Finance Research Letters 15 (2015) 31–40

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Fama–MacBeth two-pass regressions: Improving risk premia estimates

Jushan Bai[a], Guofu Zhou[b,*]

[a] Columbia University, United States
[b] Washington University in St. Louis, United States

ARTICLE INFO

Article history:
Received 21 April 2015
Accepted 6 August 2015
Available online 12 August 2015

JEL classification:
C21
C11
G12
G11

Keywords:
Cross-section
Fama–MacBeth
Risk premia
Asset pricing models

ABSTRACT

In this paper, we provide the asymptotic theory for the widely used Fama and MacBeth (1973) two-pass risk premia estimates in the usual case of a large number of assets. We demonstrate analytically and using simulations that the standard OLS and GLS estimators can contain large bias when the time series sample size is small, but our proposed OLS and GLS estimators can reduce the bias significantly.

1. Introduction

A fundamental problem in finance is to explain cross-sectional differences in asset expected returns. The CAPM of Sharpe (1964), Lintner (1965), and Black (1972) and various factor models as well as intertemporal/consumption models show that expected returns should be a linear function of asset betas with respect to economic fundamentals. There is a large literature that examines this linear asset pricing relationship (see, e.g., Jagannathan et al., 2010, for a survey). One of the most

* Corresponding author at: Olin School of Business, Washington University, St. Louis, MO 63130, United States. Tel.: +1 314 935 6384.

E-mail address: zhou@wustl.edu (G. Zhou).

http://dx.doi.org/10.1016/j.frl.2015.08.001
1544-6123/

widely used econometric methodologies is the Fama and MacBeth (1973) two-pass regression. This procedure is used not only in asset pricing, but also in many other areas of finance and accounting.[1] Shanken (1992) provides an early study on its econometric properties. Recently, Shanken and Zhou (2007), Kan et al. (2013), among others, provide further analytical results and simulation evidence. However, except for Shanken (1992), almost all of the existing studies have focused on the case when the times series sample size, T, is larger than the number of assets, N. But in practice, most of the applications occur in the case of $N > T$.

In this paper, we provide the asymptotic theory and simulation evidence for the widely used FM OLS (Fama–MacBeth ordinary least squares) estimator when $N > T$. We show that the convergence of the FM OLS estimator depends crucially on T. This is contrary to the widely held belief that the larger the N, the greater accuracy the risk premium estimator. We demonstrate that this is not true. First, there is an order of $1/T$ bias for the estimator. Second, the asymptotic variance has two components, one of which is of order $1/T$ and the other is of $1/(NT)$. As a result, even if N is large, the first term will dominate the estimation error, and this term remains important even when N goes to infinity.

Because of the relatively small sample sizes used in empirical studies, we propose new OLS and GLS risk premia estimators that account for the bias caused by small T. As expected, the new estimators perform well in simulations for common sample sizes of $T = 60$ and $T = 120$. The reason that many or most cross-section regressions are run at these levels of relatively small T is the concern of parameter stability over time.

This article is organized as follows. In Section 2, we review the FM regression, and then, in Section 3, we present the asymptotic theory for the OLS estimator. In Section 4, we study the GLS estimator. In Section 5, we provide simulation evidence on the usefulness of the bias adjusted estimators. In Section 6, we conclude.

2. Fama–MacBeth regressions

In this section, we review the standard FM regression and the associated OLS estimator.

We assume that asset returns are governed by a multi-factor model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \;\; t = 1, \ldots, T, \tag{1}$$

where

R_{it} = the return on asset i in period t $(1 \le i \le N)$,
f_{jt} = the realization of the j-th factor in period t $(1 \le j \le K)$,
ϵ_{it} = the disturbances or random errors,
N = the number of assets, and T is the number of time-series observations.

In vector and matrix notation, we write the above model as

$$R_t = \alpha + \beta_1 f_{1t} + \cdots \beta_K f_{kt} + \epsilon_t = \alpha + B f_t + \epsilon_t,$$

where $R_t = (R_{1t}, \ldots, R_{Nt})'$ is the N-vector of asset returns; $\beta_1, \ldots, \beta_K$ are N-vectors of the multiple-regression betas; $B = (\beta_1, \beta_2, \ldots, \beta_K)$ is an $N \times K$ matrix; $f_t = (f_{1t}, \ldots, f_{Kt})'$, and $\alpha = (\alpha_1, \ldots, \alpha_N)'$. Let $\epsilon = (\epsilon_1, \ldots, \epsilon_T)$ be the $N \times T$ matrix of errors.

