, adding all of the CRR, LAB, MOM, and REV factors to the FF model increases the rank of the beta matrix by one or two. This result, together with the result that adding the CRR and LAB factors or the MOM and REV factors to the FF model can increase the rank of the beta matrix by one, implies that the CRR+LAB factors and the MOM+REV factors have information about at most two different latent factors that cannot be identified by the FF factors alone. However, the extra explanatory power of the CRR, LAB, MOM, and REV factors for portfolio returns is quite low. When the factors are added to the FF model, the adjusted *R*-square increases by 2% or less. The only exception is the case with the 25 Size and Momentum portfolios, in which adding all of the factors to the FF model increases the adjusted *R*-square by at most 8.3%.

Overall, the FF model is the only model that generates full-column beta matrices for all of the five portfolio sets we investigate. Most of the individual factors out of the CRR, MOM, REV, and LAB factors fail to identify the latent factor that cannot be explained by the FF factors alone. Use of multiple empirical factors can help identify additional latent factors. However, it is important to note that the corresponding beta matrix is likely to fail to have full column rank.

Panels B and C of Table 5 report the estimation results from two subsample periods. The main results from Panel B and Panel C are the same as those from Panel A. For both subsample periods, the estimated rank of the FF beta matrix is three for every portfolio set, and adding the CRR, LAB, MOM, and REV factors to the FF model increases the rank of the beta matrix at most by two. However, some observations from Panels B and C are also worth noting here.

[12] For example, the rank of the beta matrix increases by one when we add the MP, UI, and UTS factors to the FF model with the 25 Size and B/M portfolios and when the UI, DEI, and UPR factors are added to the FF model with the 25 Size and Momentum portfolios.

First, adding the two REV factors to the FF model often increases the rank of the beta matrix, especially for the second subsample period (from January 1982 to December 2011, Panel C). Adding both the MOM and REV factors to the FF model can increase the rank of the beta matrix by two. It appears that the REV factors have become more informative for true latent factors in more recent years. Second, the explanatory power of the CRR factors for portfolio returns has decreased over time. For the second subsample period (Panel C), the adjusted R-squares from the regressions with the CRR factors alone are smaller than 1% for all five sets of portfolio returns.

The MBICD estimation results reported in Table 5 show that the demeaned beta matrices corresponding to the FF factors have the rank of three for all of the five sets of portfolio returns and over both the entire sample and the two subsample periods. Intriguingly, the demeaned beta matrices corresponding to the CRR factors have a rank of zero, which implies that the latent factors correlated with them may have constant betas over different portfolio returns. For most of the portfolio sets analyzed, the demeaned beta matrices corresponding to the MOM+REV factors have a rank of one. When we use all of the twelve empirical factors (FF, CRR, LAB, MOM, and REV), the MBICD estimates are often smaller than the MBIC estimates by one, indicating that the betas corresponding to a latent factor may be constant over different portfolios.

4.2. Results from Monthly Individual Stock Returns

In this subsection, we report the estimation results obtained using monthly individual stock returns and the same empirical factors used in the previous subsection. We use the data from January 1952 to December 2011. Again, whenever we estimate a model using the JW factors, we use the data from March 1959. As in the previous subsection, we also divide the entire sample period into two 30-year subsample periods. In order to make sure we use a sufficiently large number of time series observations for each stock, we only choose those with at least two thirds of T ($2T/3$), where T is the sample period. The number of individual stocks for the entire sample period is 614. The numbers of individual stocks for the two subsample periods are 781 and 2,268. The individual return data cover large numbers of cross-sectional units, which are often greater than the numbers of time series observations. Our simulation results indicate that the MBIC and MBICD estimators are appropriate for the analysis of such data. Since the data are unbalanced, we use the cross-sectional average of the time series observations for the T in the MBIC and MBICD estimators.

The estimation results are reported in Table 6. We find that the beta matrix corresponding to the FF factors has full column rank for the entire sample period and the first subsample period. For the second subsample period, we find that the FF beta matrix has deficient rank. This result could be explained by the fact that the explanatory power of the empirical factors has dramatically decreased in the second subsample period: their explanatory power during the second subsample period is almost half of their power during the first subsample period. Weak factors are hard to detect. Thus, the rank estimates are likely to be downward biased ones when some factors have very weak explanatory power.

Adding the MOM, REV, CRR, or JW factors individually to the FF model does not increase the rank of the beta matrix for any sample period. Adding both the MOM and REV factors to the FF model increases the rank of the beta matrix by one for the entire sample period. In contrast, adding the CRR or/and JW factors to the FF model does not increase the rank for any sample period. Adding all of the MOM, REV, CRR, and LAB factors to the FF model increases the beta rank by one when the data over the entire sample period are used. The twelve empirical factors we consider (FF, MOM, REV, CRR, and LAB) appear to be correlated with four latent factors for the entire sample and the first subsample periods and with three latent factors for the second subsample period. The estimated ranks of beta matrices are smaller for individual stocks than those for portfolios, especially for the second subsample period. Again, this may be related to the fact that the empirical factors have weaker explanatory power for monthly individual stock returns than for portfolio returns. For example, the twelve empirical factors together explain at most 33.6% of the variation in individual monthly stock returns, while they explain at least 60.0% and often more than 75% of the variation in monthly portfolio returns.

Overall, the results in Table 6 are consistent with the notion that one of the four latent factors that was important in earlier years may have become less important in more recent years. This may have happened because idiosyncratic risks of individual stocks have increased over time. The results in Tables 5 and 6 support this scenario. Table 6 indicates that the explanatory power of the twelve empirical factors for individual stock returns is substantially weaker for the second subsample period. In contrast, Table 5 shows that the explanatory power of the same empirical factors for portfolio returns has been only mildly decreasing over time.

Similarly to Table 5, Table 6 also shows that the MBICD estimates are often smaller than the MBIC estimates by one during the entire sample and first subsample periods. Even for individual stock returns, some latent factors appear to have constant or near-constant betas. However, for the second subsample period, the MBICD estimates are all the same as the MBIC estimates, indicating that the latent factors that we identified during the second subsample period have non-constant betas.

4.3. Results from Quarterly Returns

For sensitivity analysis, we re-estimate the above factor models using quarterly data. The quarterly portfolio returns used are again the 25 Size and B/M portfolios, 30 Industrial portfolios, the combination of the 25 Size and B/M portfolios and the 30 Industry portfolios, the 25 Size and Momentum portfolios, and the 100 Size and B/M portfolios. The quarterly individual stock returns consist of the same 614 individual stocks used in the monthly analysis for the entire sample period. We use the data from the first quarter of 1952 to the fourth quarter of 2011 (T = 240). The results from the estimation with the quarterly returns are presented in Table 7.

The estimation results are very similar to those from the analyses of monthly portfolio returns (Panel A of Table 5) and individual stock returns (Table 6). For the quarterly portfolio returns, we again find that the FF beta matrices have full column rank for all of the five portfolio sets. The MBICD estimation results suggest that the demeaned beta matrices corresponding to the FF factors also have a rank of three for every portfolio set we consider. Adding the MOM factor to the FF model increases the rank of the beta matrix to four only for the 25 Size and Momentum portfolios, while adding the REV or MOM+REV factors does increase the rank by one for the portfolio sets other than the 100 Size and B/M portfolios. It appears that the REV factors are more informative for the analysis of quarterly portfolio returns, while the MOM factor is more informative for monthly portfolio returns. Adding the CRR or JW factors, or adding both the CRR and JW factors to the FF model, increases the rank of the beta matrix only by one for the portfolios sorted by industry (the 30 Industrial portfolios and the combination of the 25 Size and B/M and 30 Industry portfolios). When we add the CRR and JW factors to the model with the FF and MOM+REV factors, the rank of the beta matrix increases to five for the portfolios sorted by industry and to four for the other portfolios. It appears that the CRR+LAB factors are more informative for portfolios sorted by industry than for other portfolios. For the 100 Size and B/M portfolios, adding either the MOM+REV or CRR+LAB factors to the FF model does not increase the rank of the beta matrix, while adding all of the empirical factors together increases the rank by one. The extra explanatory power of the CRR, LAB, MOM, and REV factors for quarterly portfolio returns is quite low. When the factors are added to the FF model, the adjusted R-squares increase by less than 2% percent for four of the five portfolio sets. For the 25 Size and Momentum portfolios, the adjusted R-square increases by 6.2%.

For quarterly individual stock returns, we find a deficient rank from the FF beta matrix. Adding any of the MOM, REV, CRR, and JW factors to the FF model yields the beta matrix with a rank of three. When the CRR+LAB factors are added to the FF model, we find the rank of the corresponding beta matrix increases to four. Overall, we find evidence that there are four (linear combinations of) latent factors correlated with the empirical factors we consider to explain the quarterly individual stock returns.

Next, we consider seven factor models discussed in Lewellen, Nagel, and Shanken (2010). The models are the CAPM and the CCAPM and the models of Lettau and Ludvigson (2001, LL); Yogo (2006, Y); Santos and Veronesi (2006, SV); Li, Vassalou, and Xing (2006, LVX); and Lustig and Van Nieuwerburgh (2004, LVN). Except for the CAPM factor (VW), all of the empirical factors used by these models are observed only quarterly. We refer to all these factors as *quarterly macro factors*. We also consider the quarterly FF, MOM, REV, CRR, and LAB factors for comparison.

The estimation results with quarterly portfolio returns and quarterly individual stock returns are reported in Table 8.[13] The main results from the MBIC estimation with quarterly portfolio returns are the following. First, for every macro factor model, the beta matrix has a rank of one. This result implies that the two-pass estimation could not successfully identify the risk premiums related to the quarterly macro factors. The adjusted R-squares for the models with no VW component (CCAPM, LL, and LVX) are very small (smaller than 4%), while the adjusted R-squares from the models with the VW component (Y and SV) are very similar to those from the CAPM.

Second, when the LL factors are added to the FF model, the rank of the beta matrix increases by one for three of the five portfolio sets (the 30 Industrial portfolios, the 25 Size and B/M plus 30 Industrial portfolios, and the 25 Size and Momentum portfolios). When the SV or LVX factors are added to the FF model, the rank of the beta matrix increases by one only for the 25 Size and Momentum portfolios. When we add all of the macro quarterly factors to the FF model, the rank of the beta matrix increases by one for the three portfolio sets, while the rank does not increase at all for the other two portfolio sets. This result indicates that the quarterly

[13] We do not report the estimation results for the LVN model because the time series data on the MYMO (housing collateral ratio) factor are available only up to the first quarter of 2005. From the estimation with the data up to the first quarter of 2005, we found that the beta matrices corresponding to the LVN three factors have the rank of one for all of the five portfolio sets and individual stock returns. In addition, adding the three LVN factors to the FF model does not increase the rank of the beta matrix.

macro factors are correlated with at most one single latent factor that cannot be identified by the FF factors alone.

Third, the additional latent factor identified by the quarterly macro factors appears to be also correlated with the CRR+LAB factors. When the MOM+REV factors are added to the model with the quarterly macro factors and the FF factors, the rank of the beta matrix increases by one for every portfolio set except for the 25 Size and Momentum portfolios. However, when the CRR+LAB factors are added to the same model, the rank of the beta matrix does not increase for any set of portfolios. These results show that the latent factor captured by the MOM+REV factors is different from the latent factor captured by the macro quarterly factors. However, the CRR+LAB and the quarterly macro factors appear to capture the same latent factor that cannot be identified by the FF factors alone.

