Testing Beta-Pricing Models
Using Large Cross-Sections*

Valentina Raponi  Cesare Robotti  Paolo Zaffaroni

March 27, 2017

Abstract

Building on the Shanken (1992) estimator, we develop a methodology for estimating and testing beta-pricing models when a large number of assets $N$ is available but the number of time-series observations is fixed, possibly very small. We show empirically that our large-$N$ framework poses a serious challenge to common empirical findings regarding estimated risk premia and validity of beta-pricing models, especially when estimated over a short time window. We generalize our theoretical results to the more realistic case of unbalanced panels. The practical relevance of our findings is confirmed via Monte Carlo simulations.

Keywords: beta-pricing models; ex-post risk premia; two-pass cross-sectional regression; large-$N$ asymptotics; specification test; unbalanced panel.

JEL classification: C12, C13, G12.

*Valentina Raponi, Imperial College Business School, e-mail: v.raponi13@imperial.ac.uk; Cesare Robotti, Imperial College Business School, e-mail: c.robotti@imperial.ac.uk; Paolo Zaffaroni, Imperial College Business School, e-mail: p.zaffaroni@imperial.ac.uk. We gratefully acknowledge comments from Adrian Buss, Fernando Chague, Victor DeMiguel, Francisco Gomes, Cam Harvey, Andrew Karolyi, Ralph Koijen, Lőrinc Pástor, Tarun Ramadorai, Krishna Ramaswamy, Olivier Scaillet, Jay Shanken, Pietro Veronesi, Grigory Vilkov, Guofu Zhou, and especially Raman Uppal, and seminar participants at CORE, Imperial College London, Luxembourg School of Finance, University of Georgia, University of Southampton, Toulouse School of Economics, Tinbergen Institute, the 2015 Meetings of the Brazilian Finance Society, the CFE 2015, the NBER-NSF 2016. An earlier version of this paper was circulated with the title “Ex-Post Risk Premia and Tests of Multi-Beta Models in Large Cross-Sections”.

1
1 Introduction

Traditional methodologies for estimating risk premia require a large time-series sample size, \( T \), and a fixed number of securities, \( N \). Tens of thousands of stocks are traded every day in financial markets, providing an extremely rich information set to estimate asset pricing models.\(^1\) Although we have about a hundred years of financial data available, because of structural breaks such as the two world wars, the oil shock crises and various financial crises, typically short time-series are used in practice.\(^2\)

Under these circumstances, the (asymptotic) distribution, and the related standard errors, of any traditional estimator of betas and risk premia provide a poor approximation to its finite-sample distribution, adversely affecting the inference of any asset pricing model. The alternative approach of increasing the time-series frequency leads to other complications.\(^3\) Therefore, it is important to have a methodology that allows us to undertake statistically correct inference on risk premia and the validity of beta-pricing models, when the number of individual assets, \( N \), is large but the number of time-series observations, \( T \), is limited.

Our main contribution in this paper is to develop such a formal methodology, built on the large-\( N \) estimator proposed by Shanken (1992). Our paper is close to Gagliardini, Ossola, and Scaillet (2016) in the sense that both papers provide a formal methodology for inference, namely estimation and testing, of beta-pricing models. However, while their work is developed in a joint-asymptotics setting, where both \( T \) and \( N \) must diverge, our analysis is derived when \( T \) is fixed.\(^4\) Petersen (2009) documents the dangers of using the inappropriate methodology, in particular the wrong standard errors, in finance applications. In Proposition\(^2\) below, we show that the consequences of using standard errors designed for a small cross-section, when there are a large number of assets, are even more serious. This outcome is particularly pronounced for the Fama-MacBeth (1973) \( t \)-ratios, but it can also arise when one modifies them with the Shanken (1992) correction, which is theoretically appropriate in a large-\( T \), fixed-\( N \), environment.

---

\(^1\) For example, one can obtain the returns for 18,474 US stocks in December 2013 (source: CRSP), half of which are actively traded.

\(^2\) For example, if one wishes to use only post-2009 observations, due to the most recent financial crisis, fewer than one hundred monthly time-series observations are available.

\(^3\) For instance, it induces a substantial increase of noisiness in the data. It also leads to unequal spacing of the observations and market micro-structure effects, which, although fully understood nowadays, greatly complicate the statistical theory of any adopted estimator. Finally, it raises some computational difficulties when aligning times-series extracted from different, high-frequency, data sets.

\(^4\) A detailed comparison with Gagliardini at al. (2016) is provided below, in our review of the related literature.
We demonstrate the usefulness of our methodology by means of an empirical analysis. The three prominent beta-pricing specifications that we consider are the Capital Asset Pricing Model (CAPM), the three-factor Fama and French (1993) model, and the recently proposed five-factor Fama and French (2015) model. We find significant pricing ability for all the factors for each of the three models, even when using a relatively short time window of three years. In contrast, the same risk premia appear often insignificantly different from zero when using the traditional approach. In terms of asset pricing test, our methodology tends to reject the CAPM even when using a short time window, in contrast to the traditional large-\( T \) approach of Gibbons, Ross and Shanken (1989).

Estimation of risk premia in a fixed-\( T \) environment implies that traditional estimation methods do not exhibit the desired statistical properties, for example consistency. However, one could estimate a related quantity denoted the “ex-post” risk premia by Shanken (1992): this equals the conventional (ex-ante) risk premia plus the unexpected outcomes of the factors, corresponding to the beta-pricing model.\(^5\) To motivate our analysis, we now explain how the ex-post risk premium is a parameter with several attractive properties: it is an unbiased estimator of the (ex-ante) risk premium, and the beta-pricing model is still linear in the ex-post risk premia under the assumption of correct model specification. Moreover, we show that the corresponding ex-post pricing errors can be used to derive a new test, for testing the correct specification of any beta-pricing model. Finally, inference on the ex-post risk premia delivers, as a by-product, accurate information on the portion of the ex-ante risk premia unrelated to the factors’ expected value, as explained in detail in Section 3.

Using the arguments of Litzenberger and Ramaswamy (1979), Shanken (1992) proposes an estimator of the ex-post risk premia and shows that it is asymptotically unbiased when only \( N \) diverges.\(^6\) Building on this work, we provide a formal methodology for estimating and testing beta-pricing models in a large-\( N \) environment. In particular, first, we demonstrate a uniqueness result for the Shanken estimator: it is the only element of a class of OLS bias-adjusted estimators that

\(^5\)Obviously as \( T \) gets large any discrepancy between the conventional risk premia and the ex-post risk premia dissipates because the factors’ sample mean converge to their expected value.

\(^6\)Noticeably, Shanken (1992) demonstrates this result without imposing a rigid structure on the \( N \times N \) covariance matrix of the equity returns’ residuals, such as constant variances or null cross-correlations, as assumed by Litzenberger and Ramaswamy (1979). Among other interesting features, this estimator does not require any additional information besides a panel of asset returns and a sample of factors’ realizations, unlike other approaches such as instrumental-variable estimation. Moreover, it can be used when either traded and non-traded factors are used in the beta-pricing model.
does not require preliminary estimation of the bias-adjustment. Second, we establish the asymptotic properties of the estimator for large \( N \) and fixed \( T \), namely \( \sqrt{N} \)-consistency and asymptotic normality. Our technical assumptions are mild and easily verifiable, in particular permitting a degree of cross-correlation among returns akin to the one typically postulated in the Arbitrage Pricing Theory (APT) of Ross (1976). Third, we derive an explicit expression for the estimator’s asymptotic covariance matrix, showing how it can be used to construct correctly-sized confidence intervals for the risk premia estimates. Fourth, we generalize our results to the case of unbalanced panels. This extension is useful particularly when only a small number of time-series observations is available, making it extremely costly to eliminate observations solely to achieve a balanced panel, which leads to unnecessarily large confidence intervals. Fifth, we present Monte Carlo simulations to corroborate our theoretical results. In particular, we highlight how, when \( N \) is much larger than \( T \), inference based on the traditional CSR OLS \( t \)-ratios can be severely misleading, even when accounting for the correct large-\( T \) standard errors. In contrast, the \( t \)-ratios based on our large-\( N \) standard errors are correctly sized. Our theory can be used to provide the theoretical foundation for the risk premia and characteristics’ CSR estimator of Chordia, Goyal, and Shanken (2015), an extension of the Shanken (1992) estimator that allows for firms’ characteristics, together with time-varying betas.

In addition to estimation, we provide a new test for the validity of the asset pricing restrictions and characterize its distribution for large \( N \) and fixed \( T \), under the null hypothesis that the model is correctly specified. Our test is applicable to both balanced and unbalanced panels. All other tests proposed in the literature, without exception, require large \( T \). Importantly, our test does not lead to under- or over-rejections (is correctly sized), and is able to discriminate whether the beta-pricing model is correctly specified (has power), despite being built on the ex-post pricing errors, which are, necessarily, contaminated by the unexpected outcomes of the factors. Finally, because our test is designed for large \( N \), it mitigates the concerns raised by Barillas and Shanken (2016) and Harvey, Liu and Zhu (2016) regarding the conventional way in which beta-model specification tests have been used with the data, as discussed in the next section.

The rest of the paper is organized as follows. Section 2 reviews the literature and Section 3 surveys the two-pass CSR methodology. Here we formally demonstrate the possible dangers arising

---

\footnote{This convenient feature avoids, for example, any pre-testing biases and, at the same time, it does not require sacrificing data for preliminary estimation of the bias-adjustment.}
from using the traditional CSR OLS t-ratios when \( N \) is large. Then, we establish a uniqueness result for the Shanken estimator. Section 4 presents the empirical application while Section 5 presents the asymptotic analysis when \( N \to \infty \) and \( T \) is fixed, both in terms of estimation and testing. Section 6 summarizes our conclusions. Our assumptions and the proofs of our theorems are collected in appendixes. The Internet Appendix (IA) contains further material: Section IA.1 illustrates the finite-\( N \) sample properties of the Shanken estimator and of the associated test through a set of Monte Carlo simulations; Section IA.2 shows how our methodology can be extended to accommodate unbalanced panels; Section IA.3 contains a set of figures associated with the empirical application of the Fama and French (2015) five-factor model.

2 Literature Review

The traditional empirical methodology for exploring beta-pricing models is asymptotically valid when \( T \) is large and \( N \) is fixed: it entails estimation of asset betas, which represent systematic risk measures, by means of time-series factor model regressions, followed by estimation of risk premia via a CSR of observed average returns on the estimated betas. For example, in the empirical strategy developed by Fama and MacBeth (1973) to analyse the CAPM, a CSR is estimated each month, with inference ultimately based on the time-series mean and standard error of the monthly risk premia estimates. The formal econometric analysis of such two-pass methodology was first provided by Shanken (1992), who shows how the asymptotic standard errors of the second-pass CSR OLS and CSR generalized least squares (CSR GLS) risk premia estimators are influenced by estimation error in the first-pass betas, requiring the well-known Shanken’s correction. Following the seminal contributions of Litzenberger and Ramaswamy (1979) and Shanken (1992), other methods have been recently proposed in the empirical asset pricing literature to take advantage of the increasing availability of large cross-sections of individual securities. Using a joint-asymptotics setting, namely when both \( N \) and \( T \) diverge, Gagliardini et al. (2016) derive the limiting properties of a bias-adjusted estimator of the risk premia. Just like us, Gagliardini

---

8See also the related paper by Black, Jensen, and Scholes (1972). Jagannathan and Wang (1998) relax the conditional homoskedasticity assumption in Shanken (1992) and derive expressions for the asymptotic variances of the OLS and GLS estimators that are valid under fairly general distributional assumptions. Shanken and Zhou (2007) and Kan, Robotti, and Shanken (2013) provide a unifying treatment of the two-pass methodology in the presence of global (or fixed) model misspecification.
et al. (2016) need a bias adjustment, even though they require $T$ to diverge.\footnote{Recall that in the traditional analysis of the CSR OLS estimator, where $T$ diverges and $N$ is fixed, no bias adjustment is required.} Their parameter of interest, the difference between the ex-ante risk premia and the factors’ expected value, can be derived from the ex-post risk premia by netting out the sample mean of the factors (an observed quantity). Finally, while Gagliardini et al. (2016) assume random betas, in our analysis we prefer to keep them non-random. This is, for us, mostly a convenience assumption: we show that allowing for randomness of the betas in a large-$N$ environment does not change our theoretical results; see Section 5.2, where we clarify the required identification assumption.

Bai and Zhou (2015) investigate the joint-asymptotics of a modified OLS CSR estimator of the (ex-ante) risk premia. Kim and Skoulakis (2014) propose a $\sqrt{N}$-consistent estimator of the ex-post risk premia in a two-pass CSR setting, employing the so-called regression calibration approach used in errors-in-variables models, building on Jagannathan, Skoulakis, and Wang (2010).\footnote{Kim and Skoulakis (2014) estimator can be interpreted as providing a nice interpretation of the Shanken estimator: the two estimators are almost identical, the only difference being that in Kim and Skoulakis (2014) the first- and second-pass regressions are evaluated on non-overlapping time periods.} Finally, Jegadeesh and Noh (2014) and Pukthuanthong, Roll, and Wang (2014) propose instrumental-variable estimators of the ex-post risk premia, exploiting the assumed independence across time of the returns data.\footnote{Besides the classical econometric challenges associated with the choice of weak instruments, leading to instability of the estimates and unreliable inference, these instrumental-variable approaches require additional information, such as a larger $T$ (twice or three times as large, depending on the method proposed) and, obviously, valid instruments, in order to obtain possibly the same statistical accuracy of the Shanken (1992) estimator. Moreover, Jegadeesh and Noh (2014), and Pukthuanthong et al. (2014) crucially require stochastic independence across time of the returns data, which permits an ingenious construction of their instruments. Stochastic independence across time is also necessary for Kim and Skoulakis (2014). In contrast, it can be shown that the Shanken (1992) estimator retains its asymptotic properties even when the data are not independent across time, and in fact an arbitrary degree of serial dependence, of the returns data, can be allowed for. This is an immediate consequence of the large-$N$, fixed-$T$, approach.}

Regarding tests of the asset pricing no-arbitrage restriction, extending the classical test of Gibbons et al. (1989), Pesaran and Yamagata (2012) propose a number of tests of the asset pricing no-arbitrage restriction: their setup accommodates only traded factors, and the feasible versions of their tests require joint asymptotics. Gagliardini et al. (2016) derive the asymptotic distribution of an asset pricing specification test, under joint asymptotics, but, like us, allow for factors that are not necessarily returns of traded portfolios.
3 Two-Pass Methodology

A beta-pricing model seeks to explain cross-sectional differences in expected asset returns in terms of asset betas computed relative to the model’s systematic economic factors. Let \( f_t = [f_{1t}, \ldots, f_{Kt}]' \) be a \( K \)-vector of observed factors at time \( t \) and let \( R_t = [R_{1t}, \ldots, R_{Nt}]' \) be an \( N \)-vector of test asset returns at time \( t \). Assume that asset returns are governed by the following factor model:

\[
R_{it} = \alpha_i + \beta_i f_{it} + \cdots + \beta_i K f_{Kt} + \epsilon_{it} = \alpha_i + \beta_i' f_t + \epsilon_{it},
\]

where \( i \) indicates the \( i \)th stock and \( t \) refers to time, \( \alpha_i \) is a scalar representing the asset specific intercept, \( \beta_i = [\beta_{i1}, \ldots, \beta_{iK}]' \) is a vector of multiple regression betas of asset \( i \) with respect to the \( K \) factors, and the \( \epsilon_{it} \)s are the model true residuals. In matrix notation, we can write the above model as

\[
R_t = \alpha + B f_t + \epsilon_t, \quad t = 1, \ldots, T,
\]

where \( \alpha = [\alpha_1, \ldots, \alpha_N]' \), \( B = [\beta_1, \ldots, \beta_N]' \), and \( \epsilon_t = [\epsilon_{1t}, \ldots, \epsilon_{Nt}]' \).

For a better understanding of the notion of ex-post risk premia, let \( \bar{R}_i = \frac{1}{T} \sum_{t=1}^{T} R_{it}, \bar{R} = [\bar{R}_1, \ldots, \bar{R}_N]' \), \( \bar{\epsilon} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_{it} \) and \( \bar{f} = \frac{1}{T} \sum_{t=1}^{T} f_t \). Averaging \( \frac{1}{T} \sum_{t=1}^{T} R_{it} \) over time, imposing the exact pricing condition (see (A.5) in Appendix A), and noting that \( \mathbb{E}[R_t] = \alpha + B \mathbb{E}[f] \) by (2), yield

\[
\bar{R} = X \Gamma^{P} + \bar{\epsilon},
\]

where we define \( \Gamma^{P} = [\gamma_0, \gamma_1']' \), with \( \gamma_0 \) representing the zero-beta rate, and where

\[
\gamma_{1}^{P} = \gamma_1 + \bar{f} - \mathbb{E}[f],
\]

with \( \gamma_1 \) representing the (ex-ante) risk premia corresponding to the factors \( f_t \), whose population mean is defined as \( \mathbb{E}[f] \).