Like most studies, we maintain the standard assumption that the disturbances are independent of the common factors and that the disturbances are independent and identically distributed (iid) over time with mean zero and a nonsingular residual covariance matrix $\Sigma = E(\epsilon_t \epsilon_t')$, although this assumption can be relaxed with more complex formulas. The iid assumption is consistent with the behavior of stock returns data which have little correlations, and is also consistent with the underlying theoretical factor models which often are one-period models. Pedagogically, the iid assumption simplifies both notations and discussions, making it easy for readers at large to understand and apply the results. In what follows, we denote $E(\epsilon_{it}^2) = \sigma_{ii}^2$, and we also allow $E(\epsilon_{it}\epsilon_{jt}) \neq 0$, but the cross-sectional correlation must be weak so that $(NT)^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{it} = O_p(1)$, i.e., the residuals are stochastically bounded.

[1] Applications of the procedure in recent years can be found in well over 9,690 papers that cite Fama and MacBeth (1973), as complied by Google Scholar.

The asset pricing hypothesis underlying the standard two-pass procedure is

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1 + \cdots + \gamma_K \beta_K, \tag{2}$$

where $E[R_t]$ is the N-vector of expected returns on the assets, and $\gamma_1, \ldots, \gamma_1$ are the risk premia. In the first stage of the two-pass procedure, estimates of the betas are obtained by applying OLS to Eq. (1) for each asset. Let $\hat{B} = (\hat{\beta}_1, \ldots, \hat{\beta}_K)$ be the resulting $N \times K$ matrix of the OLS beta estimates. In the second stage, one runs a cross-sectional regression of R_t on $\hat{X} = [1_N, \hat{B}]$ in each period t to get an estimator of $\Gamma = (\gamma_0, \gamma_1, \ldots, \gamma_K)'$,

$$\widehat{\Gamma}_t = (\hat{X}'\hat{X})^{-1}\hat{X}'R_t. \tag{3}$$

The average,

$$\widehat{\Gamma} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\Gamma}_t = (\hat{X}'\hat{X})^{-1}\hat{X}'\overline{R}, \tag{4}$$

is taken as the final estimator of Γ, where $\overline{R}$ is an N-vector of sample means of the asset returns. Eq. (4) provides the widely used FM OLS risk premia estimator by using the FM two-pass procedure.

3. Asymptotic distribution of OLS

In this section, we provide the asymptotic distribution of the OLS estimator, the bias-corrected estimator and the associated asymptotic covariances matrix which is useful for computing the standard errors and the t-tests.

Let $F = (f_1, f_2, \ldots, f_T)'$ represent a $T \times K$ matrix of factor observations and $\dot{F} = F - 1_T\overline{f}'$ be the demeaned data with $\overline{f} = \frac{1}{T}\sum_{s=1}^{T} f_s = F'1_T/T$ and $1_T = (1,1,\ldots,1)'$ $(T \times 1)$. We assume f_t is stationary so that $E(\overline{f}) = E(f_t)$, which will be denoted as $E(f)$. For notational simplicity, we partition Γ as

$$\Gamma = \begin{bmatrix} \gamma_0 \\ \gamma \end{bmatrix},$$

where $\gamma = (\gamma_1, \ldots, \gamma_K)'$. Then we have

Theorem 1. *Under the model structure of Section 2, the two-pass OLS estimator of Γ has the following asymptotic representation for large N,*

$$\begin{aligned}\widehat{\Gamma} - \Gamma = {} & \Omega^{-1}\begin{bmatrix} 0 \\ -\overline{\sigma}^2(\dot{F}'\dot{F}/T)^{-1}\gamma/T \end{bmatrix} + \left[I_{k+1} - \Omega^{-1}\begin{bmatrix} 0 & 0 \\ 0 & \overline{\sigma}^2(\dot{F}'\dot{F}/T)^{-1}/T \end{bmatrix}\right]\begin{bmatrix} 0 \\ \overline{f} - Ef \end{bmatrix} \\ & + \frac{1}{NT}\Omega^{-1}\begin{bmatrix} -1_N' \in \dot{F}(\dot{F}\dot{F}/T)^{-1}\gamma + 1_N' \in 1_T \\ -B' \in \dot{F}(\dot{F}'\dot{F}/T)^{-1}\gamma + B' \in 1_T \end{bmatrix} + o_p(\frac{1}{\sqrt{NT}}),\end{aligned} \tag{5}$$

where

$$\Omega = \frac{1}{N}\begin{bmatrix} 1_N'1_N & 1_N'B \\ B'1_N & B'B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \overline{\sigma}^2(\dot{F}'\dot{F}/T)^{-1}/T \end{bmatrix}, \quad \overline{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_{ii}^2.$$

There are three terms in the asymptotic representation. The first term is a bias term whose order of magnitude is $1/T$. This bias can be removed through a bias correction. The second is a zero-mean random variable with order of $O_p(T^{-1/2})$ because $\overline{f} - E(f) = O_p(T^{-1/2})$. The third term also has a zero mean, but its order of magnitude is $O_p((NT)^{-1/2})$. This follows from, for example, $1_N \in 1_T = \sum_{i=1}^{N}\sum_{t=1}^{T} \in_{it}$, and $(N^{-1}T^{-1})\sum_{i=1}^{N}\sum_{t=1}^{T} \in_{it} = O_p((NT)^{-1/2})$. Under large N, this term is dominated by $\overline{f} - E(f) = O_p(T^{-1})$. We keep this term so that the result is still valid when T is large or fixed. The last two terms contribute to the variation of the OLS estimator.