Fourth, Tables 7 and 8 show that for any portfolio set, adding the CRR+LAB and MOM+REV factors and the nine quarterly macro factors to the FF model increases the adjusted *R*-square by no more than 6.2% (= 92.8% - 86.5%, for the 25 Size and Momentum portfolios). In contrast, adding the nine quarterly macro factors to the model with the FF, MOM, REV, CRR, and LAB factors increases the adjusted *R*-square at most by 0.5% (= 67.2% - 66.7%, for the 30 Industrial portfolios). It appears that the quarterly macro factors have only a limited explanatory power for portfolio returns.

Fifth, the MBICD estimation results from monthly and quarterly portfolio returns show similar patterns. For the models without the VW factor (the CCPAM, LL, and LVX models), the MBICD estimates are all equal to zero.[14] This indicates that the betas corresponding to the quarterly macro factors have little cross-sectional variation. For the models using the three FF factors and the quarterly macro factors, the MBICD estimates are equal to four for the 25 Size and B/M plus 30 Industrial portfolios but equal to three for all other portfolio sets. When we add the quarterly macro factors to the model with the FF, MOM, REV, CRR, and LAB factors, the MBICD estimates are equal to five for the 25 Size and B/M plus 30 Industrial portfolios but equal to four for other portfolio sets.

Some findings from the quarterly individual stock returns are different from those from the quarterly portfolio returns. First, the MBIC estimates of the beta matrices corresponding to the CCAPM and LL models equal zero, while the MBIC rank estimate for the SV model is two. Second, adding the quarterly macro factors individually to the FF model does not increase the

[14]Using the data up to the second quarter of 2005, we found that the demeaned beta matrix of the LV model also has zero rank.

beta rank. Third, adding all the quarterly macro factors to the FF model increases rank by one. This implies that the macro quarterly factors are correlated with a latent factor that cannot be identified by the FF factors alone. This latent factor is different from the latent factor correlated with the MOM+REV, CRR, and JW factors. Fourth, when all of the empirical factors (the FF, CRR, MOM+REV, JW, and quarterly macro factors) are used, the rank of the beta matrix equals five. This finding is consistent with the notion that there are five latent factors that are important to explain the quarterly individual stock returns.

4.4. Summary of Empirical Results

The main results from our analysis of actual return data can be summarized as follows. First, the twenty-six empirical factors we have considered can identify at most five latent factors in the U.S. stock returns. Second, our estimation results suggest that the CAPM factor (VW) and the two factors (SMB and HML) added by FF are correlated with three linearly independent latent factors, at least in portfolio returns. Third, the MOM and REV factors appear to be correlated with one additional latent factor that cannot be identified by the three FF factors alone. Fourth, there is some evidence that some linear combinations of macro factors may be correlated with another latent factor not explained by the FF and MOM+REV factors. Fifth and finally, the FF model is the only multifactor model that persistently generates full-column rank beta matrices for portfolio returns. For individual stock returns, the FF beta matrices have ranks of two or three depending on sample periods and data frequencies.

5. Concluding Remarks

In this paper, we have estimated the ranks of the beta matrices corresponding to many empirical factors that have been used in the literature. The estimation method used in this paper can be used for the data with a large number of asset returns. We develop two modified versions of the BIC estimator proposed by Cragg and Donald (1997). The estimators can be used for both data with large and small numbers of asset returns, as long as the number of time series observations on each return is sufficiently large. Our simulation results provide promising evidence for the estimators. Recent studies have shown that some factor loadings may have little cross-sectional variations. Our estimation method can also be used to detect such loadings.

Our empirical analysis shows that the three factors in the augmented CAPM of Fama and French (1993, FF) have strong explanatory power for the U.S. individual stock returns and portfolio returns. The three factors appear to be correlated with three linearly independent latent factors, at least in portfolio returns. If the number of linearly independent latent factors can be viewed as representing the number of different sources of risk, our results suggest that the co-movement in the U.S. stock returns is driven by at least three different types of risk. We also have considered many other empirical factors that have been used in the literature. These factors appear to have some power to identify one or two additional latent factors that the three FF factors alone are not able to identify. However, their additional explanatory power for returns is limited, conditional on the FF factors.

Bai and Ng (2002) and Onatski (2010) have developed the estimation methods that can estimate the number of latent factors without using empirical factors. Using their methods, Bai and Ng (2002) and Onatski (2012) have found evidence that two latent factors drive the U.S. individual stock returns. In contrast, our results suggest two or three more latent factors in the individual stock returns. These contradictive results can be explained as follows. The estimation results of Bai and Ng (2002) and Onatski (2012) are obtained from principal components analysis of stock returns. As Onatski (2012) noted, principal components analysis may underestimate the number of latent factors and provide poor estimates of latent factors when some latent factors have only weak explanatory power for response variables. An advantage of our rank estimation method is that it utilizes the information from empirical factors to identify the number of latent factors. Our simulation and empirical results show that estimating the rank of the beta matrix is a more powerful way to identify the presence of weak factors.

Appendix

Let $F = (f_1, f_2, ..., f_T)'$, and let X and E be the $T \times N$ matrices of individual returns and their error components, respectively. Also, let $P_N = N^{-1} 1_N 1_N'$ and $Q_N = I_N - P_N$, where 1_N is the $N \times 1$ vector of ones. Define P_T and Q_T using the $T \times 1$ vector of ones (1_T). Then, $Q_N 1_N = 0_{N \times 1}$, $Q_N P_N = 0_{N \times N}$, $Q_N Q_N = Q_N$, and $P_N P_N = P_N$. The Q_T and P_T matrices have the same properties. With this notation, model (3) can be written as $X = 1_T \alpha' + F\mathrm{B}' + \mathrm{E}$. If we pre-multiply this matrix equation by Q_T, we have $Q_T X = Q_T F \mathrm{B}' + Q_T \mathrm{E}$. In addition,

$$\hat{\mathrm{B}} = \hat{\Sigma}_{xf} \hat{\Sigma}_{ff}^{-1} = \frac{X' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1} = \mathrm{B} + \frac{\mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1}. \qquad \text{(A1)}$$

Under Assumption B, $rank(Q_N B) = r^d$. Thus, there exist $N \times r^d$ and $K \times r^d$ matrices, A_d and S_d, such that $Q_N \mathrm{B} = A_d S_d'$ and $rank(A_d) = rank(S_d) = r^d$. Let $P(A_d) = A_d (A_d' A_d)^{-1} A_d'$ and $Q(A_d) = I_N - P(A_d)$. Similarly, there exist $N \times r$ and $K \times r$ matrices, A and S, such that $\mathrm{B} = AS'$, $rank(A) = rank(S) = r$, and S is a finite matrix. With this notation, we can rewrite (A1) as

$$\begin{aligned} Q_N \hat{\mathrm{B}} &= A_d S_d' + [Q(A_d) + P(A_d)] \frac{Q_N \mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1} \\ &= A_d \left[S_d' + (A_d' A_d)^{-1} A_d' \frac{Q_N \mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1} \right] + Q(A_d) \frac{Q_N \mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1} \qquad \text{(A2)} \\ &\equiv A_d \hat{S}_d' + Q(A_d) \frac{Q_N \mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \hat{\mathrm{B}} &= A \left[S' + (A'A)^{-1} A' \frac{\mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1} \right] + Q(A) \frac{\mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1} \\ &\equiv A \hat{S}' + Q(A) \frac{\mathrm{E}' Q_T F}{T} \left(\frac{F' Q_T F}{T} \right)^{-1}. \end{aligned} \qquad \text{(A3)}$$

The following three Lemmas are used to prove Theorem 1.

Lemma 1: Under Assumptions A – D, we have the following results:

(i) $\left\| \frac{A_d}{\sqrt{N}} \right\| = O_p(1)$; (ii) $\left\| \frac{(A_d' A_d)^{-1} A_d' Q_N \mathrm{E}' Q_T F}{\sqrt{T}} \right\| = O_p(1)$; (iii) $\left\| \frac{Q(A_d) Q_N \mathrm{E}' Q_T F}{\sqrt{NT}} \right\| = O_p(1)$;

(iv) $\left\|\frac{A}{\sqrt{N}}\right\| = O_p(1)$; (v) $\left\|\frac{(A'A)^{-1}A'\mathrm{E}'Q_T F}{\sqrt{T}}\right\| = O_p(1)$; $\left\|\frac{Q(A)\mathrm{E}'Q_T F}{\sqrt{NT}}\right\| = O_p(1)$.

***Proof of Lemma*1**: We here prove only (i), (ii), and (iii) because (iv), (v), and (vi) can be proven in the same way by dropping Q_N and replacing A_d and S_d with *A* and *S*, respectively. By Assumption B,

$$\left\|\frac{Q_N \mathrm{B}}{\sqrt{N}}\right\| = \sqrt{trace\left(\frac{\mathrm{B}'Q_N\mathrm{B}}{N}\right)} = \sqrt{O(1)} = O(1) .$$

Thus, (i) holds because S_d is a finite matrix, and, therefore,

$$\left\|\frac{A_d}{\sqrt{N}}\right\| = \left\|\frac{Q_N \mathrm{B} S_d (S_d' S_d)^{-1}}{\sqrt{N}}\right\| \le \left\|\frac{Q_N \mathrm{B}}{\sqrt{N}}\right\| \left\|S_d \left(S_d' S_d\right)^{-1}\right\| = O(1) . \qquad \text{(A4)}$$

By Assumptions A, C and D,

$$\left\|\frac{F'1_T}{T}\right\| = \left\|\bar{f}\right\| = O_p(1) ;$$

$$\left\|\frac{\mathrm{E}'1_T}{\sqrt{NT}}\right\|^2 = \frac{\Sigma_{i=1}^N \left(\Sigma_{t=1}^T \varepsilon_{it}\right)^2}{NT} = \frac{1}{N}\Sigma_{i=1}^N \left(\frac{1}{\sqrt{T}}\Sigma_{t=1}^T \varepsilon_{it}\right)^2 = O_p(1) ;$$

$$\left\|\frac{\mathrm{E}'F}{\sqrt{NT}}\right\|^2 = \frac{1}{NT}\Sigma_{i=1}^N \Sigma_{t=1}^T \Sigma_{s=1}^T \varepsilon_{it}\varepsilon_{is} f_t' f_s = O_p(1) .$$

Therefore, we have

$$\begin{aligned} &\left\|\frac{Q_N \mathrm{E}' Q_T F}{\sqrt{NT}}\right\| \\ &= \frac{1}{\sqrt{NT}}\left\|\mathrm{E}'F - \frac{1}{T}\mathrm{E}'1_T 1_T' F - \frac{1}{N}1_N 1_N' \mathrm{E}'F + \frac{1}{NT}1_N 1_N' \mathrm{E}'1_T 1_T' F\right\| \\ &\le \frac{1}{\sqrt{NT}}\left\|\mathrm{E}'F\right\| + \frac{1}{\sqrt{NT}}\left\|\frac{1}{T}\mathrm{E}'1_T 1_T' F\right\| + \frac{1}{\sqrt{NT}}\left\|\frac{1}{N}1_N 1_N'\right\|\left\|\mathrm{E}'F\right\| + \frac{1}{\sqrt{NT}}\left\|\frac{1}{N}1_N 1_N'\right\|\left\|\frac{1}{T}\mathrm{E}'1_T 1_T' F\right\| \\ &= 2\left\|\frac{\mathrm{E}'F}{\sqrt{NT}}\right\| + 2\left\|\frac{1}{T\sqrt{NT}}\mathrm{E}'1_T 1_T' F\right\| \le 2\left\|\frac{\mathrm{E}'F}{\sqrt{NT}}\right\| + 2\left\|\frac{1_T'\mathrm{E}}{\sqrt{NT}}\right\|\left\|\frac{1_T' F}{T}\right\| = O_p(1). \end{aligned} \qquad \text{(A5)}$$