Definition (4) applies regardless of whether traded or non-traded factors are considered, although for traded factors, (4) reduces to \( \gamma_{1}^{P} = \bar{f} - \gamma_0 1_K \), where \( 1_K \) is a \( K \times 1 \) vector of ones, because in this case \( \gamma_1 = \mathbb{E}(f) - \gamma_0 1_K \); see Shanken (1992), Section 1.1, for derivation of this result. It follows that, with traded factors, only the scalar parameter \( \gamma_0 \) needs to be estimated, with possible efficiency gains. By (3), expected returns are still linear in the asset betas conditional on the factor outcomes through the quantity \( \gamma_{1}^{P} \) which, in turn, depends on the sample mean realization \( \bar{f} - \mathbb{E}[f] \). For this
reason, the random coefficient vector $\gamma_P$ is referred to, accordingly, as the vector of ex-post risk premia. Obviously, $\gamma_P$ is random but is unbiased for the (ex-ante) risk premia $\gamma_1$ for any $T$. The notion of ex-post risk premia $\gamma_P$ is meaningful only conditional on the factors’ realizations: when $T \to \infty$, $\bar{f}$ will converge to $E(f)$ and thus the ex-post and ex-ante risk premia will coincide but $\Gamma = (\gamma_0, \gamma_1)'$ and $\Gamma_P$ will differ, in general, for any finite $T$.

Therefore, the importance of the ex-post risk premia naturally emerges when studying estimation of beta factor models with large $N$ and fixed, especially small, $T$. However, the ex-post risk premia is relevant more broadly. In fact, given that $\bar{f}$ is observed and constant for fixed $T$, any valid estimator of $\gamma_P$ provides, as a by-product, a valid estimator of the parameter $\nu = \gamma_1 - E(f) = \gamma_P - \bar{f}$. In other words, by solving the estimation of the ex-post risk premia, one obtains an estimator for $\nu$, namely the portion of the (ex-ante) risk premia non-linearly related to the factors. This is precisely the quantity studied in Gagliardini et al. (2016). Note that $\nu$ is identically zero when all factors are excess returns of traded assets but, obviously, in this case the risk premia can be readily estimated with the factors’ sample mean.\(^\text{13}\)

Notice that (3) cannot be used to estimate the ex-post risk premia $\Gamma_P$ because $X$ is unobserved. For this reason, the popular two-pass CSR method first obtains estimates of the betas of each asset $i$, $\hat{\beta}_i$, by running the following multivariate regression for every $i = 1, \ldots, N$:

$$R_i = \alpha_i 1_T + F\beta_i + \epsilon_i,$$

where $R_i = [R_{i1}, \ldots, R_{iT}]'$ is a time-series of returns on asset $i$, $\epsilon_i = [\epsilon_{i1}, \ldots, \epsilon_{iT}]'$ is the associated vector of idiosyncratic residuals, $F = [f_1, \ldots, f_T]'$ is the $T \times K$ matrix of factors, and $1_T$ is a $T \times 1$ vector of ones. This is the first pass. Define $\tilde{F} = \left( I_T - \frac{1_T 1_T'}{T} \right) F = F - 1_T \bar{f}'$, where $I_T$ is the $T \times T$ identity matrix. One obtains the OLS estimator as

$$\hat{\beta}_i = \beta_i + (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\epsilon_i,$$  

or, in matrix form,

$$\hat{B} = R'\tilde{F}(\tilde{F}'\tilde{F})^{-1} = B + \epsilon'P,$$  

\(^\text{12}\)To gauge an idea of the magnitude, the relative difference between the ex-post and the ex-ante risk premia is estimated to be about 2%, with a standard deviation of 0.8% ($T = 36$) and 0.4% ($T = 120$), with respect to the market return factor. Similar figures are obtained with respect to the other factors considered in the empirical application of Section 4.

\(^\text{13}\)Interestingly, Gagliardini et al. (2016) show that their estimator for $\nu$ is $\sqrt{NT}$-consistent. Although their estimator differs from the Shanken estimator, their rate coincides exactly with the $\sqrt{N}$-rate of convergence of the Shanken estimator when $T$ is fixed.
where \( \hat{B} = [\hat{\beta}_1, \ldots, \hat{\beta}_N]' \), \( R = [R_1, \ldots, R_N] \) and \( \epsilon = [\epsilon_1, \ldots, \epsilon_N] \) are \( N \times K \) (the first) and \( T \times N \) (the second and third) matrices, respectively, and \( P = \hat{F}(\hat{F}'\hat{F})^{-1} \). The corresponding \( T \times N \) matrix of OLS residuals is denoted by \( \hat{\epsilon} = [\hat{\epsilon}_1, \ldots, \hat{\epsilon}_N] \) = \( R' - 1_T \hat{R}' - \hat{B}\hat{F}' \). We then run a single CSR of the sample mean vector \( \hat{R} \) on \( \hat{X} = [1_N, \hat{B}] \) to estimate the risk premia. This is the second pass. However, notice that we have two different feasible representations of (3), namely

\[
\bar{R} = \hat{X}\Gamma + \eta, \tag{8}
\]

with residuals \( \eta = (\hat{\epsilon} + B(\bar{f} - E[f]) - (\hat{X} - X)\Gamma) \), and

\[
\tilde{R} = \bar{X}\Gamma^P + \eta^P, \tag{9}
\]

with residuals \( \eta^P = (\hat{\epsilon} - (\hat{X} - X)\Gamma^P) \). The CSR OLS estimator applied to either (8) or (9) yields obviously the same quantity:

\[
\hat{\Gamma} = \begin{bmatrix}
\hat{\gamma}_0 \\
\hat{\gamma}_1
\end{bmatrix} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R}. \tag{10}
\]

However, as Shanken (1992) points out, \( \hat{\Gamma} \) cannot be used as a consistent estimator of the ex-ante risk premia \( \Gamma \) in (8) for a fixed \( T \). The reason is that \( \bar{f} \) does not converge in probability to \( E[f] \) unless \( T \to \infty \). Although Bai and Zhou (2015) conjecture that the impact of the term \( \bar{f} - E[f] \) is small in practice, we follow Shanken (1992) and conduct our analysis based on the representation (9) where the ex-post risk premia \( \Gamma^P \) represent the parameter of interest. Because the innovation \( \eta^P \) in (9) does not contain the term \( B(\bar{f} - E[f]) \) as opposed to the innovation \( \eta \) in (8), the bias term is less severe now. However Shanken (1992) and Bai and Zhou (2015), among others, show that the CSR OLS \( \hat{\Gamma} \) is still biased and inconsistent for \( \Gamma^P \) when \( T \) is fixed. Nevertheless, Shanken (1992) shows that this bias of the CSR OLS estimator can be corrected as follows. Denote the trace operator by \( \text{tr}(\cdot) \) and a \( K \)-dimensional vector of zeros by \( 0_K \). In addition, let \( \hat{\sigma}^2 = \frac{1}{N(T - K - 1)}\text{tr}(\hat{\epsilon}'\hat{\epsilon}) \). Then, the bias-adjusted estimator of Shanken (1992) is given by

\[
\hat{\Gamma}^* = \begin{bmatrix}
\hat{\gamma}_0^* \\
\hat{\gamma}_1^*
\end{bmatrix} = (\hat{\Sigma}_X - \hat{\Lambda})^{-1}\hat{X}'\hat{R}, \tag{11}
\]

where

\[
\hat{\Sigma}_X = \hat{X}'\hat{X}/N, \tag{12}
\]

and

\[
\hat{\Lambda} = \begin{bmatrix}
0 & 0_K' \\
0_K & \hat{\sigma}^2(\hat{F}'\hat{F})^{-1}
\end{bmatrix}. \tag{13}
\]
The formula for the Shanken estimator $\hat{\Gamma}^*$ shows a multiplicative bias-adjustment through the term $(\hat{\Sigma}_X - \hat{\Lambda})^{-1}$. This prompts us to understand the analogies of the Shanken estimator $\hat{\Gamma}^*$ with the more conventional class of additive bias-adjusted CSR OLS estimators. To this end, it is useful to consider the following property satisfied by the CSR OLS estimator, obtained by a suitably rewriting of Bai and Zhou, Theorem 1, (2015):

$$\hat{\Gamma} = \Gamma^P + \left( \frac{\hat{X}'\hat{X}}{N} \right)^{-1} \begin{bmatrix} 0 & 0' \ K \\ 0_K & -\hat{\sigma}^2 (\hat{F}'\hat{F})^{-1} \end{bmatrix} \Gamma^P + O_p\left( \frac{1}{\sqrt{N}} \right) = \Gamma^P - \left( \frac{\hat{X}'\hat{X}}{N} \right)^{-1} \hat{\Lambda} \Gamma^P + O_p\left( \frac{1}{\sqrt{N}} \right).$$

This formula immediately suggests an easy way to construct an additive bias-adjusted estimator for $\Gamma^P$, such as

$$\hat{\Gamma}^{bias-adj} = \hat{\Gamma} + \left( \frac{\hat{X}'\hat{X}}{N} \right)^{-1} \hat{\Lambda} \Gamma^{prelim},$$

where $\hat{\Gamma}^{prelim}$ defines an arbitrary preliminary estimator for $\Gamma^P$. Although this class of estimators $\hat{\Gamma}^{bias-adj}$ appears appealing, it requires the availability of a preliminary estimator, here denoted by $\hat{\Gamma}^{prelim}$. A possible approach is to impose that the preliminary estimator $\hat{\Gamma}^{prelim}$ and the resultant bias-adjusted estimator $\hat{\Gamma}^{bias-adj}$ coincide exactly, and seek the conditions required for uniqueness of the corresponding solution. It turns out that such an approach is not only meaningful but its (unique) solution is precisely the Shanken estimator $\hat{\Gamma}^*$ in (11), which, therefore, is the unique additive bias-adjusted CSR OLS estimator that does not require preliminary estimation of the model. This is formalized in the following proposition, the proof of which is reported in Appendix C.

**Proposition 1** Assume that $\hat{\Sigma}_X - \hat{\Lambda}$ is non-singular. Then, the Shanken estimator $\hat{\Gamma}^*$ in (11) is the unique solution of the linear system of equations:

$$\hat{\Gamma}^* = \hat{\Gamma} + \left( \frac{\hat{X}'\hat{X}}{N} \right)^{-1} \hat{\Lambda} \hat{\Gamma}^*.$$

Somewhat surprisingly, the Shanken estimator does not suffer from the so-called curse of dimensionality, despite involving potentially an infinite number of parameters, viz. the elements of the residuals covariance matrix $\Sigma$, as $N$ diverges. It remains to verify whether the Shanken estimator $\hat{\Gamma}^*$ exhibits desirable (asymptotic) statistical properties. This is studied in Section 5 where we provide a formal asymptotic analysis of $\hat{\Gamma}^*$. As a final precaution, note that the errors-in-variables

\[\text{Bai and Zhou (2015) propose to use the CSR OLS } \hat{\Gamma} \text{ itself as the preliminary estimator, plugging it into the above formula in place of } \hat{\Gamma}^{prelim}. \text{ This is justified only when } T \to \infty, \text{ unlike the framework considered here.}\]
(EIV) correction in (11) entails subtracting the estimated covariance matrix of the beta estimation errors from $\hat{B}'\hat{B}$ leading to the bias-correction. However, it is possible that this EIV correction will overshoot making the matrix $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)$ nearly singular or even not positive definite for a given $N$. To deal with the possibility that the estimator will occasionally produce extreme results, in the simulation and empirical sections of the paper we multiply the matrix $\hat{\Lambda}$ by a scalar $k$ ($0 \leq k \leq 1$), effectively implementing a shrinkage estimator. If $k$ is zero, we get back the CSR OLS estimator $\hat{\Gamma}$, whereas if $k$ is one, we obtain the Shanken estimator $\hat{\Gamma}^\star$. The choice of the shrinkage parameter $k$ should be based on the eigenvalues of the matrix $\left(\hat{\Sigma}_X - k\hat{\Lambda}\right)$. Starting from $k = 1$, if the minimum eigenvalue of this matrix is negative and/or the condition number (i.e., the ratio of the maximum and minimum eigenvalues) of this matrix is bigger than 20, then we lower $k$ by an arbitrarily small amount. In our empirical application we set this amount equal to 0.05 and add a further condition for applying the shrinkage, namely when the relative change (in absolute value) between the Shanken and the OLS estimators is bigger than 100%. We iterate this procedure until the minimum eigenvalue is positive and the condition number becomes smaller than 20. In our simulation experiments, we find that this shrinkage estimator is virtually unbiased, leading to $k = 1$. This is mainly due to the fact that in our simulations we encounter very rare instances in which $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)$ is not positive definite.

Although finding a suitable bias correction of the CSR OLS estimator is extremely desirable, establishing the asymptotic distribution of such bias-corrected estimator, and in particular its appropriate standard errors, is not less important. To demonstrate this, we study the behaviour of the traditional $t$-ratios, designed for a large $T$ environment, when in fact a very large cross-section $N$ of asset returns is used. The results, presented in Proposition 2 below, are striking. When using the Fama-MacBeth standard errors, the $t$-ratio, corresponding to each risk premium, is arbitrarily large (in absolute value). Interestingly, using the Shanken’s correction, the degree of over-rejection is less severe although still present: the $t$-ratio, corresponding to any particular element of $\gamma_1$, equals the standartized sample mean of the corresponding factor, plus a bias term of the Shanken’s correction. The two parts that make the Shanken’s correction, the $\hat{c}$ term is irrelevant when $N$ is large whereas taking into account the factor’s variability, through $1_{(k>0)}\hat{\sigma}_k^2/T$, ensures that the $t$-ratio does not diverge (in probability) as $N \to \infty$. This holds when evaluating $t$-ratios with respect to the elements of the ex-ante risk premia $\gamma_1$. When,
As \( T \) diverges, convergence to a standard normal is re-obtained but when evaluating the \( t \)-ratio for a finite \( T \) and a large \( N \), the bias term can be substantial. Instead, the \( t \)-ratio corresponding to the zero beta rate \( \gamma_0 \) diverges to infinity (in absolute value) for both approaches. Summarizing, our result shows that a methodology designed for large \( T \) is likely to lead to severe over-rejections, invalidating any inference on the beta-pricing model. Our Monte Carlo simulations corroborate this finding\(^{19}\).

**Proposition 2** Under Assumptions 1–6 (listed in Appendix A), in particular under the assumption of correct model specification (Assumption A.5), as \( N \to \infty \):

(i) the \( t \)-ratios, constructed using the the Fama-MacBeth (1973) methodology, satisfy

\[
|t_{FM}(\hat{\gamma}_0)| = \left| \frac{(\hat{\gamma}_0 - \gamma_0)}{SE_{0FM}} \right| \to_p \infty \quad \text{and} \quad |t_{FM}(\hat{\gamma}_{1k})| = \left| \frac{(\hat{\gamma}_{1k} - \gamma_{1k})}{SE_{FM}} \right| \to_p \infty,
\]

for every \( k = 1, \ldots, K \), where \( SE_{FM} = \left( \hat{\Sigma}^{-1} \right)^{1/2}, \hat{\Sigma} \) denotes the \( j \)th column (or row), for \( j = 1, \ldots, J \), of the identity matrix \( I_J \), \( \gamma_1 = (\gamma_{11}, \ldots, \gamma_{1j}, \ldots, \gamma_{1K})' \) and \( \hat{\Sigma} = \hat{\epsilon}'/T - K - 1 \) is the OLS residual covariance matrix.

(ii) the \( t \)-ratios, constructed using the EIV correction of Shanken (1992), satisfy

\[
|t_{EIV}(\hat{\gamma}_0)| = \left| \frac{(\hat{\gamma}_0 - \gamma_0)}{SE_{EIV}} \right| \to_p \infty \quad \text{and} \quad |t_{EIV}(\hat{\gamma}_{1k})| = \left| \frac{(\hat{\gamma}_{1k} - \gamma_{1k})}{SE_{EIV}} \right| \to_p \left| \frac{\hat{f}_k - E(f_k)}{\hat{\sigma}_k/\sqrt{T}} - \frac{\hat{\epsilon}_{k,K} A^{-1} C \gamma_{1k}^p}{\hat{\sigma}_k/\sqrt{T}} \right|,
\]

for every \( k = 1, \ldots, K \), where \( SE_{EIV} = \left( \frac{1}{\hat{\sigma}_k^2} \right)^{1/2}, \hat{\sigma}_k \) denotes the \( k \)th element of \( \hat{\Sigma} / T \), and \( A = \Sigma - \mu_\beta \mu_\beta' + C \) with \( C = \sigma^2(\bar{f}/\hat{f})^{-1} \), where \( \mu_\beta, \Sigma_\beta, \sigma^2 \) are defined in Assumptions 1 and 5(i).

**Proof:** See Appendix C and Lemmas 1 to 4 in Appendix B.

We also investigate the limiting statistical properties of a new specification test based on the ex-post sample pricing errors,

\[
\hat{\epsilon}^P = (\hat{\epsilon}_{1}^P, \ldots, \hat{\epsilon}_{\gamma}^P)' = \hat{R} - \hat{X},
\]

(14)

instead, one seeks the \( t \)-ratios centred around the elements of the ex-post risk premia \( \gamma^p \), the factor’s variability term must be excluded from the standard errors (see Shanken (1992), Theorem 1(ii)) and the \( t \)-ratios will diverge. We derived the exact limiting distribution of the \( t \)-ratios and details are available upon request.

\(^{19}\)See Section IA.1.2 of the Internet Appendix.
by looking at the limiting behaviour of the (plain) average of the squared ex-post pricing errors, \(N^{-1} \sum_{i=1}^{N} (\hat{e}_i^P)^2\), suitably normalized to converge to a normal distribution as \(N \to \infty\).

Two recent advances challenged the conventional way in which beta-pricing model specification tests have been used with the data.