Theorem 1 says that the OLS estimator must be asymptotically consistent for large N and large T. It does not matter which of them is larger. That is, N can be much larger than T and vice versa, provided that both are large. However, no matter how large N is, and even with a bias correction of the first term, the second term on the right-hand side of the theorem says that the consistency to Γ requires $\overline{f} - Ef \to^{p} 0$, which is not possible theoretically unless we assume T is large. However, the impact of this term is usually small in practice.

In particular, Theorem 1 invalidates the common belief that one can get an accurate estimate of the risk premia using a great number of assets even when there is a small time series sample size (say $T = 60$). This is because the accuracy is critically depends on T as stated in Theorem 1. We can also understand this point intuitively. The risk premia imply a time series relation because the expected returns are about the future returns and a test of this relation requires time series observations. Therefore, statistically, a relatively large time series sample size is needed to assess the relation due to volatilities of the returns. Interestingly, in contrast, when we estimate the factor realizations using the principle components method, as shown by Connor and Korajczyk (1988), we can estimate the realizations as accurate as possible (up to a rotation) as long as the number of assets is enough, even with a time series sample size of one at any time t. This is because here the cross-section relation is sufficient to identify the factors that affect the returns cross-sectionally.

Our case of interest in this paper is when N can be much larger than T, as often in applications. In this case, the first term in the asymptotic expansion (the major bias term), though of order $1/T$, can be significant in small samples while the second term is negligible. Hence, we propose a bias-corrected estimator,

$$\widehat{\Gamma}^{*} = \widehat{\Gamma} + \Omega^{-1}\begin{bmatrix} 0 \\ -\overline{\sigma}^{2}(\dot{F}'\dot{F}/T)^{-1}\hat{\gamma}/T \end{bmatrix}, \tag{6}$$

where $\widehat{\Gamma} = (\hat{\gamma}_0, \hat{\gamma}')'$ is the usual OLS estimator. The new estimator adjusts the bias from the first term of the asymptotic representation arising from a relatively small T. In practice, we also replace Ω and $\overline{\sigma}^2$ by their standard sample estimates that are easy to compute. We can show that the asymptotic variance of the bias-corrected estimator is the same as $\widehat{\Gamma}$.

Based on Theorem 1, the asymptotic covariance matrix of $\widehat{\Gamma}^{*}$ when both N and T are large is

$$\Pi = \frac{1}{T}\Pi_1 + \frac{1}{NT}\Pi_2, \tag{7}$$

where

$$\Pi_1 = H\begin{bmatrix} 0 & 0 \\ 0 & \Sigma_f \end{bmatrix}H', \quad H = I_{k+1} - \Omega^{-1}\begin{bmatrix} 0 & 0 \\ 0 & -\overline{\sigma}^{2}(\dot{F}'\dot{F}/T)^{-1}/T \end{bmatrix}$$

and

$$\Pi_2 = \Omega^{-1}\begin{bmatrix} (1+\gamma'\Sigma_f^{-1}\gamma)1_N'\Sigma 1_N/N & (1-\gamma'\Sigma_f^{-1}\gamma)1_N'\Sigma B/N \\ (1-\gamma'\Sigma_f^{-1}\gamma)B'\Sigma 1_N/N & (1+\gamma'\Sigma_f^{-1}\gamma)B'\Sigma B/N \end{bmatrix}\Omega^{-1},$$

with $\Sigma_f = var(f_t)$. In practice, not only is $\widehat{\Gamma}^{*}$ of interest, its associated asymptotic standard errors are important, and can be easily computed from the above equation by plugging in the sample estimates of the parameters.

4. Asymptotic distribution of GLS

In applications, it will be useful to have alternative estimates of the risk premia. As demonstrated by Shanken (1992), the use of the generalized least-squares (GLS) in the second-step regression of the two pass yields a GLS estimator of the risk premia,

$$\widehat{\Gamma}^{GLS} = (\hat{X}'\widehat{\Sigma}^{-1}\hat{X})^{-1}\hat{X}'\widehat{\Sigma}^{-1}\overline{R}, \tag{8}$$

where $\widehat{\Sigma}$ is a consistent estimator of Σ and is assumed invertible.