By (A4) and (A5), the result (ii) holds because

$$\left\|\frac{(A_d' A_d)^{-1} A_d' Q_N \mathrm{E}' Q_T F}{\sqrt{T}}\right\| \le \left\|\left(\frac{A_d' A_d}{N}\right)^{-1}\right\| \left\|\frac{A_d}{\sqrt{N}}\right\| \left\|\frac{Q_N \mathrm{E}' Q_T F}{\sqrt{NT}}\right\| = O_p(1) .$$

Finally, (iii) holds because

$$\left\| \frac{Q(A_d)Q_N \mathrm{E}'Q_T F}{\sqrt{NT}} \right\|^2 = trace\left[\frac{F'Q_T \mathrm{E} Q_N Q(A_d) Q_N \mathrm{E}' Q_T F}{NT} \right]$$

$$\leq trace\left[\frac{F'Q_T \mathrm{E} Q_N Q_N \mathrm{E}' Q_T F}{NT} \right] = \left\| \frac{Q_N \mathrm{E}' Q_T F}{\sqrt{NT}} \right\|^2 = O_p(1).$$

Lemma 2 (Ahn and Horenstein, 2013): If A and B are $p \times p$ positive semi-definite matrices,

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

Lemma 3 (Rao, 1973, p. 68): Suppose that two matrices A and B are symmetric of order p. Then,

$$\psi_{j+j'-p}(A+B) \leq \psi_j(A) + \psi_{j'}(B), \; j + j' \geq p+1 .$$

Remark: Although the lemma in Rao (1973, p. 68) looks different from Lemma 3, they are the same. The former lemma is written with the eigenvalues ordered from the largest to smallest and the latter with the eigenvalues ordered from the smallest to the largest.

***Proof of Theorem* 1:** We will first derive the asymptotic properties of the eigenvalues of $\hat{\Sigma}_{ff}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}/N$. The eigenvalues of $\hat{\Sigma}_{ff}^{1/2}(\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}/N)\hat{\Sigma}_{ff}^{1/2}$ are the same as those of $\hat{\Sigma}_{ff}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}/N$. Thus, we consider the properties of the former eigenvalues. By (A2), we have

$$\frac{\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}}{N} = \hat{\Sigma}_{ff}^{1/2}\frac{(Q_N\hat{\mathrm{B}})'Q_N\hat{\mathrm{B}}}{N}\hat{\Sigma}_{ff}^{1/2}$$

$$= \hat{\Sigma}_{ff}^{1/2}\frac{\hat{S}_d A_d' A_d \hat{S}_d'}{N}\hat{\Sigma}_{ff}^{1/2} + \hat{\Sigma}_{ff}^{-1/2}\frac{F'Q_T\mathrm{E}Q_N Q(A_d)Q_N\mathrm{E}'Q_T F}{NT^2}\hat{\Sigma}_{ff}^{-1/2}.$$

By Lemmas 2 and 3, for $j = 1, \ldots, K$,

$$\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{S}_d A_d' A_d \hat{S}_d'}{N}\hat{\Sigma}_{ff}^{1/2}\right)$$

$$\leq \psi_j\left(\frac{\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}}{N}\right) \qquad \text{(A6)}$$

$$\leq \psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{S}_d A_d' A_d \hat{S}_d'}{N}\hat{\Sigma}_{ff}^{1/2}\right) + \psi_K\left(\hat{\Sigma}_{ff}^{-1/2}\frac{F'Q_T\mathrm{E}Q_N Q(A_d)Q_N\mathrm{E}'Q_T F}{NT^2}\hat{\Sigma}_{ff}^{-1/2}\right).$$

Note that

$$\psi_K\left(\hat{\Sigma}_{ff}^{-1/2}\frac{F'Q_T \mathrm{E} Q_N Q(A_d)Q_N \mathrm{E}'Q_T F}{NT^2}\hat{\Sigma}_{ff}^{-1/2}\right)$$

$$\begin{aligned}
&\le \Sigma_{j=1}^{K}\psi_j\left(\hat{\Sigma}_{ff}^{-1/2}\frac{F'Q_T \mathrm{E} Q_N Q(A_d)Q_N \mathrm{E}'Q_T F}{NT^2}\hat{\Sigma}_{ff}^{-1/2}\right)\\
&= trace\left(\hat{\Sigma}_{ff}^{-1/2}\frac{F'Q_T \mathrm{E} Q_N Q(A_d)Q_N \mathrm{E}'Q_T F}{NT^2}\hat{\Sigma}_{ff}^{-1/2}\right) \qquad \text{(A7)}\\
&\le \frac{1}{T}\left\|\frac{Q(A_d)Q_N \mathrm{E}'Q_T F}{\sqrt{NT}}\hat{\Sigma}_{ff}^{-1/2}\right\|^2 \le \frac{1}{T}\left\|\frac{Q(A_d)Q_N \mathrm{E}'Q_T F}{\sqrt{NT}}\right\|^2\left\|\hat{\Sigma}_{ff}^{-1/2}\right\|^2\\
&= \frac{1}{T}\times O_p(1) = O_p\left(\frac{1}{T}\right),
\end{aligned}$$

because of Lemma 1(iii) and the fact that $\left\|\hat{\Sigma}_{ff}\right\| = O_p(1)$ by Assumption A. (A6) and (A7) imply

$$\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{S}_d A_d' A_d \hat{S}_d'}{N}\hat{\Sigma}_{ff}^{1/2}\right) \le \psi_j\left(\frac{\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}}{N}\right) \le \psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{S}_d A_d' A_d \hat{S}_d'}{N}\hat{\Sigma}_{ff}^{1/2}\right) + O_p\left(\frac{1}{T}\right). \quad \text{(A.8)}$$

Since $rank(A_d\hat{S}_d') \le rank(A_d) = r^d$, for $1 \le j \le K - r^d$, $\psi_j\left(\hat{S}_d A_d' A_d \hat{S}_d' / N\right) = 0$. Thus, for 1 **≤***j***≤***K* - r^d,

$$0 \le \psi_j\left(\frac{\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}}{N}\right) \le O_p\left(\frac{1}{T}\right).$$

Lemma 1(ii) implies $p\lim_{T\to\infty}\hat{S}_d = S_d$ because

$$p\lim_{T\to\infty}\frac{(A_d'A_d)^{-1}A_d'Q_N \mathrm{E}'Q_T F}{T}\left(\frac{F'Q_T F}{T}\right)^{-1} = 0.$$

Thus, for $K - r^d + 1 \le j \le K$,

$$\begin{aligned}
\psi_j\left(\Sigma_{ff}^{1/2}\frac{\mathrm{B}'Q_N\mathrm{B}}{N}\Sigma_{ff}^{1/2}\right) &= p\lim_{T\to\infty}\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{S}_d A_d' A_d \hat{S}_d'}{N}\hat{\Sigma}_{ff}^{1/2}\right)\\
&\le p\lim_{T\to\infty}\psi_j\left(\frac{\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}}{N}\right) \le \psi_j\left(\Sigma_{ff}^{1/2}\frac{\mathrm{B}'Q_N\mathrm{B}}{N}\Sigma_{ff}^{1/2}\right),
\end{aligned}$$

which implies

$$p\lim_{T\to\infty}\psi_j\left(\frac{\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}}{N}\right) = \psi_j\left(\Sigma_{ff}^{1/2}\frac{\mathrm{B}'Q_N\mathrm{B}}{N}\Sigma_{ff}^{1/2}\right) > 0.$$

By the same procedure, we can obtain the same results for the eigenvalues of $\hat{\Sigma}_{ff}\hat{\mathrm{B}}'\hat{\mathrm{B}} / N$.

***Proof of Theorem* 2:** We here show the consistency of the MBICD estimator. Let $m(p) = (N-1-p)(K-p)$. Let us denote the MBICD estimator by $\tilde{r}_2$. Note that for all N (even if $N \to \infty$),

$$\frac{m(\tilde{r}_2)-m(r^d)}{N} = \left(1-\frac{\tilde{r}_2+1}{N}\right)(K-\tilde{r}_2)-\left(1-\frac{r^d+1}{N}\right)(K-r^d) < (>)\, 0\,, \tag{A10}$$

if $\tilde{r}_2 > (<) r^d$. For $\tilde{r}_2$ to be greater than r^d, it must be the case that $D_{2,T}(r^d)$ - $D_{2,T}(\tilde{r}_2) > 0$. Thus, $\Pr(\tilde{r}_2 > r^d) \leq \Pr[D_{2,T}(r^d) - D_{2,T}(\hat{r}_2) > 0]$, but as $T \to \infty$,

$$\begin{aligned} &\Pr\left[D_{2,T}(r^d) - D_{2,T}(\tilde{r}_2) > 0\right] \\ &\qquad = \Pr\left[T\Sigma_{j=K-\tilde{r}_2+1}^{K-r^d}\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}}{N}\hat{\Sigma}_{ff}^{1/2}\right)+\hat{\sigma}_\varepsilon^2 w(T)\frac{m(\tilde{r}_2)-m(r^d)}{N} > 0\right] \to 0, \end{aligned} \tag{A11}$$

because $T \times \Sigma_{j=k-\tilde{r}_2+1}^{K-r^d}\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}/N\right) = O_p(1)$ by Theorem 1, $\hat{\sigma}_\varepsilon^2 \to_p \sigma_\varepsilon^2$ by Assumption C(ii), and $w(T)(m(\tilde{r}_2)-m(r^d))/N \to -\infty$ by (A10). Therefore, $\Pr(\tilde{r}_2 > r^d) \to 0$ as $T \to \infty$.

Similarly, for $\tilde{r}_2 < r^d$,

$$\begin{aligned} &\Pr[D_{2,T}(r^d) - D_{2,T}(\tilde{r}_2) > 0] \\ &\qquad = \Pr\left[-\Sigma_{j=K-r^d+1}^{K-\tilde{r}_2}\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\frac{\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}}{N}\hat{\Sigma}_{ff}^{1/2}\right)+\hat{\sigma}_\varepsilon^2\frac{w(T)}{T}\frac{m(\tilde{r}_2)-m(r^d)}{N} > 0\right] \to 0, \end{aligned} \tag{A12}$$

Because $w(T)/T \to 0$, $\Sigma_{j=K-r^d+1}^{K-\tilde{r}_2}\psi_j\left(\hat{\Sigma}_{ff}^{1/2}\hat{\mathrm{B}}'Q_N\hat{\mathrm{B}}\hat{\Sigma}_{ff}^{1/2}/N\right) \to_p \Sigma_{j=K-r^d+1}^{K-\tilde{r}_2}\psi_j\left(\Sigma_{ff}^{1/2}\mathrm{B}'Q_N\mathrm{B}\Sigma_{ff}^{1/2}/N\right) > 0$. (A11) and (A12) imply that the MBICD estimator is consistent. The consistency of the MBIC estimator can be shown similarly.

References

Ahn, Seung C., and Alex R. Horenstein, 2013, Eigenvalue ratio test for the number of factors, *Econometrica* 81, 1203 – 1227.

Ahn, Seung C., Marcos F. Perez and C. Gadarowski, 2013, Two-pass estimation of risk premiums with multicollinear and near-invariant betas, *Journal of Empirical Finance* 20, 1 – 17.

