Dissertation on

Linear Asset Pricing Models

by

Na Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2011 by the
Graduate Supervisory Committee:

Seung Ahn, Chair
Jarl Kallberg
Crocker Liu

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

One necessary condition for the two-pass risk premium estimator to be consistent and asymptotically normal is that the rank of the beta matrix in a proposed linear asset pricing model is full column. I first investigate the asymptotic properties of the risk premium estimators and the related t-test and Wald test statistics when the full rank condition fails. I show that the beta risk of useless factors or multiple proxy factors for a true factor are priced more often than they should be at the nominal size in the asset pricing models omitting some true factors. While under the null hypothesis that the risk premiums of the true factors are equal to zero, the beta risk of the true factors are priced less often than the nominal size. The simulation results are consistent with the theoretical findings. Hence, the factor selection in a proposed factor model should not be made solely based on their estimated risk premiums. In response to this problem, I propose an alternative estimation of the underlying factor structure. Specifically, I propose to use the linear combination of factors weighted by the eigenvectors of the inner product of estimated beta matrix.

I further propose a new method to estimate the rank of the beta matrix in a factor model. For this method, the idiosyncratic components of asset returns are allowed to be correlated both over different cross-sectional units and over different time periods. The estimator I propose is easy to use because it is computed with the eigenvalues of the inner product of an estimated beta matrix. Simulation results show that the proposed method works well even in small samples. The analysis of US individual stock returns suggests that there are six

common risk factors in US individual stock returns among the thirteen factor candidates used. The analysis of portfolio returns reveals that the estimated number of common factors changes depending on how the portfolios are constructed. The number of risk sources found from the analysis of portfolio returns is generally smaller than the number found in individual stock returns.

ACKNOWLEDGMENTS

I am very grateful to Min for his guidance, advice and encouragement during my studies. I am also thankful to my committee members, Jarl and Crocker, for their valuable help of my research. Thanks to my family who has always been there for me.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

TWO-PASS TESTS FOR RISK PREMIUMS IN LINEAR FACTOR MODELS

1.1 Introduction

The two-pass cross-sectional regression method, developed by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973), has been widely used in testing asset pricing models relating risk premiums to betas, in particular, testing whether the beta risk of a proposed factor is priced or not. In the two-pass regression, the betas are first estimated using asset-by-asset time-series regressions, and then the risk premiums are estimated by the cross-sectional regression of the individual means of asset returns on the estimated betas. Whether the beta risk of a proposed factor is priced or not is determined by the significance of the estimated risk premium. The risk premium test statistics used are the t-test and Wald test for the null hypothesis that the risk premiums for some factors are equal to zero. The properties of the test statistics with two-pass cross-sectional regression have been well developed under the assumptions that the asset pricing model is correctly specified. The study of Shanken (1992) reveals large sample properties of the two-pass risk premium test for the correctly specified model with conditionally homoskedastic returns. Jagannathan and Wang (1998) generalize the large sample results of Shanken (1992) to the cases in which returns are conditionally heteroskedastic and/or autocorrelated. However, if the beta matrix in the asset pricing model fails to have full column rank, the two-pass risk premium test statistics of the risk premium are unreliable.

In this paper, we study the asymptotic properties of the t-test and Wald test statistics of the estimated risk premiums when the rank of beta matrix is not full column. There are generally two cases where beta matrix fails to have full column rank. The first is that some proposed factors are useless factors (following the definition in Kan and Zhang (1999b)), useless in the sense that they are not correlated with asset returns. The second is the case in which some proposed factors are multiple proxy factors for a true factor (e.g., two proxy factors for one true factor). In a proposed factor model failing to include all the relevant true factors, we can show analytically that the useless factors and the multiple proxy factors for a true factor are priced more often than they should be at the nominal size (significance level); in the meanwhile, under the null hypothesis that the risk premiums of the true factors are equal to zero, we find that the beta risk of the true factors are priced less often than the nominal size. If the proposed factor model includes all the relevant true factors, the risk premium of the problematic factors (useless factors or multiple proxy factors) will be priced less often than the nominal size. Our Monte Carlo simulation results are consistent with these theoretical findings. Hence, we could not select factors based on the relative significance of their estimated risk premiums. In response to this problem, we propose an alternative estimation of the underlying factor structure in a proposed factor model. Specifically, we propose to use linear combination of factors weighed by the eigenvector of the inner product of the beta matrix.

There is an extensive literature on the properties of asset pricing models for the cases in which models are misspecified. One form of misspecification is that the proposed factors in an asset pricing model are proxy factors for the unobservable true factors. Nawalkha (1997) points out that proxy factors could be used in place of true factors without loss of pricing accuracy. In contrast, Lewellen, Nagel and Shanken (2010) convey a different message by studying the effect of using no more than the correct number of proxy factors, which are correlated with asset returns only through the true factors. They argue that asset pricing tests using cross-sectional R^2 and pricing errors are often highly misleading, in the sense that apparently strong explanatory power (high R^2 and low pricing errors) does not indicate that the asset pricing model is correct. All these results are derived under the assumption that beta matrix has full column rank.

Another form of misspecifications is useless factors, which mean the ones independent of all the asset returns. Kan and Zhang (1999b) investigate the asymptotic properties of the two-pass estimators for a beta pricing model with only one factor, which is a useless factor. They show that the beta risk of the useless factor is more likely to be priced than it should be at the nominal size, and the increasing time series observations exacerbates the problem. Similar issues in context of stochastic discount factor models are studied by Kan and Zhang (1999a). A more related study is presented in Burnside (2010), which focuses the power of the Wald tests of rejecting the stochastic discount factor models when the covariance matrix of asset returns with proposed factors has less than full column rank.

The study in this paper contributes to the literature in the following way. First, we provide a comprehensive analysis of the two-pass t-test and Wald test statistics of the estimated risk premiums when the beta matrix fails to have full column rank. We generalize the asymptotic results of Kan and Zhang (1999b) to models containing multiple proxy factors for a true factor, useless factors, and true factors. We show that in a proposed model omitting some relevant true factors, the risk premiums of the useless factors and the multiple proxy factors for a true factor are always significant with EIV unadjusted standard errors. In the meanwhile, with the existence of either useless factors or multiple proxy factors for a true factor, the risk premiums of true factors are priced less often than the nominal size, when the EIV adjusted standard errors are used.

Second, we emphasize that it is important to check whether the corresponding beta matrix has full column rank. Moreover, we provide a consistent estimation of the underlying true factors in a proposed factor model using the eigenvector of the inner product of beta matrix.

The rest of the paper is presented as follows. Section 2 discusses the properties of the risk premium test statistics in a proposed factor model when the rank condition fails. Section 3 shows the simulation design and results. Section 4 presents the consistent estimation of the underlying factor structure in a proposed factor model. Section 5 concludes.

1.2 Model and Risk Premium Test Statistics

1.2.1 Model Setup and Two-Pass Tests

The basic asset pricing model we consider is a multifactor model in which asset returns are a linear function of k common factors:

$$R_t = \alpha + \beta_1 f_{1t} + \cdots + \beta_k f_{kt} + \varepsilon_t = \alpha + \beta f_t + \varepsilon_t,$$

Where $t = 1,\ldots,T$, $R_t = (R_{1t}, R_{2t},\ldots,R_{Nt})'$, and R_{it} is the gross return on asset i at time t, $f_t = (f_{1t},\ldots,f_{kt})'$ is a vector of k common factors, $\beta = (\beta_1,\ldots,\beta_k)$, $\beta_j = (\beta_{1j},\beta_{2j},\ldots,\beta_{Nj})'$, β_{ij} is the factor loading of asset i corresponding to factor j, $\alpha = (\alpha_1,\alpha_2,\ldots,\alpha_N)'$, α_i is the intercept of asset i, $\varepsilon_t = (\varepsilon_{1t},\varepsilon_{2t},\ldots,\varepsilon_{Nt})'$, and ε_{it} is the idiosyncratic error for asset i at time t.

For analytical convenience, we adopt the same assumptions that are used in Shanken (1992) and Kan and Zhang (1999b) for the two-pass estimators:

i) Factors are independently and identically distributed over time. That is, $f_t \sim N(0,\Sigma_f)$, for all t.

ii) Factors and idiosyncratic errors are not correlated. $E(f_t \otimes \varepsilon_s) = 0_{kN\times 1}$, for all t and s.

iii) Conditional on the factors, the idiosyncratic errors are assumed to be independent and identically distributed over time. That is, $E(\varepsilon_t\varepsilon_s' \mid f_1,\cdots,f_T) = 0_{N\times N}$, for all $t \neq s$, and $Var(\varepsilon_t \mid f_1,\cdots,f_T) = \Sigma_\varepsilon$, for any t, where Σ_ε is the unconditional variance matrix of ε_t .

Under the k-factor beta pricing model, for some scalar γ_0 and $k \times 1$ vector γ, we have

$$E(R_t) = \gamma_0 1_N + \beta\gamma,$$

where $E(R_t)$ is the $N \times 1$ vector of expected returns on the assets, γ_0 is the zero-beta returns, 1_N is a $N \times 1$ vector of ones, $\gamma = (\gamma_1, \ldots, \gamma_k)'$, and γ_j is the risk price corresponding to the risky factor j, $j = 1, \cdots, k$.

Under the assumption that $rank(\beta) = k$, the standard two-pass estimation of the risk premium $\gamma = (\gamma_1, \ldots, \gamma_k)$ is conducted in two steps. In the first step, each row of the beta matrix is estimated by the time-series regression of individual returns on common factors f_t. Let $b = (b_1, \ldots, b_k)$ be the $N \times k$ vector of estimated betas. In the second step, a cross-sectional regression of $R_t = (R_{it}, \ldots, R_{Nt})'$ on $(1_N, b)$ is run for each period t to obtain the time varying estimates of risk premium, defined as $\hat{\gamma}_t$, and the estimated risk premium over T periods is defined as $\hat{\gamma} = (1/T)\Sigma_{t=1}^{T}\hat{\gamma}_t$.

In the cross-sectional regression, we focus on the OLS and the GLS estimation of $\hat{\gamma}$. For each period t, the OLS estimate of $\hat{\gamma}_t$ is given as

$$\hat{\gamma}_t^{OLS} = (b'b)^{-1} b' R_t,$$

and the GLS estimate is given as

$$\hat{\gamma}_t^{GLS} = (b'\hat{\Sigma}_\varepsilon^{-1} b)^{-1} b' \hat{\Sigma}_\varepsilon^{-1} R_t,$$

where $\hat{\Sigma}_\varepsilon$ is a consistent estimation of the covariance matrix of the idiosyncratic errors Σ_ε.

The t-test statistic for the null hypothesis $H_0 : \gamma_j = 0,\ j = 1,\ldots,k$ is given:

$$t(\hat{\gamma}_j) = \frac{\hat{\gamma}_j}{s(\hat{\gamma}_j)/\sqrt{T}}.$$

Using the Frisch-Waugh Theorem (Frisch and Waugh (1993)), we have the mean of the estimated risk premium of factor $j,\ j = 1,\cdots,k$, given as

$$\hat{\gamma}_j^{OLS} = (b_j' M_{-j} b_j)^{-1} b_j' M_{-j} \bar{R},$$

where

$$M_{-j} = I_N - b_{-j}(b_{-j}' b_{-j})^{-1} b_{-j}';$$

$$b_{-j} = (b_1,\cdots,b_{j-1},b_{j+1},\ldots,b_k),$$

and $\bar{R} = (1/T)\Sigma_{t=1}^T R_t$; the OLS standard error of the estimated risk premium $\hat{\gamma}_j$ is given as

$$s^2(\hat{\gamma}_j)^{OLS} = (b_j' M_{-j} b_j)^{-1} b_j' M_{-j} \hat{V} M_{-j} b_j (b_j' M_{-j} b_j)^{-1},$$

where $\hat{V} = 1/(T-1)\Sigma_{t=1}^T (R_t - \bar{R})(R_t - \bar{R})'$ is the estimated covariance matrix of cross-sectional asset returns. Given $R_t - \bar{R} = \beta(f_t - \bar{f})' + (\varepsilon_t - \bar{\varepsilon})$, we have $\hat{V} = \hat{\beta}\hat{\Sigma}_f \hat{\beta}' + \hat{\Sigma}_\varepsilon$, where $\hat{\Sigma}_f$ is a consistent estimation of the covariance matrix of the factors Σ_f.

Using the GLS estimation, we have the mean of the estimated risk premium as

$$\hat{\gamma}_j^{GLS} = (b_j'\hat{\Sigma}_\varepsilon^{-1/2} M_{-j}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} b_j)^{-1} b_j' \hat{\Sigma}_\varepsilon^{-1/2} M_{-j}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} \overline{R},$$

where

$$M_{-j}^{GLS} = I_N - \hat{\Sigma}_\varepsilon^{-1/2} b_{-j} (b_{-j}' \hat{\Sigma}_\varepsilon^{-1} b_{-j})^{-1} b_{-j}' \hat{\Sigma}_\varepsilon^{-1/2}.$$

The GLS standard deviation of estimated risk premium is given as

$$s^2(\hat{\gamma}_j)^{GLS} = (b_j'\hat{\Sigma}_\varepsilon^{-1/2} M_{-j}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} b_j)^{-1} \times b_j'\hat{\Sigma}_\varepsilon^{-1/2} M_{-j}^{GLS} \hat{V} M_{-j}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} b_j (b_j'\hat{\Sigma}_\varepsilon^{-1/2} M_{-j}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} b_j)^{-1}.$$

The Wald test for the joint hypothesis that, for simplicity, $H_0 : \gamma_1 = \gamma_2 = 0$ is as follows:

$$W(\hat{\gamma}_{12}) = \hat{\gamma}_{12}'[Cov(\hat{\gamma}_{12})/T]^{-1} \hat{\gamma}_{12},$$

where $\hat{\gamma}_{12} = (\hat{\gamma}_1, \hat{\gamma}_2)'$. The mean of the OLS estimated risk premium can be calculated as

$$\hat{\gamma}_{12}^{OLS} = (\hat{\gamma}_1^{OLS}, \hat{\gamma}_2^{OLS})' = (b_{12}' M_{-12} b_{12})^{-1} b_{12}' M_{-12} \overline{R},$$

where

$$M_{-12} = I_N - b_{-12}(b_{-12}' b_{-12})^{-1} b_{-12}';$$

$$b_{-12} = (b_3, \ldots, b_k).$$

The OLS covariance matrix of the estimated risk premium is given as

$$Cov(\hat{\gamma}_{12})^{OLS} = (b_{12}' M_{-12} b_{12})^{-1} b_{12}' M_{-12} \hat{V} M_{-12} b_{12} (b_{12}' M_{-12} b_{12})^{-1},$$

where $\hat{V}$ is defined the same as above.

Using the GLS estimation, we have the estimated risk premium as

$$\hat{\gamma}_{12}^{GLS} = (b_{12}'\hat{\Sigma}_\varepsilon^{-1/2} M_{-12}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} b_{12})^{-1} b_{12}' \hat{\Sigma}_\varepsilon^{-1/2} M_{-12}^{GLS} \hat{\Sigma}_\varepsilon^{-1/2} \overline{R},$$

where

$$M_{-12}^{GLS} = I_N - \hat{\Sigma}_{\varepsilon}^{-1/2} b_{-12} (b'_{-12} \hat{\Sigma}_{\varepsilon}^{-1} b_{-12})^{-1} b'_{-12} \hat{\Sigma}_{\varepsilon}^{-1/2}.$$

The GLS estimated covariance matrix is given as:

$$C\text{ov}(\hat{\gamma}_{12})^{GLS} = (b'_{12} \hat{\Sigma}_{\varepsilon}^{-1/2} M_{-12}^{GLS} \hat{\Sigma}_{\varepsilon}^{-1/2} b_{12})^{-1} \times b'_{12} \hat{\Sigma}_{\varepsilon}^{-1/2} M_{-12}^{GLS} \hat{V} M_{-12}^{GLS} \hat{\Sigma}_{\varepsilon}^{-1/2} b_{12} (b'_{12} \hat{\Sigma}_{\varepsilon}^{-1/2} M_{-12}^{GLS} \hat{\Sigma}_{\varepsilon}^{-1/2} b_{12})^{-1},$$

where all the parameters are defined the same as above.

Since betas are estimated with errors in the first step regression, following Shanken (1992), we can adjust the Error-In-Variable (EIV) problem using the correct covariance matrix:

$$Cov(\hat{\gamma})_{EIV} = (1 + \hat{\gamma} \hat{\Sigma}_f^{-1} \hat{\gamma})(Cov(\hat{\gamma}) - \hat{\Sigma}_f) + \hat{\Sigma}_f,$$

where $\hat{\gamma}$ and $Cov(\hat{\gamma})$ can be estimated using OLS and GLS estimation, respectively, and we define the corresponding estimated EIV adjusted covariance matrix as $Cov(\hat{\gamma})_{EIV}^{OLS}$ and $Cov(\hat{\gamma})_{EIV}^{GLS}$. So the EIV adjusted t-test and Wald test statistics are the same as above except substituting the variance/covariance matrix with the EIV adjusted variance/covariance matrix.

1.2.2 Test Statistics when Rank Condition Fails

The validity of the t-test and Wald test statistics of risk premiums could be shown if the rank condition $rank(\beta) = k$ holds. However, if the rank condition fails, the inferences from the t-test and Wald test statistics are unreliable. In this subsection, we will derive the properties of the risk premium test statistics when the rank condition fails. There are generally two cases when the rank of beta matrix is less than full column. The first case is that some proposed factors are useless factors, which are not correlated with asset returns. The other case is some proposed factors are multiple proxy factors for a true factor.

Whether or not the failure of the full rank condition causes serious problems depends on whether the proposed model includes all the relevant true factors. If the proposed model omits some relevant true factors, then the useless factors and multiple proxy factors for a true factor might be priced more often than they should be at the nominal size. We consider three representative cases for a proposed *k*-factor model with $rank(\beta) < k$, from Case 1 to Case 3, and we name these models as under-identified *k*-factor models.

If the proposed factor model includes all the relevant true factors and, in addition, includes useless factors or multiple proxy factors for a true factors, it is less likely to find the problematic factors (useless factors or multiple proxy factors for a true factor) are priced. We name these models as fully-identified *k*-factor models, and we consider an example in Case 4.

Case 1: A proposed *k*-factor model omits some true factors and one of the proposed factor is a useless factor, for example, f_{1t}, where $f_{1t} \sim N(0, \Sigma_{f_1})$, and f_{1t} is correlated with neither asset returns nor other factors. In this case, $rank(\beta) = k - 1$. This is a generalized case of Kan and Zhang (1999b), where they suppose that the model has only one factor, which is a useless factor.

For case 1, we first study the asymptotic properties of the risk premium estimator for the useless factor f_{1t} in Lemma 1.

Lemma 1: Under Case 1, the estimated risk premium of the useless factor f_{1t} has the asymptotic property that $\hat{\gamma}_1 / \sqrt{T}$ is a random variable, with OLS and GLS estimation.

The proof of Lemma 1 is in the appendix. This is the key property that we use study the *t*-test statistics. The asymptotic properties of the *t*-statistic for testing the null hypothesis that the risk premium of the useless factor is equal to zero are given in Proposition 1.

Proposition 1: Under Case 1, when $rank(\beta) = k - 1$ in a proposed under-identified k factor model where one factor is a useless factor, the EIV unadjusted OLS and GLS estimated t-statistics of testing the null hypothesis that the risk premium of the useless factor is equal to zero goes to infinity as $T \to \infty$. Based on the EIV adjusted OLS or GLS standard error, the risk premium of the useless factor is still priced more often than it should be at the nominal size.