First, Barillas and Shanken (2016) show that, when comparing beta-pricing models with traded factors only, the choice of the test assets is completely irrelevant with respect to many popular criteria.\(^{20}\) Although we do not consider model comparison explicitly, our test statistic can be used for that purpose. However, the irrelevance result of Barillas and Shanken (2016) will not apply to our framework, regardless of whether the beta-pricing model under consideration includes trading or non-trading factors only. This is a consequence of the large-\(N\) approach: the sum of the (plain) squared pricing errors, necessarily employed in our test statistic, does not satisfy the irrelevance result, as opposed to the sum of the GLS-weighted squared pricing error as established in Barillas and Shanken (2016), building upon the insights of Lewellen, Nagel and Shanken (2010). In fact, although desirable, because for instance it allows re-packaging (see footnote 20 for more details), the GLS-weighting is not feasible in our context when \(N\) is larger than \(T\). In particular, the inverse of the residual covariance matrix \(\Sigma = E(\epsilon_t \epsilon_t')\) cannot be estimated unless extremely strong parametric restrictions are imposed, such as when \(\Sigma = \sigma^2 I_N\). On the other hand, if no structure is imposed, one can only rely on the sample covariance matrix, which has rank \(T\), smaller than \(N\), and thus invertibility fails.\(^{21}\)

Second, Harvey et al (2016) point out that, because of extensive data mining, with hundreds of research papers analyzing the same cross-sectional data of (portfolio) returns, the usual criteria to assess whether one factor is priced in a cross section of asset returns, namely a \(t\)-ratio greater than 2, is misleading. Based on multiple testing arguments, they show that instead a cut off greater than 3 is required, with this cut off bound to increase over time, as more empirical investigations are

\(^{20}\)In particular, Barillas and Shanken (2016) establish this irrelevance result with respect to three criteria for model comparison: the GRS test statistics, the likelihood ratio with certain elliptical distributions, and the GLS CSR \(R^2\).

\(^{21}\)Some insights can be derived by comparing the limit of the feasible \(N^{-1} \hat{e}_e' \hat{e}_e\) with the unfeasible \(N^{-1} \hat{e}_e' \Sigma^{-1} \hat{e}_e\) as \(N \to \infty\), the latter relying on the unknown \(\Sigma\). Some calculations show that (abstracting from smaller-order terms) one obtains \(N^{-1} \hat{e}_e' + \sigma^2 Q' Q\) and \(N^{-1} \hat{e}_e' \Sigma^{-1} \hat{e}_e + Q' Q\) for large \(N\), for the feasible and unfeasible case respectively, where \(e = (e_1, \ldots, e_N)' = E(R_i) - \gamma_0 1_N - B \gamma_1\) denotes the vector of ex-ante pricing errors and \(Q = 1_T / T - \tilde{F}(\tilde{F}' \tilde{F})^{-1} \gamma_{1e}\). Therefore, the un-weighted sum of the squared ex-post pricing errors will depend on \(\Sigma\), and, hence, the irrelevance result fails. Instead, the weighted (and unfeasible) sum of the squared ex-post pricing errors will be independent of \(\Sigma\), including in particular the second term \(Q' Q\). In this case, the irrelevance result holds but, of course, as discussed, weighting by \(\Sigma^{-1}\) is unfeasible in a fixed \(T\), large \(N\), environment. Details are available upon request.
carried out. Our approach could mitigate this severe problem in three dimensions: first, it requires using a panel of data on individual returns, in particular when \( N \) is much larger than \( T \). Given that the usual methodologies cannot be used for this type of data set, namely large \( N \) and small \( T \), as described in the Introduction, necessarily the risk of data mining is drastically reduced. Moreover, unlike the Fama and French portfolios data sets, a specific data set of individual asset returns repeatedly examined in empirical asset pricing literature simply does not exist. Finally, the typical size of the standard errors for our risk premia estimates, based on the available cross-sections of thousands of individual stocks, implies highly significant t-ratios.

4 Empirical Analysis

In this section, we empirically estimate the risk premia associated with some prominent beta-pricing models, using individual stock return data, and investigate their performance. This demonstrates how the empirical results obtained using our large-\( N \) methodology, illustrated in the previous section, can differ, even dramatically, from the results obtained with the more traditional large-\( T \) methodologies. The confidence intervals and the p-values are based on the theoretical analysis of our methodology presented in Section 5.

We consider three linear beta-pricing models: (i) the CAPM, (ii) the three-factor model of Fama and French (1993, FF3), and (iii) the five-factor model recently proposed by Fama and French (2015, FF5). The five factors entering these empirical specifications are the market excess return (\( mkt \)), the return difference between portfolios of stocks with small and large market capitalizations (\( smb \)), the return difference between portfolios of stocks with high and low book-to-market ratios (\( hml \)), the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios (\( rmw \)), and the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios (\( cma \)). The data on the above factors is available from from Kenneth French’s website. We use monthly data on individual stocks from the CRSP database, available from January 1966 to December 2013. We carry out the empirical analysis using balanced panel with three different time windows of, respectively, three- , six- and ten-year (i.e., \( T=36, 72, \) and 120, respectively). For each of these time windows, we estimate each of the above beta-pricing models by rolling the window one month at the time. In this way, we obtain time-series of estimated risk premia and of the test statistic
based on overlapping time windows of fixed length $T$. After filtering the data, we obtain an average number of approximately 2,800 stocks for the three-year periods, 1,900 for the six-year period, and 1,200 for the 10-year period.\footnote{Specifically, we download monthly stock data from January 1966 to December 2013 from the CRSP database and apply two filters in the selection of stocks. First, we require that a stock has a Standard Industry Classification (SIC) code (we adopt the 49 industry classifications listed on Kenneth French’s website). Second, we keep a stock in our sample only for the months in which its price is at least three dollars. The resulting dataset consists of 3,435 individual stocks and we randomly select 3,000 of them.}

We choose to conduct our analysis over relatively short time spans in order to assess the performance of our large-$N$ approach. To this end, we compare our results with the ones obtained using the conventional large-$T$ approach. This is done for both the risk premia estimates and the specification test statistic. In particular, in terms of estimation, we compare the risk premia estimates based on the Shanken estimator $\hat{\Gamma}^* = \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \hat{X}'\hat{R}/N$ with the estimates based on the OLS estimator $\hat{\Gamma} = \left(\hat{\Sigma}_X\right)^{-1} \hat{X}'\hat{R}/N$, defined in (11) and (10), respectively. For both estimators, we also derive their corresponding 95% confidence intervals. However, whereas for the Shanken estimator we use the large-$N$ standard errors derived in Theorem 2 below, for the OLS estimator we adopt the standard errors of Theorem 1(ii) in Shanken (1992), which are valid for large $T$ only. Hereafter, we will refer to the former confidence intervals as the large-$N$ intervals, while the latter OLS intervals will be denoted as the large-$T$ confidence intervals. Finally, recall that computation of $\hat{\Gamma}^*$ requires occasionally to set a shrinkage parameter $k < 1$ by means of an iterative procedure, as discussed in Section 2.

The results are reported in a series of figures and one table.\footnote{Tables with estimates and confidence intervals are available upon request.} Using a time window of three years ($T = 36$ monthly observations), the top panel of Figure I shows the rolling-window estimates for the risk premium of the market excess return ($\gamma_{mkt}$), together with the corresponding 95% confidence intervals, derived using both the large-$T$ and the large-$N$ approach. Using a sample of approximately $N = 2,800$ stocks, we obtain 541 time-series estimates of the market risk premium, from January 1968 to January 2013. The blue line represents the Shanken estimates, $\hat{\gamma}^*_{mkt}$, while the OLS estimates, $\hat{\gamma}_{mkt}$, are represented by the red line. The light blue band represents the 95% confidence interval using the large-$N$ approach, whereas the orange band refers to the large-$T$ intervals.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Figure I about here}
\end{figure}
A striking feature emerges. The large-\(T\) confidence intervals include the zero value, implying that we cannot reject (at 95% confidence level) the hypothesis of zero risk premium of the market excess return for most dates in the sample. The only few exceptions are observed between 1975–1981, and 1999–2001. In contrast, our large-\(N\) intervals seldom include zero, except for the period around 1975, as well as between 1982 and 1983. More in general, the large-\(T\) approach tends to favour the hypothesis of constant risk premia, whereas our large-\(N\) results point toward a significant time-variation in risk premia. The second and third panels of Figure I report the results for rolling windows of \(T=72\) (six years) and \(T=120\) (10 years), respectively. As expected, the large-\(T\) confidence intervals get narrower, although still sizeably wider than the large-\(N\) intervals, yielding risk premia estimates that are now often significant. Finally, the large-\(N\) estimates appear systematically larger than the corresponding large-\(T\) estimates for most dates, especially when the rolling window is longer. This is the result of the systematic (negative) bias affecting the OLS estimator. At the same time, for a long rolling window, the time-variation in estimated risk premia appear much less prominent, as expected.

Figures II.a, II.b, II.c report the rolling window estimates for the risk premia of the market excess return \((\gamma_{mkt})\), the \(smb\) return \((\gamma_{smb})\), and the \(hml\) return \((\gamma_{hml})\), respectively, obtained by estimating the FF3 model. The results are aligned with what is observed for the CAPM: whereas the large-\(N\) results tend to show estimated risk premia statistically different from zero, the large-\(T\) approach gives the opposite indications. This divergence is attenuated when the estimation (rolling) window is longer across time. Interestingly, for the \(smb\) risk premia, even the large-\(N\) often leads to estimated risk premia non-statistically different from zero especially for the largest estimation window of 10 years. For the other two factors, \(mkt\) and \(hml\), the corresponding large-\(N\) risk premia estimates appear to be non-zero in most cases, regardless of the size of the estimation window.24

Table I contains the percentage difference between the Shanken estimator and the OLS estimator.

\(^{24}\)Figures III.a, III.b, III.c, III.d, and III.e, reported in Section IA.3 of the Internet Appendix, illustrate the estimation results for the FF5 model. The results are analogous to the previous cases for the first three common factors \((mkt, smb, hml)\). For the additional two factors \((rmw\) and \(cma)\), the corresponding large-\(T\) risk premia estimates are non-significant for most periods, regardless of the estimation window. In contrast, the large-\(N\) procedure shows that the estimates of these risk premia are, in fact, significantly different from zero for about half of the cases, the evidence being somewhat stronger for the \(rmw\) risk premium.
for the various risk premia corresponding to the three beta-pricing models, averaged across the time-series of the two set of estimates. Panel A shows that the bias-correction of the Shanken estimator for the market risk premium $\gamma_{mkt}$, when estimating the CAPM, is extremely sizeable, accounting for about 65% when $T = 36$, and diminishing, as expected, to 27% when $T = 120$. Panel B shows that, when estimating the FF3 model, the discrepancy between the two estimators is very sizeable for the $hml$ risk premium, ranging from 52% to 31%, and less pronounced for the market and $smb$ risk premia, going from approximately 15% to 10%. Finally, with respect to estimation of FF5, Panel C shows that the bias-correction discrepancy between the two estimators is very sizeable for the $cma$ risk premium, ranging from 43% to 33%. For the market, $hml$, $smb$ and $rmw$ risk premia the discrepancy ranges from approximately 15% to 10%, without even diminishing as $T$ increases for $hml$ and $rmw$.

Table I about here

Summarizing, we document a sizeable difference between the results of our large $N$ approach and the results of the conventional large $T$ approaches. This outcome is a combination of two elements, the extremely small standard errors associated with a very large $N$ and the bias-correction of the Shanken estimator that leads to an increment, on average of about 20% but sometimes up to 50%, of the risk premia estimates over the OLS estimator.

We finally consider the performance of our specification test $S^*$, formally defined in (42) in Section 5.3 below. We report only the results for the CAPM, in Figure IV and Figure V. We compare the p-values of $S^*$, based on its asymptotic distribution for large $N$ (derived in Theorem 3 below), with the p-values associated with the Gibbons, Ross, and Shanken (1989) test (hereafter GRS), which is a common testing procedure valid for large $T$ and fixed (moderate) $N$. For both ours and the GRS test, the null hypothesis is $H_0: \epsilon_i = 0$ for every asset $i$, namely that the beta-pricing model is correctly specified. The black line in Figure IV denotes the time-series of p-values associated with our test statistic $S^*$, for the time windows of three years (top panel), six years (middle panel) and 10 years (bottom panel), respectively. A black line below the dotted line (that indicates the value of 0.05) means that the CAPM, estimated over that particular sample period, is rejected at 5% level. Figure IV clearly shows that our test rejects the validity of the CAPM for the

\footnotetext[25]{The results are available for the FF3 and FF5 model upon request.}
The large majority of the data periods, even for the shortest time window of \( T = 36 \). At the same time, as expected, rejection of the CAPM happens more frequently as the time window increases from \( T=36 \) to \( T=120 \). Given the availability of time-series of p-values, one could cast this analysis in a multiple testing framework. In particular, implementing the procedure outlined in Barras, Scaillet and Wermers (2010), the test statistic \( S^* \) turns out to be (jointly) significant 65\%, 89\%, and 96\% of the time for samples sizes of \( T = 36, 72, 120 \), respectively. Hence, the CAPM appears to be significantly and systematically rejected, not simply due to sampling variability, in most cases even for moderate sample sizes. Details are available upon request.

Figure IV about here

Figure V reports, instead, the p-value corresponding to the GRS test applied to 25 portfolios, as opposed to the thousands of individual stocks used by our test. Indeed, in order to facilitate the comparison, based on our data of individual asset returns, we have constructed 25 time-series of portfolios, by taking simple average of the appropriate number of stocks.\(^{26}\) The result is, again, rather striking: in contrast to our large-\( N \) test, the GRS test is almost always unable to reject the CAPM at 5\% when considering the shortest time window of three years of data (\( T = 36 \)). The CAPM will be rejected about half of the time for the six years window and will almost always be rejected for the long time window of 10 years.

Figures V about here

5 Asymptotic Analysis

The analysis in this section assumes that \( N \to \infty \) and \( T \) is fixed. We first establish the limiting distribution of the Shanken bias-adjusted estimator \( \hat{\Gamma}^* \) and explain how its asymptotic covariance matrix can be consistently estimated. We then characterize the limiting behavior of our test \( S^* \) of the asset pricing restriction. Our assumptions are described in Appendix A. The empirical analysis, presented in Section 4, relies on these theoretical results.

\(^{26}\)For instance, when one considers \( T = 36 \), each portfolio is made by approximately 110 stocks.
5.1 Asymptotic Distribution of the Shanken Estimator

In this subsection, we study the asymptotic distribution of $\hat{\Gamma}^*$, under the assumption that the model is correctly specified, namely that exact no-arbitrage holds (Assumption 4 in Appendix A). The empirical applications presented in Figures I, II, and III of Section 4, in particular the confidence intervals for the risk premia estimates based on the Shanken estimator, use these theoretical results.

Let $\Sigma_X = \left[ \begin{array}{c} \mu^\prime \\ \Sigma_{\beta} \end{array} \right]$, $\sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2$, $U_\epsilon = \lim_{N \to \infty} \frac{1}{N} \sum_{i,j=1}^N E \left[ \text{vec}(\epsilon_i \epsilon_i^\prime - \sigma_i^2 I_T) \text{vec}(\epsilon_j \epsilon_j^\prime - \sigma_j^2 I_T) \right]$, $M = I_T - D(D^\prime D)^{-1} D'$, where $I_T$ is a $T \times T$ identity matrix, $D = [1_T, F]$, $Q = \frac{1}{T} - P \gamma_1^P$, and $Z = (Q \otimes P) + \frac{\text{vec}(M)}{T-K-1} P^P P'$, the operator vec$(\cdot)$ stacks all the columns of the matrix argument into a vector, and where all the limits are finite by our assumptions, as $N \to \infty$. In the following theorem, we provide the rate of convergence and the limiting distribution of $\hat{\Gamma}^*$.

**Theorem 1**

(i) Under Assumptions 1–5 (listed in Appendix A),

$$\hat{\Gamma}^* - \Gamma^P = O_P \left( \frac{1}{\sqrt{N}} \right).$$  \hfill (15)

(ii) Under Assumptions 1–6 (listed in Appendix A),

$$\sqrt{N} \left( \hat{\Gamma}^* - \Gamma^P \right) \overset{d}{\to} N \left( 0_{K+1}, V + \Sigma_X^{-1} W \Sigma_X^{-1} \right),$$  \hfill (16)

where

$$V = \sigma^2 \left[ 1 + \gamma_1^P \left( \hat{F}^\prime \hat{F}/T \right)^{-1} \gamma_1^P \right] \Sigma_X^{-1} \quad \text{(17)}$$

and

$$W = \begin{bmatrix} 0 & 0^K \\ 0_K & Z' \tilde{U} Z \end{bmatrix}.$$  \hfill (18)

**Proof**: See Appendix C and Lemmas 1 to 5 in Appendix B.

Note that the expression in (16) for the asymptotic covariance matrix is very simple and, moreover, it has a very neat interpretation. The first term of this asymptotic variance, $V$, accounts for the estimation error in the betas, and it is essentially identical to the large $T$ expression of the correct asymptotic covariance matrix associated with the CSR OLS estimator, derived by
Shanken (1992), Theorem 1(ii): part $\frac{\sigma^2}{T} \Sigma_X^{-1}$ is the classical CSR OLS covariance matrix, which one would obtain if the betas were observed, whereas the term $c = \gamma^P \left( \hat{F}^t \hat{F} / T \right)^{-1} \gamma^P$ is an asymptotic adjustment for EIV, with $c \sigma^2 \Sigma_X^{-1}$ being the corresponding overall EIV contribution of the asymptotic covariance matrix. As Shanken (1992) points out, the EIV adjustment reflects the fact that the variability of the estimated betas is directly related to residual variance, $\sigma^2$, and inversely related to factor variability $\left( \hat{F}^t \hat{F} / T \right)^{-1}$. The last term of the asymptotic covariance, $\Sigma_X^{-1} W \Sigma_X^{-1}$, arises because of the bias adjustment that characterises exclusively $\hat{\Gamma}^*$ (but not $\hat{\Gamma}$), which also vanishes when $T \to \infty$. In addition, the $W$ matrix also accounts for the cross-sectional variation in the residual variances of asset returns through $U_\epsilon$.