Ahn, Seung C., Young H. Lee, and Peter Schmidt, 2013, Panel data models with multiple time-varying individual effects, *Journal of Econometrics* 174, 1 – 14.

Ang, Andrew, Jun Liu, and Krista Schwarz, 2010, Using stocks or portfolios in tests of factor models, Working paper, Columbia University.

Bai, Jushan, and Serena Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191 – 221.

Bai, Jushan, and Serena Ng, 2002, Evaluating latent and observed factors in macroeconomics and finance, *Journal of Econometrics*, 131, 507 – 537.

Burnside, Craig, 2010, Identification and inference in linear stochastic discount factor models, NBER Working Paper 16634.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57 – 82.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, The Econometrics Of Financial Markets, (Princeton University Press, Princeton, New Jersey).

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

Chen, Nai-Fu, Richard Roll, and Stephen A., Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383 – 403.

Cochrane, John H., 2011, Presidential address: discount rates, *Journal of Finance,* 4, 1047 – 1108.

Connor, Gregory, and Robert A. Korajzcyk, 1989, An intertemporal equilibrium beta pricing model, *The Review of Financial Studies* 2, 373 – 392.

Connor, Gregory, and Robert A. Korajzcyk, 1993, A test for the number of factors in an approximate factor model, *Journal of Finance* 48, 1263 – 1291.

Cragg, John G., and Stephen G. Donald, 1997, Inferring the rank of a matrix, *Journal of Econometrics* 76, 223 – 250.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3 – 56.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: empirical tests, *Journal of Political Economy* 81, 607 – 636.

Harvey, Campbell, Yan Liu and Heqing Zhu, 2013, . . . and the Cross-Section of Expected Returns, working paper.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected return, *Journal of Finance* 51, 3 – 53.

Kan, Raymond, and Chu Zhang, 1999a, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203 – 235.

Kan, Raymond, and Chu Zhang, 1999b, GMM tests of stochastic discount factor models with useless factor, *Journal of Financial Economics* 54, 103 – 127.

Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and two-pass cross-section regression methodology, *The Journal of Finance*, online.

Kleibergen, Frank, and Richard Paap, 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97 – 126.

Lewellen, Jonathan, Stephan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175 – 194.

Lettau, Martin, and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238 – 1287.

Li, Qing, Maria Vassalou, and Yuhang Xing, 2006, Sector investment growth rates and the cross section of equity returns, *Journal of Business* 79, 1637 – 1665.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economic and Statistics* 47, 13 – 37.

Liu, Laura X., and Lu Zhang, 2008, Momentum profits, factor pricing, and macroeconomic risk, *Review of Financial Studies* 21, 2417 – 2448.

Lustig, Hanno N., and Stijn Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *Journal of Finance* 60, 1167 – 1219.

Merton, Robert, 1972, An intertemporal capital asset pricing model, *Econometrica* 41, 867 – 887.

Mossin, Jan, 1966, Equilibrium in a capital asset market, *Econometrica* 35, 768 – 783.

Onatski, Alexei, 2010, Determining the number of factors from empirical distribution of eigenvalues, *Review of Economic and Statistics* 92, 1004 – 1016.

Onatski, Alexei, 2012, Asymptotics of the principal components estimator of large factor models with weakly influential factors, *Journal of Econometrics* 168, 244 – 258.

Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642 – 685.

Rao, Calyampudi R., 1973, Linear statistical inference and its applications, 2nd eds., (John Wiley & Sons, New York, New York).

Robin, Jean-Marc, and Richard J. Smith, 2000, Tests of rank, *Econometric Theory* 16, 151 – 175.

Roll, Richard, 1977, A critique of the asset pricing theory's tests, Part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129 – 176.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341 – 360.

Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1 – 44.

Sharpe, William, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425 – 442.

Treynor, Jack, 1999, Toward a theory of market value and risky assets, Unpublished manuscript, Final version in Asset Pricing and Portfolio Performance; Robert A. Korajczyk, ed., London: Risk Books, 15 – 22.

Zhou, Guofu, 1995, Small sample rank tests with applications to asset pricing, *Journal of Empirical Finance* 2, 71 – 93.

Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539 – 580.

Table 1: TC Estimation of Beta Matrix from Five Empirical Factors

Reported are the percentages (%) of correct estimation by the TC1 estimator from 1,000 simulated data sets. The percentages (%) of under- and over-estimation by each estimator are reported in parentheses (•,•). Data are drawn by $x_{it} = \Sigma_{j=1}^{K} \beta_{ij} f_{jt} + \varepsilon_{it}$; $\varepsilon_{it} = \phi(\xi_{1i} h_{1t} + \xi_{2i} h_{2t}) / (\xi_{1i}^2 + \xi_{2i}^2)^{1/2} + (1-\phi^2)^{1/2} v_{it}$, where K = 5, $\phi = 0.2$, and f_{jt} (j = 1, ... , K), $h_{j't}$, $\xi_{j'i}$ ($j' = 1, 2$) and v_{it} are all randomly drawn from $N(0,1)$. For the beta matrix B, we draw an $N \times r$ random matrix B_g such that its first column equals the vector of ones and the entries in the other columns are drawn from $N(0,1)$. We also draw a random $K \times K$ positive definite matrix, compute the first r orthonormalized eigenvectors of the matrix, and set a $K \times r$ matrix C using the eigenvectors. $\mathrm{B} = \mathrm{B}_g \Lambda^{1/2} C'$, where Λ = $diag(\lambda_1, ..., \lambda_r)$. The rank of the beta matrix equals one or three (r = 1, 3). TC1 refers to the Testing Criterion estimator of Cragg and Donald (1997), which is computed under the assumption that the ε_{it} are *i.i.d.* over time (but cross-sectionally correlated).

			Panel I: K = 5, r = 1, $\lambda_1 = \lambda$			Panel II: K = 5, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=60	N =25	TC1	5.1% (0.0, 94.9)	4.2% (0.0, 95.8)	3.5% (0.0, 96.5)	44.0% (7.3, 48.7)	38.1% (1.6, 60.3)	31.5% (0.1, 68.4)
	N =50	TC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)
T=120	N =25	TC1	50.0% (0.1, 50.0)	47.3% (0.0, 52.7)	45.7% (0.0, 54.3)	69.0% (8.1, 22.9)	69.8% (2.1, 28.1)	67.4% (0.0, 32.6)
	N =50	TC1	0.1% (0.0, 99.9)	0.1% (0.0, 99.9)	0.1% (0.0, 99.9)	8.7% (0.0, 91.3)	7.3% (0.0, 92.7)	6.7% (0.0, 93.3)
	N=100	TC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)
T=240	N =25	TC1	78.9% (0.0, 21.1)	78.1% (0.0, 21.9)	77.8% (0.0, 22.2)	87.8% (0.8, 11.4)	87.2% (0.0, 12.8)	86.5% (0.0, 13.5)
	N =50	TC1	26.7% (0.0, 73.3)	26.4% (0.0, 73.6)	26.1% (0.0, 73.9)	55.0% (0.0, 45.0)	52.7% (0.0, 47.3)	51.6% (0.0, 48.4)
	N =100	TC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.8% (0.0, 99.2)	0.6% (0.0, 99.4)	0.4% (0.0, 99.6)
	N=200	TC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0, 100)

Table 2: BIC Rank Estimation of Beta Matrix from Five Empirical Factors

Reported are the percentages (%) of correct estimation by the BIC1 and MBIC estimators from 1,000 simulated data sets. The percentages (%) of under- and over-estimation by each estimator are reported in parentheses (•,•). Data are drawn by $x_{it} = \Sigma_{j=1}^{K} \beta_{ij} f_{jt} + \varepsilon_{it}$; $\varepsilon_{it} = \phi(\xi_{1i} h_{1t} + \xi_{2i} h_{2t}) / (\xi_{1i}^2 + \xi_{2i}^2)^{1/2} + (1-\phi^2)^{1/2} v_{it}$, where K = 5, $\phi = 0.2$, and f_{jt} (j = 1, ... , K), $h_{j't}$, $\xi_{j'i}$ ($j' = 1, 2$) and v_{it} are all randomly drawn from $N(0,1)$. For the beta matrix B, we draw an $N \times r$ random matrix B_g such that its first column equals the vector of ones and the entries in the other columns are drawn from $N(0,1)$. We also draw a random $K \times K$ positive definite matrix, compute the first r orthonormalized eigenvectors of the matrix, and set a $K \times r$ matrix C using the eigenvectors. $\mathrm{B} = \mathrm{B}_g \Lambda^{1/2} C'$, where Λ = $diag(\lambda_1, ..., \lambda_r)$. The rank of the beta matrix equals one or three (r = 1, 3). BIC1 refers to the Bayesian Information Criterion (BIC) estimator of Cragg and Donald (1997), which is computed under the assumption that the ε_{it} are *i.i.d.* over time (but cross-sectionally correlated). MBIC refers to the BIC estimator computed under the assumption that the ε_{it} are *i.i.d.* over both time and individual response variables. The criterion function of the MBIC estimator uses $T^{0.2}$ instead of ln(T).

			Panel I: K = 5, r = 1, $\lambda_1 = \lambda$			Panel II: K = 5, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=60	N=25	BIC1	55.6% (41.4, 4.8)	72.7% (22.5, 4.8)	87.5% (4.6, 7.9)	2.3% (97.7, 0.0)	11.9% (87.9, 0.2)	45.2% (54.5, 0.3)
		MBIC	33.9% (65.9, 0.2)	70.1% (29.7, 0.2)	97.1% (2.7, 0.2)	1.9% (98.1, 0.0)	13.0% (86.9, 0.1)	56.8% (43.0, 0.2)
	N=30	BIC1	70.4% (15.1, 14.5)	73.3% (6.1, 20.6)	73.1% (1.0, 25.9)	13.3% (86.5, 0.2)	31.9% (67.6, 0.5)	66.3% (30.7, 3.0)
		MBIC	37.5% (62.4, 0.1)	72.4% (27.5, 0.1)	97.7% (2.1, 0.2)	1.8% (98.2, 0.0)	15.9% (84.1, 0.0)	65.9% (34.1, 0.0)
	N =36	BIC1	43.5% (1.8, 64.7)	38.1% (0.3, 61.6)	32.3% (0.0, 67.7)	43.1% (54.1, 2.8)	64.3% (29.8, 5.9)	81.7% (5.7, 12.6)
		MBIC	35.9% (64.1, 0.0)	74.9% (25.0, 0.1)	98.7% (1.2, 0.1)	2.3% (97.7, 0.0)	17.6% (82.4, 0.0)	71.6% (28.4, 0.0)
	N =40	BIC1	14.3% (0.0, 85.7)	12.2% (0.0, 87.8)	9.1% (0.0, 90.9)	60.2% (25.7, 14.1)	64.9% (10.7, 24.4)	62.5% (1.3, 36.2)
		MBIC	36.1% (63.9, 0.0)	74.7% (25.3, 0.0)	99.4% (92.3, 0.0)	1.5% (98.5, 0.0)	18.1% (81.9, 0.0)	76.0% (23.9, 0.1)
	N =50	BIC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	7.3% (0.0, 92.7)	4.6% (0.0, 95.4)	3.0% (0.0, 97.0)
		MBIC	38.3% (61.7, 0.0)	77.9% (22.1, 0.0)	99.3% (0.7, 0.0)	1.5% (98.5, 0.0)	21.2% (78.8, 0.0)	82.0% (18.0, 0.0)
	N =100	BIC1	0.0% (100, 0.0)	0.4% (99.6, 0.0)	5.5% (94.5, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBIC	37.0% (63.0, 0.0)	88.2% (11.8, 0.0)	99.9% (0.1, 0.0)	7.8% (92.2, 0.0)	28.0% (72.0, 0.0)	93.5% (6.5, 0.0)
	N=200	BIC1	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.4% (99.6, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBIC	37.7% (62.3, 0.0)	90.6% (9.4, 0.0)	100% (0.0, 0.0)	13.0% (86.9, 0.1)	38.1% (61.9, 0.0)	98.3% (1.7, 0.0)

Table 2 continued...