Proposition 1 is similar to the result of Kan and Zhang (1999b), but obtained under a more generalized setting, in which we include a useless factor and true factors in the proposed *k*-factor model. For this case, the EIV unadjusted *t*-statistics are not credible, because one will always find the useless factors are priced even when large samples are used. We define it as an over-rejection problem, when the null hypothesis that the risk premium of a factor is equal to zero is rejected more than it should be at the nominal size. With EIV adjusted standard errors, the over rejection problem still exist when *t*-test is performed, but the properties of the OLS t-statistic are different from those using GLS estimation. The difference is shown in the proof of Proposition 1.

Proposition 1 is derived for the cases with only one useless factor. If the proposed factor model contains more than one useless factor, we can not make the strong conclusion that over rejection problems of useless factors with EIV adjusted standard errors always exist. This point is illustrated in Case 2. We also consider the case of multiple proxy factors for a true factor in Case 3. Since the properties of the *t*-tests and Wald tests are similar under these two cases, we derive the results of these two cases together.

Case 2: A proposed *k*-factor model omits some true factors but includes two useless factors, say, f_{1t} and f_{2t}, where $f_{1t} \sim N(0, \Sigma_{f_1})$, $f_{2t} \sim N(0, \Sigma_{f_2})$, and $(f_{1t}, f_{2t})' \sim N(0, \Sigma_{f_{12}})$. Factors f_{1t} and f_{2t} are not correlated with either asset returns or other factors. In this case, $rank(\beta) = k - 2$.

Case 3: A proposed *k*-factor model omits some true factors but includes two proxy factors, f_{1t} and f_{2t}, for a true factor. Consider a general form that

$f_t^* = c_1 f_{1t} + c_2 f_{2t} + u_t$, where $u_t \mid (f_{1t}, \cdots, f_{kt}) \sim N(0, \Sigma_u)$, f_t^* is a true factor but not in the proposed factor model, $(f_{1t}, f_{2t})' \sim N(0, \Sigma_{f_{12}})$, and f_{1t} and f_{2t} are not correlated with either asset returns or other factors. In this case, $rank(\beta) = k - 1$.

For Case 2 and Case 3, we first study the asymptotic properties of estimated risk premiums for the two factors f_{1t} and f_{2t} in the Lemma 2, where f_{1t} and f_{2t} stand for either two useless factors or two proxy factors for a true factor.

Lemma 2: Under Case 2 and Case 3, the estimated risk premiums for factors, f_{1t} and f_{2t}, have the asymptotic property that $\hat{\gamma}_1 / \sqrt{T}$ and $\hat{\gamma}_2 / \sqrt{T}$ are two random variables, with OLS and GLS estimation.

The proof of Lemma 2 is in the appendix. Since we have two factors in the proposed factor model with the estimated risk premiums converging to infinite, the properties of the EIV adjusted t-statistics are different from those in Case 1. The asymptotic properties of the *t*-statistics for testing the null hypothesis that the risk premium of factor f_{1t} or f_{2t} is equal to zero and the Wald test statistics for the joint hypothesis that the risk premiums of factors, f_{1t} are f_{2t}, are both equal to zero are given in Proposition 2.

Proposition 2: Under Case 2 and Case 3, if $rank(\beta) \leq k-1$ in a proposed under-identified k factor model where two factors, f_{1t} are f_{2t}, are two useless factors or two proxy factors for a true factor, the EIV unadjusted OLS and GLS estimated t-statistics and Wald statistics of testing the single and joint null hypothesis that the risk premiums of the factors f_{1t} are f_{2t} are equal to zero goes to infinity as $T \to \infty$. Based on the EIV adjusted OLS and GLS estimated covariance matrix, the risk premiums of the factors f_{1t} are f_{2t} might still be priced more often than they should be at the nominal size.

For Case 2 and Case 3, the EIV unadjusted *t*-statistics and Wald statistics are not credible, because one will always find two useless factors or two proxy factors for a true factor are priced, even when the risk premium of the true factor is equal to zero. But with EIV adjusted variance matrix, we can not make the strong conclusions that the *t*-statistics and Wald statistics will always reject the null hypothesis that the risk premiums of the useless factors or two proxy factors are equal to zero more often than the nominal size, using either OLS or GLS estimations. The results in Proposition 2 are weaker than those in Propostion 1. In Proposition 1, we can show that the EIV adjusted risk premium of one useless factor will always be priced more often than it should at the nominal size.

For more general cases in which $0 < rank(\beta) \leq k-1$, the results from Proposition 2 still hold. The EIV unadjusted t-statistics or Wald test statistics of testing the null hypothesis that the risk premiums of useless factors or multiple proxy factors for a true factor are equal to zero go to infinity as $T \to \infty$. Based on the EIV adjusted estimated covariance matrix, the risk premiums of the useless factors or multiple proxy factors might still be priced more often than they should be at the nominal size.

Now let us consider the properties of t-test statistics of one proposed true factors in the under-identified *k*-factor model with $rank(\beta) < k$. For example, under Case 1, Case 2, or Case 3, suppose f_{kt} is a true factor and $\beta_k \neq 0_{N\times 1}$. Based on the Central Limit Theorem, we have the OLS and GLS estimated $\hat{\gamma}_k$ converges to γ_k. Since the rank of beta matrix is not of full column, there exist either useless factors or multiple proxy factors for a true factor. We have at least one estimated risk premium converging to infinite. The asymptotic properties of the *t*-statistic for testing the null hypothesis that the risk premium of the true factor f_{kt} is equal to zero are given in Proposition 3.

Proposition 3: Under Case 1, Case 2, and Case 3, if $rank(\beta) < k$ in a proposed under-identified k factor model where exist either useless factors or multiple proxy factors for a true factor, under the null hypothesis that the risk premium of a proposed true factor is equal to zero, the EIV adjusted OLS and GLS estimated t-statistics tend to rejected less than its nominal size.

Proposition 3 shows the EIV adjusted *t*-test tends to reject the null hypothesis that the risk premium of a proposed true factor is priced less often than it should be at the nominal size. However, this problem does not happen using EIV unadjusted standard error. Similar analysis could be applied to other true factors. Proposition 3 further demonstrates the importance of the full rank condition. If rank is not full column, we not only tend to accept the problematic factors (useless factors or multiple proxy factors for a true factor), but also reject the true factors.

Case 4: A proposed fully-indentified *k*-factor model contains all the relevant true factors and, in addition, useless factors or multiple proxy factors for a true factor. In this case, $rank(\beta) \le k-1$.

The difference between models containing all the relevant true factors and those omitting some true factors lies in the second step cross-sectional regression of risk premium. When the proposed model contains all the relevant true factors, we can see, in Lemma 3, that the properties of the estimated risk premiums for the useless factors and multiple proxy factors are different from those in Lemma 1 and Lemma 2.

Lemma 3: Under Case 4, the estimated risk premium of the factor f_{1t}, which is either a useless factor or one of the multiple proxy factors for a true factor, has the asymptotic property that $\hat{\gamma}_1$ is a random variable, with OLS and GLS estimation.

The proof of Lemma 3 is in the appendix. When the proposed factor model contains all the relevant true factors, the estimated risk premiums of the useless factors or multiple proxy factors for a true factor do not converge to infinite. This is the main difference between Case 4 and the previous three cases.

Then the asymptotic properties of the *t*-statistic for testing the null hypothesis that the risk premium of the factor f_{1t} is equal to zero are given in Proposition 4.

Proposition 4: Under Case 4, if $rank(\beta) < k$ in a proposed fully-identified k factor model where exists either useless factors or multiple proxy factors for a true factor, the EIV adjusted OLS and GLS estimated square of t-statistics of testing the null hypothesis that the risk premium of a useless factor or one of the multiple proxy factors is equal to zero is stochastically dominated by a χ_1^2-distributed random variable.

Proposition 4 states that with EIV adjusted standard error, we will find that useless factors or multiple proxy factors for a true factor with a zero risk premium are priced less often than the nominal size. This means the problems caused by useless factors or multiple proxy factors for a true factor are less harmful in a fully-identified factors model than in an under-identified model.

In practice, it is very hard to incorporate all the relevant true factors. Hence, it is important to check whether the corresponding beta matrix has full column rank.

1.3 Simulations

The objective of our Monte Carlo experiments is to evaluate the finite sample properties of t-test statistics in the models where the rank of beta matrix is not full column. Since we do not know the data generating process for the actual asset returns, we use the simulated returns with the same mean and variance as those from the actual data. Furthermore, to control the factor structure in proposed factor models, we also generate proposed factors based on the average of the estimated means and variances of actual Fama-French three factors. The real return data in our consideration are the monthly returns of Fama-French 25 portfolios during the period 1970 and 2004. We conduct the two-pass t-tests using 1000 simulations.

Specifically, the base specification is given as follows. We generate the $T \times 4$ matrix of factors $f = (f_1, f_2, f_3, f_4)$, and each factor $f_j = (f_{j1}, \cdots, f_{jT})$, $j = 1, \cdots, 4$, is drawn from $N(u_f, \sigma_f^2)$, where u_f and σ_f^2 are the average of the mean and variance estimated from Fama-French three factors, and we choose $f^* = (f_2 + f_3)/2$. The simulated returns are obtained in the following equation:

$$r = \alpha + f^* \beta^{*\prime} + f_4 \beta_4' + \varepsilon ,$$

where ε is a $T \times N$ matrix with each element drawn from $N(0, \sigma_\varepsilon^2)$, where σ_ε^2 is the variance of the estimated error terms from regressing real returns on the Fama-French three factor model; $\alpha = \gamma_0 1_N + (\gamma^* - \overline{f}^*)\beta^* + (\gamma_4 - \overline{f}_4)\beta_4$, where $\overline{f}^* = 1/T\Sigma_{t=1}^T f_t^*$, $\overline{f}_4 = 1/T\Sigma_{t=1}^T f_{4t}$, $\gamma^* = 0$, and γ_4 is the average of the estimated risk premiums from Fama-French three factor models; also we generate the $N \times 1$ matrix β_4 and β^* from $N(u_\beta, \sigma_\beta^2)$, where σ_β^2 are the average of the variance of estimated beta matrix from regressing real returns on Fama-French three factors. We choose the value of γ_0 and u_β to generate data mimicking the actual returns as much as we can.

The two-pass t-tests are conducted on the different subsamples of proposed factors $f = (f_1, f_2, f_3, f_4)$, where f_1 is useless factor, f_2 and f_3 are two proxy factors for the true factor f^*, and f_4 is a true factor. The significance levels considered are 1%, 5%, and 10%, respectively. If the model is correctly specified, under the null hypothesis, the percentage of rejecting the null hypothesis should be equal to the significance level. The sample sizes contain all the combinations of cross-sectional observation $N = \{10, 25, 100, 200\}$ and the time-series observation $T = \{100, 300, 500, 1000\}$. These combinations allow us to fix one dimension and study the effect of the other dimension.

In Table 1, we report the probability of rejecting the null hypothesis that $\gamma_i = 0$, for $i = 1, 2, 3$, based on the subsample of the proposed factors (f_1, f_2, f_3). This is the case where the model of estimation does not contain all the relevant true factors. Panel A of Table 1 reports the EIV unadjusted OLS estimated t-test statistics of the risk premiums of the three proposed factors. Given that we generate factor f_1 as a useless factor, and factors (f_2, f_3) are two proxy factors for a true factor with risk premium $\gamma^* = 0$, the rejection rate of the null hypothesis that the risk premium is equal to zero should be equal to the significance level. However, we can see that the t-tests over-reject the null hypotheses for the useless factor and multiple proxy factors for a true factor. Now consider the effects of the sample size on the t-test statistics. The larger the number of time series observations, the more likely we will find that the risk premiums of the useless factor and two proxy factors for a true factor are incorrectly significant. Using large number of time series observations increases probability of rejecting the null hypothesis of the problematic factors. Given the number of time series observations T, the larger the cross-sectional observations, the more likely to reject null hypothesis for the problematic factors. Panel B of Table 1 reports the results for the t-tests with EIV adjusted standard errors. Similar to the EIV unadjusted results in Panel A, there are over-rejection problems related to the risk premiums of useless factor and the two proxy factors for a true factor, especially when T is large. In the small samples, especially when N is small, the t-tests with EIV adjusted standard errors are much less likely to reject the incorrect null hypothesis than those without EIV adjusted errors.

Table 1: Test statistics for a useless factor and multiple proxy factors for a true factor in an under-identified factor model

Panel A Test statistics from OLS unadjusted standard errors										
significance		1%			5%			10%		
N	T	r1	r2	r3	r1	r2	r3	r1	r2	r3
10	100	0.003	0.002	0.004	0.027	0.022	0.020	0.084	0.068	0.064
	300	0.023	0.020	0.019	0.181	0.180	0.174	0.352	0.331	0.331
	500	0.106	0.113	0.113	0.391	0.385	0.382	0.549	0.546	0.539
	1000	0.445	0.469	0.472	0.639	0.632	0.639	0.714	0.712	0.711
25	100	0.002	0.001	0.000	0.024	0.007	0.009	0.059	0.038	0.039
	300	0.019	0.011	0.010	0.218	0.207	0.207	0.486	0.449	0.447
	500	0.154	0.136	0.138	0.581	0.554	0.558	0.697	0.684	0.689
	1000	0.651	0.677	0.673	0.766	0.799	0.803	0.817	0.836	0.842
100	100	0.000	0.000	0.000	0.003	0.000	0.000	0.011	0.001	0.003
	300	0.007	0.005	0.004	0.193	0.187	0.192	0.648	0.600	0.606
	500	0.155	0.131	0.133	0.754	0.754	0.755	0.827	0.832	0.837
	1000	0.833	0.833	0.831	0.910	0.900	0.901	0.924	0.917	0.921
200	100	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.001
	300	0.000	0.001	0.002	0.219	0.173	0.176	0.707	0.669	0.670
	500	0.135	0.105	0.103	0.825	0.795	0.797	0.883	0.851	0.849
	1000	0.884	0.843	0.847	0.927	0.902	0.902	0.940	0.924	0.925
Panel B Test statistics from EIV adjusted errors										
significance		1%			5%			10%		
N	T	r1	r2	r3	r1	r2	r3	r1	r2	r3
10	100	0.000	0.000	0.000	0.001	0.000	0.000	0.005	0.007	0.007
	300	0.000	0.000	0.000	0.003	0.000	0.001	0.018	0.017	0.018
	500	0.000	0.000	0.000	0.009	0.005	0.005	0.049	0.050	0.048
	1000	0.000	0.000	0.000	0.015	0.022	0.025	0.084	0.108	0.107
25	100	0.000	0.000	0.000	0.005	0.001	0.001	0.019	0.005	0.008
	300	0.000	0.000	0.000	0.015	0.007	0.006	0.061	0.038	0.042
	500	0.000	0.000	0.000	0.019	0.020	0.023	0.161	0.128	0.137
	1000	0.002	0.004	0.003	0.069	0.081	0.082	0.286	0.300	0.298
100	100	0.000	0.000	0.000	0.001	0.000	0.000	0.005	0.000	0.000
	300	0.000	0.001	0.001	0.008	0.007	0.006	0.047	0.053	0.053
	500	0.001	0.003	0.003	0.029	0.033	0.030	0.230	0.235	0.233
	1000	0.007	0.006	0.006	0.155	0.150	0.151	0.513	0.530	0.527
200	100	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.001
	300	0.000	0.000	0.000	0.001	0.005	0.004	0.077	0.042	0.045
	500	0.002	0.000	0.000	0.044	0.021	0.023	0.258	0.234	0.232
	1000	0.008	0.009	0.009	0.194	0.170	0.170	0.629	0.592	0.597

Note: The results reported in the table are the percentage from 1000 simulations of rejecting the null hypothesis that the risk premium of each factor is equal to zero. If the model is correctly specified, under the null hypothesis, the percentage should be equal to the significance level.

To further investigate the properties of t-test statistics with the useless factor and the two proxy factors for a true factor separately, we conduct two independent simulations with the existing of one kind of the problematic factors. First, we keep the same data generating process as the base specification, defined in the beginning of the simulation, but consider the proposed factor model with only two factors, (f_2, f_3), which are two proxy factors for a true factor. The results are reported in Table 2. Since we omit one true relevant factor f_4 in the estimation, we can see that the two proxy factors for a true factor are priced more often than they should be at the nominal size. Again the EIV adjusted t-test statistics over reject the null hypothesis, and the large sample size T even worsens the over rejection problem. This table tells us again that if the model omits some relevant true factors, the risk premiums of the multiple proxy factors for a true factor will be significant, even when the risk premium of the true factor is zero. This over rejection problem is severe when the sample size T is large.

Second, we modify the data generating process with $\gamma^* = \gamma_4$, and use only (f_1, f_4) as proposed factors. In this case, we have a proposed two factor model containing one useless factor, one true factor, and omitting one true factor f^* with a positive risk premium γ^*. Kan and Zhang (1999a) study a similar problem with stochastic discount factor model, and they find that the estimated risk premium of a true factor is priced less often than that of a useless factor.

Table 2: Test statistics for multiple proxy factors for a true factor in an under-identified factor model

		OLS standard errors					
significance		1%		5%		10%	
N	T	r2	r3	r2	r3	R2	r3
10	100	0.003	0.003	0.016	0.017	0.048	0.045
	300	0.015	0.012	0.218	0.215	0.433	0.448
	500	0.164	0.154	0.576	0.580	0.728	0.724
	1000	0.685	0.684	0.798	0.792	0.842	0.842
25	100	0.000	0.001	0.005	0.006	0.030	0.025
	300	0.011	0.009	0.246	0.238	0.587	0.595
	500	0.175	0.175	0.729	0.739	0.828	0.827
	1000	0.822	0.811	0.889	0.891	0.910	0.917
100	100	0.000	0.000	0.000	0.000	0.002	0.004
	300	0.002	0.001	0.198	0.212	0.754	0.755
	500	0.153	0.150	0.866	0.864	0.914	0.921
	1000	0.910	0.910	0.949	0.949	0.955	0.954
200	100	0.000	0.000	0.000	0.000	0.001	0.001
	300	0.001	0.001	0.193	0.189	0.832	0.833
	500	0.115	0.114	0.897	0.898	0.931	0.928
	1000	0.929	0.927	0.959	0.954	0.967	0.970
		EIV Adjusted errors					
significance		1%		5%		10%	
N	T	r2	r3	r2	r3	R2	r3
10	100	0.000	0.000	0.000	0.000	0.009	0.008
	300	0.000	0.000	0.001	0.001	0.037	0.037
	500	0.000	0.000	0.011	0.013	0.092	0.096
	1000	0.000	0.000	0.067	0.070	0.258	0.251
25	100	0.000	0.000	0.001	0.004	0.008	0.007
	300	0.000	0.000	0.012	0.009	0.052	0.046
	500	0.000	0.000	0.036	0.039	0.237	0.237
	1000	0.006	0.005	0.176	0.181	0.531	0.536
100	100	0.000	0.000	0.000	0.000	0.000	0.001
	300	0.000	0.000	0.011	0.012	0.071	0.076
	500	0.005	0.005	0.080	0.083	0.375	0.375
	1000	0.040	0.039	0.312	0.312	0.711	0.708
200	100	0.000	0.000	0.000	0.000	0.000	0.001
	300	0.000	0.000	0.010	0.011	0.074	0.077
	500	0.001	0.001	0.097	0.100	0.379	0.383
	1000	0.089	0.089	0.359	0.355	0.786	0.790

Note: The results reported in the table are the percentage from 1000 simulations of rejecting the null hypothesis that the risk premium of each factor is equal to zero. If the model is correctly specified, under the null hypothesis, the percentage should be equal to the significance level.