To conduct statistical inference, we need a consistent estimator of the asymptotic covariance matrix $V + \Sigma_X^{-1} W \Sigma_X^{-1}$. Let $M^{(2)} = M \odot M$, where $\odot$ denotes the Hadamard product operator. In addition, define

$$\hat{\sigma}_4 = \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{\epsilon}_{it}^4,$$

and let

$$\hat{Z} = (\hat{Q} \odot \mathcal{P}) + \frac{\text{vec}(M)}{T - K - 1} \hat{\gamma}^* \mathcal{P}^t \mathcal{P},$$

with

$$\hat{Q} = \frac{1}{T} - \mathcal{P} \hat{\gamma}^*.$$

The following theorem provides a consistent estimator of the asymptotic covariance matrix of the $\hat{\Gamma}^*$ estimator.

**Theorem 2** Under Assumptions 1–5 (listed in Appendix A), we have

$$\hat{V} + (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \hat{W} (\hat{\Sigma}_X - \hat{\Lambda})^{-1} \stackrel{p}{\to} V + \Sigma_X^{-1} W \Sigma_X^{-1},$$

where

$$\hat{V} = \frac{\hat{\sigma}_2^2}{T} \left[ 1 + \hat{\gamma}_1^* \left( \hat{F}^t \hat{F} / T \right)^{-1} \hat{\gamma}_1^* \right] (\hat{\Sigma}_X - \hat{\Lambda})^{-1},$$

$$\hat{W} = \begin{bmatrix} 0 & 0_K \\ 0_K & \hat{Z}^t U_\epsilon \hat{Z} \end{bmatrix},$$

and $\hat{U}_\epsilon$ is a consistent plug-in estimator of $U_\epsilon$ described in Appendix D.
**Proof:** See Appendix C and Lemmas 1 to 6 in Appendix B.

A remarkable feature of the above result is that a consistent estimate of the asymptotic covariance matrix of \( \hat{\Gamma}^* \) could be obtained, while leaving the residual covariance matrix \( \Sigma \) unspecified. In fact, with \( \Sigma \) having in general \( N(N + 1)/2 \) distinct elements and our asymptotic theory only allowing \( N \to \infty \), it follows that consistent estimation of \( \Sigma \) is completely unfeasible, a phenomenon known in econometrics as the curse-of-dimensionality. However, a key feature of the Shanken estimator is that it depends on \( \Sigma \) only through \( \sum_{i=1}^{N} \sigma_i^2 / N \). Moreover, its asymptotic covariance matrix depends on the average \( \sum_{i,j=1}^{N} \sigma_{ij} / N \). Our large \( N \) asymptotic theory shows how both quantities can be estimated consistently, unlike for the individual covariances \( \sigma_{ij} \). This crucial feature of the Shanken estimator is in sharp contrast to the GLS estimator, which requires to estimate consistently every element of \( \Sigma \).

### 5.2 Consequences of random loadings \( \beta_i \)

We now discuss the consequences of allowing for random \( \beta_i \). Consider at first the case in which these, although random, are mutually independent from any other cross-sectional characteristic of the individual asset returns. In this case, no consequences arise in terms of the asymptotic properties of the Shanken estimator \( \hat{\Gamma}^* \). The only, marginal, changes involve Assumptions 1 and 4 in Appendix A. In particular, equations (A.1) and (A.2) in Assumption 1 must be stated in terms of convergence in probability, instead of the conventional convergence valid for non-random sequences; equation (A.5) in Assumption 4 must be replaced by \( E[R_t | X] = X \Gamma \). All the other assumptions are unchanged, except that now (A.16) involves random \( \beta_i \). Then, by easy calculations,

\[
\lim \text{Var}\left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( C_T^t \otimes \left( \frac{1}{\beta_i} \right) \right) \epsilon_i \right) = \lim \frac{1}{N} \sum_{i,j=1}^{N} E\left( C_T^t \otimes \left( \frac{1}{\beta_i} \right) \right) \epsilon_i \epsilon_j' \left( C_T \otimes \left( \frac{1}{\beta_j} \right) \right)
\]

\[
= (C_T^t C_T) \lim \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 E\left( \left( \frac{1}{\beta_i} \right) \left( \frac{1}{\beta'_i} \right) \right) + (C_T^t C_T) \lim \frac{1}{N} \sum_{i,j=1}^{N} \sigma_{ij} E\left( \left( \frac{1}{\beta_i} \right) \left( \frac{1}{\beta_j} \right) \right)
\]

\[
= (C_T^t C_T) \sigma^2 \Sigma_X.
\]

In fact, the second term on the right-hand side of (27) converges to zero under our assumptions, given that \( E[\| \beta_i \beta'_j \|] \leq E(\beta_i' \beta_i)^{1/2} E(\beta'_j \beta_j)^{1/2} \leq C < \infty \) and \( \sum_{i,j=1}^{N} |\sigma_{ij}| = o(N) \). Expression (28)
coincides exactly with the asymptotic covariance matrix of Theorem 1, which holds for non-random \( \beta_i \).

Consider now the case in which the \( \beta_i \) are potentially cross-sectionally correlated with the \( \epsilon_i \). We argue that such covariance structure could not be identified based on the OLS estimators \( \hat{\beta}_i \) and \( \hat{\epsilon}_i \), either for finite or arbitrarily large \( N \), implying that the possibility of cross-correlation should be ruled out. In fact, inspecting the proof of Theorem 1, it turns out that the asymptotic distribution of \( \sqrt{N}(\hat{\Gamma}^* - \Gamma^*) \) depends, among others, on \( N^{-\frac{1}{2}} \sum_{i=1}^N \beta_i \epsilon_i'Q \), where we recall that \( Q = \frac{1}{T} - P\gamma_1^P \).

Setting the \( K \)-dimensional vector \( \delta_i = E(\beta_i \epsilon_i) = \text{Cov}(\beta_i, \epsilon_i) \) implies that

\[
E(\beta_i \epsilon_i') = \Delta_i = \delta_i 1_T', \tag{29}
\]

where the second equality follows by the identically-distributed assumption of the \( \epsilon_{it} \) across time (see Assumption 2 in Appendix A). Then

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^N \beta_i \epsilon_i'Q = \frac{1}{\sqrt{N}} \sum_{i=1}^N (\beta_i \epsilon_i' - \Delta_i)Q + \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta_i Q.
\]

The first term on the right-hand side of the last equation will converge to a normal distribution, by a simple generalization of Assumption 6 (iii). Given (29), the bias term can be rewritten as

\[
\sqrt{N}^{-1} \sum_{i=1}^N \Delta_i Q = \sqrt{N}^{-1} \sum_{i=1}^N \delta_i 1_T'Q = \sqrt{N}^{-1} \sum_{i=1}^N \delta_i, \text{ because } 1_T'Q = 1.
\]

It is evident that in order to avoid an asymptotic bias for \( \hat{\Gamma}^* \), although each \( \delta_i \) could be non-zero, their average must satisfy \( \sqrt{N}^{-1} \sum_{i=1}^N \delta_i = o(1) \), including the case \( \delta_i = 0_K \). We now illustrate how this restriction is naturally asked for when considering the OLS-based estimator of \( \delta_i \).

In particular, the OLS-based estimator of \( N^{-1} \sum_{i=1}^N \Delta_i = N^{-1} \sum_{i=1}^N E(\beta_i \epsilon_i') \) will be \( N^{-1} \sum_{i=1}^N \hat{\Delta}_i \) with \( \hat{\Delta}_i = \hat{\beta}_i \hat{\epsilon}_i' \). However, by easy calculations, one can show that \( \hat{\epsilon}_i \) and \( Q \) are orthogonal for any finite \( T \) (and \( N \)). Set \( M_A = I - A(A'A)^{-1}A' \) for any generic, full row-rank, \( A \), implying that \( M_A \) is the matrix that projects onto the space orthogonal to the space spanned by the columns of \( A \). Then

\[
\hat{\epsilon}_i'Q = \epsilon_i'MQ = \epsilon_i'M\hat{F}M_{1_T}Q = \epsilon_i'M\hat{F}_{1_T}P\gamma_1^P = \epsilon_i'M\hat{F}P\gamma_1^P = 0,
\]

since \( M_{1_T}1_T = 0 \) and \( M\hat{F} \hat{F} = 0 \), where we use the identity \( M = M\hat{F}M_{1_T} = M_{1_T}M\hat{F} \), where using our notation \( M = M_D \) with \( D = (1_T, F) \). Therefore, the estimated bias term \( N^{-1} \sum_{i=1}^N \hat{\Delta}_i Q \) is identically zero for any finite \( N \). On the other hand, even without the post-multiplication by \( Q \), as
\[ N \text{ diverges,} \]
\[
\frac{1}{N} \sum_{i=1}^{N} \hat{\Delta}_i = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \hat{\epsilon}_i = \frac{1}{N} \sum_{i=1}^{N} (\hat{\beta}_i + P' \epsilon_i) \hat{\epsilon}_i M \]
\[
\rightarrow_p \delta 1_T M + \sigma^2 P' M = 0, \quad (31)
\]
regardless of whether \( \delta = \lim N^{-1} \sum_{i=1}^{N} \delta_i \) is zero or not, because both \( 1_T M = 0 \) and \( P' M = 0 \).

Therefore, both the finite \( N \) and the large \( N \) arguments strongly suggest that the assumption,
\[ \delta_i = \text{Cov}(\beta_i, \epsilon_{it}) = 0_K, \]
is not avoidable in the large \( N \) environment or, alternatively, the slightly more general assumption,
\[ \sqrt{N}^{-1} \sum_{i=1}^{N} \delta_i = o(1). \]

5.3 Limiting Distribution of the Specification Test

In this section, we are interested in deriving the properties of our test for the validity of the beta-pricing model, based on the Shanken estimator. The p-values presented in Figures IV and V of Section 4 are based on the limiting distribution of Theorem 3 below.

The null hypothesis underlying the asset pricing restriction can be formulated as
\[ H_0 : e_i = 0 \quad \text{for every } i = 1, 2, \ldots, \]
where \( e_i = E[R_{it}] - \gamma_0 - \beta_i' \gamma_1 \) is the pricing error associated with asset \( i \). The null hypothesis \( H_0 \) easily follows by simply rewriting Assumption 4. Let \( X_i = [1, \beta_i'] \), \( \hat{X}_i = [1, \hat{\beta}_i'] \), and denote by \( \hat{e}_i^P \) the ex-post sample pricing error for asset \( i \). Then, we have
\[
\hat{e}_i^P = \bar{R}_i - \hat{X}_i \hat{\Gamma}^* \quad (33)
\]
\[
= e_i + Q' \epsilon_i - \hat{X}_i \left( \hat{\Gamma}^* - \Gamma^P \right). \quad (34)
\]
It follows that
\[ \hat{e}_i^P \xrightarrow{p} e_i + Q' \epsilon_i = e_i^P. \quad (35) \]
Equation (35) shows that even when the ex-ante pricing error, \( e_i \), is zero, \( \hat{e}_i^P \) will not converge in probability to zero. This is a consequence of the fact that, when \( T \) is fixed, \( Q' \epsilon_i \) will not converge to zero even under the null of zero ex-ante pricing errors. This is the price that we have to pay.
when \( N \) is large and \( T \) is fixed. Nonetheless, a test of \( H_0 \) with good size and power properties can be developed. Since we estimate \( \Gamma^P \) via OLS cross-sectional regressions, we propose a test based on the sum of the squared ex-post sample pricing errors, that is,

\[
\hat{Q} = \frac{1}{N} \sum_{i=1}^{N} (\hat{e}_i^P)^2.
\]  

(36)

Consider the centered statistic

\[
S = \sqrt{N} \left( \hat{Q} - \frac{\hat{\sigma}^2}{T} \left( 1 + \hat{\gamma}_1^*(\hat{F}'\hat{F}/T)^{-1}\hat{\gamma}_1^* \right) \right).
\]  

(37)

The following theorem provides the limiting distribution of \( S \) under \( H_0 : e_i = 0 \) for all \( i \). The centering is asked for by (35). In fact, one can easily verify that by considering the population ex-post pricing errors \( e_i^P \) one obtains:

\[
\frac{1}{N} \sum_{i=1}^{N} (e_i^P)^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2 + Q'(\frac{1}{N} \sum_{i=1}^{N} e_i e_i')Q + o_p(1) = \frac{1}{N} \sum_{i=1}^{N} e_i^2 + \sigma^2 Q'Q + o_p(1).
\]  

(38)

Therefore, even under \( H_0 : e_i = 0 \) for all \( i \), the average of the squared (population) ex-post pricing errors \( e_i^P \) one obtains:

\[
\frac{1}{N} \sum_{i=1}^{N} e_i^2 + \sigma^2 Q'Q + o_p(1)
\]

Theorem 3  
Under Assumptions 1–6 (listed in Appendix A), implying that \( H_0 : e_i = 0 \) holds for all \( i \), we have

\[
S \xrightarrow{d} N(0, V),
\]  

(39)

where \( V = Z_Q'U\epsilon Z_Q \) and \( Z_Q = (Q \otimes Q) - \frac{\text{vec}(M)}{T-K-1}Q'Q \).

Proof: See Appendix C and Lemmas 1 to 5 in Appendix B.

The expression for the asymptotic variance of the test in (39) is rather simple. This variance can be consistently estimated by replacing \( Q \) with \( \hat{Q} \) and \( U\epsilon \) with \( \hat{U}\epsilon \). Specifically, using Theorem 2 and Lemma 6 in Appendix B, we have

\[
\hat{Z}_Q'\hat{U}\epsilon \hat{Z}_Q \xrightarrow{p} Z_Q'U\epsilon Z_Q,
\]  

(40)

where

\[
\hat{Z}_Q = (\hat{Q} \otimes \hat{Q}) - \frac{\text{vec}(M)}{T-K-1}\hat{Q}'\hat{Q}.
\]  

(41)
Then, under $H_0$, it follows that
\[ S^* = \frac{S}{(\tilde{Z}_Q' \tilde{U}_Q \tilde{Z}_Q)^{1/2}} \xrightarrow{d} N(0,1). \] (42)

It turns out that our test will have power when $e_i^2$ is greater than zero for the majority of the assets.\footnote{To be precise, the pricing errors $\epsilon_i$ can be zero for only a number $N_0$ of assets, such that $N_0/N \rightarrow 0$ as $N \rightarrow \infty$. This condition allows $N_0$ to diverge, although not too fast.}

In Section IA.1 of the Internet Appendix, we undertake a Monte Carlo simulation experiment calibrated to real data in order to determine whether our test possesses desirable size and power properties.

Our test is immune to re-packaging, in particular, its distribution under the null hypothesis of correct model specification will not depend on an arbitrary linear combination of the test assets. Typically, to allow for re-packaging one should standardize the pricing errors by $\Sigma^{-1}$, which in our context will be unfeasible anyway due to having a fixed $T$. See Lewellen, Nagel and Shanken (2010) for more formal arguments.\footnote{Consider a (non-singular) linear combination of the test assets, such as $AR_t$ for an arbitrary non-singular matrix $A$, implying that the residuals satisfy $AC_t$. Then, the ex-post pricing errors would become $Ae^P$ and their squared sum $e^P A' A e^P$, depending on the choice of the matrix $A$. Instead, the weighted sum, by the inverse of the residual covariance matrix, becomes $e^P A'(A\Sigma A')^{-1} A e^P = e^P A'(A')^{-1} \Sigma^{-1} A^{-1} A e^P = e^P \Sigma^{-1} e^P$, regardless of the chosen $A$. However, although changes in $A$ lead to corresponding changes in $\sigma^2$ and $e^P$, our test statistic $S^*$ continues to have zero mean (and unit variance), and, in fact, the same asymptotic distribution, for any arbitrary $A$. The power properties of $S^*$, viz. the behaviour under the alternative hypothesis, will be affected by changes in $A$.}

6 Conclusion

This paper is concerned with estimation of risk premia and testing of beta-pricing models when data is available for a large cross-section of securities $N$ but only for a limited number of time periods. Because in this context the CSR OLS estimator of the risk premia is asymptotically biased and inconsistent, the focus of the paper is on the bias-adjusted estimator of the ex-post risk premia proposed by Shanken (1992). In terms of estimation, we demonstrate that the Shanken estimator exhibits desirable properties, such as $\sqrt{N}$-consistency and asymptotic normality, as $N$ diverges. In terms of testing, building on the pricing errors stemming from the Shanken estimator, we propose a new test of the no-arbitrage asset pricing restriction and establish its asymptotic distribution (assuming that the restriction holds) as $N$ diverges. Finally, we show how our results can be extended to deal with the more realistic case of unbalanced panels, allowing us to take...
advantage of the large cross-sections of stocks existing only for certain time periods. Monte Carlo simulations corroborate our theoretical finding, both in terms of estimation and in terms of testing for the asset pricing restriction.