In the case when $T > N$ or much larger, the standard sample covariance matrix can be used as $\widehat{\Sigma}$. However, our interest is in the case when $N > T$. While the sample covariance matrix is still a consistent estimator of Σ (under large N and T), it is not invertible. Clearly certain structures have to be imposed on Σ to obtain a consistent and yet invertible estimator. Under certain factor structure assumption on Σ, Bai (2011), Fan et al. (2013), among others, provide estimators of Σ that are invertible. With any of these estimators, the above GLS estimator can be implemented in practice.

For the GLS estimator, we have an asymptotic expression similar to that of the OLS, i.e.,

Theorem 2. *Under some regularity conditions, the GLS estimator has the asymptotic representation for large N,*

$$\widehat{\Gamma}^{GLS} - \Gamma = \Phi^{-1}\begin{bmatrix} 0 \\ -(\dot{F}'\dot{F}/T)^{-1}\gamma/T \end{bmatrix} + \left[I_{k+1} - \Phi^{-1}\begin{bmatrix} 0 & 0 \\ 0 & (\dot{F}'\dot{F}/T)^{-1}/T \end{bmatrix}\right]\begin{bmatrix} 0 \\ \bar{f} - Ef \end{bmatrix}$$
$$+ \frac{1}{NT}\Phi^{-1}\begin{bmatrix} -1_N'\Sigma^{-1}\in \dot{F}(\dot{F}\dot{F}/T)^{-1}\gamma + 1_N'\Sigma^{-1}\in 1_T \\ -B'\Sigma^{-1}\in \dot{F}(\dot{F}'\dot{F}/T)^{-1}\gamma + B'\Sigma^{-1}\in 1_T \end{bmatrix} + o_p\left(\frac{1}{\sqrt{NT}}\right), \tag{9}$$

where

$$\Phi = \frac{1}{N}\begin{bmatrix} 1_N'\Sigma^{-1}1_N & 1_N'\Sigma^{-1}B \\ B'\Sigma^{-1}1_N & B'\Sigma^{-1}B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & (\dot{F}'\dot{F}/T)^{-1}/T \end{bmatrix}.$$

The GLS estimator also has a bias of order $1/T$, characterized by the first term. In comparison to Theorem 1, there are only two differences in the asymptotic expression. First, the term $\overline{\sigma}^2$ disappears in the GLS case because it is taken care of by the weighting matrix now. Second, the earlier Ω is replaced now by Φ. Other than these differences, the asymptotic interpretation is the same as before. In particular, we can define the adjusted GLS estimator similarly. The asymptotic covariance matrix is analogous to (7) with obvious adjustment for Π_1. For Π_2, we replace Ω by Φ, and we replace Σ by Σ^{-1}.

5. Simulations

To see whether the adjusted OLS and GLS make any difference, we provide a comparison of the FM estimator with our new ones in simulations.

Following Shanken and Zhou (2007), we set $\gamma_0 = 0.0833\%$ and $\gamma_1 = 0.6667\%$ and calibrate the model parameters based on data of Fama and French (1993) 25 portfolios when $N = 25$. When $N > 25$, we randomly select parameters around the previous ones. Then, based on the calibrated parameters, we can simulate 10,000 data sets from the standard market model,

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \in_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

for various T and N.

Table 1 reports the results. Consider first the case of $T = 60$. As expected, the OLS estimates do not change much as the number of assets goes from $N = 500$ to $N = 4000$. The percentage errors are about −20%. So the OLS estimates tend to underestimate the true risk premium with sizable errors. The adjusted OLS (denoted as OLS-Adj in the table) performs clearly much better. It reduces the bias to about only 5% in absolute value. The GLS estimator also underestimates the true premium, and only performs marginally better than the OLS estimator. In contrast, the adjusted GLS performs well, and its relatively small bias is similar to the adjusted OLS method.

Consider next the case of $T = 120$. As the sample size increases, the bias of both the OLS and GL estimates decline by nearly 50%, but they are still about 10% off the true value. In contrast, the adjusted estimates are very close to the true value. Overall, our simulation evidence suggests that the adjusted estimators are valuable for conducting proper inference when the time series sample is relatively small.

Table 1
The table reports the average risk premium estimate, its percentage error, and root-mean-square error (all in percent) over 10,000 simulated data sets. The data-generating process is the standard market model,

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T,$$

with normally distributed residuals and factors, and the asset pricing restrictions are

$$H_0: \quad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length, and N is the number of assets. The true value for the factor risk premium is $\gamma_1 = 0.6667\%$ and the zero-beta intercept is $\gamma_0 = 0.0833\%$. The estimation methods are the Fama–MacBeth OLS and the GLS as well as their adjustments accounting for the small sample size T.