Table 3: Test statistics for a useless factor and a true factor in an under-identified factor model

		OLS standard errors					
significance		1%		5%		10%	
N	T	r1	r4	r1	r4	r1	r4
10	100	0.011	0.000	0.087	0.008	0.196	0.061
	300	0.184	0.078	0.454	0.377	0.590	0.575
	500	0.459	0.347	0.638	0.633	0.712	0.768
	1000	0.704	0.736	0.777	0.869	0.818	0.913
25	100	0.009	0.000	0.060	0.000	0.190	0.010
	300	0.268	0.034	0.653	0.400	0.770	0.710
	500	0.655	0.352	0.798	0.814	0.847	0.924
	1000	0.810	0.894	0.864	0.974	0.887	0.989
100	100	0.000	0.000	0.021	0.000	0.169	0.000
	300	0.388	0.000	0.811	0.351	0.869	0.869
	500	0.827	0.289	0.912	0.967	0.936	0.998
	1000	0.905	0.996	0.928	1.000	0.942	1.000
200	100	0.000	0.000	0.014	0.000	0.178	0.000
	300	0.462	0.001	0.871	0.280	0.912	0.918
	500	0.851	0.239	0.915	0.991	0.932	1.000
	1000	0.939	1.000	0.958	1.000	0.963	1.000
		EIV Adjusted errors					
significance		1%		5%		10%	
N	T	r1	r4	r1	r4	r1	r4
10	100	0.000	0.000	0.003	0.002	0.025	0.042
	300	0.000	0.038	0.039	0.288	0.170	0.502
	500	0.000	0.225	0.083	0.547	0.307	0.701
	1000	0.010	0.574	0.252	0.785	0.484	0.854
25	100	0.000	0.000	0.011	0.000	0.049	0.008
	300	0.002	0.015	0.144	0.357	0.479	0.673
	500	0.019	0.277	0.406	0.777	0.722	0.906
	1000	0.114	0.840	0.668	0.957	0.847	0.982
100	100	0.000	0.000	0.003	0.000	0.041	0.000
	300	0.013	0.000	0.304	0.320	0.797	0.851
	500	0.101	0.253	0.796	0.958	0.931	0.997
	1000	0.417	0.989	0.916	1.000	0.938	1.000
200	100	0.000	0.000	0.001	0.000	0.030	0.000
	300	0.023	0.001	0.416	0.256	0.902	0.908
	500	0.176	0.219	0.847	0.991	0.932	1.000
	1000	0.494	1.000	0.956	1.000	0.962	1.000

Note: The results reported in the table are the percentage from 1000 simulations of rejecting the null hypothesis that the risk premium of each factor is equal to zero. If the model is correctly specified, under the null hypothesis, the percentage should be equal to the significance level.

Table 3 shows the t-test statistics in the two-pass estimation in the beta pricing model process the same properties as those in the stochastic discount factor model Kan and Zhang (1999a). When the model does not include all the relevant true factors, the risk premium of a useless factor is priced more often than it should be at the nominal size. As *T* increases, the over rejection problem becomes even severer. With the EIV adjusted t-tests, the over rejection problem with the useless factor still exists, in the meanwhile, the null hypothesis for the risk premium of the true factor f_4 is priced less often, given that the true risk premium is larger than zero. In the sample with small T, we find that the useless factor f_1 is priced more often than the true factor f_4.

In the last part of simulations, we consider the case that the factor model we use contains all the relevant true factors. We use the data generating process from the base specification, and the proposed factors include all the four factors $f = (f_1, f_2, f_3, f_4)$. The results are reported in Table 4. We can see that in Table 4 there is no over rejection with the useless factor or two proxy factors for a true factor, once all the relevant factors are included. Furthermore, the EIV adjusted t-tests statistics are more likely to be smaller than the size of the test. This is consistent with the results in the proposition 4.

Table 4: Test statistics in a fully-identified factor model with a useless factor and multiple proxy factors for a true factor

Panel A Test statistics from OLS unadjusted standard errors									
Significance		1%				5%			
N	T	r1	r2	r3	r4	r1	r2	r3	r4
10	100	0.010	0.008	0.006	0.000	0.041	0.038	0.035	0.002
	300	0.008	0.006	0.005	0.017	0.053	0.038	0.040	0.293
	500	0.007	0.010	0.011	0.188	0.054	0.044	0.046	0.760
	1000	0.008	0.011	0.010	0.911	0.043	0.047	0.046	0.991
25	100	0.004	0.002	0.003	0.000	0.040	0.018	0.022	0.000
	300	0.008	0.005	0.005	0.001	0.046	0.044	0.041	0.304
	500	0.010	0.005	0.007	0.199	0.051	0.033	0.036	0.972
	1000	0.014	0.004	0.004	1.000	0.049	0.040	0.039	1.000
100	100	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000
	300	0.002	0.002	0.002	0.000	0.016	0.013	0.013	0.251
	500	0.006	0.002	0.002	0.128	0.032	0.013	0.014	1.000
	1000	0.012	0.006	0.005	1.000	0.037	0.041	0.043	1.000
200	100	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000
	300	0.000	0.000	0.000	0.000	0.013	0.002	0.002	0.200
	500	0.001	0.000	0.000	0.090	0.020	0.009	0.010	1.000
	1000	0.008	0.002	0.002	1.000	0.038	0.022	0.020	1.000
Panel B Test statistics from EIV adjusted errors									
significance		1%				5%			
N	T	r1	r2	r3	r4	r1	r2	r3	r4
10	100	0.000	0.000	0.000	0.000	0.003	0.001	0.001	0.001
	300	0.000	0.000	0.000	0.007	0.002	0.000	0.000	0.233
	500	0.000	0.000	0.000	0.134	0.001	0.002	0.002	0.708
	1000	0.000	0.000	0.000	0.866	0.003	0.004	0.004	0.977
25	100	0.000	0.000	0.000	0.000	0.019	0.006	0.011	0.000
	300	0.000	0.001	0.001	0.001	0.018	0.018	0.017	0.297
	500	0.001	0.001	0.001	0.191	0.028	0.013	0.014	0.971
	1000	0.002	0.000	0.000	1.000	0.023	0.012	0.012	1.000
100	100	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000
	300	0.001	0.002	0.002	0.000	0.012	0.013	0.010	0.248
	500	0.003	0.002	0.001	0.127	0.025	0.012	0.011	1.000
	1000	0.008	0.004	0.004	1.000	0.031	0.035	0.032	1.000
200	100	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000
	300	0.000	0.000	0.000	0.000	0.012	0.002	0.002	0.198
	500	0.001	0.000	0.000	0.090	0.018	0.008	0.008	1.000
	1000	0.006	0.002	0.002	1.000	0.034	0.016	0.016	1.000

Note: The results reported in the table are the percentage from 1000 simulations of rejecting the null hypothesis that the risk premium of each factor is equal to zero. If the model is correctly specified, under the null hypothesis, the percentage should be equal to the significance level.

1.4 Consistent Estimation of Factor Structure

From the above analysis and simulations, we can see that with non-full rank betas, the t-tests are not credible. Hence we can not select the factors based on the relative significance of their estimated risk premium. In order to obtain the underling factor structure in a proposed factor model, we need to use eigenvector from the estimated beta matrix to form the linear combinations of the proposed factors.

Now consider a generalized model $R = FB^{0\prime} + E$, where R is a $T \times N$ matrix of asset returns, F is a $T \times k$ matrix of proposed factors, B^0 is a $N \times k$ matrix of true factor loadings, and E is a $T \times N$ matrix of idiosyncratic errors. For any $N \times k$ beta matrix, we can rewrite it as $B^0 = A^0 C^{0\prime}$, where A^0 and C^0 are $N \times r$ and $k \times r$ matrix, respectively, $rank(B^0) = rank(C^0) = r$, and $r \leq k$.

The model could be rewritten as

$$R = FB^{0\prime} + E = (FC^0)A^{0\prime} + E,$$

where $B^0 = A^0 C^{0\prime}$. For any estimated $N \times k$ beta matrix, we can also rewrite it as $\hat{B} = AC'$, where A and C are $N \times r$ and $k \times r$ matrix, respectively. Hence, we have $R = F\hat{B}' + E = (FC)A' + E$. To find the consistent estimation of C, note that

$$vec(\hat{B} - AC') = vec(\hat{B}) - vec(AC') = vec(\hat{B}) - (C \otimes I_N)vec(A).$$

Consider the following minimization problem:

$$\min\,[vec(\hat{B}) - (C \otimes I_N)vec(A)]'\Pi[vec(\hat{B}) - (C \otimes I_N)vec(A)].$$

Suppose $\Pi = I_{Nk}$, then the minimization problem equals $\min \Sigma_{i=1}^{N} \Sigma_{j=1}^{k} (\hat{B}_{ij} - A_i C_j)^2$.

The solution is given by $\tilde{C}\,(k \times r)$, where $\tilde{C}$ is $\sqrt{k}$ times the eigenvectors corresponding to the first r largest eigenvalues of the $k \times k$ matrix $\hat{B}'\hat{B}$. We claim that $\tilde{C}$ is a consistent estimation of a linear transformation of C^0, and hence $F\tilde{C}$ is a consistent estimation of a linear transformation of real FC^0. We summarize the results in Proposition 6, and the proof is shown in the appendix.

Proposition 6: In a generalized model $R = FB^{0\prime} + E = (FC^0)A^{0\prime} + E$, where $rank(B^0) = rank(C^0) = r$, define $\tilde{C}$ is $\sqrt{k}$ times the eigenvectors corresponding to the first r largest eigenvalues of the $k \times k$ estimated matrix $\hat{B}'\hat{B}$. Then $\tilde{C}$ is a consistent estimation of a linear transformation of C^0, and $F\tilde{C}$ is a consistent estimation of a linear transformation of real FC^0.

1.5 Conclusion

In this paper, we study the properties of the t-test and Wald test statistics of risk premiums when the beta matrix in the proposed asset pricing model is not of full column rank. There are generally two cases where the full rank condition fails. The first is that some proposed factors are useless factors, which are not correlated with asset returns. The second is the case in which some proposed factors are multiple proxy factors for a true factor. In a factor model omitting some relevant true factors, with proposed factors in either of the above two cases, we can show analytically that the useless factor is priced more often than it should at the nominal size, and the same problem might happen to the multiple proxy factors for a true factor; in the meanwhile, we find that the risk premiums related to true factors in the under-identified factor models are tend to be priced less often than the nominal size with EIV adjusted standard errors. If the proposed factor model includes all the relevant true factors, the risk premiums of the problematic factors will be priced less often than they should be at the nominal size with the EIV adjusted standard errors. Our Monte Carlo simulation results are consistent with the theoretical findings.

Moreover, if the beta matrix from the proposed model fails to have full column rank, we propose that a consistent estimation of a linear transformation of true factors can be obtained by using the linear combinations of the proposed factors weighted by the eigenvectors of the inner product of estimated beta matrix.

CHAPTER 2

DETERMINING THE RANK OF THE BETA MATRIX IN LINEAR ASSET PRICING MODELS[1]

2.1 Introduction

Jack Treynor (1962), William Sharpe (1964), John Lintner (1965) and Jan Mossin (1966) developed the Capital Asset Pricing Model (CAPM). The model laid out the foundations of modern asset pricing theory. Since the advent of the CAPM, it has become an important question whether a small number of economic or financial variables can capture the sources of non-diversifiable risk. If the answer is affirmative, then the variables should be priced and the information contained in them is crucial for the agents' portfolio strategies.

Determining whether a factor is priced or not became more important with the development of multifactor asset pricing models, like Merton's Intertemporal CAPM (1972) and the Arbitrage Price Theory (APT) of Ross (1976). These multifactor models tell us that if there exist multiple (r) factors determining non-diversifiable sources of risks, then the factors should properly price the risky assets. However, these models do not tell us what the factors are.

In the empirical asset pricing literature many time-series variables have been proposed as possible risk factors (see Chapter 6 of Campbell, Lo and MacKinlay (1997), Chen, Roll, and Ross (1986), and Fama and French (1992)), which we call factor-candidate variables. Several important questions arise with

---

1 This Chapter is written with Seung Ahn and Alex Horenstein.

respect to these factor candidates. Which ones should be included in the pricing equation? Are they capturing different risk sources? By estimating the rank of the beta matrix, we can answer these questions. If we add one factor which does not explain asset returns, we add a column of zero to the corresponding beta matrix, and the rank will not increase. If we add one factor which captures the same risk as the existing factors, we add a column of betas that can be spanned by the existing betas, and the rank will not increase. Hence, by choosing factors that increase the rank of the beta we will find the ones that capture different risk sources.

Estimating the rank of beta matrix is also a necessary condition for the two-pass (TP) risk premium estimation. The two-pass estimation developed by Fama and MacBeth (1973) has been widely used to estimate the risk premium of each factor-candidate variable. Using this method, the betas of candidate variables are first estimated using asset-by-asset time-series regressions, and then the risk premiums related to the variables are estimated by the cross sectional regression of the mean asset returns on the estimated betas. Whether a factor-candidate variable is priced or not is determined by the significance of the estimated risk premium.

An important condition for the consistency of the TP estimator is that the matrix of the true beta values has full columns. However, there are two cases in which the beta matrix may fail to have full columns. The first case is the true betas related to a factor are all zeros. Kan and Zhang (1999b) name such a factor "useless" factor. For a one-factor model in which the factor is useless, Kan and

Zhang (1999b) have investigated the asymptotic properties of the TP estimator. The useless factor cannot be priced; that is, the premium of the useless factor should be undefined. However, Kan and Zhang show that the estimated coefficient of an undefined risk premium is asymptotically significant when using the TP estimator. This happens because the estimated betas are not zeros although the true betas are. The second case is when relevant factors are not the factor-candidate variables themselves, but rather a few linear combinations of them. For such cases, the true beta matrix is not full column, but the estimated matrix may appear to be of full column. Accordingly, some TP premium estimates could falsely appear to be statistically significant, although the corresponding premiums are in fact undefined. Thus, when using the two-pass estimation method researchers need to check the rank of the beta matrix before continuing the second pass cross sectional regression.

This paper proposes a new estimation method, called the Threshold estimation for the rank of the beta matrix in an approximate factor model. We allowed the idiosyncratic error terms for individual observations to be both auto and cross-sectional correlated. Specifically, we estimate the rank using the eigenvalues of the inner product of the estimated beta matrix. The Threshold method produces consistent estimations as the time series dimension T goes to infinity. For the number of cross sectional units (*N*) the only requirement is to be greater than or equal to the number of factor candidates used.

A few papers in the literature have also considered the estimation methods for the rank of a matrix. Zhou (1995) proposes a Wald test in samples with small

N to test the hypothesis of a given rank. Cragg and Donald (1997) provide the tests for the rank of a matrix based on a minimum chi-squared criterion. Robin and Smith (2000) consider the tests based on certain estimated characteristic roots, and show that the limiting distributions of the test statistics are a weighted sum of independent chi-square variables. Kleibergen and Paap (2006) propose a rank statistic using a consistent estimator of the unrestricted matrix, and the proposed rank statistic has a standard χ^2 limiting distribution. However, all these methods are applicable only to data with small *N*. When *N* is large, too many parameters need to be estimated. This is very restrictive for asset prcing applications in which the number of cross-sectional observations, *N*, is usually large.

A method closely related to our method is proposed by Connor and Korajzcyk (1993). Their method is designed to be appropriate for the analysis of the data with large *N* and relatively small *T* observations. Autocorrelation is not allowed for the idiosyncratic components of stock returns. For such data, the number of relevant factors is estimated by evaluating whether adding one more factor results in a significant decrease in the sum of the squares of estimated error terms. To use this sequential method, one needs to determine the order of the factor variables to be tested in an arbitrary matter. In contrast, the Threshold method we propose requires looser restrictions in data. In addition, no ordering of the factors is necessary.

Estimating the rank of the beta matrix is also related to estimating the number of factors. They are related in the sense that the number of the common factors in return data equals to the rank of the beta matrix corresponding to the factors. Bai and Ng (2002), Onatski (2010), and Ahn and Horenstein (2009) have developed formal statistical procedures to estimate the number of the true factors in approximate factor models. Our approach is different from their approaches in one important aspect. Our Threshold method is for the case in which the factor-candidate variables are available, while their methods are designed for the cases in which factor-candidate variables are not observed. Our interest is not to estimate the number of all common factors in asset return data, but to estimate the number of relevant factors contained in observed factor-candidate variables. For this purpose, we estimate the number of relevant factors using the estimated betas corresponding to the candidate variables.

The Threshold estimator we propose possesses several good properties. First, its consistency does not require any particular restriction on the relation between N and T. Its consistency only requires data with large T. Second, the Threshold estimator allows idiosyncratic error terms to have weak time-series and cross-sectional dependence. Third, it has power to detect the weak factors which have only limited explanatory power. Fourth, it can be applied to the zero factor case. Finally, our simulation exercises indicate that the Threshold estimator has good finite sample properties.

Application of the Threshold estimation is conducted first on the US individual stock returns. We confirm that all of the Fama-French (1993) three factors have explanatory power. In contrast, only one or two among the five factors of Chen, Roll, and Ross (1986) have explanatory power. When we combine the three factors of Fama-French (FF) together with the five factors of

Chen, Roll, and Ross (CRR) we find that a factor not captured by FF is captured by CRR. Furthermore, we find that momentum and reversal factors (MOM) capture a source of risk not captured by either FF or CRR. Similarly, the two factors proposed by Chen, Novy-Marx, and Zhang (2010, CNZ) capture an additional source missed by all the other factors. We find evidence for six factors in US individual stock returns among the thirteen factor candidates used. When we use Industrial Portfolio returns, results remain the same. However, when we use portfolios that are better diversified such as the ones sorted on characteristics like Size and Book to Market, the FF factors seem to be enough to capture all the common sources of risk among the thirteen factor candidates, except for the 100 Size and Book to Market portfolios in which an extra factor appears when adding the CNZ factors. Overall, our analysis of portfolio returns reveals that the estimated number of common factors changes depending on how the portfolios are constructed. The rank of the beta matrix found from the analysis of portfolio returns is generally smaller than the one found in individual stock returns, except for the industry portfolios. This result suggests that some industry specific factors disappear when well diversified portfolios are used.

The rank estimation proposed in the paper has two implications for the asset pricing literature. First, it emphasizes the over-identification problem, where all the available factors may be simply throw into the asset pricing models. The rank estimation produces the number of independent sources of commovement that we should include from all the factor candidates when searching for priced risk premiums. The estimator works very well even when some important factors are not included in the set of factor candidates since we allow for a factor structure in the residuals. Another implication is that the rank estimation method is free of the debate whether or not firm characteristics are priced risk factors. Since we use the double demeaned data set, we exclude the effect of firm characteristics. If priced, the risk sources captured by estimating the rank of the beta matrix can only be systematic risk.

The rest of this paper is presented as follows. Section 2 introduces the factor model we investigate and the assumptions imposed on it. Section 3 derives the asymptotic properties of the Threshold estimator. Simulation results are reported in section 4. Section 5 shows the application to the Fama-French three factors, the five factors of Chen, Roll, and Ross (1986), three factors that capture momentum profits and the IA and ROA factors from Chen, Novy-Marx, and Zhang (2010). Concluding remarks follow in section 6. All of the proofs are given in the appendix.