We apply our large-$N$ methodology to empirically investigate the performance of some prominent asset pricing specifications using individual stock return data, namely a monthly data set of about 3,500 individual stocks, from the Center for Research in Security Prices (CRSP), from January 1966 until December 2013. We consider three beta-pricing models: the CAPM, the three-factor model of Fama and French (1993), and the five-factor model recently proposed by Fama and French (2015). The results are striking: for all the beta-pricing models under consideration, even for a short time window (three years), our methodology shows clearly that all the estimated risk premia are statistically significant across time, with few exceptions. Our results are completely at odds with the results obtained using the traditional approach, which relies on Shanken’s (1992) correct standard errors when $T$ is large, in which the estimated risk premia turn out to be often non-significantly different from zero. Technically speaking, our methodology only requires $T$ to be slightly larger than the number of factors corresponding to the beta-pricing model under consideration. This allows for exploiting a considerably large cross-sectional size $N$. For example, any $T > K$ suffices when estimating a beta-pricing model based on $K$ observed factors. Likewise, in terms of testing the validity of some specific beta-pricing model, our large-$N$ test is able to reject the CAPM at conventional significance levels even for the short time window of three years ($T = 36$ months) whereas, using the same data suitably grouped, we are unable to reject the CAPM using the Gibbons et al. (1989) methodology.

Appendix A: Assumptions

All the limits are taken for $N \to \infty$. In addition, the expectation operator used throughout these appendixes has to be understood as conditional of $F$.

**Assumption 1.** (loadings) As $N \to \infty$,

$$\frac{1}{N} \sum_{i=1}^{N} \beta_i \to \mu_{\beta},$$

(A.1)

---

29 This outcome is a combination of two elements, the extremely small standard errors associated with a large-$N$ and the bias-correction that leads to an increment, up to 50%, of the risk premia estimates.
\[
\frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' \rightarrow \Sigma_\beta,
\]
(A.2)

where \(\Sigma_\beta\) is a finite symmetric and positive-definite matrix.

The first part of Assumption 1 states that the limiting cross-sectional average of the betas exists, while the second part states that the limiting cross-sectional average of squared betas exists and is a symmetric and positive-definite matrix. To simplify the exposition of our theory, we are not assuming the \(\beta_i\) to be random. See Gagliardini et al. (2016) for a beta-pricing model with random betas. By construction, this implies that the \(\beta_i\) are cross-sectionally unrelated to any other characteristics of the returns’ distribution, in particular to the returns’ idiosyncratic innovations \(\epsilon_{it}\). In Section 3.2 we discuss the consequences of relaxing this assumption.

**Assumption 2.** (residuals; see Shanken (1992), Assumption 1) Assume that the vector \(\epsilon_t\) is independently and identically distributed (i.i.d.) over time with

\[
E[\epsilon_t | F] = 0_N
\]

and

\[
\text{Var}[\epsilon_t | F] = \begin{pmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN}
\end{pmatrix} = \Sigma \quad \text{is of rank } N,
\]

(A.4)

where \(F = [f_1, \ldots, f_T]'\) is the \(T \times K\) matrix of factors, \(0_N\) is a \(N\)-vector of zeros, and \(\sigma_{ij}\) denotes the \((i,j)\)th element of \(\Sigma\), with \(i, j = 1, \cdots, N\). Hereafter we denote \(\sigma_{ii}^2 = \sigma_i^2\).

The i.i.d. assumption over time is common to many studies, including in particular Shanken (1992). Noticeably, our large \(N\) asymptotic theory permits the \(\epsilon_{it}\), in principle, to be arbitrarily correlated across time. However, the i.i.d. assumption could be relaxed only at the cost of more cumbersome derivations and formulations. Conditions (A.3) and (A.4) are implied if the factors \(f_t\) and the innovations \(\epsilon_s\) are mutually independent for any \(s, t\). Noticeably, (A.4) is not imposing any structure on the elements of \(\Sigma\). In particular, we are not imposing the returns \(R_{it}\) to be uncorrelated across assets or to exhibit the same variance. Instead, we allow for a substantial degree of heterogeneity in the cross-section of stock returns. Although the expression for \(\Sigma\) is here left unspecified, obviously our asymptotic theory limits the degree of cross-correlation between the residuals \(\epsilon_{it}\) (see Assumption 5 below). Essentially, we will require that the sum of the \(|\sigma_{ij}|\) across
every row (or column) of \( \Sigma \) is bounded, that is \( \sup_{1 \leq j \leq N} \sum_{i=1}^{N} |\sigma_{ij}| \leq C < \infty \). This condition is slightly stronger than (and thus implies) the corner-stone assumption of the APT which requires \( \Sigma \) to have bounded maximum eigenvalue (see Chamberlain and Rothschild, 1983).

Regarding the observed factors, we require very minimal assumptions because our asymptotic analysis holds conditional on the factors realization \( F \), with a fixed \( T \). The following is the only assumption we make.

**Assumption 3. (factors)** Assume that \( E[f_t] = E[f] \) does not vary over time. Moreover, for every \( T \geq K \),

\[
\hat{F}'\hat{F} \quad \text{is of full rank,}
\]

where \( \hat{F} = (I_T - \frac{1_T1_T'}{T})F = F - 1_T\bar{f}' \) and \( I_T, 1_T \) are the \( T \times T \) identity matrix and the \( T \)-dimensional column vector of ones, respectively, implying that \( \bar{f} \) is the sample mean of the \( f_t \).

Finally, to close the asset pricing model, one must postulate the form of no-arbitrage required. All the results developed in this paper assume that exact pricing holds, which, in view of the constant mean assumption for \( f_t \), can be expressed as follows:

**Assumption 4. (exact pricing)**

\[
E[R_t] = X\Gamma, \tag{A.5}
\]

where \( X = [1_N, B] \) is assumed to be of full column rank for every \( N \), \( 1_N \) is the \( N \)-dimensional column vector of ones, and \( \Gamma = [\gamma_0, \gamma_1'] \) is a vector consisting of the zero-beta rate \( (\gamma_0) \) and ex-ante risk premia \( (\gamma_1) \) associated to the \( K \) factors.

When the model is misspecified, Assumption 4 is not satisfied and the \( N \)-vector of pricing errors, \( e = E[R_t] - X\Gamma \), will be different from the \( N \)-dimensional zero vector. Throughout this paper, we will only explore non-zero pricing errors when providing a sufficient condition for statistical power of our asset pricing test \( S^* \).

**Assumption 5. (Idiosyncratic component)** We require

\[ (i) \quad \frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) = o \left( \frac{1}{\sqrt{N}} \right) \tag{A.6} \]

with \( 0 < \sigma^2 < \infty \).
(ii) \[
\sum_{i,j=1}^{N} |\sigma_{ij}| \mathbb{1}_{\{i\neq j\}} = o(N), \tag{A.7}
\]
where \(\sigma_{ij} = E[\epsilon_{it}\epsilon_{jt}]\) and \(\mathbb{1}_{\{\cdot\}}\) denotes the indicator function.

(iii) \[
\frac{1}{N} \sum_{i=1}^{N} \mu_{4i} \rightarrow \mu_4 \tag{A.8}
\]
with \(0 < \mu_4 < \infty\) and \(\mu_{4i} = E[\epsilon_{it}^4]\).

(iv) \[
\frac{1}{N} \sum_{i=1}^{N} \sigma_{4i}^4 \rightarrow \sigma_4 \tag{A.9}
\]
with \(0 < \sigma_4 < \infty\).

(v) \[
\sup_{i} \mu_{4i} \leq C < \infty \tag{A.10}
\]
for a generic constant \(C\).

(vi) \[
E[\epsilon_{it}^3] = 0. \tag{A.11}
\]

(vii) \[
\frac{1}{N} \sum_{i=1}^{N} \kappa_{4,\cdot\cdot\cdot\cdot} \rightarrow \kappa_4 \tag{A.12}
\]
with \(0 \leq |\kappa_4| < \infty\), where \(\kappa_{4,\cdot\cdot\cdot\cdot} = \kappa_4(\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it})\) denotes the fourth-order cumulant of the random variables \(\{\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it}\}\).

(viii) For every \(3 \leq h \leq 8\), all the mixed cumulants of order \(h\) are such that
\[
\sup_{i_1} \sum_{i_2, \ldots, i_h=1}^{N} |\kappa_{h,i_1i_2\ldots i_h}| = o(N) \tag{A.13}
\]
for at least one \(i_j\) \((2 \leq j \leq h)\) different from \(i_1\).
Assumption 5 essentially describes the cross-sectional behavior of the model disturbances. Assumption 5(i) limits the cross-sectional heterogeneity of the return conditional variances. Assumption 5(ii) implies that the conditional correlation among asset returns is sufficiently weak. In particular, Assumptions 5(ii) and 5(v) imply that \( \sup_i \sum_j^N |\sigma_{ij}| \leq C < \infty \), which in turn implies that the maximum eigenvalue of the conditional covariance of asset returns is bounded. The latter is the most common assumption in factor pricing models such as the APT (see, e.g., Chamberlain and Rothschild, 1983). In Assumption 5(iii), we simply assume the existence of the limit of the conditional fourth moment average across assets. In Assumption 5(iv), the magnitude of \( \sigma_4 \) reflects the degree of cross-sectional heterogeneity of the conditional variance of asset returns. Assumption 5(v) is a bounded fourth moment condition uniform across assets, which implies that \( \sup_i \sigma_i^2 \leq C < \infty \). Assumption 5(vi) is a convenient symmetry assumption, but it is not strictly necessary for our results. It could be relaxed at the cost of a more cumbersome notation. Assumption 5(vii) allows for non-Gaussianity of asset returns when \(|\kappa_4| > 0\). For example, this assumption is satisfied when the marginal distribution of asset returns is a Student \( t \) with degrees of freedom greater than four. However, when estimating the asymptotic covariance matrix of the Shanken estimator one needs to set \( \kappa_4 = 0 \) merely for identification purposes, as indicated below (cf Lemma 6 in Appendix B).

**Assumption 6.** (Convergence in distribution)

(i) \[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i \xrightarrow{d} \mathcal{N} \left( 0_T, \sigma^2 I_T \right).
\] (A.14)

(ii) \[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec} (\epsilon_i' \epsilon_i' - \sigma^2_i I_T) \xrightarrow{d} \mathcal{N} (0_{I_2}, U_{\epsilon}).
\] (A.15)

(iii) For a generic \( T \)-dimensional column vector \( C_T \),

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( C_T' \otimes \left( \frac{1}{\beta_i} \right) \right) \epsilon_i \xrightarrow{d} \mathcal{N} (0_{K+1}, V_c),
\] (A.16)

where \( V_c = c \sigma^2 \Sigma_X \) and \( c = C_T' C_T \). In particular, \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (C_T' \otimes \beta_i) \epsilon_i \xrightarrow{d} \mathcal{N} (0_K, V_c^\dagger) \), where \( V_c^\dagger = c \sigma^2 \Sigma_\beta \).

Primitive conditions for Assumption 6 can be derived but they will substantially raise the level of technicality of the proofs. Details are available upon request.
Appendix B: Lemmas

Lemma 1

(i) Under Assumptions 2 to 5, we have
\[
\hat{\sigma}^2 - \sigma^2 = O_p \left( \frac{1}{\sqrt{N}} \right).
\]  
(B.1)

(ii) In addition, under Assumption 6, we have
\[
\sqrt{N} (\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N \left( 0, \sigma^2 \right).
\]  
(B.2)

Proof

(i) Rewrite \(\hat{\sigma}^2 - \sigma^2\) as
\[
\hat{\sigma}^2 - \sigma^2 = \left( \hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right) + \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 - \sigma^2 \right)
\]  
(B.3)

by Assumption 5(i). Moreover,
\[
\hat{\sigma}^2 - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = \frac{\text{tr} (M\epsilon'\epsilon)}{N(T-K-1)} - \frac{\text{tr} (M)}{T-K-1} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2
\]  
(B.4)

As for the second term on the right-hand side of (B.4), we have
\[
\frac{\text{tr} (\epsilon'\epsilon') - T \sum_{i=1}^{N} \sigma_i^2}{N(T-K-1)} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it}^2 - \sigma_i^2)}{N(T-K-1)}
\]  
(B.5)

As for the first term on the right-hand side of (B.4), we have
\[
\frac{\text{tr} \left( P \left( \sum_{i=1}^{N} \sigma_i^2 I_T - \epsilon'\epsilon' \right) \right)}{N(T-K-1)} = \frac{\sum_{t=1}^{T} d_{it} (D'D)^{-1} D' \left( \sum_{i=1}^{N} \sigma_i^2 t_i - \sum_{i=1}^{N} \epsilon_i \epsilon_{it} \right)}{N(T-K-1)}
\]  
(B.6)
where \( t \) is a \( T \)-vector with 1 in the \( t \)-th position and zeros elsewhere, \( d_t \) is the \( t \)-th row of \( D \), and \( p_t = d_t (D' D)^{-1} D' \). Since (B.6) has zero mean, we only need to consider its variance to determine the rate of convergence. We have

\[
\text{Var} \left( \frac{\sum_{t=1}^{T} p_t \left( \sum_{i=1}^{N} \sigma_i^2 t_i - \sum_{i=1}^{N} \epsilon_i \epsilon_{it} \right)}{N(T-K-1)} \right) = \frac{1}{N^2(T-K-1)^2} E \left[ \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t \left( \sigma_i^2 t_i - \epsilon_i \epsilon_{it} \right) \left( \sigma_j^2 t_s - \epsilon_j \epsilon_{js} \right)' p_s' \right]
\]

\[
= \frac{1}{N^2(T-K-1)^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t E \left[ \left( \sigma_i^2 t_i - \epsilon_i \epsilon_{it} \right) \left( \sigma_j^2 t_s - \epsilon_j \epsilon_{js} \right)' p_s' \right]. \quad (B.7)
\]

Moreover, we have

\[
E \left[ \left( \sigma_i^2 t_i - \epsilon_i \epsilon_{it} \right) \left( \sigma_j^2 t_s - \epsilon_j \epsilon_{js} \right)' \right] = E \left[ \sigma_i^2 \sigma_j^2 t_i t_s' + \epsilon_i \epsilon_j' \epsilon_{it} \epsilon_{is} - \sigma_i^2 t_i t_s' \epsilon_j \epsilon_{js} - \sigma_j^2 \epsilon_{it} \epsilon_{is} t_s' \right]
\]

\[
= \begin{cases} 
\mu_{4t} t_i t_i' + \sigma_i^4 (I_T - 2 t_i t_i') & \text{if } i = j, \ t = s \\
\kappa_{4,ij} t_i t_i' + \sigma_i^2 (I_T + t_i t_i') & \text{if } i \neq j, \ t = s \\
\sigma_i^4 t_i t_i' & \text{if } i = j, \ t \neq s \\
\sigma_{ij}^2 t_i t_i' & \text{if } i \neq j, \ t \neq s.
\end{cases} \quad (B.8)
\]
It follows that
\[
\text{Var} \left( \frac{\sum_{t=1}^{T} p_t \left( \sum_{i=1}^{N} \sigma_i^2 t_t - \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right)}{N(T-K-1)} \right) = \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} p_t \left( \mu_{4i} t_t' + \sigma_i^4 (I_T - 2t_t') \right) p_t' + \frac{1}{N^2(T-K-1)^2} \sum_{i,j \neq j}^{N} p_{i} t_t' \sum_{s \neq s}^{N} p_{s} t_s' p_s' + \frac{1}{N^2(T-K-1)^2} \sum_{i,j \neq j}^{N} \sigma_i^2 \sum_{t \neq s} \sum_{s \neq s} p_{i} t_t' \sum_{t \neq s} p_{s} t_s' p_s' = O \left( \frac{1}{N^2} \right) \]  

by Assumptions 5(ii), 5(iii), 5(iv), and 5(viii), which implies that the first term on the right-hand side of (B.4) is \( O_p \left( \frac{1}{\sqrt{N}} \right) \). Putting the pieces together concludes the proof of part (i).

(ii) Using Assumption 5(i) and the properties of the vec operator, we can write \( \sqrt{N}(\hat{\sigma}^2 - \sigma^2) \) as
\[
\sqrt{N}(\hat{\sigma}^2 - \sigma^2) = \frac{1}{T-K-1} \text{vec}(M)' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec} \left( \epsilon_i \epsilon_i' - \sigma_i^2 I_T \right) + o(1). \tag{B.10}
\]

The desired result then follows from using Assumption 6(ii). This concludes the proof of part (ii).