			Panel I: K = 5, r = 1, $\lambda_1 = \lambda$			Panel II: K = 5, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=120	N=25	BIC1	26.8% (72.1, 0.0)	65.3% (34.7, 0.0)	98.1% (1.9, 0.0)	0.1% (99.9, 0.0)	9.6% (90.4, 0.0)	60.9% (39.1, 0.0)
		MBIC	77.3% (22.7, 0.0)	97.5% (2.5, 0.0)	100% (0.0, 0.0)	20.1% (79.9, 0.0)	64.5% (35.5, 0.0)	95.5% (4.5, 0.0)
	N=50	BIC1	81.6% (18.2, 0.3)	98.3% (1.4, 0.3)	99.7% (0.0, 0.3)	18.4% (81.6, 0.0)	68.6% (31.4, 0.0)	97.7% (2.3, 0.0)
		MBIC	89.5% (10.5, 0.0)	99.8% (0.2, 0.0)	100% (0.0, 0.0)	33.1% (66.9, 0.0)	86.9% (13.1, 0.0)	99.7% (0.3, 0.0)
	N=60	BIC1	95.3% (3.8, 0.0)	98.9% (0.0, 1.1)	98.6% (0.0, 1.4)	46.0% (53.9, 0.1)	88.8% (11.0, 0.2)	99.4% (0.3, 0.3)
		MBIC	92.2% (7.8, 0.0)	99.8% (0.2, 0.0)	100% (0.3, 0.0)	37.7% (62.3, 0.0)	92.5% (7.5, 0.0)	99.8% (0.2, 0.0)
	N=75	BIC1	75.3% (0.0, 24.7)	73.8% (0.0, 26.2)	72.6% (0.0, 27.4)	88.5% (8.8, 2.7)	96.0% (0.4, 3.6)	95.2% (0.0, 4.8)
		MBIC	93.4% (6.6, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	44.7% (55.3, 0.0)	94.4% (5.6, 0.0)	100% (100, 0.0)
	N=80	BIC1	43.0% (0.0, 57.0)	42.0% (0.0, 58.0)	40.7% (0.0, 59.3)	88.4% (2.3, 9.3)	88.1% (0.0, 11.9)	86.4% (0.0, 13.6)
		MBIC	93.2% (6.8, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	46.9% (53.1, 0.0)	96.0% (4.0, 0.0)	100% (0.0, 0.0)
	N=100	BIC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	3.7% (0.0, 96.3)	3.2% (0.0, 96.8)	2.7% (0.0, 97.3)
		MBIC	96.0% (4.0, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	52.9% (47.1, 0.0)	97.8% (2.2, 0.0)	100% (0.0, 0.0)
	N=200	BIC1	0.1% (99.9, 0.0)	8.3% (91.7, 0.0)	69.0% (31.0, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	7.0% (93.0, 0.0)
		MBIC	98.3% (1.7, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	68.5% (31.5, 0.0)	99.9% (0.0, 0.1)	100% (0.0, 0.0)

Table 2 continued...

			Panel I: K = 5, r = 1, $\lambda_1 = \lambda$			Panel II: K = 5, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .010 R^2=.010	λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .010 R^2=.029	λ = .020 R^2=.057	λ = .030 R^2=.083
T=240	N=25	BIC1	2.0% (98.0, 0.0)	62.4% (37.6, 0.0)	97.6% (2.4, 0.0)	0.0% (100, 0.0)	8.5% (91.5, 0.0)	53.0% (47.0, 0.0)
		MBIC	54.4% (41.6, 0.0)	99.8% (0.2, 0.0)	100% (0.0, 0.0)	6.9% (93.1, 0.0)	76.0% (24.0, 0.0)	96.2% (3.8, 0.0)
	N=50	BIC1	7.0% (93.0, 0.0)	92.7% (7.3, 0.0)	100% (0.0, 0.0)	0.1% (99.9, 0.0)	35.9% (64.1, 0.0)	90.9% (9.1, 0.0)
		MBIC	69.4% (30.6, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	10.3% (89.7, 0.0)	95.3% (4.7, 0.0)	99.8% (0.2, 0.0)
	N=100	BIC1	68.8% (31.2, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	9.6% (0.0, 0.0)	96.4% (3.6, 0.0)	100% (0.0, 0.0)
		MBIC	79.2% (20.8, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	16.2% (83.8, 0.0)	99.8% (0.2, 0.0)	100% (0.0, 0.0)
	N=120	BIC1	93.6% (6.4, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	40.0% (0.0, 0.0)	99.8% (0.2, 0.0)	100% (0.0, 0.0)
		MBIC	81.6% (18.4, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	20.0% (80.0, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)
	N=145	BIC1	99.4% (0.1, 0.5)	99.4% (0.0, 0.6)	99.4% (0.0, 0.6)	92.2% (7.8, 0.0)	99.9% (0.0, 0.1)	99.9% (0.0, 0.1)
		MBIC	84.5% (15.5, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	22.9% (77.1, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)
	N=160	BIC1	93.7% (0.0, 6.3)	92.8% (0.0, 7.2)	92.6% (0.0, 7.4)	98.8% (1.0, 0.2)	99.4% (0.0, 0.6)	99.4% (0.0, 0.6)
		MBIC	83.2% (16.8, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	24.4% (75.6, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)
	N=200	BIC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	10.3% (0.0, 89.7)	8.5% (0.0, 91.5)	7.9% (0, 92.1)
		MBIC	89.4% (10.6, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	27.1% (72.9, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)

Table 3: BIC Estimation of Beta Matrix from Ten Empirical Factors

Reported are the percentages (%) of correct estimation by the BIC1 and MBIC estimators from 1,000 simulated data sets. The percentages (%) of under- and over-estimation by each estimator are reported in parentheses (•,•). Data are drawn by $x_{it} = \Sigma_{j=1}^{K}\beta_{ij} f_{jt} + \varepsilon_{it}$; $\varepsilon_{it} = \phi(\xi_{1i}h_{1t} + \xi_{2i}h_{2t}) / (\xi_{1i}^2 + \xi_{2i}^2)^{1/2} + (1-\phi^2)^{1/2} v_{it}$, where K = 10, $\phi = 0.2$, and f_{jt} (j = 1, ... , K), $h_{j't}$, $\xi_{j'i}$ ($j' = 1, 2$) and v_{it} are all randomly drawn from $N(0,1)$. For the beta matrix B, we draw an $N \times r$ random matrix B_g such that its first column equals the vector of ones and the entries in the other columns are drawn from $N(0,1)$. We also draw a random $K \times K$ positive definite matrix, compute the first r orthonormalized eigenvectors of the matrix, and set a $K \times r$ matrix C using the eigenvectors. $\mathrm{B} = \mathrm{B}_g \Lambda^{1/2} C'$, where Λ = $diag(\lambda_1,...,\lambda_r)$. The rank of the beta matrix equals one or three (r = 1, 3). BIC1 refers to the Bayesian Information Criterion (BIC) estimators of Cragg and Donald (1997), which is computed under the assumption that the ε_{it} are *i.i.d.* over time (but cross-sectionally correlated). MBIC refers to the BIC estimator computed under the assumption that the ε_{it} are *i.i.d.* over both time and individual response variables. The criterion function of the MBIC estimator uses $T^{0.2}$ instead of ln(T).

			Panel I: K = 10, r = 1, $\lambda_1 = \lambda$			Panel II: K = 10, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=60	N=25	BIC1	59.3% (11.2, 29.5)	55.8% (6.8, 37.4)	52.0% (1.7, 46.3)	12.5% (86.6, 0.9)	27.9% (70.5, 1.6)	53.6% (39.3, 7.1)
		MBIC	31.1% (68.4, 0.5)	63.3% (35.9, 0.8)	92.6% (6.1, 1.3)	1.7% (98.3, 0.0)	9.8% (90.2, 0.0)	47.1% (52.8, 0.1)
	N=30	BIC1	23.9% (0.8, 75.3)	19.1% (0.3, 80.6)	13.7% (0.0, 86.3)	48.6% (42.8, 8.6)	57.0% (26.1, 16.9)	57.5% (8.9, 33.6)
		MBIC	32.4% (67.2, 0.4)	66.9% (32.5, 0.6)	95.3% (3.6, 1.1)	0.2% (99.8, 0.0)	10.3% (89.7, 0.0)	55.7% (44.2, 0.1)
	N=36	BIC1	0.6% (0.0, 99.4)	0.2% (0.0, 99.8)	0.2% (0.0, 99.8)	33.9% (3.1, 63.0)	22.2% (1.3, 76.5)	12.6% (0.1, 87.3)
		MBIC	34.5% (65.3, 0.2)	69.6% (30.0, 0.4)	96.8% (2.6, 0.6)	2.1% (97.9, 0.0)	14.0% (86.0, 0.0)	65.3% (34.7, 0.0)
	N=40	BIC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	5.4% (0.1, 94.5)	2.6% (0.0, 97.4)	1.0% (0.0, 99.0)
		MBIC	33.1% (66.7, 0.2)	72.3% (27.4, 0.3)	98.3% (1.4, 0.3)	1.0% (99.0, 0.0)	14.2% (85.8, 0.0)	68.9% (31.0, 0.1)
	N=50	BIC1	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)
		MBIC	33.0% (67.0, 0.0)	76.0% (24.0, 0.0)	98.9% (1.1, 0.0)	0.9% (99.1, 0.0)	13.8% (86.2, 0.0)	75.4% (24.6, 0.0)
	N=100	BIC1	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBIC	35.3% (64.7, 0.0)	85.2% (14.8, 0.0)	100% (0.0, 0.0)	0.5% (99.5, 0.0)	25.0% (75.0, 0.0)	90.7% (9.3, 0.0)
	N=200	BIC1	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBIC	34.0% (66.0, 0.0)	89.6% (10.4, 0.0)	100% (0.0, 0.0)	0.4% (99.6, 0.0)	35.5% (64.5, 0.0)	97.5% (2.5, 0.0)

Table 3 continued...