2.2 Model and Assumptions

We begin by defining an approximate factor model as the one considered by Chamberlain and Rothschild (1983) and Bai and Ng (2002). Let x_{it} be the response variable for the i^{th} cross-section unit at time t, where $i = 1, 2, \cdots, N$, and $t = 1, 2, \cdots, T$. Explicitly, x_{it} can be the (excess) return on asset i at time t. The response variables x_{it} depend on the individual effect α_i, the time effect δ_t and the k factor-candidate variables in $f_t = (f_{1t}, f_{2t}, ..., f_{kt})'$. That is,

(1) $x_{it} = \alpha_i + \delta_t + \beta_i' f_t + \varepsilon_{it}$

where $\beta_i = (\beta_{i1}, \beta_{i2}, \cdots, \beta_{ik})'$ is the beta vector for cross section unit *i*. The product $\beta_i' f_t$ is the common component of x_{it}, and the ε_{it} are idiosyncratic components or idiosyncratic risks.[2]

Our interest for model (1) is to estimate rank of the beta matrix B, where $\mathrm{B} = (\beta_1, \beta_2, ..., \beta_N)'$. However, because of the presence of the time effects δ_t, we are unable to estimate β_i. Instead we can estimate the demeaned betas, $\dot{\beta}_i = \beta_i - \overline{\beta}$, where $\overline{\beta} = N^{-1}\Sigma_{i=1}^{N}\beta_i$. Use of the demeaned beta estimated instead of the raw beta estimates does not cause any technical problem. As long as any β_{ij} is varying over different cross-section units, $rank(\mathrm{B}) = rank(\mathrm{B}^d)$, where $\mathrm{B}^d = (\dot{\beta}_1, \dot{\beta}_2, ..., \dot{\beta}_N)'$. In addition, the rank of B^d matters more than the rank of B for the two-pass regression, because the risk premiums corresponding to the factors in f_t are estimated by the cross-section regression of the individual mean of x_{it} ($\overline{x}_i = T^{-1}\Sigma_{t=1}^{T} x_{it}$) on one and β_i. If any beta in β_i is constant over *i*, the risk premiums are undefined. The premiums are identified only if the demeaned beta matrix B^d has full column.

The demeaned betas can be estimated by estimating the following double demeaned model,

---

2 In this model, we consider only the case of time invariant betas. Our method can be easily extended to the case of time-varying betas since the rank estimation is based on the estimated beta matrix.

(2) $\ddot{x}_{it} = \dot{\beta}_i' \dot{f}_t + \ddot{\varepsilon}_{it}$,

where $\ddot{x}_{it} = x_{it} - \overline{x}_t - \overline{x}_i + \overline{x}$, $\dot{f}_t = f_t - \overline{f}$, $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \overline{\varepsilon}_t - \overline{\varepsilon}_i + \overline{\varepsilon}$, $\overline{x}_t = N^{-1}\Sigma_{i=1}^{N} x_{it}$,

$\overline{x}_i = T^{-1}\Sigma_{t=1}^{T} x_{it}$, $\overline{x} = (NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T} x_{it}$, $\overline{f} = (\Sigma_{t=1}^{T} f_t)/T$, and $\overline{\varepsilon}_t$, $\overline{\varepsilon}_i$, and $\overline{\varepsilon}$ are similarly defined. For each time period t, model (2) can be written as

$$\underset{(T\times 1)}{\ddot{x}_{i\cdot}} = \underset{(T\times k)}{\dot{F}} \underset{(k\times 1)}{\dot{\beta}_i} + \underset{(T\times 1)}{\ddot{\varepsilon}_{i\cdot}},$$

where $\ddot{x}_{i\cdot} = (\ddot{x}_{i1}, \ddot{x}_{i2}, \cdots, \ddot{x}_{iT})'$, $\ddot{\varepsilon}_{i\cdot}$ is similarly defined, and $\dot{F} = (\dot{f}_1, \dot{f}_2, \cdots, \dot{f}_T)'$. For all data, we have

$$\underset{(T\times N)}{\ddot{X}} = \underset{(T\times k)}{\dot{F}} \underset{(k\times N)}{\mathrm{B}^{d\prime}} + \underset{(T\times N)}{\ddot{\mathrm{E}}},$$

where $\ddot{X} = (\ddot{x}_{1\cdot}, \ddot{x}_{2\cdot}, \cdots, \ddot{x}_{N\cdot})$, , and $\ddot{\mathrm{E}} = (\ddot{\varepsilon}_{1\cdot}, \ddot{\varepsilon}_{2\cdot}, \cdots, \ddot{\varepsilon}_{N\cdot})$. Then, the demeaned beta matrix B^d can be estimated by the OLS estimator $\hat{\mathrm{B}}^d = \ddot{X}\dot{F}(\dot{F}'\dot{F})^{-1}$.

In what follows, we use $\lambda_j(A)$ to denote the j^{th} largest eigenvalue of a matrix A, and the norm of A is denoted by $\|A\| = [tr(A'A)]^{1/2}$. We define c as a generic positive constant. With this notation, we make the following assumptions:

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

Assumption B (betas): (i) $\|\beta_i\| \le c$ for all $i = 1, 2, \cdots, N$. (ii) $\mathrm{B}^{d\prime}\mathrm{B}^d / N$ is positive semi-definite and $rank(\mathrm{B}^d) = rank(\mathrm{B}^{d\prime}\mathrm{B}^d) = r \le k$ for all $N \ge r$. (iii) If $N \to \infty$, $\mathrm{B}^{d\prime}\mathrm{B}^d / N \to \Sigma_\beta$, where Σ_β is finite.

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

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

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

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

The four assumptions are a subset of the assumptions used in Bai and Ng (2002) and Ahn and Horenstein (2009). Assumption A implies that the factors should be stationary. Assumption B(i) ensures that each factor loading does not explode. Assumption B(ii) allows that the rank of B^d to be smaller than the number of the variables in f_t. Assumption B (iii) implies that for the cases where N is large, $\mathrm{B}^{d\prime}\mathrm{B}^d / N$ is asymptotically finite. That is, the explanatory power of each factor increases at the rate of *N*. The estimators we propose below do not require large *N*. Under Assumption B (iii), the estimators are consistent regardless of the size of *N*. Under Assumption B, we treat the betas as fixed constants. We can easily relax this assumption, but at the cost of more notation.

Assumption C allows weak time-series correlations and does not impose any restrictions on the cross-sectional correlation among the error terms ε_{it}. Our asymptotic results obtained below depend not on the covariance among the errors, but on the dependence between the errors and factors. Assumption C implies that $\Sigma_{t=1}^{T}\varepsilon_{it} / \sqrt{T}$ is a bounded random variable for all *i*. This assumption is weaker than Assumption C of Bai and Ng (2002):

$$\frac{1}{NT}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}\left|E(\varepsilon_{it}\varepsilon_{js})\right| < c .$$

Assumption D implies that the random vectors $\Sigma_{t=1}^{T}\varepsilon_{it} f_t / \sqrt{T}$ are bounded. This assumption is required for the consistency of the ordinary least squares (OLS) estimator of B^{d}. Assumption D is essentially the same assumption as Assumption D of Bai and Ng (2002).

Furthermore, Assumption D allows the errors ε_{it} to have a factor structure. To see why, consider a simple case in which the ε_{it} have an one-factor structure: $\varepsilon_{it} = \xi_i g_t$ where $E(g_t) = 0$, $E(g_t f_t) = 0$, $E(|g_t|^4) < c$, and $T^{-1}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}E(g_s g_t f_t' f_s) < c$ for all *t*, and $|\xi_i| < c$ for all *i*. For this case, the random variable $\Sigma_{t=1}^{T} g_t f_t / \sqrt{T}$ is bounded. Thus, we can easily show that Assumption C holds. In addition,

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

Thus, Assumption D holds. Given that the ε_{it} can have a factor structure, estimating the rank of B^d is not equivalent to estimating the number of all of the common factors in response variables. The rank of B^d is the maximum number of the common components in response variables among the factor candidate variables f_t. Hence, the rank estimation method works well even when the factor candidates do not include all the common underlying factors. The missing information is captured in the error terms with a factor structure.

2.3 Rank Estimation Using Eigenvalues

The Threshold estimator we propose below uses the eigenvalues of $\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^d / N$. So, we begin this section by studying the asymptotic properties of the eigenvalues. Below, we use the notation $\hat{\mu}_{NT,j} = \lambda_j(\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^d / N)$ where j indicates that $\hat{\mu}_{NT,j}$ is the j^{th} largest eigenvalue of the matrix $\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^d / N$. The following theorem presents the asymptotic properties of the eigenvalues.

Theorem 1: Under assumption A – D, (i) $p\lim_{T\to\infty}\hat{\mu}_{NT,j} > 0$ for $0 < j \le r$; and (ii) $\tilde{\mu}_{NT,j} = O_p(T^{-1})$, for $0 \le r < j \le k$.

Theorem 1 shows that the first $r > 0$ largest eigenvalues of $\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^{d} / N$ 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 matrix B^{d}, r. Notice that the asymptotic properties of the eigenvalues do not require $N \to \infty$. Theorem 1 holds for any fixed number *N*. Therefore, the estimator we propose below does not require large *N*.

The following theorem defines the consistent estimator that we call "Threshold" estimator.

Theorem 2 (Threshold Estimator): For a given threshold function $g(T) > 0$ such that $g(T) \to 0$ and $Tg(T) \to \infty$ as $T \to \infty$, define $\hat{r}_{TH} = \#\{1 \leq j \leq k : \hat{\mu}_{NT,j} > g(T)\}$, where $\#\{\cdot\}$ is the cardinality of a set . Then, under Assumptions A – D, $\lim_{T\to\infty} \Pr(\hat{r}_{TH} = r) = 1$.

The result of Theorem 2 is quite intuitive. Observe that $g(T)$ converges to zero at a lower rate than the last $(k-r)$ eigenvalues of $\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^{d} / N$ do. The first r eigenvalues converge to positive numbers. Accordingly, for sufficiently large T, the value of $g(T)$ is most likely to be smaller than the first r eigenvalues and larger than the rest of the eigenvalues. The threshold estimation procedure is similar to the methods suggested by Bai and Ng (2002) to estimate the number of unobservable common factors in an approximate factor model with a large number of response variables.

Note that we can also use the Threshold estimator proposed in Theorem 2 for the cases in which (i) the data is not generated by a factor model and/or (ii) all the factor candidates are useless. We will call this situation "no-factor" case. For such a case, $r = 0$.

A possible pitfall of the threshold estimator is that there are many possible choices for $g(T)$. Whenever a function is an appropriate choice for $g(T)$, so is a finite multiple of the function. If T is large, the estimation results would be insensitive to the choice of $g(T)$. However, for the data with relatively small T, the estimation result could change depending on the choice of $g(T)$. The optimal choice of the threshold function $g(T)$ may depend on the data generating processes. In the following paragraph we propose a specific function for $g(T)$ which provides reliable estimates for many different data generating processes we have considered in our Monte Carlo experiments.

Let $\hat{\sigma}^2 = [(N-1)(T-1)]^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\ddot{e}_{it}^2$, where the $\ddot{e}_{it}$ are the OLS residuals from the regression of the double demeaned model (2). The estimator $\hat{\sigma}^2$ is a consistent estimator of $\mathrm{var}(\varepsilon_{it})$. Also, let $R^2 = 1-[\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\ddot{e}_{it}^2]/[\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\ddot{x}_{it}^2]$ be the R-square from the OLS regression of model (2). Then, the threshold function we suggest to use for the Threshold estimator is given by:

(3) $$g(d,T) = \frac{d\hat{\sigma}^2}{T^d},$$

where $d = 1 - R^2$ for $0.3 \le 1 - R^2 \le 0.8$, $d = 0.3$ for $1 - R^2 < 0.3$, and $d = 0.8$ for $1 - R^2 > 0.8$.

The function $g(d,T)$ is designed to be a non-decreasing function of R^2 for sufficiently large T. Specifically, for $T > 28$, $g(d,T)$ is a monotonically decreasing function of d. Because d is a non-increasing function of R^2, $g(d,T)$ is an increasing (specifically, non-decreasing) function of R^2. The use of $g(d,T)$ is motivated by our findings from Monte Carlo simulations: when the data are generated by weak factors (that have low explanatory power), smaller threshold values are needed to better estimate the rank of B^d. Since $g(d,T)$ should satisfy the two conditions given in Theorem 2, we limit the range of d to be [0.3, 0.8]. The choice of the range is somewhat arbitrary. However, this range is the best choice we have found from simulations.

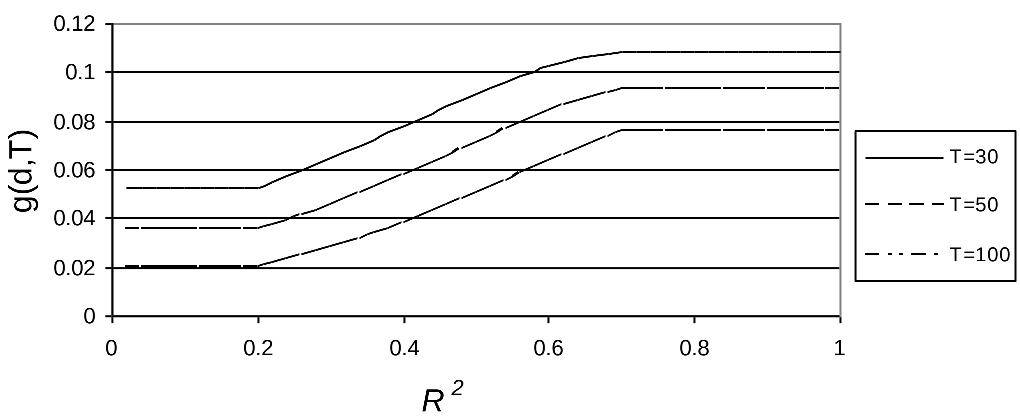

Figure 1. The value of *g(d,T)* with different R_square and T

The property of $g(d,T)$ could be stated from Figure 1. When R^2 is low, factors have low explanatory power, we need a small value of $g(d,T)$ to detect the weak factors. When R^2 is high, factors are stronger and a relative larger threshold function is needed. When T increases, the last $(k-r)$ eigenvalues of $\hat{B}^{d\prime}\hat{B}^{d} / N$ converges to zero faster, and a smaller $g(d,T)$ is needed.

2.4 Simulations

2.4.1 The Basic Simulations

Our simulation data are drawn by the same model used in Bai and Ng (2002) and Ahn and Horenstein (2009):

$$x_{it} = \alpha_i + \delta_t + \Sigma_{j=1}^{k} \beta_{ij} f_{jt} + u_{it}\,;\ u_{it} = v_{it}\sqrt{\frac{1-\rho^2}{1+2J\delta}},$$

where $v_{it} = \rho v_{i,t-1} + \xi_{it} + \Sigma_{h=\max(i-J,1)}^{i-1} \delta v_{ht} + \Sigma_{h=i+1}^{\min(i+J,N)} \delta v_{ht}$, and the ξ_{it} $(1 \le i \le N)$ and the factor candidate variables f_{jt} are randomly drawn from $N(0,1)$. In this setup, the variance of u_{it} is roughly equal to one.

For simplicity, we set $\alpha_i = 0$ for all i, and $\delta_t = 0$ for all t. The beta matrix B is drawn by the following way. We draw a $N \times r$ random matrix A, each entry of which is $N(0,1)$. 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.[3] Then, we set $\mathrm{B} = A\Lambda^{1/2}C'$, where

3 We first generate a $N \times k$ matrix M whose entries are drawn from N(0,1), and then compute the r eigenvectors of $M'M$.

$\Lambda = diag(\lambda_1, ..., \lambda_r)$. This setup is equivalent to the case in which the true factors are $f_t^* = \Lambda^{1/2} C' f_t$ with $Var(f_t^*) = \Lambda$ and the beta matrix corresponding to $f_t^* = (f_{1t}^*, ..., f_{rt}^*)'$ is A.

The parameter Λ controls the signal to noise ratio of each of the true factors (SNR, ratio of the variances of a factor and the idiosyncratic error, u_{it}). When the j^{th} true factor, f_{tj}^*, has the variance of $\lambda_j = 1/r$, its SNR equals $1/r$, where $r \geq 1$. In case of $r = 0$, we present the table separately. For benchmark simulations, we use $\lambda_j = 1/r$, for $1 \leq j \leq r$. In other simulations we try different λ_j's.

For the error terms, we consider four cases: (i) the cases with *i.i.d.* errors ($\rho = J = \delta = 0$), (ii) with both cross-sectional and auto-correlated errors ($\delta = 0.2$, $\rho = 0.5$, $J = 8$), (iii) with only cross-sectional correlated errors ($\rho = 0$), and (iv) with only auto-correlated errors ($J = \delta = 0$). For each case, we try 25 different combinations of *N* and *T*, where *N*, *T* $\in$ {50, 100, 200, 500, 1000}. 1,000 samples are drawn for each combination of *N* and *T*.

Tables 5 – 7 report the results from our benchmark simulations ($\lambda_1 = ... = \lambda_r = 1/r$).

Table 5 shows the estimation results from the cases with *i.i.d.* idiosyncratic errors and both cross-sectional and auto-correlated errors. Specifically, for the correlated error cases, we set $\rho = 0.5$, $\delta = 0.2$, $J = 8$.

Table 5

Results of Threshold Estimation from the Simulated Data with *I.I.D.* Errors and Both Cross- and Auto-Correlated Errors

T	N	$\rho = J = \delta = 0$						$\delta = 0.2, \rho = 0.5, J = 8$					
		$k=3$			$k=5$			$k=3$			$k=5$		
		$r=1$	$r=2$	$r=3$	$r=1$	$r=3$	$r=5$	$r=1$	$r=2$	$r=3$	$r=1$	$r=3$	$r=5$
50	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.04]	3.00 [0.00]	5.00 [0.06]	1.01 [0.10]	2.00 [0.03]	3.00 [0.00]	1.06 [0.24]	3.00 [0.06]	5.00 [0.04]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.05]	2.00 [0.00]	3.00 [0.00]	1.03 [0.17]	3.00 [0.03]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.03]	2.00 [0.00]	3.00 [0.00]	1.00 [0.05]	3.00 [0.03]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
100	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.05]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
200	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
500	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
1000	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]

Note: Data are generated with $\lambda_j = 1/r$ for $1 \le j \le r$. The value reported in each cell is the mean of the rank estimates from 1,000 simulations, and the value in the bracket is the standard deviation of the estimates.

Table 6

Results from the Simulated Data with only Cross-Correlated errors and Only Auto-Correlated Errors

T	N	$\delta=0.2, \rho=0, J=8$						$\delta=0, \rho=0.5$					
		$k=3$			$k=5$			$k=3$			$k=5$		
		$r=1$	$r=2$	$r=3$	$r=1$	$r=3$	$r=5$	$r=1$	$r=2$	$r=3$	$r=1$	$r=3$	$r=5$
50	50	1.00 [0.08]	2.00 [0.03]	3.00 [0.00]	1.08 [0.27]	3.00 [0.08]	5.00 [0.03]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.01 [0.08]	3.00 [0.00]	5.00 [0.06]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.02 [0.13]	3.00 [0.03]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.05]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
100	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
200	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.03]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
500	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
1000	50	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	100	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	200	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	500	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]
	1000	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]

Note: Data are generated with $\lambda_j = 1/r$ for $1 \le j \le r$. The value reported in each cell is the mean of the rank estimates from 1,000 simulations, and the value in the bracket is the standard deviation of the estimates.

All factors have the SNRs of $1/r$, where $r \geq 1$. The Threshold estimator performs very well, even in the case of small sample size (e.g., $T = 50$).
For every case, the mean of the rank estimates is almost equal to the true rank. Also, only for a few cases, the standard deviation of the estimates is larger than zero. The results with correlated errors are not noticeably different from those with *i.i.d.* errors.

Tables 6 shows the results from the cases of cross-sectional correlation only ($\rho = 0$, $\delta = 0.2$, $J = 8$) and auto-correlation only ($\rho = 0.5, \delta = 0.2$). The factors are generated with $\lambda_j = 1/r$, for $1 \leq j \leq r$. For all cases, the Threshold estimator performs very well even if T is small.