**Lemma 2** Let
\[
\Lambda = \begin{bmatrix} 0 & 0' \cr 0_K & \sigma^2 (F'F)^{-1} \end{bmatrix}. \tag{B.11}
\]

(i) Under Assumptions 1 to 5, we have
\[
\hat{X}' \hat{X} = O_p(N). \tag{B.12}
\]

In addition, under Assumption 6, we have
\[ \hat{\Sigma}_X \overset{p}{\to} \Sigma_X + \Lambda, \tag{B.13} \]

33
and

\[(\hat{X} - X)'(\hat{X} - X) \overset{p}{\to} \Lambda.\]  

(B.14)

**Proof**

(i) Consider

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

(B.15)

Then, we have

\[\hat{B}'1_N = \sum_{i=1}^{N} \hat{\beta}_i = \sum_{i=1}^{N} \beta_i + \mathcal{P}' \sum_{i=1}^{N} \epsilon_i.\]  

(B.16)

Under Assumptions 2 to 5,

\[
\text{Var} \left( \sum_{t=1}^{T} \sum_{i=1}^{N} \epsilon_{it} (f_t - \bar{f}) \right) = \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} (f_t - \bar{f})(f_s - \bar{f})'E[\epsilon_{it}\epsilon_{js}] \\
\leq \sum_{t=1}^{T} \sum_{i,j=1}^{N} (f_t - \bar{f})(f_t - \bar{f})'|\sigma_{ij}| \\
= O \left( N\sigma^2 \sum_{t=1}^{T} (f_t - \bar{f})(f_t - \bar{f})' \right) = O(NT). \]  

(B.17)

Using Assumption 1, we have

\[\hat{B}'1_N = O_p \left( N + \left( \frac{N}{T} \right)^{\frac{1}{2}} \right) = O_p(N).\]  

(B.18)

Next, consider

\[
\hat{B}'\hat{B} = \sum_{i=1}^{N} \hat{\beta}_i \hat{\beta}_i' \\
= \sum_{i=1}^{N} (\beta_i + \mathcal{P}'\epsilon_i) (\beta_i' + \epsilon_i'\mathcal{P}) \\
= \sum_{i=1}^{N} \beta_i \beta_i' + \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} \\
+ \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \left( \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P}. \]  

(B.19)
By Assumption 1,
\[ \sum_{i=1}^{N} \beta_i \beta_i' = O(N). \]  
(B.20)

Using similar arguments as for (B.17),
\[ \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \beta_i' \right) = O_p \left( \left( \frac{N}{T} \right)^{\frac{1}{2}} \right) \]  
(B.21)

and
\[ \left( \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} = O_p \left( \left( \frac{N}{T} \right)^{\frac{1}{2}} \right). \]  
(B.22)

For \( \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} \), consider its central part and take the norm of its expectation. Using Assumptions 2 to 5,
\[
\left\| E \left[ \tilde{F}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \tilde{F} \right] \right\|
\]
\[
= \left\| E \left[ \sum_{t,s=1}^{T} \sum_{i=1}^{N} (f_t - \bar{f})(f_s - \bar{f})' \epsilon_{it} \epsilon_{is} \right] \right\|
\]
\[
\leq \sum_{t,s=1}^{T} \sum_{i=1}^{N} \left\| (f_t - \bar{f})(f_s - \bar{f})' \epsilon_{it} \epsilon_{is} \right\| E \left[ |\epsilon_{it} \epsilon_{is}| \right]
\]
\[
= \sum_{i=1}^{T} \sum_{i=1}^{N} \left\| (f_t - \bar{f})(f_t - \bar{f})' \sigma_i^2 \right\|
\]
\[
= O \left( N \sigma^2 \sum_{i=1}^{T} \left\| (f_t - \bar{f})(f_t - \bar{f})' \right\| \right) = O(NT). \]  
(B.23)

Then, we have
\[ \mathcal{P}' \left( \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} = O_p \left( \frac{N}{T} \right) \]  
(B.24)

and
\[ \hat{B}' \hat{B} = O_p \left( N + \left( \frac{N}{T} \right)^{\frac{1}{2}} + \frac{N}{T} \right) = O_p(N). \]  
(B.25)

This concludes the proof of part (i).

(ii) Using part (i) and under Assumption 2 to 6, we have
\[ N^{-1} \hat{B}' 1_N = \frac{1}{N} \sum_{i=1}^{N} \beta_i + O_p \left( \frac{1}{\sqrt{N}} \right) \]  
(B.26)
and

\[ N^{-1} \hat{B}' \hat{B} = \frac{1}{N} \sum_{i=1}^{N} \beta_i \beta_i' + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \hat{\beta}_i' + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) \mathcal{P} + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} \]

\[ + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \beta_i' \right) + \mathcal{P}' \left( \frac{1}{N} \sum_{i=1}^{N} \beta_i \epsilon_i' \right) \mathcal{P} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i \hat{\beta}_i' + \sigma^2 \mathcal{P}' \mathcal{P} + O_p \left( \frac{1}{\sqrt{N}} \right) + o \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{N}} \right). \]

(B.27)

Assumption 1 concludes the proof of part (ii).

(iii) Note that

\[ \frac{(\hat{X} - X)'(\hat{X} - X)}{N} = \frac{1}{N} \left[ 0' \right. \left( \hat{B} - B \right)'] \left[ 0_N, \ (\hat{B} - B) \right] \]

\[ = \left[ \begin{array}{c} 0 \\ 0_K \\ \mathcal{P}' \epsilon^2 \mathcal{P} \end{array} \right], \]

where \( 0_N \) is an \( N \)-vector of zeros. As in part (ii) we can write

\[ \frac{\epsilon' \epsilon'}{N} = \frac{1}{N} \sum_{i=1}^{N} (\epsilon_i \epsilon_i' - \sigma_i^2 I_T) + \left( \frac{1}{N} \sum_{i=1}^{N} (\sigma_i^2 - \sigma^2) I_T + \sigma^2 I_T \right). \]

(B.29)

Assumptions 5(i) and 6(ii) conclude the proof since

\[ \mathcal{P}' \frac{\epsilon' \epsilon'}{N} \mathcal{P} = \sigma^2 \mathcal{P}' \mathcal{P} + O_p \left( \frac{1}{\sqrt{N}} \right) + o \left( \frac{1}{\sqrt{N}} \right). \]

(B.30)

Lemma 3

(i) Under Assumptions 1 to 5, we have

\[ X' \hat{\epsilon} = O_p \left( \sqrt{N} \right). \]

(B.31)
(ii) In addition, under Assumption 6, we have

\[
\frac{1}{\sqrt{N}} X' \epsilon \xrightarrow{d} N(0_{K+1}, V). \tag{B.32}
\]

Proof

(i) We have

\[
X' \epsilon = \frac{1}{T} \sum_{t=1}^{T} \left[ \begin{array}{c} 1' \cr B' \end{array} \right] \epsilon_t \tag{B.33}
\]

and

\[
\text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} 1' \epsilon_t \right) = \frac{1}{T^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[\epsilon_{it} \epsilon_{js}] \leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} |\sigma_{ij}| = O \left( \frac{NT}{T^2} \sigma^2 \right) = O \left( N \right). \tag{B.34}
\]

Moreover, using Assumptions 1 and 5(ii),

\[
\text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} B' \epsilon_t \right) = \frac{1}{T^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[\epsilon_{it} \epsilon_{js}] |\beta_i | |\beta'_j | \leq \frac{1}{T^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} |\beta_i | |\beta'_j | \sigma_{ij} = O \left( \frac{NT}{T^2} \sigma^2 \right) = O \left( N \right). \tag{B.35}
\]

Putting the pieces together, \( X' \epsilon = O_p \left( \sqrt{N} \right) \). This concludes the proof of part (i).

(ii) We have

\[
\frac{1}{\sqrt{N}} X' \epsilon = \frac{1}{\sqrt{N}} X' \epsilon \frac{1}{T} \sum_{t=1}^{T} \left[ \begin{array}{c} 1' \cr B' \end{array} \right] \epsilon_t = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{1'}{T} \otimes \left[ \begin{array}{c} 1' \cr \beta_i \end{array} \right] \right) \epsilon_i. \tag{B.36}
\]

Assumption 6(iii) concludes the proof of part (ii).
Lemma 4

(i) Under Assumptions 2 to 5, we have

\[ (\hat{X} - X)'X\Gamma^P = O_p(\sqrt{N}). \]  \hfill (B.37)

(ii) In addition, under Assumption 6, we have

\[ \frac{1}{\sqrt{N}}(\hat{X} - X)'X\Gamma^P \xrightarrow{d} N(0_{K+1}, K), \]  \hfill (B.38)

where

\[ K = \sigma^2 \begin{bmatrix} 0 & 0' \\ 0_K & \Gamma^P\Sigma_X\Gamma^P(\tilde{F}'\tilde{F})^{-1} \end{bmatrix}. \]  \hfill (B.39)

Proof

(i) We have

\[ (\hat{X} - X)'X\Gamma^P = \left[ \begin{array}{c} 0' \\ \mathcal{P}'\epsilon \end{array} \right] X\Gamma^P. \]  \hfill (B.40)

Using similar arguments as for (B.17) concludes the proof of part (i).

(ii) Using the properties of the vec operator

\[
\frac{1}{\sqrt{N}}(\hat{X} - X)'X\Gamma^P = \frac{1}{\sqrt{N}} \left[ \begin{array}{c} 0 \\ \mathcal{P}'\epsilon 1_N \\ \mathcal{P}'\epsilon B \end{array} \right] \left[ \begin{array}{c} \gamma_0 \\ \gamma_1^P \end{array} \right] \\
= \frac{1}{\sqrt{N}} \left[ \begin{array}{c} 0 \\ \mathcal{P}'\epsilon X\Gamma^P \end{array} \right] \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ \Gamma^P \left[ \begin{array}{c} 0' \\ 1, \beta_i \end{array} \right] \otimes \mathcal{P}' \right] \epsilon_i. \]  \hfill (B.41)

Using Assumption 6(iii) concludes the proof of part (ii).

Lemma 5

(i) Under Assumptions 2 to 5, we have

\[ (\hat{X} - X)'\tilde{\epsilon} = O_p(\sqrt{N}). \]  \hfill (B.42)

(ii) In addition, under Assumption 6, we have

\[ \frac{1}{\sqrt{N}}(\hat{X} - X)'\tilde{\epsilon} \xrightarrow{d} N(0_{K+1}, \mathcal{W}). \]  \hfill (B.43)
Proof

(i) 
\[ (\hat{X} - X)'\hat{\epsilon} = \begin{bmatrix} 0 \\ \mathcal{P}'\hat{\epsilon} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathcal{P}'\epsilon' \mathcal{T} \end{bmatrix} \]
\[ = \begin{bmatrix} \mathcal{P}' \left( \epsilon' - \sum_{i=1}^{N} \sigma_i^2 I_T \right) + \left( \sum_{i=1}^{N} \sigma_i^2 - N\sigma^2 \right) I_T \end{bmatrix} \mathcal{T} \]
\[ = O_p(\sqrt{N}) \quad \text{(B.44)} \]

by Assumption 5.

(ii) 
\[ \frac{1}{\sqrt{N}}(\hat{X} - X)'\hat{\epsilon} = \frac{1}{\sqrt{N}} \begin{bmatrix} 0' \mathcal{T} \\ \mathcal{P}' \end{bmatrix} \left( \epsilon' - \sum_{i=1}^{N} \sigma_i^2 I_T \right) + \left( \sum_{i=1}^{N} \sigma_i^2 - N\sigma^2 \right) I_T \mathcal{T} \]
\[ = \begin{bmatrix} \mathcal{T} \\ \mathcal{T} \mathcal{T} \otimes \mathcal{P}' \end{bmatrix} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \text{vec}(\epsilon_i' - \sigma_i^2 I_T) + o(1). \quad \text{(B.45)} \]

The \( o(1) \) term in (B.45) is due to Assumption 5(i). Using Assumption 6(ii) concludes the proof of part (ii).

Lemma 6 Under Assumption 5 and the identification assumption \( \kappa_4 = 0 \), we have 
\[ \hat{\sigma}_4 \xrightarrow{p} \sigma_4. \quad \text{(B.46)} \]

Proof

We need to show that (i) \( E(\hat{\sigma}_4) \rightarrow \sigma_4 \) and (ii) \( \text{Var}(\hat{\sigma}_4) = O \left( \frac{1}{N} \right) \).

(i) By Assumptions 5(iv), 5(vi), and 5(vii), we have
\[ E \left[ \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} \epsilon_{it}^4 \right] = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} E \left[ \epsilon_{it}^4 \right] \]
\[ = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{s_1,s_2,s_3=1}^{s_4} m_{ts_1}m_{ts_2}m_{ts_3}m_{ts_4} E \left[ \epsilon_{is_1} \epsilon_{is_2} \epsilon_{is_3} \epsilon_{is_4} \right] \]
\[ = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} \kappa_{4,iii} \sum_{s=1}^{T} m_{ts}^4 + 3 \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} \sigma_i^4 \left( \sum_{s=1}^{T} m_{ts}^2 \right)^2 \]
\[ + \sigma_4 \sum_{t=1}^{T} \sum_{s=1}^{T} m_{ts}^4 + 3\sigma_4 \sum_{t=1}^{T} \left( \sum_{s=1}^{T} m_{ts}^2 \right)^2, \quad \text{(B.47)} \]
where $\hat{\epsilon}_{it} = i_t' M \epsilon_i$ and $M = [m_{ts}]$ for $t, s = 1, \ldots, T$. Note that

$$\sum_{s=1}^{T} m_{ts}^2 = ||m_t||^2 = i_t'Mi_t$$

$$= i_t(I_T - D(D'D)^{-1}D')i_t$$

$$= 1 - \text{tr} (D(D'D)^{-1}D'i_ti_t')$$

$$= 1 - \text{tr} (Pi_ti_t')$$

$$= 1 - p_{tt}$$

$$= m_{tt}, \quad (B.48)$$

where $p_{tt}$ is the $(t, t)$-element of $P$. Then, we have

$$\sum_{t=1}^{T} \left( \sum_{s=1}^{T} m_{ts}^2 \right)^2 = \sum_{t=1}^{T} m_{tt}^2 = \text{tr} \left( M^{(2)} \right). \quad (B.49)$$

By setting $\kappa_4 = 0$, it follows that

$$E [\hat{\sigma}_4] \to \sigma_4. \quad (B.50)$$

This concludes the proof of part (i).
(ii) As for the variance of \( \hat{\sigma}_4 \), we have

\[
\text{Var}\left( \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it}^4 \right) = \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \text{Cov}\left( \hat{e}_{it}^4, \hat{e}_{js}^4 \right)
\]

\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1,u_2,v_1,v_2} \sum_{u_3,u_4,v_3,v_4=1}^{T} m_{tu_1} m_{tu_2} m_{tv_1} m_{tv_2} m_{su_1} m_{su_2} m_{sv_1} m_{sv_2}
\]

\[
\times \text{Cov}\left( \epsilon_{iu_1} \epsilon_{iu_2} \epsilon_{iu_3} \epsilon_{iu_4}, \epsilon_{jv_1} \epsilon_{jv_2} \epsilon_{jv_3} \epsilon_{jv_4} \right)
\]

\[
= \frac{1}{N^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \sum_{u_1,u_2,v_1,v_2} \sum_{u_3,u_4,v_3,v_4=1}^{T} m_{tu_1} m_{tu_2} m_{tv_1} m_{tv_2} m_{su_1} m_{su_2} m_{sv_1} m_{sv_2}
\]

\[
\times \left( \kappa_8 \left( \epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4} \right) + \sum \kappa_6 \left( \epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2} \right) \text{Cov} \left( \epsilon_{jv_3}, \epsilon_{jv_4} \right) \right)
\]

\[
+ \sum \kappa_4 \left( \epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{jv_1}, \epsilon_{jv_2} \right) \kappa_4 \left( \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_3}, \epsilon_{jv_4} \right)
\]

\[
+ \sum \kappa_4 \left( \epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{jv_1}, \epsilon_{jv_2} \right) \text{Cov} \left( \epsilon_{iu_3}, \epsilon_{iu_4} \right) \text{Cov} \left( \epsilon_{jv_3}, \epsilon_{jv_4} \right)
\]

\[
+ \sum \text{Cov} \left( \epsilon_{iu_1}, \epsilon_{iu_2} \right) \text{Cov} \left( \epsilon_{iu_3}, \epsilon_{jv_1} \right) \text{Cov} \left( \epsilon_{iu_4}, \epsilon_{jv_2} \right) \text{Cov} \left( \epsilon_{jv_3}, \epsilon_{jv_4} \right)
\]

(B.51)

where \( \kappa_4 (\cdot), \kappa_6 (\cdot), \) and \( \kappa_8 (\cdot) \) denote the fourth-, sixth-, and eighth-order mixed cumulants, respectively. By \( \sum_{(\nu_1,\nu_2,\ldots,\nu_k)} \) we denote the sum over all possible partitions of a group of \( K \) random variables into \( k \) subgroups of size \( \nu_1, \nu_2, \ldots, \nu_k \), respectively. As an example, consider \( \sum_{(6,2)} \). \( \sum_{(6,2)} \) defines the sum over all possible partitions of the group of eight random variables \( \{ \epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4} \} \) into two subgroups of size six and two, respectively. Moreover, since \( E [\epsilon_{it}] = E [\epsilon_{it}^3] = 0 \), we do not need to consider further partitions in the above relation.\(^{30}\) Then, under Assumptions 5(i), 5(ii), 5(v), and 5(viii), it follows that

\[
\text{Var}\left( \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{e}_{it}^4 \right) = O \left( \frac{1}{N} \right)
\]

(B.52)

and \( \text{Var} \left( \hat{\sigma}_4 \right) = O \left( \frac{1}{N} \right) \). This concludes the proof of part (ii).

\(^{30}\) According to the theory on cumulants (Brillinger, 1975), evaluation of \( \text{Cov} \left( \epsilon_{iu_1} \epsilon_{iu_2} \epsilon_{iu_3} \epsilon_{iu_4}, \epsilon_{jv_1} \epsilon_{jv_2} \epsilon_{jv_3} \epsilon_{jv_4} \right) \) requires considering the indecomposable partitions of the two sets \( \{ \epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4} \}, \{ \epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4} \} \), meaning that there must be at least one subset that includes an element of both sets.
Appendix C: Proofs of Theorems

Proof of Proposition 1

Consider the class of additive bias-adjusted estimators $\hat{\Gamma}^{bias-adj}$ for $\Gamma^P$

$$\hat{\Gamma}^{bias-adj} = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\Lambda \hat{\Gamma}_{prelim} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{R} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\Lambda \hat{\Gamma}_{prelim},$$

where $\hat{\Gamma}_{prelim}$ denotes any preliminary $\sqrt{N}$-consistent estimator of $\Gamma^P$. Simply imposing the restriction that $\hat{\Gamma}^{bias-adj} = \hat{\Gamma}_{prelim}$, and, rearranging, one gets

$$\left[I_{K+1} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\left[\begin{array}{cc} 0 & 0'_{K,1} \\ 0 & \hat{\sigma}^2(F'F)^{-1} \end{array}\right]\right] \hat{\Gamma}^{bias-adj} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{R},$$

which implies

$$\hat{\Gamma}^{bias-adj} = \left(\hat{\Sigma}_X - \Lambda\right)^{-1}\hat{X}'\hat{R} = \hat{\gamma}^*,$$

that is, one obtains the modified estimator of Shanken (1992).