			Panel I: K = 10, r = 1, $\lambda_1 = \lambda$			Panel II: K = 10, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=120	N=25	BIC1	17.8% (82.2, 0.0)	51.1% (48.9, 0.0)	94.4% (5.6, 0.0)	0.0% (100, 0.0)	3.2% (96.8, 0.0)	42.3% (57.7, 0.0)
		MBIC	65.7% (34.2, 0.0)	94.0% (5.9, 0.1)	99.9% (0.0, 0.1)	9.6% (90.4, 0.0)	48.5% (51.5, 0.0)	91.9% (8.1, 0.0)
	N=50	BIC1	82.3% (15.3, 2.4)	96.3% (0.8, 2.9)	96.6% (0.0, 3.4)	23.8% (76.2, 0.0)	69.6% (30.3, 0.1)	98.1% (1.4, 0.5)
		MBIC	81.6% (18.4, 0.0)	99.6% (0.4, 0.0)	100% (0.0, 0.0)	24.1% (75.9, 0.0)	83.7% (16.3, 0.0)	99.6% (0.4, .0.0)
	N=60	BIC1	76.2% (2.1, 21.7)	74.9% (0.2, 24.9)	73.6% (0.0, 26.4)	62.0% (36.8, 1.2)	90.1% (6.3, 3.6)	93.4% (0.0, 6.6)
		MBIC	86.0% (14.0, 0.0)	99.6% (0.4, 0.0)	100% (0.0, 0.0)	29.3% (70.7, 0.0)	87.4% (12.6, 0.0)	100% (0.0, .0.0)
	N=75	BIC1	7.3% (0.0, 92.7)	5.8% (0.0, 94.2)	5.0% (0.0, 95.0)	42.4% (1.1, 56.5)	33.7% (0.2, 66.1)	28.1% (0.0, 71.9)
		MBIC	88.8% (11.2, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	37.1% (62.9, 0.0)	93.4% (6.6, 0.0)	100% (0.0, .0.0)
	N=80	BIC1	0.4% (0.0, 99.6)	0.3% (0.0, 99.7)	0.4% (0.0, 99.6)	11.8% (0.0, 88.1)	7.6% (0.0, 92.4)	5.7% (0.0, 94.3)
		MBIC	91.7% (8.3, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	36.9% (63.1, 0.0)	94.7% (5.3, 0.0)	100% (0.0, .0.0)
	N=100	BIC1	0.0% (0.0, 100)	0.0% (0.0, 100)	100% (0.0, 0.0)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)
		MBIC	92.8% (7.2, 0.0)	100% (0.0, 0.0)	99.9% (0.1, 0.0)	44.3% (55.7, 0.0)	97.6% (2.4, 0.0)	100% (0.0, 0.0)
	N=200	BIC1	0.1% (99.9, 0.0)	1.4% (98.6, 0.0)	38.8% (61.2, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.7% (99.3, 0.0)
		MBIC	97.5% (2.5, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	64.5% (35.5, 0.0)	99.7% (0.3, 0.0)	100% (0.0, 0.0)

Table 3 continued...

			Panel I: K = 10, r = 1, $\lambda_1 = \lambda$			Panel II: K = 10, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .010 R^2=.010	λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .010 R^2=.029	λ = .020 R^2=.057	λ = .030 R^2=.083
T=240	N=25	BIC1	0.8% (99.2, 0.0)	37.7% (62.3, 0.0)	92.1% (7.9, 0.0)	0.0% (100, 0.0)	2.5% (97.5, 0.0)	30.6% (69.4, 0.0)
		MBIC	40.7% (59.3, 0.0)	99.5% (0.5, 0.0)	100% (0.0, 0.0)	2.9% (97.1, 0.0)	63.3% (36.7, 0.0)	93.3% (6.7, 0.0)
	N=50	BIC1	4.4% (95.6, 0.0)	86.1% (13.9, 0.0)	100% (0.0, 0.0)	0.0% (100, 0.0)	25.9% (74.1, 0.0)	86.4% (13.6, 0.0)
		MBIC	58.4% (41.6, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	5.0% (95.0, 0.0)	93.1% (6.9, .0.0)	99.7% (0.3, 0.0)
	N=100	BIC1	69.2% (30.8, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	9.9% (90.1, 0.0)	97.9% (2.1, 0.0)	100% (0.0, 0.0)
		MBIC	74.8% (25.2, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	13.0% (87.0, 0.0)	99.8% (0.2, .0.0)	100% (0.0, 0.0)
	N=120	BIC1	94.3% (5.7, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	42.8% (52.2, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)
		MBIC	77.5% (22.5, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	14.3% (85.7, 0.0)	100% (0.0, .0.0)	100% (0.0, 0.0)
	N=145	BIC1	89.5% (0.0, 10.5)	87.9% (0.0, 12.1)	87.5% (0.0, 12.5)	94.9% (3.9, 1.2)	97.7% (0.0, 2.3)	97.5% (0.0, 2.5)
		MBIC	77.7% (22.3, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	15.7% (84.3, 0.0)	100% (0.0, .0.0)	100% (0.0, 0.0)
	N=160	BIC1	30.5% (0.1, 69.4)	28.5% (0.0, 71.5)	27.4% (0.0, 72.6)	75.2% (0.1, 24.7)	69.6% (0.0, 30.4)	66.5% (0.0, 33.5)
		MBIC	81.5% (18.5, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	19.9% (80.1, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)
	N=200	BIC1	0.0% (0.0, 100)	0.0% (0.0, 0.0)	0.0% (0.0, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (0.0, 0.0)
		MBIC	87.4% (12.6, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	25.3% (74.7, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)

Table 4: BIC Rank Estimation of Demeaned Beta Matrix

Reported are the percentages (%) of correct estimation by the BICD1 and MBICD estimators from 1,000 simulated data sets. The percentages (%) of under- and over-estimation by each estimator are reported in parentheses (•,•). Data are drawn by $x_{it} = \Sigma_{j=1}^{K} \beta_{ij} f_{jt} + \varepsilon_{it}$; $\varepsilon_{it} = \phi(\xi_{1i} h_{1t} + \xi_{2i} h_{2t}) / (\xi_{1i}^2 + \xi_{2i}^2)^{1/2} + (1-\phi^2)^{1/2} v_{it}$, where $\phi = 0.2$, and f_{jt} (j = 1, ... , K), $h_{j't}$, $\xi_{j'i}$ ($j' = 1, 2$) and v_{it} are all randomly drawn from $N(0,1)$. For the beta matrix B, we draw an $N \times r$ random matrix B_g such that its first column equals the vector of ones and the entries in the other columns are drawn from $N(0,1)$. We also draw a random $K \times K$ positive definite matrix, compute the first r orthonormalized eigenvectors of the matrix, and set a $K \times r$ matrix C using the eigenvectors. $\mathrm{B} = \mathrm{B}_g \Lambda^{1/2} C'$, where $\Lambda = diag(\lambda_1, ..., \lambda_r)$. The rank of the beta matrix equals three (r = 3). BICD1 refers to the Bayesian Information Criterion (BIC) estimators of Cragg and Donald (1997) for the demeaned beta matrix, which is computed under the assumption that the ε_{it} are *i.i.d.* over time. MBICD refers to the BICD estimator computed under the assumption that the ε_{it} are *i.i.d.* over both time and individual response variables. The criterion function of the MBICD estimator uses $T^{0.2}$ instead of ln(T).

			Panel I: K = 5, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$			Panel II: K = 10, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=60	N=25	BICD1	13.0% (86.5, 0.5)	30.3% (69.0, 0.7)	64.3% (34.0, 1.7)	34.0% (62.5, 3.5)	49.1% (43.7, 7.2)	63.5% (20.3, 16.2)
		MBICD	8.5% (91.5, 0.0)	29.1% (70.7, 0.2)	72.9% (26.7, 0.4)	6.9% (93.1, 0.0)	24.7% (75.1, 0.0)	65.7% (33.6, 0.7)
	N=30	BICD1	35.2% (63.4, 1.4)	55.1% (41.1, 3.8)	77.9% (14.1, 8.0)	53.4% (17.9, 28.7)	50.1% (9.8, 40.1)	42.6% (3.1, 54.3)
		MBICD	10.0% (90.0, 0.0)	35.3% (64.7, 0.0)	81.2% (18.8, 0.0)	7.3% (92.7, 0.0)	28.4% (71.6, 0.0)	73.5% (26.4, 0.1)
	N=36	BICD1	61.8% (24.5, 13.7)	66.7% (11.9, 21.4)	68.9% (1.6, 29.5)	13.1% (0.5, 86.4)	9.1% (0.2, 90.7)	5.3% (0.0, 94.7)
		MBICD	8.0% (92.0, 0.0)	37.4% (62.5, 0.1)	84.6% (15.3, 0.1)	8.0% (91.9, 0.1)	32.6% (67.3, 0.1)	78.8% (21.0, 0.2)
	N=40	BICD1	60.2% (25.7, 14.1)	49.4% (2.0, 48.6)	39.9% (0.1, 60.0)	1.0% (0.0, 99.0)	0.6% (0.0, 99.4)	0.3% (0.0, 99.7)
		MBICD	7.9% (92.1, 0.0)	39.5% (60.5, 0.0)	87.3% (12.7, 0.0)	7.1% (92.8, 0.1)	33.5% (66.4, 0.1)	83.1% (16.6, 0.0)
	N=50	BICD1	0.6% (0.0, 99.4)	0.7% (0.0, 99.3)	0.3% (0.0, 99.7)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)
		MBICD	9.7% (90.3, 0.0)	45.2% (54.8, 0.0)	91.6% (8.4, 0.0)	5.9% (94.1, 0.0)	35.2% (64.8, 0.0)	87.7% (12.2, 0.1)
	N=100	BICD1	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBICD	8.6% (91.4, 0.0)	54.3% (45.7, 0.0)	97.2% (2.8, 0.0)	7.3% (92.7, 0.0)	52.7% (47.3, 0.0)	96.5% (3.5, 0.0)
	N=200	BICD1	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBICD	6.5% (93.5, 0.0)	62.8% (37.2, 0.0)	99.4% (0.6, 0.0)	6.4% (93.6, 0.0)	62.9% (37.1, 0.0)	99.0% (1.0, 0.0)

Table 4 continued...

			Panel I: K = 5, r = 3,$\lambda_1 = \lambda_2 = \lambda_3 = \lambda$			Panel II: K = 10, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .020 R^2=.020	λ = .030 R^2=.029	λ = .050 R^2=.048	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .050 R^2=.130
T=120	N=25	BICD1	3.4% (96.6, 0.0)	25.8% (72.2, 0.0)	73.9% (26.1, 0.0)	2.4% (97.6, 0.0)	14.2% (85.8, 0.0)	61.9% (38.1, 0.0)
		MBICD	39.1% (60.9, 0.0)	76.5% (23.5, 0.0)	96.9% (3.1, 0.0)	26.3% (73.7, 0.0)	66.0% (34.0, 0.0)	95.7% (53.8, 0.0)
	N=50	BICD1	41.2% (58.8, 0.0)	83.1% (16.9, 0.0)	99.2% (0.8, 0.0)	49.7% (50.0, 0.3)	83.8% (15.7, 0.5)	98.7% (0.5, 0.8)
		MBICD	56.0% (44.0, 0.0)	93.0% (7.0, 0.0)	99.9% (0.1, 0.0)	45.9% (54.1, 0.0)	91.1% (8.9, 0.0)	99.8% (0.2, 0.0)
	N=60	BICD1	67.9% (31.6, 0.5)	95.2% (4.2, 0.6)	99.3% (0.0, 0.7)	76.5% (18.6, 4.9)	88.9% (3.1, 8.0)	89.7% (0.0, 10.3)
		MBICD	58.3% (41.7, 0.0)	96.3% (3.7, 0.0)	100% (0.0, 0.0)	52.0% (48.0, 0.0)	93.9% (6.1, 0.0)	100% (0.0, 0.0)
	N=75	BICD1	88.3% (2.9, 8.8)	88.7% (0.1, 11.2)	88.3% (0.0, 11.7)	25.0% (0.3, 74.7)	20.7% (0.1, 79.2)	18.2% (0.0, 81.8)
		MBICD	65.8% (34.2, 0.0)	97.2% (2.8, 0.0)	100% (0.0, 0.0)	61.0% (39.0, 0.0)	97.5% (2.5, 0.0)	100% (0.0, 0.0)
	N=80	BICD1	76.6% (0.4, 23.0)	73.9% (0.0, 26.1)	71.7% (0.0, 28.3)	3.9% (0.0, 96.1)	3.0% (0.0, 97.0)	2.2% (0.0, 97.8)
		MBICD	67.4% (32.6, 0.0)	98.7% (1.3, 0.0)	100% (0.0, 0.0)	59.7% (40.3, 0.0)	97.7% (2.3, 0.0)	100% (0.0, 0.0)
	N=100	BICD1	0.4% (0.0, 99.6)	0.4% (0.0, 99.6)	0.3% (0.0, 99.7)	0.0% (0.0, 100)	0.0% (0.0, 100)	0.0% (0.0, 100)
		MBICD	71.5% (28.5, 0.0)	99.1% (0.9, 0.0)	100% (0.0, 0.0)	68.8% (31.2, 0.0)	99.3% (0.7, 0.0)	100% (0.0, 0.0)
	N=200	BICD1	0.0% (100, 0.0)	0.2% (99.8, 0.0)	33.7% (66.3, 0.0)	0.0% (100, 0.0)	0.0% (100, 0.0)	10.2% (89.8, 0.0)
		MBICD	86.3% (13.7, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	84.2% (15.8, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)

Table 4 continued...