Methods	$N = 500$	$N = 750$	$N = 1000$	$N = 1500$	$N = 2000$	$N = 4000$
Panel A: T = 60						
OLS	0.5379	0.5311	0.5382	0.5243	0.5327	0.5167
	−19%	-20%	-19%	-21%	-20%	-22%
	(0.5032)	(0.4897)	(0.4936)	(0.4837)	(0.4942)	(0.4816)
OLS-Adj	0.6353	0.6391	0.6399	0.6358	0.6350	0.6323
	−5%	−4%	−4%	−5%	−5%	−5%
	(0.5750)	(0.5662)	(0.5669)	(0.5607)	(0.5676)	(0.5603)
GLS	0.5588	0.5480	0.5548	0.5428	0.5506	0.5373
	−16%	−18%	−17%	−19%	−17%	−19%
	(0.5160)	(0.4979)	(0.5016)	(0.4929)	(0.5041)	(0.4920)
GLS-Adj	0.6398	0.6425	0.6432	0.6406	0.6380	0.6376
	−4%	−4%	−4%	−4%	−4%	−4%
	(0.5781)	(0.5669)	(0.5670)	(0.5630)	(0.5689)	(0.5634)
Panel B: T = 120						
OLS	0.5931	0.5860	0.5805	0.5874	0.5840	0.5933
	−11%	−12%	−13%	−12%	−12%	−11%
	(0.3842)	(0.3791)	(0.3802)	(0.3818)	(0.3808)	(0.3822)
OLS-Adj	0.6525	0.6543	0.6496	0.6568	0.6538	0.6595
	−2%	−2%	−3%	−1%	−2%	−1%
	(0.4151)	(0.4136)	(0.4147)	(0.4179)	(0.4161)	(0.4169)
GLS	0.6058	0.5972	0.5903	0.5981	0.5967	0.6042
	−9%	−10%	−11%	−10%	−10%	−9%
	(0.3883)	(0.3829)	(0.3826)	(0.3851)	(0.3846)	(0.3856)
GLS-Adj	0.6544	0.6561	0.6509	0.6581	0.6559	0.6611
	−2%	−2%	−2%	−1%	−2%	−1%
	(0.4145)	(0.4137)	(0.4137)	(0.4170)	(0.4156)	(0.4162)

6. Conclusion

The Fama–MacBeth two-pass estimation method has been widely used in finance and accounting to examine various factors and their pricing power in the cross-section of asset returns. In contrast to the widely held belief that the number of assets alone determines the accuracy of the risk premia estimates, we show that the time series sample size is critical too. Moreover, we provide adjusted OLS and GLS estimators of the risk premia that reduce the biases significantly which are caused by relatively small time series sample sizes.

Our work here raises many interesting issues. First, it will be useful to provide simple adjusted estimators to account for more general processes with possibly more complex panel structures, and to analyze new model specification tests. Second, extensive simulations may be conducted to examine the small (time series) sample behavior of both the estimators and tests. Third, a comprehensive empirical applications of the proposed methods may be carried out to see to what extent the new econometric tools alter the conclusions of existing studies. These important issues are beyond the scope of this article, and we leave them for future research.

Acknowledgement

We are grateful to seminar participants at Shanghai Advanced Institute of Finance, Southwestern University of Finance and Economics and Washington University in St Louis, conference participants at the 2014 the International Symposium on Financial Engineering and Risk Management, and especially to an anonymous referee for insightful and detailed comments that have substantial improved the paper.

Appendix A

A.1. Proof of Theorem 1

For model (1), the first pass estimator $\hat{B}$ for B satisfies

$$\hat{B} - B = \in \dot{F}(\dot{F}\dot{F})^{-1}. \tag{A1}$$

where $\in = (\in_1, \ldots, \in_T)$ is $N \times T$.

From $R_t = \alpha + Bf_t + \in_t = \alpha + \hat{B}f_t + \in_t + (B - \hat{B})f_t$, we have $\overline{R} = \alpha + \hat{B}\overline{f} + (B - \hat{B})\overline{f} + \overline{\in}$. Together with

$$\alpha = 1_N\gamma_0 + B(\gamma - Ef),$$

we have

$$\overline{R} = 1_N\gamma_0 + B(\gamma - Ef) + \hat{B}\overline{f} + (B - \hat{B})\overline{f} + \overline{\in} = 1_N\gamma_0 + \hat{B}(\gamma + \overline{f} - Ef) + (B - \hat{B})(\gamma + \overline{f} - Ef) + \overline{\in}.$$

From $\widehat{\Gamma} = (\hat{X}'\hat{X})^{-1}\hat{X}'\overline{R}$ with $\hat{X} = [1_N, \hat{B}]$, it follows that

$$\widehat{\Gamma} = \Gamma + \begin{bmatrix} 0 \\ \overline{f} - Ef \end{bmatrix} + (\hat{X}'\hat{X})^{-1}\left[\hat{X}'(B - \hat{B})(\gamma + \overline{f} - Ef) + \hat{X}'\overline{\in}\right].$$