Table 7 shows the results for the cases in which all factors are weak with the same SNRs. The left part of the table reports the results from the cases with *i.i.d.* errors, while the right part presents the results from the cases with both cross-sectionally and auto-correlated errors. For small T (T = 60), the Threshold estimator does not perform well when the SNRs of the factors are as low as 0.025. But it works well in the cases with the SNRs larger than 0.05. For the case in which $T = 100$, the Threshold estimator performs very well even in the cases with the SNRs of 0.025. The estimation results from the data simulated with *i.i.d.* errors are more reliable than those from the data with correlated errors, especially when T is small and factors are weak. In fact, we can add one more dimension of the SNR to the threshold function. If the weak factors defined as important factors need SNRs at least larger than 1/5, we can adjust the threshold function with the simulated data to make our estimation capturing al the factors with SNRs larger than 1/5.

Table 7

Results from the Simulated Data with Weak Factors

T	*SNR*	$\delta = \rho = J = 0$						$\delta = 0.2, \rho = 0.5, J = 8$					
		$k=3$			$k=5$			$k=3$			$k=5$		
		$r=1$	$r=2$	$r=3$	$r=1$	$r=3$	$r=5$	$r=1$	$r=2$	$r=3$	$r=1$	$r=3$	$r=5$
60	0.025	1.00 [0.15]	1.93 [0.26]	2..74 [0.44]	1.22 [0.42]	2.94 [0.30]	4.45 [0.55]	1.13 [0.35]	1.96 [0.28]	2.79 [0.42]	1.56 [0.56]	2.99 [0.39]	4.45 [0.57]
	0.05	1.01 [0.12]	2.00 [0.06]	3.00 [0.00]	1.26 [0.44]	3.02 [0.15]	4.99 [0.08]	1.17 [0.38]	2.04 [0.19]	3.00 [0.00]	1.64 [0.55]	3.13 [0.34]	4.99 [0.08]
	0.1	1.02 [0.12]	2.00 [0.04]	3.00 [0.00]	1.28 [0.45]	3.00 [0.04]	5.00 [0.00]	1.18 0.38	2.03 [0.18]	3.00 [0.00]	1.67 [0.55]	3.03 [0.18]	5.00 [0.00]
	0.2	1.01 [0.10]	2.00 [0.00]	3.00 [0.00]	1.19 [0.39]	3.00 [0.00]	5.00 [0.00]	1.14 [0.35]	2.01 [0.08]	3.00 [0.00]	1.50 [0.52]	3.00 [0.05]	5.00 [0.00]
	0.3	1.00 [0.05]	2.00 [0.00]	3.00 [0.00]	1.06 [0.23]	3.00 [0.00]	5.00 [0.00]	1.06 [0.25]	2.00 [0.04]	3.00 [0.00]	1.28 [0.45]	3.00 [0.00]	5.00 [0.00]
	0.5	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.07]	3.00 [0.00]	5.00 [0.00]	1.02 [0.12]	2.00 [0.00]	3.00 [0.00]	1.08 [0.26]	3.00 [0.00]	5.00 [0.00]
100	0.025	1.00 [0.00]	1.99 [0.06]	2.98 [0.13]	1.01 [0.11]	2.99 [0.08]	4.93 [0.24]	1.04 [0.20]	2.01 [0.11]	2.99 [0.08]	1.34 [0.48]	3.03 [0.18]	4.96 [0.20]
	0.05	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.01 [0.11]	3.00 [0.03]	5.00 [0.00]	1.05 [0.22]	2.01 [0.12]	3.00 [0.00]	1.37 [0.49]	3.06 [0.24]	5.00 [0.00]
	0.1	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.01 [0.11]	3.00 [0.00]	5.00 [0.00]	1.05 [0.22]	2.01 [0.11]	3.00 [0.00]	1.38 [0.49]	3.01 [0.11]	5.00 [0.00]
	0.2	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.01 [0.11]	3.00 [0.00]	5.00 [0.00]	1.04 [0.20]	2.00 [0.04]	3.00 [0.00]	1.30 [0.46]	3.00 [0.00]	5.00 [0.00]
	0.3	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.05]	3.00 [0.00]	5.00 [0.00]	1.02 [0.13]	2.00 [0.00]	3.00 [0.00]	1.12 [0.33]	3.00 [0.00]	5.00 [0.00]
	0.5	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.00 [0.00]	3.00 [0.00]	5.00 [0.00]	1.00 [0.00]	2.00 [0.00]	3.00 [0.00]	1.01 [0.10]	3.00 [0.00]	5.00 [0.00]

Note: All simulated data are drawn with $N = 100$. The value reported in each cell is the mean of the estimated ranks from 1,000 simulations, and the value in the bracket is the standard deviation of the estimates.

Table 8 is designed to investigate the performances of the Threshold estimator when both weak and strong factors coexist. As in table 7, the left part of the table reports the results from the cases with *i.i.d.* errors, while the other part presents the results from the cases with cross-sectionally and auto-correlated errors. We conduct the test with two different factor-candidates models, both of them with $k = 3$. In each of these models we study three different possible SNRs for the weak factor. In one model we construct a factor structure with $r = 2$, where the first true factor is strong with λ_1 fixed at one and the second true factor is weak with three different λ_2 values: λ_2 = 0.1, 0.2, and 0.3. In the other model we study the case with $r = 3$ where the first two true factors are strong with $\lambda_1 = \lambda_2 = 1$, the last one is weak with three different λ_3 values: λ_3 = 0.1, 0.2, and 0.3. From the table, we can see that the Threshold estimator performs very well in small samples even if the weak factor's SNR is ten times smaller than the SNRs of the strong ones ($\lambda_2 = 0.1$, in the first model and $\lambda_3 = 0.1$, in the second model). The structure of the error terms does not show significant difference in the results.

Table 8

Results from the Simulated Data with Strong and Weak Factors

N	T	$\delta = \rho = J = 0$						$\delta = 0.2, \rho = 0.5, J = 8$					
		$r = 2, \lambda_1 = 1$			$r = 3, \lambda_1 = \lambda_2 = 1$			$r = 2, \lambda_1 = 1$			$r = 3, \lambda_1 = \lambda_2 = 1$		
		λ_2=0.1	λ_2=0.2	λ_2=0.3	λ_3=0.1	λ_3=0.2	λ_3=0.3	λ_2=0.1	λ_2=0.2	λ_2=0.3	λ_3=0.1	λ_3=0.2	λ_3=0.3
100	40	1.99 [0.10]	2.00 [0.00]	2.00 [0.00]	2.91 [0.30]	3.00 [0.00]	3.00 [0.00]	1.99 [0.08]	2.00 [0.04]	2.00 [0.04]	2.96 [0.20]	3.00 [0.00]	3.00 [0.00]
	60	2.00 [0.04]	2.00 [0.00]	2.00 [0.00]	2.93 [0.25]	3.00 [0.00]	3.00 [0.00]	1.99 [0.03]	2.00 [0.00]	2.00 [0.00]	2.98 [0.15]	3.00 [0.00]	3.00 [0.00]
	100	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.98 [0.14]	3.00 [0.00]	3.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.99 [0.05]	3.00 [0.00]	3.00 [0.00]
	150	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.04]	3.00 [0.00]	3.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]
	200	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]
	300	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]
	500	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]
	1000	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]	2.00 [0.00]	2.00 [0.00]	2.00 [0.00]	3.00 [0.00]	3.00 [0.00]	3.00 [0.00]

Note: Data are generated with three factor candidate variables ($k = 3$). The value reported in each cell is the mean of the rank estimates from 1,000 simulations, and the value in the bracket is the standard deviation of the estimates.

Table 9 is designed to investigate the performances of the Threshold estimator for the data generated without true factors. That is, all of the factor candidate factors used for Table 9 are "useless." We consider the cases with different numbers of useless factors. Table 9 shows that the Threshold estimator correctly detects the cases in which all factors are useless, if the number of factor candidate variables is small (e.g., $k = 1$), or T is large, or errors are only weakly correlated. When the errors are highly correlated, the estimator has relatively low power to detect useless factors unless T is sufficiently large.

Table 9

Threshold Estimation Results for the Data Simulated without Factors

N	T	$\delta = \rho = J = 0$			$\delta = 0.2, \rho = 0.5, J = 8$		
		$k=1$	$k=3$	$k=5$	$k=1$	$k=3$	$k=5$
100	50	0.00 [0.00]	0.14 [0.35]	0.81 [0.51]	0.05 [0.21]	0.48 [0.52]	1.12 [0.60]
	100	0.00 [0.00]	0.01 [0.04]	0.05 [0.21]	0.01 [0.10]	0.16 [0.39]	0.59 [0.54]
	200	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.05]	0.05 [0.22]	0.23 [0.43]
	500	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.01 [0.10]	0.03 [0.18]
	1000	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.03]	0.01 [0.08]
200	50	0.00 [0.00]	0.09 [0.29]	0.72 [0.51]	0.02 [0.15]	0.40 [0.50]	1.07 [0.55]
	100	0.00 [0.00]	0.00 [0.03]	0.01 [0.10]	0.00 [0.04]	0.08 [0.28]	0.38 [0.50]
	200	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.01 [0.08]	0.05 [0.22]
	500	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	1000	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]

Data are generated with the factors with the SNRs of zero ($r = 0$). The value reported in each cell is the mean of the rank estimates from 1,000 simulations, and the value in the bracket is the standard deviation of the estimates.

Our simulation results can be summarized as follows. First the Threshold estimator provides quite reliable inferences on the rank of the beta matrix even if the sample size is small. The SNR of each factor, the degrees of correlations among the errors, and the number of cross section units do not substantially influence the performances of the estimators. Second, the Threshold estimator can be used to check the possibility of all factor candidates' being "useless." The Threshold estimator is relatively less precise, if the number of the factor candidates analyzed is too large, or if the errors are highly correlated. However, it performs reasonably well even under such cases if the number of the time series observations is sufficiently large.

2.4.2 The Additional Comparison

In this subsection, we analyze the choice of the threshold function (TH) proposed in this paper. We compare its performance with different threshold candidates under three simulation setups.

We consider two sets of threshold candidates for the Monte Carlo exercise. The first set of threshold candidates we consider come from three penalty functions discussed in Bai and Ng (2002). These are AIC_1, BIC_3 and PC_1:

$$AIC_1 = \hat{\sigma}^2 (\frac{2}{T});$$

$$BIC_3 = \hat{\sigma}^2 \left(\frac{(N+T-k)\ln(NT)}{NT} \right);$$

$$PC_1 = \hat{\sigma}^2 \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right).$$

None of these three functions satisfy the two convergence rate required by Theorem 2. Although $AIC_1 \to 0$ as $T \to \infty$, it fails the second convergence condition since $T \times AIC_1$ does not converge to infinite as $T \to \infty$; BIC_3 and PC_1 do not converge to zero as $T \to \infty$.

The second set of threshold functions we use for comparison satisfy the two convergence rates in Theorem 2 but take a different form than TH. They are listed as F_1, F_2, and F_3:

$$F_1 = \hat{\sigma}^2 \frac{\ln T}{T};$$

$$F_2 = \hat{\sigma}^2 \frac{1}{T^{2/5}};$$

$$F_3 = \hat{\sigma}^2 \frac{1}{T^d},$$

where *d* is defined in the same way as in the threshold function (TH) we proposed. F_1, F_2, and F_3 satisfy the two convergence rates, but they do not (fully) include the potential effect of R^2. We expect them to perform well when factors are strong, but not in the case of weak factors.

The comparison of our threshold function (TH) and the above six candidates are conducted under three simulation setup: i) data generated with *i.i.d.* errors, ii) both cross-sectional and auto-correlated errors, and iii) weak factors. For each case, we set the number of factor candidates equal to 5 (*k=5*) and try different combinations of *N* and *T* where $N, T \in \{50,100\}$. 1000 samples are drawn for each combination of *N* and *T*.

The simulation results are reported in Table 10.

Panel A of Table 10 shows the results from the case with *i.i.d.* errors and the signal-to-noise ratios (SNRs) of all factors take the benchmark level of *1/r*. Comparing with the threshold estimation (TH) we propose, we can see that AIC_1 tends to overestimate the rank of the beta matrix, especially when the true rank is small. The overestimation decreases as *T* increases. BIC_3 underestimates the rank when the true rank is large, especially the full rank case (*r* = 5). We can see the underestimation still exists when *N* and *T* are as large as 100. There is a slightly underestimation problem with PC_1 and F_1 in the full rank case, and the underestimation decreases as long as *T* is large. For functions F_2 and F_3 we observe an underestimation problem when the true rank is large.

Panel B of Table 10 shows the results with auto and cross-correlated errors with the SNRs of all the candidate factors taking the benchmark level of *1/r*. We set $\rho = 0.5$, $\delta = 0.2$, and $J = 8$. The correlated errors increase the values of rank estimation from all thresholds. Hence for the factor, like AIC_1, with overestimation problem in the case of *i.i.d.* errors, the problem gets worse. The threshold function we propose, TH, and also PC_1 and F_1 have a slight overestimation problem when the true rank is small. The tests so far consist of data with all the factors' SNRs taking the benchmark level of *1/r*. Among all the candidate threshold functions we find that F_1 performs almost as well as TH.

Table 10
Comparison of the estimation results using proposed threshold function (TH) and other threshold functions

T	*N*	TH			AIC1			BIC3			PC1			F1			F2			F3		
		r=1	r=3	r=5	r=1	r=3	r=5	r=1	r=3	r=5	r=1	r=3	r=5	r=1	r=3	r=5	r=1	r=3	r=5	r=1	r=3	r=5
Panel A: Results of different threshold estimations from simulated data with I.I.D errors																						
50	50	1	3	5	1.32	3.05	5	1	2.16	1.43	1	3	4.47	1	3	4.98	1	2.84	2.85	1	2.99	3.65
		[0.04]	[0.00]	[0.06]	[0.47]	[0.22]	[0.00]	[0.00]	[0.63]	[0.70]	[0.00]	[0.04]	[0.57]	[0.03]	[0.00]	[0.13]	[0.00]	[0.38]	[0.71]	[0.00]	[0.11]	[0.52]
	100	1	3	5	1.24	3.04	5	1	2.84	2.24	1	3	4.99	1	3	5	1	2.97	3.16	1	3	4.15
		[0.00]	[0.00]	[0.00]	[0.43]	[0.09]	[0.00]	[0.00]	[0.37]	[0.78]	[0.00]	[0.00]	[0.11]	[0.00]	[0.00]	[0.00]	[0.00]	[0.18]	[0.73]	[0.00]	[0.03]	[0.51]
100	50	1	3	5	1.07	3.07	5	1	2.45	1.71	1	3	4.76	1	3	5	1	2.98	3.57	1	3	4.68
		[0.00]	[0.00]	[0.00]	[0.25]	[0.08]	[0.00]	[0.00]	[0.56]	[0.67]	[0.00]	[0.00]	[0.44]	[0.00]	[0.00]	[0.00]	[0.00]	[0.15]	[0.66]	[0.00]	[0.00]	[0.47]
	100	1	3	5	1.02	3	5	1	3	3.49	1	3	5	1	3	5	1	2.99	4.09	1	3	4.99
		[0.00]	[0.00]	[0.00]	[0.13]	[0.03]	[0.00]	[0.00]	[0.06]	[0.67]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.03]	[0.62]	[0.00]	[0.00]	[0.12]
Panel B: Results of different threshold estimations from simulated data with both cross- and auto-correlated errors																						
50	50	1.06	3	5	1.68	3.22	5	1	2.72	2.43	1.01	3	4.86	1.09	3.1	5	1	2.98	3.79	1	2.99	3.68
		[0.24]	[0.06]	[0.04]	[0.54]	[0.41]	[0.00]	[0.00]	[0.46]	[0.77]	[0.09]	[0.00]	[0.36]	[0.28]	[0.09]	[0.00]	[0.00]	[0.16]	[0.71]	[0.03]	[0.10]	[0.51]
	100	1.03	3	5	1.64	3.19	5	1	2.95	3.02	1.01	3	5	1.03	3.01	5	1	2.99	3.84	1	3	4.14
		[0.17]	[0.03]	[0.00]	[0.53]	[0.39]	[0.00]	[0.00]	[0.22]	[0.81]	[0.07]	[0.00]	[0.07]	[0.16]	[0.07]	[0.00]	[0.00]	[0.08]	[0.70]	[0.00]	[0.00]	[0.50]
100	50	1	3	5	1.49	3.11	5	1	2.86	2.75	1	3	4.96	1.01	3	5	1	3	4.36	1	3	4.61
		[0.03]	[0.06]	[0.00]	[0.52]	[0.31]	[0.00]	[0.00]	[0.35]	[0.71]	[0.00]	[0.00]	[0.19]	[0.08]	[0.00]	[0.00]	[0.00]	[0.03]	[0.60]	[0.00]	[0.00]	[0.50]
	100	1	3	5	1.38	3.08	5	1	3	4.12	1	3	5	1	3	5	1	3	4.59	1	3	4.98
		[0.00]	[0.00]	[0.00]	[0.50]	[0.28]	[0.00]	[0.00]	[0.03]	[0.63]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.52]	[0.00]	[0.00]	[0.15]
Panel C: Results of different threshold estimations from simulated data with weak factors																						
50	50	1	3.01	4.24	1.32	3.03	4.21	1	0.01	0	1	1.78	0.11	1	2.87	1.65	1	0.32	0	1	2.97	3.23
		[0.04]	[0.07]	[0.53]	[0.47]	[0.19]	[0.060]	[0.00]	[0.10]	[0.00]	[0.00]	[0.63]	[0.32]	[0.03]	[0.34]	[0.67]	[0.00]	[0.48]	[0.00]	[0.00]	[0.17]	[0.51]
	100	1	3	4.79	1.24	3.03	4.74	1	0.02	0	1	2.63	0.25	1	2.98	1.55	1	0.19	0	1	3	3.67
		[0.00]	[0.05]	[0.41]	[0.43]	[0.18]	[0.45]	[0.00]	[0.14]	[0.00]	[0.00]	[0.50]	[0.44]	[0.00]	[0.14]	[0.68]	[0.00]	[0.40]	[0.00]	[0.00]	[0.03]	[0.50]
100	50	1	3	4.96	1.07	3	4.96	1	0.01	0	1	1.97	0.02	1	3	2.82	1	0.61	0	1	3	4.71
		[0.00]	[0.00]	[0.21]	[0.25]	[0.06]	[0.19]	[0.00]	[0.07]	[0.00]	[0.00]	[0.61]	[0.15]	[0.00]	[0.03]	[0.66]	[0.00]	[0.56]	[0.00]	[0.00]	[0.00]	[0.47]
	100	1	3	5	1.02	3	5	1	0.14	0	1	2.93	0.23	1	3	3.14	1	0.45	0	1	3	4.99
		[0.00]	[0.00]	[0.00]	[0.13]	[0.03]	[0.00]	[0.00]	[0.35]	[0.00]	[0.00]	[0.25]	[0.44]	[0.00]	[0.00]	[0.69]	[0.00]	[0.53]	[0.00]	[0.00]	[0.00]	[0.12]

However, since F_1 does not include the explanatory power from the factor candidates, we do not expect it to perform well in the presence of weak factors.

Panel C shows the case with *i.i.d.* errors and weak factors. We define a weak factors as one with signal-to-noise ratio equal to $1/r^2$. Note that in the case of $r = 1$, the results are the same as those in the Panel A. The SNRs decrease as r increases. We find that BIC_3, PC_1, and F_2 have no power to identify weak factors when the true rank equals to three or five. F_1 can not identify the full rank case even in large samples. In comparison, AIC_1 and F_3 perform well in large samples, but not as well as TH.

Overall, we have shown that the threshold function (TH) we proposed in this paper is the most robust threshold, providing quite reliable inferences on the rank of the estimated beta matrix in both small and large samples and in all the studied scenarios.

In addition, we also compare the performance of our threshold estimation with the estimation methods considered in Cragg and Donald (1997).

The simulations are conducted in two models, named large model and small model. In the large model, there are 45 cross-sectional observations and 17 independent variables, which mean that we have $N = 45$ and $k = 17$. In the small model, we have $N = 10$ and $k = 6$. All the independent variables and the error terms are generated as i.i.d. $N(0,1)$.