			Panel I: K = 5, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$			Panel II: K = 10, r = 3, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$		
			λ = .010 R^2=.029	λ = .020 R^2=.057	λ = .030 R^2=.083	λ = .010 R^2=.029	λ = .020 R^2=.057	λ = .030 R^2=.083
T=240	N=25	BICD1	0.1% (99.9, 0.0)	20.5% (79.5, 0.0)	67.2% (32.8, 0.0)	0.0% (100, 0.0)	8.6% (91.4, 0.0)	48.4% (51.6, 0.0)
		MBICD	18.9% (81.1, 0.0)	85.0% (15.0, 0.0)	97.0% (3.0, 0.0)	11.1% (88.9, 0.0)	74.8% (25.2, 0.0)	96.0% (4.0, 0.0)
	N=50	BICD1	0.8% (99.2, 0.0)	57.1% (42.9, 0.0)	94.3% (5.7, 0.0)	0.3% (99.7, 0.0)	46.0% (54.0, 0.0)	91.9% (8.1, 0.0)
		MBICD	27.8% (72.2, 0.0)	97.7% (2.3, 0.0)	99.9% (0.1, 0.0)	19.5% (80.5, 0.0)	95.8% (4.2, .0.0)	99.9% (0.1, 0.0)
	N=100	BICD1	26.1% (73.9, 0.0)	98.1% (1.9, 0.0)	100% (0.0, 0.0)	30.0% (70.0, 0.0)	99.0% (1.0, 0.0)	100% (0.0, 0.0)
		MBICD	38.3% (61.7, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	34.4% (65.6, 0.0)	99.9% (0.1, .0.0)	100% (0.0, 0.0)
	N=120	BICD1	64.4% (35.6, 0.0)	99.9% (0.1, 0.0)	100% (0.0, 0.0)	72.5% (27.5, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)
		MBICD	42.8% (57.2, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	36.0% (64.0, 0.0)	100% (0.0, .0.0)	100% (0.0, 0.0)
	N=145	BICD1	96.8% (3.2, 0.0)	99.8% (0.0, 0.2)	99.8% (0.0, 0.2)	95.6% (1.1, 2.4)	96.4% (0.0, 3.6)	96.0% (0.0, 4.0)
		MBICD	46.7% (53.3, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	43.1% (56.9, 0.0)	100% (0.0, .0.0)	100% (0.0, 0.0)
	N=160	BICD1	98.2% (0.3, 1.5)	98.1% (0.0, 1.9)	98.1% (0.0, 1.9)	62.6% (0.0, 37.4)	56.8% (0.0, 43.2)	54.8% (0.0, 45.2)
		MBICD	45.6% (54.4, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	44.8% (55.2, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)
	N=200	BICD1	1.7% (0.0, 98.3)	1.4% (0.0, 98.6)	1.3% (0.0, 98.7)	0.0% (0.0, 100)	0.0% (100, 0.0)	0.0% (100, 0.0)
		MBICD	52.8% (47.2, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)	50.1% (49.9, 0.0)	100% (0.0, 0.0)	100% (0.0, 0.0)

Table 5: Results from Estimation with Five Different Sets of Monthly Portfolio Returns

Reported are the MBIC estimates of the ranks of beta matrices from five different sets of U.S. stock portfolio returns. The MBICD estimates of the ranks of demeaned beta matrices are in parentheses. The individual rows of the table report MBIC and MBICD estimation results and the adjusted *R*-squares from portfolio-by-portfolio time series regressions obtained using different sets of empirical factors (with the numbers of empirical factors used in parentheses). FF, CRR, MOM, REV, and JW, respectively, refer to the three Fama-French factors (FF: VW, SMB, and HML), the five Chen-Roll-Ross macroeconomic factors (CRR: MP, UI, DEI, UTS, and UPR), the momentum factor (MOM), the short-term and long-term reversal factors (REV), and the three Jagannathan and Wang factors (JW: VW, UPR, and LAB). We do not report the estimation results using the LIQ factors. The data on the LIQ factors are only available from December 1969 to December 2008.

Panel A: Estimation with the sample period from January 1952 to December 2011 (T = 720)*										
Empirical Factors (K)	25 Size and B/M Port.		30 Industrial Port.		25 Size and B/M + 30 Industrial Port.		25 Size and Momentum Port.		100 Size and B/M Port.**	
	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)
FF (3)	91.0%	3 (3)	61.5%	3 (3)	73.8%	3 (3)	83.4%	3 (3)	77.6%	3 (3)
MOM and REV (3)	9.8%	2 (1)	7.1%	2 (1)	8.2%	2 (1)	19.4%	2 (1)	8.9%	2 (1)
CRR (5)	1.3%	1 (0)	1.0%	1 (0)	1.1%	1 (0)	1.9%	1 (0)	1.2%	1 (0)
JW (3)	73.8%	1 (1)	58.5%	2 (1)	64.8%	1 (1)	72.5%	2 (2)	64.0%	1 (1)
FF and MOM (4)	91.1%	3 (3)	62.1%	4 (4)	74.1%	4 (4)	91.3%	4 (3)	77.8%	4 (3)
FF and REV (5)	91.1%	3 (3)	61.8%	3 (3)	73.9%	3 (3)	91.3%	4 (3)	77.7%	3 (3)
FF, MOM and REV (6)	91.1%	3 (3)	62.2%	4 (4)	74.2%	4 (4)	91.4%	4 (4)	77.9%	4 (3)
FF and CRR (8)	91.1%	4 (4)	61.8%	4 (4)	73.9%	4 (4)	83.7%	4 (3)	77.7%	3 (3)
FF, JW (5)	91.6%	3 (3)	62.2%	4 (3)	74.3%	4 (4)	83.7%	3 (3)	79.5%	3 (3)
FF, CRR and JW (9)	91.6%	4 (4)	62.3%	4 (4)	74.4%	4 (4)	83.7%	3 (3)	79.6%	3 (3)
FF, MOM, REV, CRR, and JW (12)	91.7%	4 (4)	63.0%	5 (4)	74.9%	5 (5)	91.7%	4 (4)	79.8%	5 (4)

*T = 634 for the estimation with the LAB factor of Jagannathan and Wang.

**Since data are unbalanced, we use the average of the time series observations on individual portfolios (T = 717; T = 632 if the LAB factor is used).

Table 5 continued...

Panel B: Estimation results for the subsample period: January 1952 – December 1981 (T = 360)*										
Empirical Factors (K)	25 Size and B/M Port.		30 Industrial Port.		25 Size and B/M + 30 Industrial Port.		25 Size and Momentum Port.		100 Size and B/M Port.**	
	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)
FF (3)	91.5%	3 (3)	70.1%	3 (3)	79.7%	3 (3)	88.2%	3 (3)	76.9%	3 (3)
MOM and REV (3)	10.7%	2 (1)	8.7%	2 (1)	9.6%	2 (0)	17.1%	2 (1)	9.4%	2 (1)
CRR (5)	8.5%	1 (0)	6.4%	1 (0)	7.3%	1 (0)	9.0%	1 (0)	7.2%	1 (0)
JW (3)	76.3%	1 (1)	66.9%	1 (1)	71.1%	1 (1)	76.0%	1 (1)	66.3%	1 (1)
FF and MOM (4)	91.6%	3 (3)	70.5%	4 (4)	79.9%	4 (4)	93.1%	4 (3)	76.9%	3 (3)
FF and REV (5)	91.6%	4 (4)	70.5%	4 (3)	79.9%	4 (4)	89.1%	3 (3)	70.0%	3 (3)
FF, MOM and REV (6)	91.6%	4 (4)	70.7%	4 (4)	80.1%	4 (4)	93.2%	4 (3)	70.0%	3 (3)
FF and CRR (8)	91.6%	3 (3)	70.3%	3 (3)	79.8%	3 (3)	88.4%	3 (3)	76.9%	3 (3)
FF, Jagannathan and Wang (5)	92.7%	3 (3)	72.1%	3 (3)	81.4%	3 (3)	88.7%	3 (3)	80.8%	3 (3)
FF, CRR and JW (9)	92.7%	4 (4)	72.2%	4 (3)	81.5%	4 (3)	88.9%	3 (3)	80.8%	3 (3)
FF, MOM, REV, CRR, and JW (12)	92.8%	4 (4)	72.8%	4 (3)	81.8%	4 (4)	93.8%	5 (4)	81.0%	3 (3)

*T = 274 for the estimation with the LAB factor of Jagannathan and Wang.
**Since data are unbalanced, we use the average of the time series observations on individual portfolios (T = 358; T = 273 if the LAB factor is used).

Table 5 continued...

Panel C: Estimation results for the subsample period: January 1982 – December 2011 (T = 360)										
Empirical Factors (K)	25 Size and B/M Port.		30 Industrial Port.		25 Size and B/M + 30 Ind. Port.		25 Size and Momentum Port.		100 Size and B/M Port.	
	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)
FF (3)	91.0%	3 (3)	58.6%	3 (3)	71.2%	3 (3)	80.4%	3 (3)	79.1%	3 (3)
MOM and REV (3)	12.4%	2 (1)	8.8%	1 (1)	10.2%	1 (1)	24.7%	2 (2)	11.7%	2 (1)
CRR (5)	0.4%	1 (0)	0.2%	0 (0)	0.3%	0 (0)	0.7%	1 (0)	0.4%	1 (0)
JW (3)	72.3%	1 (1)	54.3%	1 (1)	61.3%	1 (1)	70.4%	1 (1)	62.8%	1 (1)
FF and MOM (4)	91.1%	4 (3)	59.4%	4 (4)	71.7%	4 (4)	90.6%	4 (3)	79.4%	4 (3)
FF and REV (5)	91.1%	4 (4)	58.7%	3 (3)	71.3%	3 (3)	81.2%	4 (3)	79.3%	4 (4)
FF, MOM and REV (6)	91.2%	5 (4)	59.6%	4 (3)	71.9%	4 (4)	90.8%	5 (4)	79.7%	5 (4)
FF and CRR (8)	91.1%	3 (3)	59.0%	4 (3)	71.5%	4 (4)	80.9%	4 (3)	79.2%	3 (3)
FF, JW (5)	91.0%	3 (3)	58.9%	4 (4)	71.4%	4 (4)	80.8%	3 (3)	79.1%	3 (3)
FF, CRR and JW (9)	91.1%	3 (3)	59.0%	4 (3)	71.5%	4 (4)	80.9%	4 (3)	79.2%	3 (3)
FF, MOM, REV, CRR, and JW (12)	91.3%	4 (4)	60.0%	4 (3)	72.2%	5 (4)	90.9%	5 (4)	79.9%	5 (4)

Table 6: Results from Estimation with Monthly Individual Stock Returns

Reported are the MBIC estimates of the ranks of beta matrices using individual stock returns for the period January 1952 - December 2011 (T_{max} = 720) and two different subsample periods (T_{max} = 360). The MBICD estimates of the ranks of demeaned beta matrices are reported in parentheses. The individual rows of the table report the estimation results and the adjusted *R*-squares from portfolio-by-portfolio time series regressions obtained using different sets of empirical factors (with the numbers of empirical factors used in parentheses). FF, CRR, MOM, REV, and JW, respectively, refer to the three Fama-French factors (FF: VW, SMB, and HML), the five Chen-Roll-Ross macroeconomic factors (CRR: MP, UI, DEI, UTS, and UPR), the momentum factor (MOM), the short-term and long-term reversal factors (REV), and the three Jagannathan and Wang factors (JW: VW, UPR, and LAB). We do not report the estimation results using the LIQ factors. The data on the LIQ factors are only available from December 1969 to December 2008.