Proof of Proposition 2 All the limits hold as $N \to \infty$. Consider at first the numerator of the t statistics. By Lemma 2 (ii), Lemma 4 (i) and Lemma 5 (i) one obtains $\hat{\Gamma} = (\hat{\gamma}_0, \hat{\gamma}_1)' = (\Sigma_X + \Lambda)^{-1}\Sigma_X\Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right)$. By the block-wise formula of the inverse of the matrix

$$(\Sigma_X + \Lambda)^{-1} \Sigma_X \Gamma^P = \left(\begin{array}{cc} 1 & \mu'_{\beta} \\ \mu'_{\beta} & \Sigma_{\beta} + C \end{array}\right)^{-1} \left(\begin{array}{cc} 1 & \mu'_{\beta} \\ \mu'_{\beta} & \Sigma_{\beta} \end{array}\right) \Gamma^P = \left(\begin{array}{cc} 1 + \mu'_{\beta}A^{-1}\mu_{\beta} & -\mu'_{\beta}A^{-1} \\ -A^{-1}\mu_{\beta} & A^{-1} \end{array}\right) \Gamma^P

= \left(\begin{array}{cc} 1 & \mu'_{\beta} \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) \\ 0 & A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta}) \end{array}\right) \Gamma^P.

Then

$$(\Sigma_X + \Lambda)^{-1} \Sigma_X \Gamma^P - \Gamma = \left(\begin{array}{cc} 1 & \mu'_{\beta} \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) \\ 0 & A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta}) \end{array}\right) \Gamma^P - \Gamma$$

$$= \left(\begin{array}{cc} 0 & \mu'_{\beta} \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) \\ 0 & -\left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) \end{array}\right) \Gamma + \left(\begin{array}{cc} 1 & \mu'_{\beta} \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) \\ 0 & \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) \end{array}\right) \left(\begin{array}{c} \hat{f} - E(f) \end{array}\right).

(C.3)

Hence $\text{plim} \hat{\gamma}_0 - \gamma_0 = \mu'_{\beta} \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) + \mu'_{\beta}A^{-1}C\gamma_1^P$ and, for every $1 \leq j \leq K$, $\text{plim} \hat{\gamma}_1j - \gamma_{1j} = -\mu'_{\beta} \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right) + \gamma_1j + \left(I_K - A^{-1}(\Sigma_{\beta} - \mu_{\beta}\mu'_{\beta})\right)(\hat{f} - E(f))$. Consider now the behaviour of the denominator of the t statistics. For the (squared) Fama-MacBeth standard errors $(\bar{SE}_{k}^{FM})^2 = N^{-1}\left(I_{k+1,K+1}^\prime \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1}\frac{\hat{X}'\hat{X}}{N} - I_{k+1,K+1}^\prime \right) = O_p(N^{-1})$. In fact, by Lemma 2 (ii)
Proof of Theorem 1

(i) Starting from (11), the modified estimator of Shanken (1992) can be written as

\[
\hat{\Gamma}^* = \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \frac{\hat{X}'R}{N} = \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \frac{\hat{X}'X^P + \hat{\epsilon} - (\hat{X} - X)\Gamma^P}{N}
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left( \frac{\hat{X}'X}{N} \right)^{-1} \Gamma^P + \left( \frac{\hat{X}'X}{N} \right)^{-1} \frac{\hat{X}'N}{N} (\hat{X} - X) \Gamma^P
\]

\[
= \left( I_{K+1} - \left( \frac{\hat{X}'X}{N} \right)^{-1} \right)^{-1} \left( \left( \frac{\hat{X}'X}{N} \right)^{-1} \hat{\epsilon} - \left( \frac{\hat{X}'X}{N} \right)^{-1} \frac{\hat{X}'N}{N} (\hat{X} - X) \Gamma^P \right)
\]

\[
(C.4)
\]

Hence,

\[
\hat{\Gamma}^* - \Gamma^P = \left( \frac{\hat{X}'X}{N} - \hat{\Lambda} \right)^{-1} \left[ \frac{\hat{X}'N}{N} (\hat{X} - X) \Gamma^P + \hat{\Lambda} \Gamma^P \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \frac{\hat{X}'N}{N} (\hat{X} - X) \right] \Gamma^P
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \frac{\hat{X}'N}{N} \gamma_1^P + \hat{\gamma}_1^P \right]
\]

\[
(C.5)
\]

By Lemmas 1(i) and 2(i), \( \left( \hat{\Sigma}_X - \hat{\Lambda} \right) = O_p(1) \). In addition, Lemmas 3(i) and 5(i) imply that

\[
\frac{\hat{X}'\hat{\epsilon}}{N} = \frac{1}{N} (\hat{X} - X)'\hat{\epsilon} + \frac{1}{N} X'\hat{\epsilon},
\]

\[
= O_p \left( \frac{1}{\sqrt{N}} \right)
\]

\[
(C.6)
\]
and Assumption 6(i) implies that
\[ P' \sum_{i=1}^{N} \epsilon_i = O_p \left( \sqrt{N} \right). \]  \hfill (C.7)

Note that
\[ P' \epsilon' \frac{1}{N} P \gamma_1 P - \sigma^2 (F' F)^{-1} \gamma_1 P \]  \hfill (C.8)

can be rewritten as
\[ P' \left( \frac{\epsilon' \epsilon'}{N} - \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 I_T \right) P \gamma_1 P - \left( \sigma^2 - \sigma^2 \right) \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 - \sigma^2 \right) \left( F' F \right)^{-1} \gamma_1 P. \]  \hfill (C.9)

Assumption 6(ii) implies that
\[ P' \left( \frac{\epsilon' \epsilon'}{N} - \sum_{i=1}^{N} \sigma_i^2 I_T \right) P \gamma_1 P = O_p \left( \frac{1}{\sqrt{N}} \right). \]  \hfill (C.10)

Using Lemma 1(i) and Assumption 5(i) concludes the proof of part (i) since \( \hat{\sigma}^2 - \sigma^2 = O_p \left( \frac{1}{\sqrt{N}} \right) \) and \( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 - \sigma^2 = o \left( \frac{1}{\sqrt{N}} \right) \).

(ii) Starting from (C.5), we have
\[
\sqrt{N}(\hat{\Gamma}^* - \Gamma^p) = \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \hat{X}' \hat{\epsilon} \sqrt{N} - \left( \frac{\hat{X}' \hat{\epsilon}}{\sqrt{N}} (\hat{X} - X) \Gamma^p \right) + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \hat{X}' \hat{\epsilon} \sqrt{N} - \left[ \frac{1}{\sqrt{N}} \gamma^p \right] \left[ 0_N, \frac{\epsilon' P}{\sqrt{N}} \right] \Gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \hat{X}' \hat{\epsilon} \sqrt{N} + \frac{1}{\sqrt{N}} \left[ 0' N \gamma^p \right] \frac{\epsilon' P}{\sqrt{N}} - \frac{1}{\sqrt{N}} \left[ \frac{1}{\sqrt{N}} \gamma^p \frac{\epsilon' P}{\sqrt{N}} \right] \gamma^p + \sqrt{N} \hat{\Lambda} \Gamma^p \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \left[ \frac{1}{\sqrt{N}} \gamma^p \right] \left[ \frac{\epsilon' P}{\sqrt{N}} \right] - \frac{1}{\sqrt{N}} \gamma^p \right] + \left[ \frac{\epsilon' P}{\sqrt{N}} \frac{\epsilon' P}{\sqrt{N}} \right] + \gamma^p \left( \hat{\Gamma}^* \hat{\Gamma}^* \right)^{-1} \gamma^p
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \left[ \frac{\epsilon' P}{\sqrt{N}} \frac{\epsilon' P}{\sqrt{N}} \right] + \gamma^p \left( \hat{\Gamma}^* \hat{\Gamma}^* \right)^{-1} \gamma^p \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \left[ \frac{\epsilon' P}{\sqrt{N}} \frac{\epsilon' P}{\sqrt{N}} \right] + \gamma^p \left( \hat{\Gamma}^* \hat{\Gamma}^* \right)^{-1} \gamma^p \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} \left[ \left[ \frac{\epsilon' P}{\sqrt{N}} \frac{\epsilon' P}{\sqrt{N}} \right] + \gamma^p \left( \hat{\Gamma}^* \hat{\Gamma}^* \right)^{-1} \gamma^p \right]
\]

\[
= \left( \hat{\Sigma}_X - \hat{\Lambda} \right)^{-1} (I_1 + I_2). \]  \hfill (C.11)
Using Lemmas 1(i) and 2(ii), we have

\[
\begin{align*}
(\hat{\Sigma}_X - \hat{\Lambda}) & \xrightarrow{p} \left[ \frac{1}{\mu_\beta} \Sigma_\beta + \sigma^2 (\tilde{F}'\tilde{F})^{-1} \right] - \left[ \begin{array}{cc} 0 & 0'_{K} \\ 0 & \sigma^2 (\tilde{F}'\tilde{F})^{-1} \end{array} \right] = \Sigma_X. \quad \text{(C.12)}
\end{align*}
\]

Consider now the terms \(I_1\) and \(I_2\). Both terms have mean zero and, under Assumption 5(vi), they are asymptotically uncorrelated. Assumptions 1, 5(i), 6(i), and 6(iii) imply that

\[
\text{Var}(I_1) = E \left[ \frac{Q'}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \epsilon_j' Q \frac{Q'}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \epsilon_j' (Q \otimes \beta_j') \right] = \sigma^2 \left[ \begin{array}{c} Q' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \epsilon_j' Q \frac{Q'}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \epsilon_j' (Q \otimes \beta_j') \end{array} \right] + o(1)
\]

Next, consider \(I_2\). Since \(\mathcal{P}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^2 Q + \frac{1}{T-K-1} \text{tr} \left( \mathcal{M} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^2 \right) \mathcal{P}' \mathcal{P} \gamma_1' P = 0_K\), we have

\[
\begin{align*}
I_2 & = \mathcal{P}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^2 Q + \frac{1}{T-K-1} \text{tr} \left( \mathcal{M} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sigma_i^2 \right) \mathcal{P}' \mathcal{P} \gamma_1' P \\
& = \mathcal{P}' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Q \otimes \mu_i) I_T + \frac{1}{T-K-1} \text{tr} \left( \mathcal{M} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (Q \otimes \mu_i) I_T \right) \mathcal{P}' \mathcal{P} \gamma_1' P
\end{align*}
\]

Therefore, \(\text{Var}(I_2)\) has the following form:

\[
\text{Var}(I_2) = \left[ \begin{array}{cc} 0' & 0'_{K} \\ 0 & E [I_{22}' I_{22}'] \end{array} \right]. \quad \text{(C.15)}
\]
Under Assumptions 5(i) and 6(ii), we have

\[
E \left[ I_{22}I'_{22} \right] = E \left[ \left( Q \otimes P \right) \left( \vec{\epsilon}_i \vec{\epsilon}_i' - \sigma_i^2 I_T \right) \left( Q \otimes P \right)' \right] \\
\quad + E \left[ \left( Q \otimes P \right) \left( \vec{\epsilon}_j \vec{\epsilon}_j' - \sigma_j^2 I_T \right) \left( Q \otimes P \right)' \right] \\
\quad + E \left[ \left( Q \otimes P \right) \left( \vec{\epsilon}_i \vec{\epsilon}_i' - \sigma_i^2 I_T \right) \left( Q \otimes P \right)' \right] \\
\quad + E \left[ \left( Q \otimes P \right) \left( \vec{\epsilon}_j \vec{\epsilon}_j' - \sigma_j^2 I_T \right) \left( Q \otimes P \right)' \right] \\
\quad \times \gamma \left( Q \otimes P \right) \\
\rightarrow \left[ \left( Q \otimes P \right) + \left( Q \otimes P \right)' \right] U_{\epsilon} \left[ \left( Q \otimes P \right) + \left( Q \otimes P \right)' \right]. \tag{C.16}
\]

Defining \( Z = \left[ \left( Q \otimes P \right) + \frac{\epsilon}{\sqrt{N}} \right] U_{\epsilon} \left[ \left( Q \otimes P \right) + \frac{\epsilon}{\sqrt{N}} \right] \) concludes the proof of part (ii).

**Proof of Theorem 2**

By Theorem 1(i), \( \gamma \rightarrow \gamma \). Lemma 1(i) implies that \( \hat{\Lambda} \) is a consistent estimator of \( \Lambda \). Hence, using Lemma 2(ii), we have \( \hat{\Sigma}_X \rightarrow \Sigma_X \), which implies that \( \hat{V} \rightarrow V \). A consistent estimator of \( W \) requires a consistent estimate of the matrix \( U_{\epsilon} \), which can be obtained using Lemma 6. This concludes the proof of Theorem 2.

**Proof of Theorem 3**

We first establish a simpler, asymptotically equivalent, expression for \( \sqrt{N} \left( \frac{\epsilon}{\sqrt{N}} - \hat{\gamma} \hat{Q}' \hat{Q} \right) \). Then, we derive the asymptotic distribution of this approximation. Consider the sample ex-post pricing errors

\[
\hat{\epsilon} = \bar{R} - \hat{X}\hat{\Gamma}^*.
\]

Starting from \( \bar{R} = \hat{X}\Gamma^P + \eta^P \) with \( \eta^P = \hat{\epsilon} - (\hat{X} - X)\Gamma^P \), we have

\[
\hat{\epsilon} = \hat{X}\Gamma^P + \hat{\epsilon} - (\hat{X} - X)\Gamma^P - \hat{X}\hat{\Gamma}^* \\
= \hat{\epsilon} - \hat{X}((\hat{\Gamma}^* - \Gamma^P) - (\hat{X} - X))\Gamma^P. \tag{C.18}
\]
Then,
\[
\hat{e} P' \hat{e} P = \hat{e} \hat{e}^T + \Gamma P' (\hat{X} - X)' (\hat{X} - X) \Gamma P - 2 (\hat{\Gamma}^* - \Gamma P)' \hat{X} \hat{e} - 2 \Gamma P' (\hat{X} - X)' \hat{e} \\
+ 2 \Gamma P' (\hat{X} - X)' \hat{X} (\hat{\Gamma}^* - \Gamma P) + (\hat{\Gamma}^* - \Gamma P)' \hat{X}' \hat{X} (\hat{\Gamma}^* - \Gamma P).
\]

Note that
\[
\frac{\hat{e} \hat{e}^T}{N} = \frac{1}{T^2} \frac{P' \hat{e} \hat{e}^T P}{N} \xrightarrow{p} \frac{\sigma^2}{T}, \tag{C.19}
\]
and, by Lemma 2(iii),
\[
\Gamma P' (\hat{X} - X)' (\hat{X} - X) \Gamma P = \gamma_1 P' P \hat{e} \hat{e}^T \frac{P}{N} \xrightarrow{p} \sigma^2 \gamma_1 P' (\hat{F}' \hat{F})^{-1} \gamma_1 P. \tag{C.20}
\]

Using Lemmas 3(i) and 5(i) and Theorem 1, we have
\[
\frac{(\hat{\Gamma}^* - \Gamma P)' \hat{X} \hat{e}}{N} = \frac{(\hat{\Gamma}^* - \Gamma P)' (\hat{X} - X)' \hat{e}}{N} + \frac{(\hat{\Gamma}^* - \Gamma P)' \hat{X} \hat{e}}{N} = O_p \left( \frac{1}{N} \right) \tag{C.21}
\]
and
\[
\frac{\Gamma P' (\hat{X} - X)' \hat{e}}{N} = O_p \left( \frac{1}{\sqrt{N}} \right). \tag{C.22}
\]

In addition, using Lemmas 2(i), 2(iii), 4(i), and Theorem 1, we have
\[
\frac{\Gamma P' (\hat{X} - X)' \hat{X} (\hat{\Gamma}^* - \Gamma P)}{N} = \frac{\Gamma P' (\hat{X} - X)' (\hat{X} - X) (\hat{\Gamma}^* - \Gamma P)}{N} + \frac{\Gamma P' (\hat{X} - X)' X (\hat{\Gamma}^* - \Gamma P)}{N} \\
= O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) \tag{C.23}
\]
and
\[
\frac{(\hat{\Gamma}^* - \Gamma P)' \hat{X} (\hat{\Gamma}^* - \Gamma P)}{N} = O_p \left( \frac{1}{N} \right). \tag{C.24}
\]

It follows that
\[
\frac{\hat{e} P' \hat{e} P}{N} \xrightarrow{p} \frac{\sigma^2}{T} + \sigma^2 \gamma_1 P' (\hat{F}' \hat{F})^{-1} \gamma_1 P = \sigma^2 Q' Q. \tag{C.25}
\]