The beta matrix is generated as $C\Lambda$, where C is a $N\times k$ matrix with each element from i.i.d. $N(0,1)$, and Λ is a $k\times k$ diagonal matrix with the diagonal value of Λ^2 equal to $\{0,\ldots,0,0.21,0.24,0.32,0.41,1.81\}$. In this case, we have the rank of beta matrix $r=5$.

The number of time-series observation *T* takes the value of 128, 256, and 1024. Each experiment consisted of 2526 independent replications. We report the results from Cragg and Donald (1997) and our threshold estimation in Table 11.

The four estimation criterions considered in Cragg and Donald (1997) are AIC, BIC, MSC, and TC. One prominent features are the weak performance of all the four methods in the large model with $T=128$. In contrast, our threshold estimation points 100% to the correct rank (which is 5) when the large model is estimated with 128 observations. The serious underestimation of the rank occurs in the large sample with 256 observations when BIC and MSC are used. While with 1024 observations, the underestimation still occurs. On the opposite, AIC and TC point towards a higher rank, but the overestimation is lowered when large value of *T* is used. Our threshold estimation in the large sample produces the correct estimation of the rank with small or large time-series observations.

Comparing the performance in the small model, our threshold estimation has underestimation problem with 128 observations. When the observation T gets larger, for example, $T=1024$, the threshold estimation points to the correct estimation more than 60% of the cases, and otherwise points to a lower estimation, usually rank 4.

Table 11

Estimates of the rank r =5, frequencies (%) of different rank estimates

	Large sample					Small sample				
Rank	AIC	BIC	MSC	TC	Threshold	AIC	BIC	MSC	TC	Threshold
T=128										
0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	5.4	0.0	0.0
1	0.0	63.1	0.0	0.0	0.0	0.0	5.5	94.4	0.0	0.0
2	0.0	34.9	0.0	0.0	0.0	0.0	33.1	0.2	0.0	0.0
3	0.0	1.9	0.0	0.0	0.0	0.4	38.9	0.0	0.5	7.8
4	1.1	0.0	0.0	0.0	0.0	12.3	18.8	0.0	24.1	67.5
5	14.3	0.0	0.0	2.4	100.0	83.4	3.7	0.0	72.3	24.7
6	43.6	0.0	0.0	24.9	0.0	3.9	0.0	0.0	3.0	0.0
7	33.5	0.0	0.0	43.9	0.0					
8	6.9	0.0	0.0	24.5	0.0					
9	0.4	0.0	0.0	4.0	0.0					
10	0.0	0.0	0.0	0.2	0.0					
T=256										
0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	98.6	0.0	0.0	0.0	0.0	0.0	92.4	0.0	0.0
2	0.0	1.4	0.0	0.0	0.0	0.0	0.3	7.6	0.0	0.0
3	0.3	0.0	0.0	0.0	0.0	0.0	4.3	0.1	0.0	4.3
4	13.7	0.0	0.0	6.0	0.0	0.0	26.0	0.0	0.2	60.2
5	59.3	0.0	0.0	46.8	100.0	93.4	69.4	0.0	95.8	35.5
6	24.8	0.0	0.0	39.5	0.0	6.6	0.0	0.0	4.0	0.0
7	1.9	0.0	0.0	7.2	0.0					
8	0.0	0.0	0.0	0.5	0.0					
T=1024										
0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	38.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	58.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7
4	0.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	38.6
5	91.3	0.0	0.0	93.2	100.0	92.7	100.0	100.0	97.3	60.7
6	8.7	0.0	0.0	6.5	0.0	7.3	0.0	0.0	2.7	0.0
7	0.0	0.0	0.0	0.4	0.0					
8	0.0	0.0	0.0	0.0	0.0					

All the four methods in Cragg and Donald (1997) perform better in the small sample than in the large sample, but the good performance also need large number of time-series observations. With 128 observations, BIC and MSC perform worse. Even when the number of observation is 256, MSC still fail to point to the correct rank even once.

Overall, the threshold estimation we proposed in this paper performs relatively well, especially in the large models. To further clarify the effect of the model sizes on the threshold estimation, we also consider two middle-sized models.

In the case that $N = 28$ and $k = 12$, which values we pick between the large and small models in Cragg and Donald (1997). The other parameters stay the same as reported in Table A2. Our threshold estimation points to the correct rank more than 99% of the time with the each of the three observations of 128, 256, and 1024.

Also with the model of $N = 45$ and $k = 7$, the threshold estimation points to the correct rank 100% of the time with the each of the three observations of 128, 256, and 1024.

To sum up, we suggest using the threshold estimation in estimating the rank of the beta matrix in relatively large models ($N \geq 28$).

2.5 Application

In this section we estimate the rank of the beta matrix using different factor-candidates as regressors. More specifically, we use the three factors proposed in the model of Fama and French (1992, FF), the five factors of Chen, Roll, and Ross (1986, CRR),[4] the momentum and reversal factors (MOM) available on Kenneth French webpage: momentum, short-term reversal and long-term reversal, and the two new factors developed in Chen, Novy-Marx, and Zhang (CNZ, 2010): Investment to Asset (IA) and Return on Asset (ROA). [5]

As response variables we use the US monthly individual stock returns and portfolio returns.[6] Returns are calculated in excess over the risk free rate. The individual stock returns are downloaded from CRSP. The returns include dividends. The risk free rate is the one-month Treasury bill rate, which is available from Kenneth French's webpage. For the individual stock returns, we exclude REITs (Real Estate Investment Trusts), ADRs (American Depositary Receipts) and the stocks that do not have information for every month over a sample period. We also exclude stocks that show more than 300% excess returns in a given month since we are trying to capture common variation. Excessively high or low returns are most likely to be idiosyncratic risks. US Stock portfolio

---

4 While the FF model may be more related to the APT, the CRR model is more related to Merton's (1972) Intertemporal CAPM, in the sense that they try to find the macroeconomic (state) variables that may influence future investment opportunities. The factors proposed by CRR are industrial production (MP), unexpected inflation (UI), change in expected inflation (DEI), the term premium (UTS), and the default premium (UPR). Each of these factors is available from Laura Xiaolei Liu's webpage from January 1960 to December 2004 (http://www.bm.ust.hk/~fnliu/research.html). For detailed information on how these factors have been constructed, see Liu and Zhang (2008). The FF factors are the proxy for the market risk premium, SMB and HML.

5 We thank Long Chen for providing us the latest version of their factors.

6 We do not use the daily returns since the data of some factor candidates are only available at monthly frequency.

returns are downloaded from Kenneth French's webpage. The portfolios used are 100 portfolios based on Size and Book to Market, 25 portfolios based on Size and Book to Market, 25 portfolios based on Size and Momentum, 49 Industrial portfolios and 30 Industrial portfolios. We use monthly returns in every data set.

Response variables are always double-demeaned as suggested in the Equation (2). We also use standardized factors for the following reason. The beta values corresponding to each factor change depending on the scale of the factor. For example, if we rescale a factor by multiplying 10, the (absolute) beta values corresponding to the factor are scaled down by the order of 0.1. In this case, even if the factor has a high explanatory power, the estimated betas obtained with the rescaled factor would not reflect the factor's true explanatory power.

2.5.1 Rank Estimation Using Individual Stock Returns

The time span included in the analysis is from 1972 to 2004. We divide the individual stock returns into three samples: the entire time span (1972-2004), two subsamples (1972 – 1987 and 1988 – 2004) and three subsamples (1972 – 1978, 1979 – 1992, and 1993 – 2004). Under both subdivisions, we could fit a polynomial trend to the value weighted market portfolio to estimate the up and down cycles. We do so to examine how the estimation results may change depending on time intervals. We keep the time span T at around 100 or more since the simulation exercises show that the estimators are very accurate in this case. The number of cross-sectional observations N changes as T changes in order to maintain a balanced panel. The value of N depends on the available observations with complete data on CRSP for each sample period after the data has been cleaned.

The results from the estimation of the rank of the beta matrix for individual stock returns are shown in Table 12. Each line of the table represents a different estimated model. For each model we report the number of factor candidates used (*k*), the estimated number of factors among the factor candidates ($\hat{r}$) and the average R^2 of the regressing the response variables on the factor-candidates.

The first line of table 12 shows that the Threshold estimator predicts that the rank of the beta matrix equals three when using the three FF factors in different sample periods.

The second line of table 12 shows the results from the estimation of the five CRR factors. For any period, the estimated rank does not exceed two. This means that only one or two common sources of comovement in individual stock returns are explained by the CRR factors. This result provides strong evidence that the risk premiums of some factors in the CRR model are undefined.

Given that the CRR factors can identify one or two common factors in individual stock returns, a question we wish to answer is whether the CRR factors capture some sources of comovement that the FF factors fail to do. If the CRR factors capture different information from what the FF factors do, we could expect that the rank of the beta matrix from the joint model of CRR and FF would be equal to the sum of the ranks from the CRR and FF models separately. Indeed, the Threshold estimation results are consistent with this expectation in the entire sample and every subsample. In the third line of result in table 12 the Threshold estimation suggests that the risks captured by the CRR and FF factors are different.

Since the five CRR factors capture a common source of comovement that is not captured by the FF factors, an interesting question is which of the CRR factors contain the information missed by the FF factors. For this purpose we add to the FF factors each CRR factor individually in order to estimate the rank of the beta matrix of at most four. In unreported results we find that no individual CRR factor increases the rank of the beta matrix when combined with the FF factors. Then we use every possible combination of two CRR factors together added to the three FF factors. In this case we found that adding UI (unexpected inflation) and DEI (changes in expected inflation) increases the rank of the beta matrix to four. Results are shown in the 4^{th} line of table 12. This shows that a factor related to inflation is missed by the FF factors.

Furthermore, we analyze if momentum factors (as constructed by Kenneth French) capture a different source of risk than the Fama-French factors. Results of estimating the rank of the beta matrix of the three momentum factors and the FF factors are presented in the 5^{th} row of the table. The Threshold estimator finds strong evidence for an extra factor contained in the three momentum factors in most samples. However, if we add any one or any two possible combinations of the momentum factors to FF three factors, unreported results show that in most cases we find the rank equals three.

Table 12

Rank Estimation Results from Different Factor Models Using Individual Stock Returns

Sample Period	(T,N)	January 1972 - December 2004 (396,313)		January 1972 - December 1987 (192,816)		January 1988 - December 2004 (204,1288)		January 1972 - December 1978 (84,1384)		January 1979 - December 1992 (168,1640)		January 1993 - December 2004 (144,1855)	
	k	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2
1) FF Factors	**3**	3	*0.073*	3	*0.097*	3	*0.078*	3	*0.155*	3	*0.065*	3	*0.104*
2) CRR Factors	**5**	1	*0.018*	1	*0.035*	1	*0.026*	2	*0.068*	1	*0.038*	2	*0.037*
3) FF and CRR factors	**8**	4	*0.089*	4	*0.127*	4	*0.103*	5	*0.209*	4	*0.100*	5	*0.139*
4) FF, UI and DEI factors	**5**	4	*0.081*	4	*0.112*	4	*0.089*	4	*0.180*	4	*0.080*	4	*0.120*
5) FF and Mon factors	**6**	4	*0.090*	4	*0.118*	3	*0.103*	5	*0.198*	4	*0.089*	4	*0.135*
6) FF and CNZ factors	**5**	4	*0.087*	4	*0.110*	4	*0.098*	4	*0.183*	3	*0.081*	4	*0.127*
7) FF, Mom, UI, DEI and CNZ	**10**	6	*0.110*	6	*0.146*	6	*0.131*	7	*0.250*	6	*0.119*	7	*0.174*

Note: This table reports the estimation of the rank of the beta matrix for U.S stock portfolio returns. For every portfolio set the time span is January 1972 - December 2004 (T=396). Each line of the table represents a different estimated model. For each model we report the number of factor candidates used (k), the estimated number of factors among the factor candidates ($\hat{r}$) and the average R^2 of the regressing the response variables on the factor-candidates.

We conclude that there is evidence for a momentum factor among the three momentum factors during the period under analysis that is not captured by the FF factors when using individual stock returns.

In the 6th row of table 12 we test the rank of the beta matrix when using the three FF factors and the two new factors of CNZ and find four factors in almost every subsample.[7] This is evidence that the CNZ factors capture one dimension missed by the FF factors.

Finally, the last row of the table show the results of using the ten factor-candidates that seem to contain different information together: the three FF factors, UI and DEI from CRR, the three momentum factors and the two CNZ factors. The table shows that there is evidence for at least six factors among the 10 factor candidates.

However, an open and important question is whether we need to use individual stock returns or portfolio returns to estimate the beta matrix in order to perform asset pricing tests[8]. For example, imagine a hypothetical situation in which half of the sample of the individual stock returns have betas of 0.5 with respect to a factor and the other half have betas of -0.5. In this case the factor will add a dimension to the rank of the beta matrix when using individual stock returns, but this factor will disappear in properly diversified portfolios (because the beta of the diversified portfolio with respect to the factor will be zero). In the next

---

7 Most of the time adding ROA to the FF factors is sufficient to get a rank equal to four while adding only IA never increases the estimated rank of the beta matrix. For this reason, we can conclude that ROA posses most of the information not captured by the FF factors.
8 See Ang, Liu, and Schwarz (2008).

section we estimate the rank of the beta matrix using the same factor candidates as before but using portfolio returns as response variables.

2.5.2 Rank Estimation Using Portfolio Returns

In this section we use five sets of portfolios downloaded from Kenneth French website as response variables. Since the number of portfolios is fixed in each different set, we use for every estimation the full time span from January 1972 to December 2004 (T=396). The cross-sectional dimension N equals to the number of portfolios in each set. In table 13 we report the same statistics for portfolio returns as those in the previous table for individual stock returns.

When using the FF factors we find all the time an estimated rank of three except for the 25 Size and Book to Market portfolio set where we find a rank of two. When we use the five CRR factors we find the rank equals to one or two as in the case with individual stocks. When we test together the FF factors and the CRR factors (k=8), we do not find evidence of an extra factor except for the cases of the 49 and 30 Industrial Portfolios.

A common pattern observed in table 13 is that when testing the number of factors in Industrial Portfolios the results are similar to those obtained using individual stock returns. However, once we use portfolios based on Book to Market and Size or Size and Momentum, the rank of the beta matrix is at most four. The maximum rank we find for 100 Size and Book to Market portfolios is four, and for 25 Size and Book to Market portfolios and 25 Size and Momentum portfolios is three.

Table 13

Rank Estimation Results from Different Factor Models Using Stock Portfolio Returns

		100 Size and B/M Portfolios		25 Size and B/M Portfolios		25 Mom and Size Portfolios		49 Industrial Portfolios		30 Industrial Portfolios	
	k	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2	$\hat{r}$	R^2
1) FF Factors	**3**	3	*0.42*	2	*0.66*	3	*0.33*	3	*0.12*	3	*0.12*
2) CRR Factors	**5**	1	*0.015*	1	*0.018*	2	*0.023*	1	*0.017*	1	*0.019*
3) FF and CRR factors	**8**	3	*0.43*	2	*0.67*	3	*0.34*	4	*0.14*	4	*0.14*
4) FF and Momentum	**6**	3	*0.43*	2	*0.67*	3	*0.67*	4	*0.14*	4	*0.14*
5) FF and CNZ factors	**5**	4	*0.44*	3	*0.69*	3	*0.41*	4	*0.16*	4	*0.17*
6) FF, Mom, CRR and CNZ	**13**	4	*0.46*	3	*0.70*	3	*0.70*	6	*0.19*	7	*0.20*

Note: This table reports the estimation of the rank of the beta matrix for U.S stock portfolio returns. For every portfolio set the time span is January 1972 - December 2004 (T=396). Each line of the table represents a different estimated model. For each model we report the number of factor candidates used (k), the estimated number of factors among the factor candidates ($\hat{r}$) and the average R2 of the regressing the response variables on the factor-candidate.

This is evidence that the portfolios sorted based on these characteristics are better diversified (these portfolios also show less residual variance since their R^2 is higher than the one of the Industrial portfolios). A possible explanation is the existence of industry specific factors that are diversified away when constructing portfolios based on characteristics like Size and Book to Market. This is a useful result that can clarify the discussion of whether to use portfolios or individual stock returns when testing factors and also the discussion about which type of portfolios should be used. It is known that industry portfolios tend to have positive abnormal excess returns (intercepts are significantly larger than zero). According to our result this is because the existence of industry specific factors that disappear when well diversified portfolios are used. In other words, the positive α that appears in many of the Industry Portfolios should not be considered a models' mispricing since it is exposure to a source of diversifiable risk.

Our empirical results can be summarized as follows. When using individual stock returns we find evidence for the existence of six common factors among the thirteen factor candidates used. These factors are the three FF factors, a factor related to inflation from the CRR factors, a Momentum factor and a factor captured by the new CNZ factors. When we use Industrial Portfolio returns, results remain the same. However, when we use portfolios that are better diversified such as the ones sorted on characteristics like Size and Book to Market, the FF factors seem to be enough to capture all the common sources of risk among the thirteen factor candidates, except for the 100 Size and Book to Market portfolios in which an extra factor appears when adding the CNZ factors.

2.6 Conclusions

In this paper, we have proposed a new rank estimator, called Threshold estimator, for the beta matrix from a factor model with observed factor-candidate variables. Testing whether the beta matrix has full rank is important for the two-pass estimation of the risk premiums in empirical asset pricing models. The (demeaned) beta matrix needs to have full rank. Otherwise, risk premiums are undefined. The Threshold estimator is computed easily with the eigenvalues of the inner product of an estimated beta matrix. Our simulation exercises provide promising evidence that the Threshold estimator has good finite-sample properties. Different from the existing methods, this proposed method can be used to analyze the data with a large number of cross-section units.

In our empirical investigation we find that all of the Fama-French (1993) three factors have explanatory power when using US individual stock returns as response variables, In contrast, only one or two among the five factors of Chen, Roll, and Ross (1986) have explanatory power. When we combine the three factors of Fama-French (FF) together with the five factors of Chen, Roll, and Ross (CRR) we find that a factor not captured by FF is captured by CRR. Furthermore, we find that momentum and reversal factors capture a source of risk not captured by either FF or CRR. Similarly, the two factors proposed by Chen, Novy-Marx, and Zhang (2010, CNZ) capture a source of risk missed by all the other factors. We find evidence for six factors in US individual stock returns among the thirteen factor candidates used. When we use Industrial Portfolio returns, results remain the same. However, when we use portfolios that are better diversified such as the ones sorted on characteristics like Size and Book to Market, the FF factors seem to be enough to capture all the common sources of risk among the thirteen factor candidates, except for the 100 Size and Book to Market portfolios in which an extra factor appears when adding the CNZ factors.

Bai and Ng (2002), Onatski (2010), and Ahn and Horenstein (2009) have developed the estimators for the number of factors without using factor candidate variables. Their studies have found one or two factors from US individual stock return data. In contrast, our results provide evidence of at least six factors in individual stock returns. All of the estimation methods proposed by the above three studies are based on the analysis of principal components of response variables. Ahn and Horenstein (2009) found that principal components provide poor estimates of the true factors when the true factors are weak and the idiosyncratic errors are cross sectional correlated. From their results, we can conjecture that the analysis of principal components might have limited power to detect weak factors. In contrast, the Threshold estimator proposed in this paper utilizes observed factor candidate variables. Factors need not be estimated. Thus, we can expect that the new estimator would have a higher power to detect the weak factors hidden among the factor-candidate variables. Our estimation results are consistent with this expectation.

REFERENCES

Ang, Andrew, Jun Liu, and Krista Schwarz, 2010, Using Stocks or Portfolios in Tests of Factor Models, Working paper.