Consider

$$\frac{1}{N}\hat{X}'\hat{X} = \begin{bmatrix} 1 & \frac{1}{N}1_N'\hat{B} \\ \frac{1}{N}\hat{B}'1_N & \frac{1}{N}\hat{B}'\hat{B} \end{bmatrix}.$$

For its off-diagonal element

$$\frac{1}{N}1_N'\hat{B} = \frac{1}{N}1_N'B + \frac{1}{N}1_N'(\hat{B} - B),$$

the second term is dominated by the first, and is ignorable. Next,

$$\frac{1}{N}\hat{B}'\hat{B} = \frac{1}{N}B'B + \frac{1}{N}B'(\hat{B} - B) + \frac{1}{N}(\hat{B} - B)'B + \frac{1}{N}(\hat{B} - B)'(\hat{B} - B).$$

The two middle terms on the right hand side are dominated by the first and are negligible. The last term, however, is not negligible under fixed T. In fact, its order of magnitude is $1/T$. We shall show that

$$\frac{1}{N}(\hat{B} - B)'(\hat{B} - B) = \overline{\sigma}^2(\dot{F}'\dot{F}/T)^{-1}/T + o_p(1/T). \tag{A2}$$

To see this, by (A1),

$$\frac{1}{N}(\hat{B} - B)'(\hat{B} - B) = \frac{1}{N}(\dot{F}'\dot{F})^{-1}\dot{F}'\in' \in \dot{F}(\dot{F}'\dot{F})^{-1}.$$

Under the iid assumption over t, the expected value $\frac{1}{N}E(\in' \in) = \overline{\sigma}^2 I_T$. The deviation from the expected value is negligible. This proves (A2). In summary, we have shown that

$$\frac{1}{N}\hat{X}'\hat{X} = \begin{bmatrix} 1 & \frac{1}{N}1_N'\hat{B} \\ \frac{1}{N}\hat{B}'1_N & \frac{1}{N}\hat{B}'\hat{B} \end{bmatrix} = \frac{1}{N}\begin{bmatrix} 1_N'1_N & 1_N'B \\ B'1_N & B'B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \overline{\sigma}^2(\dot{F}'\dot{F}/T)^{-1}/T \end{bmatrix} + o_p(1).$$

Next consider

$$\frac{1}{N}\left[\hat{X}'(B-\hat{B})(\gamma+\bar{f}-Ef)+\hat{X}'\bar{\epsilon}\right]=\frac{1}{N}\begin{bmatrix}1_N'(B-\hat{B})(\gamma+\bar{f}-Ef)+1_N'\bar{\epsilon}\\ \hat{B}'(B-\hat{B})(\gamma+\bar{f}-Ef)+\hat{B}'\bar{\epsilon}\end{bmatrix}. \tag{A3}$$

For the first component of (A3), using (A1) and noting $\bar{\epsilon}=\epsilon 1_T/T$, we have

$$\frac{1}{N}\left[1_N'(B-\hat{B})(\gamma+\bar{f}-Ef)+1_N'\bar{\epsilon}\right]=-\frac{1}{N}1_N'\epsilon\dot{F}\left(\dot{F}'\dot{F}\right)^{-1}\gamma+\frac{1}{N}1_N'\epsilon 1_T+o_p\left(\frac{1}{\sqrt{NT}}\right),$$

where we have used the fact that $\frac{1}{N}1_N'(B-\hat{B})(\bar{f}-Ef)$ is ignorable.

Next consider the second component of (A3). Note first that

$$\frac{1}{N}\hat{B}'(B-\hat{B})=-\frac{1}{N}(\hat{B}-B)'(\hat{B}-B)+\frac{1}{N}B'(B-\hat{B})=-\bar{\sigma}^2\left(\dot{F}'\dot{F}/T\right)^{-1}/T+\frac{1}{N}B'(B-\hat{B}),$$

where the second equality follows from (A2) (the high order term is ignored). Thus,

$$\begin{aligned}\frac{1}{N}\hat{B}'(B-\hat{B})(\gamma+\bar{f}-Ef)&=\left[\frac{1}{N}B'(B-\hat{B})-\bar{\sigma}^2\left(\dot{F}'\dot{F}/T\right)^{-1}/T\right](\gamma+\bar{f}-Ef)\\&=-\bar{\sigma}^2\left(\dot{F}'\dot{F}/T\right)^{-1}/T(\gamma+\bar{f}-Ef)+\frac{1}{N}B'\epsilon\dot{F}\left(\dot{F}'\dot{F}\right)^{-1}\gamma,\end{aligned}$$

where we have ignored the term $\frac{1}{N}B'(B-\hat{B})(\bar{f}-Ef)$. Finally,

$$\frac{1}{N}\hat{B}\bar{\epsilon}=\frac{1}{N}(\hat{B}-B)'\bar{\epsilon}+\frac{1}{N}B'\bar{\epsilon}.$$