Empirical Factors (*K*)	Jan. 1952 – Dec. 2011 (N,T_{max}) = (614,720)*		Jan. 1952 – Dec. 1981 (N,T_{max}) = (781,360)*		Jan. 1982 – Dec. 2011 (N,T_{max}) = (2268,360)***	
	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)
FF (3)	24.4%	3 (3)	31.6%	3 (3)	16.0%	2 (2)
MOM and REV (3)	3.9%	1 (1)	5.2%	2 (1)	3.9%	1 (1)
CRR (5)	0.8%	1 (0)	3.2%	1 (0)	0.7%	0 (0)
JW (3)	20.1%	1 (1)	26.4%	1 (1)	11.8%	1 (1)
FF and MOM (4)	24.9%	3 (3)	32.0%	3 (3)	16.6%	2 (2)
FF and REV (5)	24.8%	3 (3)	32.1%	3 (3)	16.4%	2 (2)
FF, MOM and REV (6)	25.2%	4 (3)	32.4%	3 (3)	17.0%	2 (2)
FF and CRR (8)	24.7%	3 (3)	31.9%	3 (3)	16.3%	2 (2)
FF, JW (5)	24.6%	3 (3)	32.5%	3 (2)	16.1%	2 (2)
FF, CRR and JW (9)	24.7%	3 (3)	32.8%	3 (2)	16.3%	2 (2)
FF, MOM, REV, CRR, and JW (12)	25.5%	4 (3)	33.6%	4 (3)	17.3%	3 (3)

*Since data are unbalanced, we use the average of the time series observations on individual stocks (*T* = 599; *T* = 555 if the LAB factor is used) for rank estimation.
**Since data are unbalanced, we use the average of the time series observations on individual stocks (*T* = 333; *T* = 262 if the LAB factor is used) for rank estimation.
***Since data are unbalanced, we use the average of the time series observations on individual stocks (*T* = 313) for rank estimation.

Table 7: Results from Estimation with Six Different Sets of Quarterly Returns

Reported are the MBIC estimates of the ranks of beta matrices from six different sets of U.S. stock portfolio returns and individual stock returns. The MBICD estimates of demeaned beta matrices are in parentheses. The individual rows of the table report the rank estimation results and the adjusted *R*-squares from portfolio-by-portfolio time regressions using different sets of empirical factors (with the numbers of empirical factors used in parentheses). FF, CRR, MOM, REV, and JW, respectively, refer to the three Fama-French factors (FF: VW, SMB, and HML), the five Chen-Roll-Ross macroeconomic factors (CRR: MP, UI, DEI, UTS, and UPR), the momentum factor (MOM), the short-term and long-term reversal factors (REV), and the three Jagannathan and Wang factors (JW: VW, UPR, and LAB). The sample period is from the first quarter of 1952 to the fourth quarter of 2011 (T = 240). We do not report the estimation results using the LIQ factors. The data on the LIQ factors are only available from December 1969 to December 2008.

Empirical Factors (K)	25 Size and B/M Port.		30 Industrial Port.		25 Size and B/M + 30 Industrial Port.		25 Size and Momentum Port.		100 Size and B/M Port.*		Individual stocks (N = 614)**	
	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)
FF (3)	92.6%	3 (3)	65.0%	3 (3)	77.0%	3 (3)	86.5%	3 (3)	81.9%	3 (3)	29.8%	2 (2)
MOM and REV (3)	17.3%	2 (1)	11.2%	1 (1)	13.9%	2 (1)	25.9%	2 (1)	15.9%	2 (1)	7.4%	1 (1)
CRR (5)	1.9%	1 (0)	2.0%	1 (0)	2.0%	1 (0)	3.0%	1 (0)	1.8%	1 (0)	2.2%	1 (0)
JW (3)	75.5%	1 (1)	60.8%	2 (1)	67.2%	1 (1)	74.7%	1 (1)	66.9%	1 (1)	24.0%	1 (1)
FF and MOM (4)	92.6%	3 (3)	65.2%	3 (3)	77.1%	3 (3)	92.5%	4 (3)	81.9%	3 (3)	30.4%	3 (3)
FF and REV (5)	92.7%	4 (4)	65.5%	4 (3)	77.3%	4 (4)	87.0%	4 (3)	82.0%	3 (3)	30.4%	3 (2)
FF, MOM and REV (6)	92.7%	4 (4)	65.7%	4 (3)	77.4%	4 (4)	92.6%	4 (3)	82.1%	3 (3)	31.1%	3 (3)
FF and CRR (8)	92.7%	3 (3)	66.0%	4 (3)	77.6%	4 (4)	86.7%	3 (3)	81.9%	3 (3)	30.6%	3 (2)
FF, JW (5)	92.6%	3 (3)	65.3%	4 (3)	77.2%	4 (3)	86.8%	3 (3)	81.9%	3 (3)	30.2%	3 (2)
FF, CRR and JW (9)	92.7%	3 (3)	66.1%	4 (3)	77.6%	4 (4)	86.8%	3 (3)	81.9%	3 (3)	30.7%	4 (3)
FF, MOM, REV, CRR, and JW (12)	92.8%	4 (4)	66.7%	5 (4)	78.1%	5 (5)	92.7%	4 (3)	82.2%	4 (4)	31.9%	4 (4)

*Since data are unbalanced, we use the average of the time series observations on individual portfolios (T = 239) for rank estimation.
**Since data are unbalanced, we use the average of the time series observations on individual stocks (T = 199) for rank estimation.

Table 8: Results from Estimation with Macro Quarterly Factors and Quarterly Returns

Reported are the MBIC estimates of the ranks of beta matrices from six different sets of U.S. stock portfolio returns and individual stock returns. The MBICD estimates of demeaned beta matrices are in parentheses. The individual rows of the table report the rank estimation results and the adjusted *R*-squares from portfolio-by-portfolio time regressions using different sets of empirical factors (with the numbers of empirical factors used in parentheses). FF, CRR, MOM, REV, and JW, respectively, refer to the three Fama-French factors (FF: VW, SMB, and HML), the five Chen-Roll-Ross macroeconomic factors (CRR: MP, UI, DEI, UTS, and UPR), the momentum factor (MOM), the short-term and long-term reversal factors (REV), and the three Jagannathan and Wang factors (JW: VW, UPR, and LAB). The additional models we consider and their empirical factors are i) the CAPM with VW; ii) the CCAPM with CG; iii) the Lettau-Ludvigson (LL) model with CAY, CG, and CAY×CG; iv) the Yogo (Y) model with VW, DCG, and NDCG; v) the Santos-Veronesi (SV) model with VW and VW×LC; and vi) the Li-Vassalou-Xing (LVX) model with DHH, DCORP, and DHCORP. We do not report the estimation results for the Lustig and Van Nieuwerburgh model because the time series data on the MYMO (housing collateral ratio) factor are available only up to the first quarter of 2005. The sample period is from the first quarter of 1952 to the fourth quarter of 2011 (T_{max} = 240).

Empirical Factors (*K*)	25 Size and B/M Port.		30 Industrial Port.		25 Size and B/M + 30 Industrial Port.		25 Size and Momentum Port.		100 Size and B/M Port.*		Individual Stocks (N = 614)**	
	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)	$\bar{R}^2$	MBIC (MBICD)
CAPM (1)	75.4%	1 (1)	60.5%	1 (1)	67.0%	1 (1)	74.5%	1 (1)	66.8%	1 (1)	23.7%	1 (1)
CCAPM (1)	2.1%	1 (0)	1.3%	1 (0)	1.7%	1 (0)	2.1%	1 (0)	1.9%	1 (0)	0.8%	0 (0)
LL (3)	2.1%	1 (0)	1.7%	1 (0)	1.9%	1 (0)	2.2%	1 (0)	1.8%	1 (0)	1.0%	0 (0)
Y (3)	75.3%	1 (1)	60.7%	1 (1)	67.0%	1 (1)	74.5%	1 (1)	66.7%	1 (1)	23.8%	1 (1)
SV (2)	75.5%	1 (1)	60.7%	1 (1)	67.1%	1 (1)	75.0%	1 (1)	66.8%	1 (1)	24.3%	2 (2)
LVX (3)	3.7%	1 (0)	2.1%	1 (0)	2.8%	1 (0)	3.8%	1 (0)	3.3%	1 (0)	2.3%	1 (0)
FF and CCAPM (4)	92.6%	3 (3)	65.0%	3 (3)	77.0%	3 (3)	86.7%	3 (3)	81.9%	3 (3)	29.9%	3 (2)
FF and LL (6)	92.6%	3 (3)	65.4%	4 (3)	77.2%	4 (3)	86.8%	4 (3)	81.9%	3 (3)	30.1%	3 (3)
FF and Y (5)	92.6%	3 (3)	65.3%	3 (3)	77.1%	3 (3)	86.6%	3 (3)	81.9%	3 (3)	30.0%	3 (2)
FF and SV (4)	92.6%	3 (3)	65.2%	3 (3)	77.1%	3 (3)	86.7%	4 (3)	81.9%	3 (3)	30.3%	3 (3)
FF and LVX (6)	92.6%	3 (3)	65.3%	3 (3)	77.1%	3 (3)	86.8%	4 (3)	81.9%	3 (3)	30.2%	2 (2)
All above together (12)	92.6%	3 (3)	65.8%	4 (3)	77.4%	4 (4)	87.1%	4 (3)	81.9%	3 (3)	31.0%	4 (4)
All above + MOM, REV (15)	92.7%	4 (4)	66.5%	5 (4)	77.9%	5 (4)	92.7%	4 (3)	82.2%	4 (4)	32.2%	5 (4)
All above + CRR, JW (18)	92.6%	3 (3)	66.5%	4 (3)	77.9%	4 (4)	87.2%	4 (3)	82.0%	3 (3)	31.7%	5 (4)
All above + MOM, REV, CRR, JW (21)	92.8%	4 (4)	7.2%	5 (4)	78.3%	5 (5)	92.8%	4 (4)	82.4%	4 (4)	32.8%	5 (5)

*Since data are unbalanced, we use the average of the time series observations on individual portfolios (T = 239) for rank estimation.
**Since data are unbalanced, we use the average of the time series observations on individual stocks (T = 199) for rank estimation.