Ahn, Seung C., and Alex R. Horenstein, 2009, Eigenvalue ratio test for the number of factors, working paper, Arizona State University.

Burnside, Craig A., 2010, Identification and Inference in Linear Stochastic Discount Factor Models, Working Paper.

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

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, Long, Robert Novy-Marx and Lu Zhang, 2010, An Alternative Three Factor Model, working paper.

Chen, Nai-Fu, Raymond Kan, and Chu Zhang, 1997, A critique of the use of t-ratios in model selection, Working Papers.

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

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.

Ferson, Wayne E., Sergei Sarkissian, and Timothy Simin, 1999, The alpha factor asset pricing model: A parable, Journal of Financial Markets 2, 49-68.

Frisch, Ragnar, and Frederick V. Waugh, 1993, Partial time regressions as compared with individual trends, Econometrica 1, 387–401.

Jagannathan, Ravi, and Zhenyu Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, The Journal of Finance 53, 1285-1309.

Jagannathan, Ravi, and Zhenyu Wang, 2002, Empirical evaluation of asset-pricing models: A comparison of the sdf and beta methods, The Journal of Finance 57, 2337-2367.

Kan, Raymond, and Chu Zhang, 1999a, Gmm tests of stochastic discount factor models with useless factors, Journal of Financial Economics 54, 103-127.

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

Kan, Raymond, and Chu Zhang, 1999c, A critique of the stochastic discount factor methodology, The Journal of Finance 54, 1221-1248.

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

Lewellen, Jonathan, Stefan Nagel, and Jan Shankan, 2010, A skeptical appraisal of asset pricing tests, Journal of Financial Economics 96, 175-194.

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 – 48.

Merton, Robert, 1972, An intertemporal capital asset pricing model, Econometrica 41, 867 – 87.

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

Nawalkha, Sanjay, 1997, A multibeta representation theorem for linear asset pricing theories, Journal of Financial Economics 46, 357-381.

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

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.

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

Shanken, Jay, 1992, On the estimation of beta-pricing models, The Review of Financial Studies 5, 1-33.

Shanken, Jay, and Guofu Zhou, 2007, Estimating and testing beta pricing models: Alternative methods and their performance in simulations, Journal of Financial Economics 84, 40-86.

Sharpe, William, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, The Journal of Finance19, 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.

APPENDIX A

PROOF FOR CHAPTER 1

Proof of Lemma 1:

Since f_{1t} is a useless factor, we have the corresponding factor loading vector $\beta_1 = 0$, based on the Law of Large Numbers and the Central Limit Theorem, we can easily show that $\sqrt{T}b_1 \sim N(0, \Sigma_\varepsilon / \Sigma_{f_1})$. We can rewrite

$$\begin{aligned}\hat{\gamma}_1^{OLS} &= (b_1' M_{-1} b_1)^{-1} b_1' M_{-1} \bar{R} \\ &= \sqrt{T}(\sqrt{T}b_1' M_{-1} \sqrt{T}b_1)^{-1} \sqrt{T}b_1' M_{-1} \bar{R},\end{aligned}$$

then $\hat{\gamma}_1^{OLS} / \sqrt{T}$ is a random variable. Following the same logic, we can show that $\hat{\gamma}_1^{GLS} / \sqrt{T}$ is also a random variable.

Proof of Proposition 1:

To show specifically that the null hypothesis that the risk premium of a useless factor is equal to zero will be rejected more often than it should be at the nominal size, we conduct the analysis with OLS estimation first, and generalize the results with GLS estimation.

With EIV unadjusted standard error, we have

$$\begin{aligned}s^2(\hat{\gamma}_1) &= (b_1' M_{-1} b_1)^{-1} b_1' M_{-1} \hat{V} M_{-1} b_1 (b_1' M_{-1} b_1)^{-1} \\ &= T(\sqrt{T}b_1' M_{-1} \sqrt{T}b_1)^{-1} \sqrt{T}b_1' M_{-1} \hat{V} M_{-1} \sqrt{T}b_1 (\sqrt{T}b_1' M_{-1} \sqrt{T}b_1)^{-1}.\end{aligned}$$

Since $\sqrt{T}b_1 \sim N(0, \Sigma_\varepsilon / \Sigma_{f_1})$, we have $s(\hat{\gamma}_1)/\sqrt{T}$ is a random variable.

The t-test statistic for the null hypothesis $H_0 : \gamma_1 = 0,$ is given:

$$t(\hat{\gamma}_1) = \frac{\hat{\gamma}_1}{s(\hat{\gamma}_1)/\sqrt{T}} = \sqrt{T}\, \frac{\hat{\gamma}_1/\sqrt{T}}{s(\hat{\gamma}_1)/\sqrt{T}} \rightarrow \infty .$$

For this case, the EIV unadjusted t-statistics are not credible, because one will always find the useless factors are priced even when large samples are used.

With the EIV adjusted standard error, following the methodology in Kan and Zhang (1999b), we have

$$t_{EIV}^2(\hat{\gamma}_1) = \frac{\hat{\gamma}_1^2}{s_{EIV}^2(\hat{\gamma}_1)/T} = \frac{T\hat{\gamma}_1^2}{(1+\hat{\gamma}'\hat{\Sigma}_f^{-1}\hat{\gamma})(b_1'M_{-1}b_1)^{-1}b_1'M_{-1}\hat{\Sigma}_\varepsilon M_{-1}b_1(b_1'M_{-1}b_1)^{-1}+\hat{\Sigma}_{f_1}}.$$

Define $d_1 = \hat{\gamma}_1^2\Sigma_{f_1}^{-1}/(1+\hat{\gamma}'\Sigma_f^{-1}\hat{\gamma}) = O_p(1)$, where $O_p(T^\alpha)$ means convergence at an exact order of T^α, then we have $\hat{\Sigma}_{f_1}/(d_1T) = O_p(1/T)$ and

$$t_{EIV}^2(\hat{\gamma}_1) - d_1\frac{T\Sigma_{f_1}(b_1'M_{-1}b_1)^2}{b_1'M_{-1}\Sigma_\varepsilon M_{-1}b_1^{-1}} \xrightarrow{p} 0.$$

Since $\hat{\Sigma}_\varepsilon \to \Sigma_\varepsilon$ and , $t_{EIV}^2(\hat{\gamma}_1)$ has the same limiting distribution as $d_1T\Sigma_{f_1}(b_1'M_{-1}b_1)^2/(b_1'M_{-1}\Sigma_\varepsilon M_{-1}b_1)$.

Now following the Proof of Proposition 6 in Kan and Zhang (1999b), given $b_1 \sim N(0, \Sigma_\varepsilon / T\Sigma_{f_1})$, we define $Z = \sqrt{T}H'\Sigma_{f_1}^{1/2}\Sigma_\varepsilon^{-1/2}b_1 \sim N(0, I_{N-k+1})$, and H is defined by the eigenvalue decomposition that $\Sigma_\varepsilon^{1/2}M_{-1}\Sigma_\varepsilon^{1/2} = H\Lambda H'$, where H is an $N \times (N-k+1)$ orthonormal matrix and $\Lambda = Diag(\lambda_1, \ldots, \lambda_{N-k+1})$ where $\lambda_1 \geq \cdots \geq \lambda_{N-k+1} > 0$ are the $N-k+1$ nonzero eigenvalue of $\Sigma_\varepsilon^{1/2}M_{-1}\Sigma_\varepsilon^{1/2}$. Hence we have

$$d_1\frac{s_1(b_1'M_{-1}b_1)^2}{b_1'M_{-1}\Sigma_\varepsilon M_{-1}b_1} = d_1\frac{(Z'\Lambda Z)^2}{Z'\Lambda^2 Z} = d_1\frac{(\sum_{i=1}^{N-k+1}\lambda_i Z_i^2)^2}{\sum_{i=1}^{N-k+1}\lambda_i^2 Z_i^2} > d_1\sum_{i=1}^{N-k+1}\frac{\lambda_i}{\lambda_1}Z_i^2.$$

Following the inequality that $\sum_{i=1}^{N-k+1}\lambda_i^2 Z_i^2 > \lambda_1(\sum_{i=1}^{N-k+1}\lambda_i Z_i^2)$, then we have $\lim t_{EIV}^2(\hat{\gamma}_1) > d_1\sum_{i=1}^{N-k+1}(\lambda_i/\lambda_1)Z_i^2$. Given the estimated $\hat{\gamma}_1^2 = O_p(T)$, and the estimated risk premium of the true factors have the property that $\hat{\gamma}_j = O_p(1)$, $j = 2,\ldots,k$. Hence, $d_1 \to 1$ as $T \to \infty$. Then we have $\lim t_{EIV}^2(\hat{\gamma}_1) > \sum_{i=1}^{N-k+1}(\lambda_i/\lambda_1)Z_i^2 > Z_1^2$ using the OLS estimation.

Using GLS estimation, it is easy to verify that we have similar results that $t(\hat{\gamma}_1)^{GLS} = \sqrt{T}(\hat{\gamma}_1^{GLS}/s(\hat{\gamma}_1)^{GLS}) \to \infty$ and $\lim t^2(\hat{\gamma}_1)_{EIV}^{GLS} > \sum_{i=1}^{N-k+1} Z_i^2 > Z_1^2$, for some well defined $N\times 1$ vector $Z \sim N(0, I_{N-k+1})$.

In the correctly specified model, $t_{EIV}^2(\hat{\gamma}_1)$ should have limiting distribution of χ_1^2. When $N > 2$, over rejection problem with both OLS and GLS estimation occurs asymptotically.

Proof of Lemma 2:

For Case 2, since f_{1t} and f_{2t} are two useless factors, we have the corresponding two factor loading vectors $\beta_1 = \beta_2 = 0$, based on the Law of Large Numbers and the Central Limit Theorem, we can show that $\sqrt{T}vec(b_1, b_2) \sim N(0, \Sigma_\varepsilon \otimes \Sigma_{f_{12}}^{-1})$. Hence, for $i = 1, 2$, we have

$$
\begin{aligned}
\hat{\gamma}_i^{OLS} &= (b_i' M_{-i} b_i)^{-1} b_i' M_{-i}\bar{R} \\
&= \sqrt{T}(\sqrt{T}b_i' M_{-i}\sqrt{T}b_i)^{-1}\sqrt{T}b_i' M_{-i}\bar{R},
\end{aligned}
$$

then $\hat{\gamma}_i^{OLS} / \sqrt{T}$ is a random variable. Following the same logic, we can show that $\hat{\gamma}_i^{GLS} / \sqrt{T}$ is also a random variable.

For Case 3, for the estimated betas, we have

$$b_1 = c_1\beta^* + O_p(T^{-1/2});$$

$$b_2 = c_2\beta^* + O_p(T^{-1/2}),$$

where β^* is the nonzero coefficient of the true factor f_t^*. Hence, $b_1 = (c_1 / c_2)b_2 + O_p(T^{-1/2})$, assuming $c_2 \neq 0$. Now define $M_{-2} = I_N - b_{-2}(b_{-2}'b_{-2})^{-1}b_{-2}'$ and $b_{-2} = (b_1, b_3, \ldots, b_k)$. Given that $M_{-2}b_1 = 0$, we have $M_{-2}b_2 = O_p(T^{-1/2})$. Then $\hat{\gamma}_2^{OLS} = (b_2'M_{-2}b_2)^{-1}b_2'M_{-2}\bar{R} = O_p(T^{1/2})$. Hence, we still have $\hat{\gamma}_2^{OLS} / \sqrt{T}$ is a random variable. Using the same logic, we can show that $\hat{\gamma}_1^{OLS} / \sqrt{T}$ is a random variable. The same results hold for GLS estimated risk premium.

Proof of Proposition 2:

The proof here is derived under Case 2, where we have two useless factors, f_{1t} and f_{2t}. Following the same proof in Proof 1, with EIV unadjusted standard error, we have OLS estimated t-test statistics $t(\hat{\gamma}_i) = \sqrt{T}(\hat{\gamma}_i / s(\hat{\gamma}_i)) \to \infty$, $i = 1, 2$. With EIV adjusted standard error, we can get $\lim t_{EIV}^2(\hat{\gamma}_i) > d_i Z_i^2$, $i = 1, 2$. However, since $\hat{\gamma}_1^2 = O_p(T)$, $\hat{\gamma}_2^2 = O_p(T)$, and the risk premiums of the true factors $\hat{\gamma}_j = O_p(1)$, $j = 3, \ldots, k$, we can get $0 < d_1 < 1$ asymptotically. Still the over-rejection problem could happen depend on the realization of $\hat{\gamma}_1$ and $\hat{\gamma}_2$, we can say one of the t-statistics of the useless factor will be larger than $\frac{1}{2}\sum_{i=1}^{N-k+1}\frac{\lambda_i}{\lambda_1}Z_i^2$.

Now consider the properties of the Wald statistics, with EIV unadjusted covariance matrix, we have the OLS estimation results of testing the joint hypothesis given as $W(\hat{\gamma}_{12}) = \hat{\gamma}_{12}{}'[C\text{ov}(\hat{\gamma}_{12})/T]^{-1}\hat{\gamma}_{12} \to \infty$, using the fact that $\hat{\gamma}_{12}/\sqrt{T}$ is a vector of two random variables.

With the EIV adjusted covariance matrix, we conduct the analysis as follows:

$$\begin{aligned} W_{EIV}(\hat{\gamma}_{12}) &= \hat{\gamma}_{12}{}'[C\text{ov}_{EIV}(\hat{\gamma}_{12})/T]^{-1}\hat{\gamma}_{12} \\ &\to \hat{\gamma}_{12}{}'\Sigma_{f_{12}}^{-1/2}\xi\Sigma_{f_{12}}^{-1/2}\hat{\gamma}_{12}/(1+\hat{\gamma}'\Sigma_f^{-1}\hat{\gamma}) \end{aligned},$$

where $\xi = T(b_{12}'M_{-12}b_{12})(b_{12}'M_{-12}\Sigma_\varepsilon M_{-12}b_{12})^{-1}(b_{12}'M_{-12}b_{12})$.

Given that $vec(b_{12}) \sim N\left(vec(\beta_{12}), T^{-1}\Sigma_\varepsilon \otimes \Sigma_{f_{12}}^{-1}\right)$, we define $y = \sqrt{T}H'\Sigma_\varepsilon^{-1/2}b_{12}\Sigma_{f_{12}}^{1/2}$, and H is defined by the eigenvalue decomposition that $\Sigma_\varepsilon^{1/2}M_{-1}\Sigma_\varepsilon^{1/2} = H\Lambda H'$, where H is an $N\times(N-k+1)$ orthonormal matrix and $\Lambda = Diag(\lambda_1,\ldots,\lambda_{N-k+1})$ where $\lambda_1 \geq \cdots \geq \lambda_{N-k+1} > 0$ are the $N-k+1$ nonzero eigenvalue of $\Sigma_\varepsilon^{1/2}M_{-1}\Sigma_\varepsilon^{1/2}$. Then we have $Cov(vec(y)) = I_{2N\times 2N}$, which means that each element of $vec(y)$ is a noncentral χ_1^2 random variable. Then

$$\begin{aligned} \xi &= T\Sigma_{f_{12}}^{1/2}(b_{12}'M_{-12}b_{12})(b_{12}'M_{-12}\Sigma_\varepsilon M_{-12}b_{12})^{-1}(b_{12}'M_{-12}b_{12})\Sigma_{f_{12}}^{1/2} \\ &= (y'\Lambda y)(y'\Lambda^2 y)^{-1}(y'\Lambda y). \end{aligned}$$

For any 2×1 matrix R, we have $R'\xi R > R'(y'\Lambda y)R/\lambda_1 > R'Rc_1$, where

$$c_1 = \min\{\Sigma_{i=1}^{N-k+1}\frac{\lambda_i}{\lambda_1}y_{1i}^2, \Sigma_{i=1}^{N-k+1}\frac{\lambda_i}{\lambda_1}y_{2i}^2, \frac{1}{2}\Sigma_{i=1}^{N-k+1}\frac{\lambda_i}{\lambda_1}(y_{1i}-y_{2i})^2\}.$$

Given that $\hat{\gamma}_{12}'\Sigma_{f_{12}}^{-1}\hat{\gamma}_{12}/(1+\hat{\gamma}'\Sigma_f^{-1}\hat{\gamma}) \to 1$, hence we have $W_{EIV}(\hat{\gamma}_{12}) > c_1$.

The same analysis can be applied to GLS estimation. We have the EIV unadjusted *t*-test statistics $t(\hat{\gamma}_i)^{GLS} = \sqrt{T}(\hat{\gamma}_i^{GLS}/s(\hat{\gamma}_i)^{GLS}) \to \infty$, for *i*=1,2, and one of EIV adjusted GLS estimation of the t-statistics of the useless factor, $t^2(\hat{\gamma}_i)_{EIV}^{GLS}$, $i=1$ or 2, will be larger than $\frac{1}{2}\sum_{i=1}^{N-k+1}Z_i^2$, for some well defined $N\times1$ vector $Z \sim N(0, I_{N-k+1})$.

For the joint hypothesis that the risk premiums of the useless factors are equal to zero, we have the EIV unadjusted Wald test statistic as

$$W^{GLS}(\hat{\gamma}_{12}) = \hat{\gamma}_{12}'[C\text{ov}^{GLS}(\hat{\gamma}_{12})/T]^{-1}\hat{\gamma}_{12} \to \infty,$$

and with the EIV adjust covariance matrix, we have $W_{EIV}^{GLS}(\hat{\gamma}_{12}) > c_2$, where

$$c_2 = \min\{\Sigma_{i=1}^{N-k+1}y_{1i}^2, \Sigma_{i=1}^{N-k+1}y_{2i}^2, \frac{1}{2}\Sigma_{i=1}^{N-k+1}(y_{1i}-y_{2i})^2\},$$

and $y=(y_1, y_2)$ are some well defined $N\times2$ matrix with $\text{var}(vec(y)) = I_{2N\times2N}$.

It is easy to verify that the properties of the test-statistics under Case 3 are the same as those with two useless factors in Case 2, since the only difference is the nonzero mean of the estimated beta matrix.

Proof of Proposition 3:

Under Case 1, Case 2, and Case 3, we will have $\hat{\gamma}\hat{\Sigma}_f^{-1}\hat{\gamma}/T$ is a random variable, because as shown in Case 1 to Case 3 that $\bar{\hat{\gamma}}_1/\sqrt{T}$ is a random variable. Also we have $s^2(\hat{\gamma}_k)$ is a random variable, and the EIV adjusted standard error of the risk premium

$$
\begin{aligned}
s^2(\hat{\gamma}_k)_{EIV} &= (1+\hat{\gamma}\hat{\Sigma}_f^{-1}\hat{\gamma})(s^2(\hat{\gamma}_k)-\hat{\Sigma}_{f_k})+\hat{\Sigma}_{f_k} \\
&= T(1/T+\hat{\gamma}\hat{\Sigma}_f^{-1}\hat{\gamma}/T)(s^2(\hat{\gamma}_k)-\hat{\Sigma}_{f_k})+\hat{\Sigma}_{f_k},
\end{aligned}
$$

then $s^2(\hat{\gamma}_k)_{EIV}/T$ is a random variable. Under the null hypothesis $H_0:\gamma_k=0$, we have

$$t_{EIV}(\bar{\hat{\gamma}}_k)=\sqrt{T}\,\frac{\hat{\gamma}_k}{s_{EIV}(\hat{\gamma}_k)}=\frac{\hat{\gamma}_k}{s_{EIV}(\hat{\gamma}_k)/\sqrt{T}}\to 0\,.$$

The analysis here holds for both the OLS and GLS estimations.