The first term on the right hand side is negligible, and the second term is equal to $\frac{1}{N}B'\bar{\epsilon}=\frac{1}{NT}B'\epsilon 1_T$. Hence, the second component of (A3) is

$$\frac{1}{N}\left[\hat{B}'(B-\hat{B})(\gamma+\bar{f}-Ef)+\hat{B}'\bar{\epsilon}\right]=-\bar{\sigma}^2\left(\dot{F}'\dot{F}/T\right)^{-1}/T(\gamma+\bar{f}-Ef)+\frac{1}{N}B'\epsilon\dot{F}\left(\dot{F}'\dot{F}\right)^{-1}\gamma+\frac{1}{NT}B'\epsilon 1_T,$$

plus a negligible high order term. Summarizing results, we obtain Theorem 1. □

A.2. Proof of Theorem 2

Similar to the OLS case, we can write $\widehat{\Gamma}$ as

$$\widehat{\Gamma}=\Gamma+\begin{bmatrix}0\\ \bar{f}-Ef\end{bmatrix}+\left(\hat{X}'\widehat{\Sigma}^{-1}\hat{X}\right)^{-1}\left[\hat{X}'\widehat{\Sigma}^{-1}(B-\hat{B})(\gamma+\bar{f}-Ef)+\hat{X}'\widehat{\Sigma}^{-1}\bar{\epsilon}\right].$$

It can be shown that

$$\frac{1}{N}\hat{X}'\widehat{\Sigma}^{-1}\hat{X}=\frac{1}{N}\begin{bmatrix}1_N'\Sigma^{-1}1_N & 1_N'\Sigma^{-1}B\\ B'\Sigma^{-1}1_N & B'\Sigma^{-1}B\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & \frac{1}{T}\left(\dot{F}'\dot{F}/T\right)^{-1}\end{bmatrix}+o_p(1).$$

Consider now

$$\frac{1}{N}\left[\hat{X}'\widehat{\Sigma}^{-1}(B-\hat{B})(\gamma+\bar{f}-Ef)+\hat{X}'\widehat{\Sigma}^{-1}\bar{\epsilon}\right]=\frac{1}{N}\begin{bmatrix}1_N'\widehat{\Sigma}^{-1}(B-\hat{B})(\gamma+\bar{f}-Ef)+1_N'\widehat{\Sigma}^{-1}\bar{\epsilon}\\ \hat{B}'\widehat{\Sigma}^{-1}(B-\hat{B})(\gamma+\bar{f}-Ef)+\hat{B}'\widehat{\Sigma}^{-1}\bar{\epsilon}\end{bmatrix}. \tag{A4}$$

We examine first the first component. Adding and subtracting the terms, we have

$$\frac{1}{N}1_N'\widehat{\Sigma}^{-1}(B-\hat{B})=\frac{1}{N}1_N'\Sigma^{-1}(B-\hat{B})+\frac{1}{N}1_N'\left(\widehat{\Sigma}^{-1}-\Sigma^{-1}\right)(B-\hat{B}).$$

It can be shown that the second term is dominated by the first (the technical details are non-trivial for this claim, see, for example, Bai and Liao, 2013). For the first term, we have

$$\frac{1}{N}1_N'\Sigma^{-1}(B-\hat{B}) = -\frac{1}{N}1_N'\Sigma^{-1}\in\dot{F}(\dot{F}\dot{F})^{-1} = \frac{1}{\sqrt{NT}}O_p(1).$$

In its product with $\gamma+\overline{f}-Ef$, we can ignore $f-Ef$. Hence we keep the term

$$\frac{1}{N}1_N'\Sigma^{-1}(B-\hat{B})\gamma = -\frac{1}{N}1_N'\Sigma^{-1}\in\dot{F}(\dot{F}\dot{F})^{-1}\gamma.$$

Next,

$$\frac{1}{N}1_N'\Sigma^{-1}\overline{\in} = \frac{1}{NT}1_N'\Sigma^{-1}\in 1_T = \frac{1}{\sqrt{NT}}O_p(1),$$

which is of the same order as the preceding one. So the first component of (A4) is equal to the summation of the preceding two terms

$$-\frac{1}{N}1_N'\Sigma^{-1}\in\dot{F}(\dot{F}\dot{F})^{-1}\gamma + \frac{1}{NT}1_N'\Sigma^{-1}\in 1_T$$

plus a term that is negligible.

Consider next the second component of (A4). We study first $\hat{B}'\widehat{\Sigma}^{-1}(B-\hat{B})$. We have

$$\hat{B}'\widehat{\Sigma}^{-1}(B-\hat{B}) = -(B-\hat{B})'\Sigma^{-1}(B-\hat{B}) + B'\Sigma^{-1}(B-\hat{B}) + \hat{B}'(\widehat{\Sigma}^{-1}-\Sigma^{-1})(B-\hat{B}). \tag{A5}$$

The last term is negligible (again the proof of this can be demanding, Bai and Liao, 2013; We can make this part as our regularity conditions. But when Σ is diagonal and is estimated based on residuals, this claim is easy to prove). Then, the first term of (A5) on the right hand side is equal to

$$-(\dot{F}'\dot{F})^{-1}\dot{F}'\in'\Sigma^{-1}\in\dot{F}(\dot{F}'\dot{F})^{-1}.$$

The $T\times T$ matrix $\in'\Sigma^{-1}\in$ has a non-zero mean. Under the iid assumption over $t, \frac{1}{N}E(\in'\Sigma^{-1}\in) = I_T$. Thus the preceding expression is equal to $(\dot{F}'\dot{F})^{-1}+o(1/T)$. Therefore, the first term of (A5), divided by N, is

$$-\frac{1}{N}(B-\hat{B})'\Sigma^{-1}(B-\hat{B})] = -(\dot{F}'\dot{F})^{-1}+o_p(1/T).$$

As a result,

$$\frac{1}{N}(B-\hat{B})'\widehat{\Sigma}^{-1}(B-\hat{B})[\gamma+f-Ef] = (\dot{F}'\dot{F})^{-1}[\gamma+f-Ef]$$

plus a term that is negligible. For the second term of (A5), we have

$$\frac{1}{N}B'\Sigma^{-1}(B-\hat{B}) = -\frac{1}{N}B'\Sigma^{-1}\in\dot{F}(\dot{F}'\dot{F})^{-1}.$$

In its product with $\gamma+\overline{f}-E(f)$, we can ignore $\overline{f}-E(f)$. In summary,

$$\frac{1}{N}\hat{B}'\widehat{\Sigma}^{-1}(B-\hat{B})[\gamma+\overline{f}-Ef] = -(\dot{F}'\dot{F})^{-1}[\gamma+\overline{f}-Ef] - \frac{1}{N}B'\Sigma^{-1}\in\dot{F}(\dot{F}'\dot{F})^{-1}\gamma$$

plus a term that is negligible.

Finally, consider

$$\frac{1}{N}\hat{B}'\widehat{\Sigma}^{-1}\overline{\in} = \frac{1}{N}B'\Sigma^{-1}\overline{\in} + \frac{1}{N}(\hat{B}-B)'\Sigma^{-1}\overline{\in} + \frac{1}{N}\hat{B}'(\widehat{\Sigma}^{-1}-\Sigma^{-1})\overline{\in}.$$

The last term is negligible. The second on the right hand side has a zero mean (no bias), even though $B-\hat{B}$ and $\overline{\in}$ are correlated. To see this, using $\frac{1}{N}E(\in'\Sigma^{-1}\in) = I_T$, we have

$$\frac{1}{N}(\dot{F}'\dot{F})^{-1}\dot{F}'[E(\in'\Sigma^{-1}\in)]1_T/T = \frac{1}{T}(\dot{F}'\dot{F})^{-1}\dot{F}'1_T,$$

but $\dot{F}1_T = 0$. This implies that the second term is also negligible. The first term contributes to the limiting distribution. Combining all of the above and rearranging terms, we obtain Theorem 2. □

References

Bai, J., 2011. Estimating high dimensional covariance matrices and its applications. Ann. Econ. Finan. 12, 199–215.
Bai, J., Liao, Y., 2013. Statistical inferences using large estimated covariances for panel data and factor models. Working paper, Columbia University.
Black, F., 1972. Capital market equilibrium with restricted borrowing. J. Business 45, 444–454.
Connor, G., Korajczyk, R., 1988. Risk and return in an equilibrium APT: application of a new test methodology. J. Finan. Econ. 21, 255–289.
Fama, E.F., MacBeth, J., 1973. Risk, returns and equilibrium: empirical tests. J. Polit. Econ. 71, 607–636.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Finan. Econ. 33, 3–56.
Fan, J., Liao, Y., Micheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. J. Roy. Stat. Soc. B 75, 603–680.
Jagannathan, R., Schaumburg, E., Zhou, G., 2010. On cross-section analysis of stock returns. Annu. Rev. Finan. Econ. 2, 49–74.
Kan, R., Robotti, C., Shanken, J., 2013. Pricing model performance and the two-pass cross-sectional regression methodology. J. Finan. 68, 617–2649.
Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. 47, 13–37.
Shanken, J., 1992. On the estimation of beta-pricing models. Rev. Finan. Stud. 5, 1–33.
Shanken, J.A., Zhou, G., 2007. Estimating and testing beta pricing models: alternative methods and their performance in simulations. J. Finan. Econ. 84, 40–86.
Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Finan. 19, 425–442.