Proof of Lemma 3:

Under Case 4, consider the OLS estimated risk premium for f_{1t}, where f_{1t} is either a useless factor or one of the multiple proxy factors for a true factor, we have $\hat{\gamma}_1^{OLS}=(b_1'M_{-1}b_1)^{-1}b_1'M_{-1}\bar{R}$. Since the proposed factor model contains all the true factors, the beta matrix $\beta_{-1}=(\beta_2,\cdots,\beta_k)$ could explain the expected returns quite well without β_1. Then, $\sqrt{T}M_{-1}\bar{R}$ is a random variable and centered at zero. Hence, we have

$$
\begin{aligned}
\hat{\gamma}_1^{OLS} &= (b_1'M_{-1}b_1)^{-1}b_1'M_{-1}\bar{R} \\
&= (\sqrt{T}b_1'M_{-1}\sqrt{T}b_1)^{-1}\sqrt{T}b_1'\sqrt{T}M_{-1}\bar{R}.
\end{aligned}
$$

Then $\hat{\gamma}_1^{OLS}$ is a random variable, which is the main difference from models omitting some true factors. The same results hold for GLS estimation.

Proof of Proposition 4:

The proof is conducted on factor 1, which is either a useless factor or one of the multiple proxy factors for a true factor. Given the proposed model contains all the true factors, we have $M_{-1}\bar{R} = O_p(1/T^{-1/2})$, and $\sqrt{T}M_{-1}\bar{R} \to 0$ as $T \to \infty$. The EIV unadjusted t-test statistic is given by:

$$t^2(\hat{\gamma}_1) = \frac{(\sqrt{T}\hat{\gamma}_1)^2}{s^2(\hat{\gamma}_1)} = \frac{T(b_1'M_{-1}\bar{R})^2}{b_1'M_{-1}\hat{V}M_{-1}b_1} \equiv (x'y)^2.$$

Following the proof of Proposition 2(B) in Kan and Zhang (1999a), we define $x = \Delta^{1/2}P'b_1 / (b_1'M_{-1}\hat{V}M_{-1}b_1)^{1/2}$ and $y = \sqrt{T}\Delta^{-1/2}P'\bar{R}$, where $P\Delta P' = M_{-1}VM_{-1}$, and Δ is the matrix of corresponding nonzero eigenvalues. Since we have

$E(x'y) = E(\sqrt{T}b_1'PP'\bar{R}) / (b_1'M_{-1}\hat{V}M_{-1}b_1)^{1/2} = E(\sqrt{T}b_1'M_{-1}\bar{R}) / (b_1'M_{-1}\hat{V}M_{-1}b_1)^{1/2} = 0$,

$Var(x'y) = (b_1'PP'VPP'b_1) / (b_1'M_{-1}VM_{-1}b_1) = b_1'M_{-1}VM_{-1}b_1 / (b_1'M_{-1}\hat{V}M_{-1}b_1) = 1$.

Then $x'y$ is a χ_1^2 variable asymptotically. Given

$$s_{EIV}^2(\hat{\gamma}_1) = (1 + \hat{\gamma}'\Sigma_f^{-1}\hat{\gamma})(s^2(\hat{\gamma}_1) - \Sigma_{f_1}) + \Sigma_{f_1} > s^2(\hat{\gamma}_1),$$

we know that $t^2{}_{EIV}(\bar{\hat{\gamma}}_1) < (x'y)^2$, which means the square of the EIV adjusted t-statistics will be stochastically dominated by a χ_1^2 distribution.

These results are derived with OLS estimation, and it also applies to GLS estimation.

Proof of Proposition 5:

Now we are going to show that $\tilde{C}$ is the consistent estimation of a linear transformation of C^0, and hence $F\tilde{C}$ is a consistent estimation of a linear transformation of real FC^0. Reconsider the minimization problem:

$\min \Sigma_{i=1}^{N}\Sigma_{j=1}^{k}(\hat{B}_{ij}-A_iC_j)^2$, with the normalization that $C'C/k=I_r$, $\tilde{C}$ is $\sqrt{k}$ times the eigenvectors corresponding to the first r largest eigenvalues of the $k\times k$ matrix $\hat{B}'\hat{B}$. Given $\tilde{C}$, define $\tilde{A}'=(\tilde{C}'\tilde{C})^{-1}\tilde{C}'\hat{B}'=\tilde{C}'\hat{B}'/k$.

Consider the minimization problem in the symmetric way, with the normalization that $A'A/N=I_r$, $\bar{A}$ is constructed as $\sqrt{N}$ times the eigenvectors corresponding to the r largest eigenvalues of the $N\times N$ matrix $\hat{B}\hat{B}'$. Given $\bar{A}$, define $\bar{C}=(\bar{A}'\bar{A})^{-1}\hat{B}'\bar{A}=\hat{B}'\bar{A}/N$.

Now, define the consistent estimator of C^0: $\hat{C}=\bar{C}(\bar{C}'\bar{C}/k)^{1/2}$. Using the mathematical identity, we have $\hat{C}=\hat{B}'\tilde{A}/N$ and $\tilde{A}=\hat{B}\tilde{C}/k$. Given that $\hat{B}=B^0+E'F(F'F)^{-1}$ and $\hat{B}'=C^0A^{0\prime}+(F'F)^{-1}F'E$, we have

$$
\begin{aligned}
\hat{C} &= \hat{B}'\tilde{A}/N=\hat{B}'\hat{B}\tilde{C}/Nk \\
&=(C^0A^{0\prime}+(F'F)^{-1}F'E)(A^0C^{0\prime}+E'F(F'F)^{-1})\tilde{C}/Nk \\
&=[C^0A^{0\prime}A^0C^{0\prime}\tilde{C}+(F'F)^{-1}F'EB^0\tilde{C}+ \\
&\quad B^{0\prime}E'F(F'F)^{-1}\tilde{C}+(F'F)^{-1}F'EE'F(F'F)^{-1}\tilde{C}]/Nk
\end{aligned}
$$

Define $H=(A^{0\prime}A^0/N)(C^{0\prime}\tilde{C}/k)$, then,

$$\|H\|\le\left\|A^{0\prime}A^0/N\right\|\left\|C^{0\prime}C^0/k\right\|^{1/2}\left\|\tilde{C}'\tilde{C}/k\right\|^{1/2}=O_p(1).$$

The norm of a matrix A is denoted by $\|A\| = [tr(A'A)]^{1/2}$. Hence,

$$\left\|\hat{C} - C^0 H\right\|^2 \le \left\|(F'F)^{-1} F'EB^0 \tilde{C} / Nk\right\|^2 + \left\|B^{0\prime} E'F(F'F)^{-1} \tilde{C} / Nk\right\|^2 + \left\|(F'F)^{-1} F'EE'F(F'F)^{-1} \tilde{C} / Nk\right\|^2$$

Given $\left\|(F'F/T)^{-1}\right\| = O_p(1)$ and $\left\|\tilde{C}/k\right\| = O_p(1)$, by Lemma 1 of AHW (2009), we have

$$\left\|\hat{C} - C^0 H\right\|^2 = O_p(1/\sqrt{T}) + O_p(1/\sqrt{T}) + O_p(1/T) = O_p(1/\sqrt{T}).$$

Hence, we can see that $\hat{C}$ is the consistent estimation of a linear transformation of C^0, and it follows that $F\tilde{C}$ is a consistent estimation of a linear transformation of real FC^0.

APPENDIX B

PROOF FOR CHAPTER 2

The following two Lemmas are used to prove Theorem 1.

Lemma 1: Under Assumption B and D, for any $k \times p$ ($p \leq k$) matrices A and G such that $\|A\| = O_p(1)$, and $\|G\| = O_p(1)$, we have two conclusions:

(i) $$\frac{1}{NT}\left| tr(A'\mathrm{B}^{d\prime}\ddot{\mathrm{E}}'\dot{F}^d G)\right| = O_p\left(T^{-1/2}\right);$$

(ii) $$\frac{1}{NT^2}\left| tr(A'\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'\dot{F}A)\right| = O_p(T^{-1}).$$

Proof: Assumption B implies

$$\begin{aligned}\left(\frac{1}{\sqrt{N}}\left\|\mathrm{B}^d\right\|\right)^2 &= \frac{1}{N}\left\|\mathrm{B} - 1_N 1_N' \mathrm{B} / N\right\|^2 \leq \frac{1}{N}\|\mathrm{B}\|^2 + \frac{1}{N}\left\|1_N 1_N' / N\right\|^2 \|\mathrm{B}\|^2 \\ &= \frac{2}{N}\|\mathrm{B}\|^2 = \frac{2}{N}\sum_{i=1}^{N}\|\beta_i\|^2 \leq 2c^2\end{aligned};$$

From Assumptions B – D, we obtain

$$\begin{aligned}\left(\frac{1}{\sqrt{NT}}\left\|\ddot{\mathrm{E}}'\dot{F}\right\|\right)^2 &= \frac{1}{NT}\left\|\ddot{\mathrm{E}}'\dot{F}\right\|^2 \\
&= \frac{1}{NT}\left\|\left(\mathrm{E}' - \frac{1}{T}\mathrm{E}'1_T 1_T' - \frac{1}{N}1_N 1_N'\mathrm{E}' + \frac{1}{NT}1_N 1_N'\mathrm{E}'1_T 1_T'\right)'\left(F - \frac{1}{T}1_T 1_T{}' F\right)\right\|^2 \\
&= \frac{1}{NT}\left\|\mathrm{E}'F - \frac{1}{T}\mathrm{E}'1_T 1_T' F - \frac{1}{N}1_N 1_N'\mathrm{E}'F + \frac{1}{NT}1_N 1_N'\mathrm{E}'1_T 1_T' F\right\|^2 \\
&= \frac{1}{NT}\|\mathrm{E}'F\|^2 + \frac{1}{NT}\left\|\frac{1}{T}\mathrm{E}'1_T 1_T' F\right\|^2 + \frac{1}{NT}\left\|\frac{1}{N}1_N 1_N'\right\|^2 \|\mathrm{E}'F\|^2 \\
&\quad + \frac{1}{NT}\left\|\frac{1}{N}1_N 1_N'\right\|^2 \left\|\frac{1}{T}\mathrm{E}'1_T 1_T' F\right\|^2 \\
&\leq \frac{2}{NT}\|\mathrm{E}'F\|^2 + \frac{2}{NT}\left\|\frac{1}{T}\mathrm{E}'1_T 1_T' F\right\|^2 \\
&\leq O_p(1) + \frac{2}{NT}\|\mathrm{E}'1_T\|^2 \left\|\frac{1}{T}1_T' F\right\|^2 = O_p(1) + 2\left(\frac{1}{NT}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\varepsilon_{it}\right\|^2\right)\left\|\bar{f}\right\|^2 = O_p(1),\end{aligned}$$

where 1_T is a $T \times 1$ vector of ones. Thus, $N^{-1/2}\left\|\mathrm{B}^d\right\| = O_p(1)$ and

$(NT)^{-1/2}\left\|\ddot{\mathrm{E}}'\dot{F}\right\| = O_p(1)$. Then, we have (i), because

$$\frac{1}{NT}\left|tr(A'\mathrm{B}^{d\prime}\ddot{\mathrm{E}}'\dot{F}G)\right| \le \frac{1}{\sqrt{T}}\|A\|\|G\|\frac{\left\|\mathrm{B}^d\right\|}{\sqrt{N}}\frac{\left\|\ddot{\mathrm{E}}'\dot{F}\right\|}{\sqrt{NT}} = O_p\left(\frac{1}{\sqrt{T}}\right)\times O_p(1) = O_p\left(\frac{1}{\sqrt{T}}\right).$$

We obtain (ii), because

$$\frac{1}{NT^2}\left|tr(A'\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'F^d A)\right| \le \frac{1}{NT^2}\|AA'\|\left\|\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'\dot{F}\right\| \le \frac{1}{T}\|A\|^2\left[\frac{1}{NT}\left\|\ddot{\mathrm{E}}'\dot{F}\right\|^2\right] = O_p\left(\frac{1}{T}\right).$$

Lemma 2: Suppose that two matrices A and B are symmetric of order p. Then,

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

Proof: See Onatski (2010) or Rao (1973, p. 68).

Proof of Theorem 1: Observe that

$$\frac{1}{NT^2}\dot{F}'\ddot{X}\ddot{X}'\dot{F} = \frac{1}{NT^2}\dot{F}'(\dot{F}\mathrm{B}^{d\prime} + \ddot{\mathrm{E}})(\mathrm{B}^d\dot{F}' + \ddot{\mathrm{E}}')\dot{F}$$
$$= \left(\frac{\dot{F}'\dot{F}}{T}\right)\frac{\mathrm{B}^{d\prime}\mathrm{B}^d}{N}\left(\frac{\dot{F}'\dot{F}}{T}\right) + \left(\frac{\dot{F}'\dot{F}}{T}\right)\frac{\mathrm{B}^{d\prime}\ddot{\mathrm{E}}'\dot{F}}{NT} + \frac{\dot{F}'\ddot{\mathrm{E}}\mathrm{B}^d}{NT}\left(\frac{\dot{F}'\dot{F}}{T}\right) + \frac{\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'\dot{F}}{NT^2}.$$

Thus, we have

$$\frac{\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^d}{N} = \frac{\mathrm{B}^{d\prime}\mathrm{B}^d}{N} + \frac{\mathrm{B}^{d\prime}\ddot{\mathrm{E}}'\dot{F}}{NT}A_T + A_T\frac{\dot{F}'\ddot{\mathrm{E}}\mathrm{B}^d}{NT} + A_T\left(\frac{\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'\dot{F}}{NT^2}\right)A_T,$$

where $A_T = (\dot{F}'\dot{F}/T)^{-1}$ and $\|A_T\| = O_p(1)$ by Assumption A.

Now, let $\hat{\Xi}^l$ be the matrix of the eigenvectors corresponding to the first $l(\leq k)$ largest eigenvalues $\hat{\mu}_{NT,1} \geq \hat{\mu}_{NT,2} \geq \cdots \geq \hat{\mu}_{NT,l}$ of $\hat{B}^{d\prime}\hat{B}^d / N$. Similarly, define Ξ^l for the matrix $B^{d\prime}B^d / N$. For any $l \leq r$, we have

$$\begin{aligned}\Sigma_{j=1}^{l}\hat{\mu}_{NT,j} &= tr\left(\frac{1}{N}\hat{\Xi}^{l\prime}\hat{B}^{d\prime}\hat{B}^d\hat{\Xi}^l\right)\\ &= tr\left(\frac{1}{N}\hat{\Xi}^{l\prime}B^{d\prime}B^d\hat{\Xi}^l\right) + 2tr\left(\frac{1}{NT}\hat{\Xi}^{l\prime}B^{d\prime}\ddot{E}'\dot{F}A_T\hat{\Xi}^l\right)\\ &\quad + tr\left(\frac{1}{NT^2}\hat{\Xi}^{l\prime}A_T\dot{F}'\ddot{E}\ddot{E}'\dot{F}'A_T\hat{\Xi}^l\right)\\ &\leq tr\left(\frac{1}{N}\Xi^{l\prime}B^{d\prime}B^d\Xi^l\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{T}\right)\\ &= \Sigma_{j=1}^{l}\lambda_j\left(\frac{B^{d\prime}B^d}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{T}\right),\end{aligned}$$

by Lemma 1, because $\left\|\hat{\Xi}^l\right\| = O_p(1)$ and $\left\|A_T\hat{\Xi}^l\right\| \leq \|A_T\|\left\|\hat{\Xi}^l\right\| = O_p(1)$. In addition,

$$\begin{aligned}\Sigma_{j=1}^{l}\hat{\mu}_{NT,j} &= tr\left(\frac{1}{N}\hat{\Xi}^{l\prime}\hat{B}^{d\prime}\hat{B}^d\hat{\Xi}^l\right) \geq tr\left(\frac{1}{N}\Xi^{l\prime}\hat{B}^{d\prime}\hat{B}^d\Xi^l\right)\\ &= tr\left(\frac{1}{N}\Xi^{l\prime}B^{d\prime}B^d\Xi^l\right) + 2tr\left(\frac{1}{NT}\Xi^{l\prime}B^{d\prime}\ddot{E}'\dot{F}A_T\Xi^l\right)\\ &\quad + tr\left(\frac{1}{NT^2}\Xi^{l\prime}A_T\dot{F}'\ddot{E}\ddot{E}'\dot{F}A_T\Xi^l\right),\\ &= \Sigma_{j=1}^{l}\lambda_j\left(\frac{B^{d\prime}B^d}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{T}\right)\end{aligned}$$

Since these two results hold for any $l \leq r$, we have

$$\hat{\mu}_{NT,j} = \lambda_j\left(\frac{B^{d\prime}B^d}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{T}\right).$$

Thus, for $1 \leq j \leq r$, we have

$$p\lim_{T\to\infty}\hat{\mu}_{NT,j}=\lambda_j\left(\frac{\mathrm{B}^{d\prime}\mathrm{B}^d}{N}\right)>0\,.$$

Next, since we have $rank(\mathrm{B}^d)=r$, we can rewrite $\mathrm{B}^d=AC'$, where A and C are $N\times r$ and $k\times r$ matrices, respectively, and $rank(A)=rank(C)=r$. Let $P(A)=A(A'A)^{-1}A'$, and $Q(A)=1-P(A)$. Using the fact that $P(A)\mathrm{B}^d=\mathrm{B}^d$ and $Q(A)\mathrm{B}^d=0$, we can easily show

$$\frac{\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^d}{N}=\frac{\hat{\mathrm{B}}^{d\prime}[P(A)+Q(A)]\hat{\mathrm{B}}^d}{N}=\frac{\hat{\mathrm{B}}^{d\prime}P(A)\hat{\mathrm{B}}^d}{N}+A_T\left(\frac{\dot{F}'\ddot{\mathrm{E}}Q(A)\ddot{\mathrm{E}}'\dot{F}}{NT^2}\right)A_T\,,$$

Thus, for $j=1,\cdots,k-r$, we have

$$\lambda_{r+j}\left(\frac{\hat{\mathrm{B}}^{d\prime}\hat{\mathrm{B}}^d}{N}\right)\le\lambda_{r+1}\left(\frac{\hat{\mathrm{B}}^{d\prime}P(A)\hat{\mathrm{B}}^d}{N}\right)+\lambda_j\left(A_T\left(\frac{\dot{F}'\ddot{\mathrm{E}}Q(A)\ddot{\mathrm{E}}'\dot{F}}{NT^2}\right)A_T\right)$$
$$\le 0+\lambda_j\left(A_T\left(\frac{\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'\dot{F}}{NT^2}\right)A_T\right)\le\frac{1}{NT^2}\left|tr(A_T'\dot{F}'\ddot{\mathrm{E}}\ddot{\mathrm{E}}'FA_T)\right|=O_p(T^{-1}),$$

where the first inequality is due to Lemma 2. Thus, for any $1\le r+1\le j'\le k$, $\hat{\mu}_{NT,j'}=O_p\left(1/T\right)$. Notice that the second part holds even for $r=0$, which is the "no-factor" case.

Proof of Theorem 2: For $1\le j\le r$, $p\lim_{T\to\infty}\hat{\mu}_{NT,j}>0$, because $rank(\mathrm{B}^{d\prime}\mathrm{B}^d/N)=r$. Since $g(T)\to 0$, $\lim_{T\to\infty}\Pr[\hat{\mu}_{NT,j}>g(T)\,|\,j\le r]=1$. For $0\le r<j\le k$, $p\lim_{T\to\infty}T\hat{\mu}_{NT,j}<\infty$. Thus, $\lim_{T\to\infty}\Pr(\hat{\mu}_{NT,j}<g(T)\,|\,0\le r<j\le k)$

$= \lim_{T\to\infty} \Pr(T\hat{\mu}_{NT,j} < Tg(T) \mid 0 \le r < j \le k) = 1$, because $Tg(T) \to \infty$ and $T\hat{\mu}_{NT,j}$

$= O_p(1)$.