Collecting terms and rewriting explicitly only the ones that are \( O_p \left( \frac{1}{\sqrt{N}} \right) \), we have
\[
\frac{\hat{e} P' \hat{e} P}{N} = \frac{\hat{e} \hat{e}^T}{N} \tag{C.26}
\]
\[
+ \frac{\Gamma P' (\hat{X} - X)' (\hat{X} - X) \Gamma P}{N}, \tag{C.27}
\]
\[
- 2 \frac{\Gamma P' (\hat{X} - X)' \hat{e}}{N}, \tag{C.28}
\]
\[
+ 2 \frac{\Gamma P' (\hat{X} - X)' (\hat{X} - X) (\hat{\Gamma}^* - \Gamma P)}{N}, \tag{C.29}
\]
\[
+ O_p \left( \frac{1}{N} \right). \tag{C.30}
\]
Consider the sum of the three terms in (C.26)–(C.28). Under Assumption 5(i), we have

\[
\frac{\hat{\epsilon}'\epsilon}{N} + \frac{\Gamma_P'(\hat{X} - X)'(\hat{X} - X)\Gamma_P}{N} - 2\frac{\Gamma_P'(\hat{X} - X)'\hat{\epsilon}}{N}
\]

\[
= \frac{1}{T} \epsilon\epsilon' \frac{1}{N} \bar{P} \bar{P}' \gamma_1^p - 2\frac{1}{T} \epsilon\epsilon' \frac{1}{N} \bar{P} \gamma_1^p
\]

\[
= \frac{1}{T} \epsilon\epsilon' \left( \frac{1}{T} \bar{P} \gamma_1^p \right) - \frac{1}{T} \epsilon\epsilon' \frac{1}{N} \bar{P} \gamma_1^p + \gamma_1^p \bar{P}' \epsilon\epsilon' \frac{1}{N} \bar{P} \gamma_1^p
\]

\[
= \frac{1}{T} \epsilon\epsilon' \frac{1}{N} \bar{Q} - \frac{Q}{N} \epsilon\epsilon' \gamma_1^p
\]

\[
= Q \epsilon\epsilon' \frac{1}{N} = Q \left( \epsilon\epsilon' - \sigma^2 I_T \right) Q + \sigma^2 Q'Q + o \left( \frac{1}{\sqrt{N}} \right),
\]

where the \( o \left( \frac{1}{\sqrt{N}} \right) \) term comes from \( (\bar{\sigma}^2 - \sigma^2)Q'Q \). As for the term in (C.29), define

\[
\left( \hat{\Sigma}_X - \bar{\Lambda} \right)^{-1} = \left[ \begin{array}{cc}
\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\
\hat{\Sigma}_{21} & \hat{\Sigma}_{22}
\end{array} \right],
\]

where every block of \( \left( \hat{\Sigma}_X - \bar{\Lambda} \right)^{-1} \) is \( O_p(1) \) by the nonsingularity of \( \Sigma_X \) and Slutsky’s theorem. Using the same arguments as for Theorem 2, we have

\[
2\frac{\Gamma_P'(\hat{X} - X)'(\hat{X} - X)(\hat{\Gamma}^* - \Gamma_P)}{N}
\]

\[
= 2 \left[ \gamma_1^p \bar{P}' \epsilon\epsilon' \frac{1}{N} \bar{P} \hat{\Sigma}_{21} \bar{P}' \epsilon\epsilon' \frac{1}{N} \bar{P} \hat{\Sigma}_{22} \right] \left[ \frac{1}{N} \epsilon\epsilon' \frac{1}{N} \bar{P} \epsilon\epsilon' \frac{1}{N} \bar{P} \bar{P}' Q + Z' \text{vec} \left( \epsilon\epsilon' - \bar{\sigma}^2 I_T \right) \right]
\]

\[
= 2\gamma_1^p \bar{P}' \left( \epsilon\epsilon' - \bar{\sigma}^2 I_T \right) \bar{P} \hat{\Sigma}_{21} \frac{1}{N} \epsilon\epsilon' Q + 2\gamma_1^p \bar{P}' \left( \epsilon\epsilon' - \sigma^2 I_T \right) \bar{P} \hat{\Sigma}_{22} \frac{B' \epsilon\epsilon' Q}{N}
\]

\[
+ 2\gamma_1^p \bar{P}' \left( \epsilon\epsilon' - \bar{\sigma}^2 I_T \right) \bar{P} \hat{\Sigma}_{22} Z' \text{vec} \left( \epsilon\epsilon' \frac{1}{N} - \bar{\sigma}^2 I_T \right)
\]

\[
+ 2\sigma^2 \gamma_1^p \bar{P}' \bar{P} \bar{P} \hat{\Sigma}_{21} \frac{1}{N} \epsilon\epsilon' Q + 2\sigma^2 \gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{22} \frac{B' \epsilon\epsilon' Q}{N}
\]

\[
+ 2\sigma^2 \gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{22} Z' \text{vec} \left( \epsilon\epsilon' \frac{1}{N} - \bar{\sigma}^2 I_T \right) + o_p \left( \frac{1}{N} \right)
\]

\[
= 2\sigma^2 \gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{21} \frac{1}{N} \epsilon\epsilon' Q + 2\sigma^2 \gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{22} \frac{B' \epsilon\epsilon' Q}{N}
\]

\[
+ 2\sigma^2 \gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{22} Z' \text{vec} \left( \epsilon\epsilon' \frac{1}{N} - \bar{\sigma}^2 I_T \right) + o_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{N} \right),
\]

where the two approximations on the right-hand side of the previous expression refer to

\[
2(\bar{\sigma}^2 - \sigma^2)\gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{21} \frac{1}{N} \epsilon\epsilon' Q + 2(\bar{\sigma}^2 - \sigma^2)\gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{22} \frac{B' \epsilon\epsilon' Q}{N}
\]

\[
+ 2(\bar{\sigma}^2 - \sigma^2)\gamma_1^p \bar{P}' \bar{P} \hat{\Sigma}_{22} Z' \text{vec} \left( \epsilon\epsilon' \frac{1}{N} - \bar{\sigma}^2 I_T \right) = o_p \left( \frac{1}{N} \right)
\]
and

\[
2\gamma_1^{P'}\mathcal{P}' \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) \mathcal{P}\hat{\Sigma}_{21} \left( \frac{1_N}{N} \epsilon Q' \right) + 2\gamma_1^{P'}\mathcal{P}' \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) \mathcal{P}\hat{\Sigma}_{22} \left( \frac{B'\epsilon Q}{N} \right) + 2\gamma_1^{P'}\mathcal{P}' \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) \mathcal{P}\hat{\Sigma}_{22} Z' vec \left( \frac{\epsilon' \epsilon}{N} - \bar{\hat{\sigma}}^2 I_T \right) = O_p \left( \frac{1}{N} \right),
\]

(C.35)

respectively. Therefore, we have

\[
\frac{\tilde{e}' e'}{N} = Q' \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) Q + \sigma^2 Q' Q
\]

\[
+2\sigma^2 \gamma_1^{P'}\mathcal{P}' \mathcal{P}\hat{\Sigma}_{21} \left( \frac{1_N}{N} \epsilon Q' \right) + 2\sigma^2 \gamma_1^{P'}\mathcal{P}' \mathcal{P}\hat{\Sigma}_{22} \left( \frac{B'\epsilon Q}{N} \right) + 2\sigma^2 \gamma_1^{P'}\mathcal{P}' \mathcal{P}\hat{\Sigma}_{22} Z' vec \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right) + o \left( \frac{1}{\sqrt{N}} \right). \tag{C.36}
\]

It follows that

\[
\frac{\tilde{e}' e'}{N} - \hat{\sigma}^2 Q' Q = Q' \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) Q - \left( \hat{\sigma}^2 Q' Q - \sigma^2 Q' Q \right)
\]

\[
+2\sigma^2 \gamma_1^{P'}\mathcal{P}' \mathcal{P}\hat{\Sigma}_{21} \left( \frac{1_N}{N} \epsilon Q' \right) + 2\sigma^2 \gamma_1^{P'}\mathcal{P}' \mathcal{P}\hat{\Sigma}_{22} \left( \frac{B'\epsilon Q}{N} \right) + 2\sigma^2 \gamma_1^{P'}\mathcal{P}' \mathcal{P}\hat{\Sigma}_{22} Z' vec \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right) + o \left( \frac{1}{\sqrt{N}} \right). \tag{C.37}
\]

Note that

\[
\hat{\sigma}^2 Q' Q - \sigma^2 Q' Q
\]

\[
= \frac{1}{T} (\hat{\sigma}^2 - \sigma^2) + \hat{\sigma}^2 \gamma_1^{P'} (\tilde{F}'\tilde{F})^{-1} \gamma_1^P - \sigma^2 \gamma_1^{P'} (\tilde{F}'\tilde{F})^{-1} \gamma_1^P
\]

\[
= \frac{1}{T} (\hat{\sigma}^2 - \sigma^2) + (\hat{\sigma}^2 - \sigma^2) \gamma_1^{P'} (\tilde{F}'\tilde{F})^{-1} \gamma_1^P + 2\sigma^2 (\gamma_1^* - \gamma_1^P)' (\tilde{F}'\tilde{F})^{-1} \gamma_1^P + O_p \left( \frac{1}{N} \right)
\]

\[
= (\hat{\sigma}^2 - \sigma^2) \left( \frac{1}{T} + \gamma_1^{P'} (\tilde{F}'\tilde{F})^{-1} \gamma_1^P \right) + 2\sigma^2 (\gamma_1^* - \gamma_1^P)' (\tilde{F}'\tilde{F})^{-1} \gamma_1^P + O_p \left( \frac{1}{N} \right)
\]

\[
= (\hat{\sigma}^2 - \sigma^2) \left( \frac{1}{T} + \gamma_1^{P'} (\tilde{F}'\tilde{F})^{-1} \gamma_1^P \right) + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P}\hat{\Sigma}_{21} \left( \frac{1_N}{N} \epsilon Q' \right) + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P}\hat{\Sigma}_{22} \left( \frac{B'\epsilon Q}{N} \right) + 2\sigma^2 \gamma_1^{P'} \mathcal{P}' \mathcal{P}\hat{\Sigma}_{22} Z' vec \left( \frac{\epsilon' \epsilon}{N} - \hat{\sigma}^2 I_T \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{N} \right) + o \left( \frac{1}{N \sqrt{N}} \right), \tag{C.38}
\]

where \(\sigma^2 (\gamma_1^* - \gamma_1^P)' (\tilde{F}'\tilde{F})^{-1} (\gamma_1^* - \gamma_1^P) + 2 (\hat{\sigma}^2 - \sigma^2) (\gamma_1^* - \gamma_1^P)' (\tilde{F}'\tilde{F})^{-1} \gamma_1^P = O_p \left( \frac{1}{N} \right)\) and \((\hat{\sigma}^2 - \sigma^2)(\gamma_1^* - \gamma_1^P)' (\tilde{F}'\tilde{F})^{-1} (\gamma_1^* - \gamma_1^P)\).
\[ \gamma_1^P y' (\tilde{F}' \tilde{F})^{-1} (\gamma_1^* - \gamma_1^P) = O_p \left( \frac{1}{N \sqrt{N}} \right). \] It follows that

\[
\frac{\hat{e}' \hat{e}}{N} - \hat{\sigma}^2 \hat{Q}' \hat{Q} = \tilde{Q}' \left( \tilde{e} \tilde{e}' - \tilde{\sigma}^2 \tilde{I}_T \right) - (\tilde{\sigma}^2 - \tilde{\sigma}^2) \left( \frac{1}{T} + \gamma_1^P (\tilde{F}' \tilde{F})^{-1} \gamma_1^P \right) + O_p \left( \frac{1}{N \sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right)
\]

where we have condensed \( O_p \left( \frac{1}{N \sqrt{N}} \right) + O_p \left( \frac{1}{N} \right) + o_p \left( \frac{1}{\sqrt{N}} \right) + o_p \left( \frac{1}{\sqrt{N}} \right) \) into the single term \( o_p \left( \frac{1}{\sqrt{N}} \right) \) for simplicity. Hence,

\[
\sqrt{N} \left( \frac{\hat{e}' \hat{e}}{N} - \hat{\sigma}^2 \hat{Q}' \hat{Q} \right) = \sqrt{N} Z'_Q \text{vec} \left( \frac{e' e}{N} - \bar{\sigma}^2 I_T \right) + o_p(1),
\]

implying that the asymptotic distribution of \( \sqrt{N} \left( \frac{\hat{e}' \hat{e}}{N} - \hat{\sigma}^2 \hat{Q}' \hat{Q} \right) \) is equivalent to the asymptotic distribution of \( \sqrt{N} Z'_Q \text{vec} \left( \frac{e' e}{N} - \bar{\sigma}^2 I_T \right) \). Finally, by Assumption 6(ii), we have

\[
\sqrt{N} Z'_Q \text{vec} \left( \frac{e' e}{N} - \bar{\sigma}^2 I_T \right) \overset{d}{\longrightarrow} \mathcal{N} \left( 0, Z'_Q U \epsilon Z_Q \right).
\]

**Appendix D: explicit form of \( U_{\epsilon} \)**

Denote by \( U_{\epsilon} \) the \( T^2 \times T^2 \) matrix

\[
U_{\epsilon} = \begin{bmatrix}
U_{11} & \cdots & U_{1t} & \cdots & U_{1T} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
U_{t1} & \cdots & U_{tt} & \cdots & U_{tT} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
U_{T1} & \cdots & U_{Tt} & \cdots & U_{TT}
\end{bmatrix}.
\]

(E.1)

Each block of \( U_{\epsilon} \) is a \( T \times T \) matrix. The blocks along the main diagonal, denoted by \( U_{tt} \), \( t = 1, 2, \ldots, T \), are themselves diagonal matrices with \( (\kappa_4 + 2\sigma_4) \) in the \( (t, t) \)-th position and \( \sigma_4 \) in
the \((s, s)\) position for every \(s \neq t\), that is,

\[
\begin{array}{c}
\downarrow \\
\text{\(t\)-th column}
\end{array}
\begin{bmatrix}
\sigma_4 & \cdots & 0 & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \sigma_4 & 0 & \cdots & 0 \\
0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & \sigma_4 \\
\end{bmatrix}.
\]

(D.2)

The blocks outside the main diagonal, denoted by \(U_{ts}, s, t = 1, 2, \ldots, T\) with \(s \neq t\), are all made of zeros except for the \((s, t)\)-th position that contains \(\sigma_4\), that is,

\[
\begin{array}{c}
\downarrow \\
\text{\(s\)-th row}
\end{array}
\begin{bmatrix}
0 & \cdots & 0 & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & \sigma_4 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 \\
\end{bmatrix}.
\]

(D.3)

Under Assumption 5 and Lemma 6 in Appendix A, it is easy to show that \(\hat{U}_t\) in Theorem 2 is a consistent plug-in estimator of \(U_t\) that only depends on \(\hat{\sigma}_4\).

References


Chordia, T., A. Goyal, and J. Shanken, 2015, Cross-sectional asset pricing with individual stocks: Betas versus characteristics, working paper, Emory University.


Table I
Discrepancy of estimated risk premia (percentage change)
The table reports the discrepancy, obtained as the percentage change of the Shanken estimator $\hat{\Gamma}^* = (\hat{\gamma}_0, \hat{\gamma}_1')'$ over the OLS estimator $\hat{\Gamma} = (\hat{\gamma}_0, \hat{\gamma}_1')'$, averaged across the time-series of estimates based on rolling samples of size $T = 36, 72, 120$, respectively. The first panel refers to the CAPM, the second panel refers to the Fama and French (1993) three-factor model (FF3), and the third panel refers to the Fama and French (2015) five-factor model (FF5).

<table>
<thead>
<tr>
<th>Bias of:</th>
<th>$T = 36$</th>
<th>$T = 72$</th>
<th>$T = 120$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: CAPM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{mkt}$</td>
<td>64.3%</td>
<td>43.9%</td>
<td>27.2%</td>
</tr>
<tr>
<td>Panel B: FF3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{mkt}$</td>
<td>13.9%</td>
<td>15.6%</td>
<td>7.3%</td>
</tr>
<tr>
<td>$\gamma_{smb}$</td>
<td>14.7%</td>
<td>12.5%</td>
<td>12.3%</td>
</tr>
<tr>
<td>$\gamma_{hml}$</td>
<td>51.6%</td>
<td>48.8%</td>
<td>31.2%</td>
</tr>
<tr>
<td>Panel C: FF5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{mkt}$</td>
<td>15.3%</td>
<td>16.8%</td>
<td>11.1%</td>
</tr>
<tr>
<td>$\gamma_{smb}$</td>
<td>13.2%</td>
<td>10.8%</td>
<td>9.7%</td>
</tr>
<tr>
<td>$\gamma_{hml}$</td>
<td>14.1%</td>
<td>19.9%</td>
<td>15.2%</td>
</tr>
<tr>
<td>$\gamma_{rmw}$</td>
<td>13.3%</td>
<td>21.1%</td>
<td>15.2%</td>
</tr>
<tr>
<td>$\gamma_{cma}$</td>
<td>43.3%</td>
<td>46.3%</td>
<td>33.0%</td>
</tr>
</tbody>
</table>
Figure I
CAPM:

Estimates and confidence intervals of market excess return (mkt) risk premium
The figure presents the estimates and the associated confidence intervals of the risk premium on the market excess return obtained by estimating the CAPM. We report the time-series of the Shanken estimator $\hat{\gamma}^*_{mkt}$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel), and 10-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}^*_{mkt}$ based on the large-N standard errors of Theorem 2. We also report the time-series of the OLS estimator $\hat{\gamma}_{mkt}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP database, monthly observations from January 1966 until December 2013.
Figure II.a

Fama and French (1993) three-factor model:

Estimates and confidence intervals of market excess return ($mkt$) risk premium

The figure presents the estimates and the associated confidence intervals of the risk premium on the market excess return obtained by estimating the Fama and French (1993) three-factor model. We report the time-series of the Shanken estimator $\hat{\gamma}_{mkt}^*$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel), and 10-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{mkt}^*$ based on the large-N standard errors of Theorem 2. We also report the time-series of the OLS estimator $\hat{\gamma}_{mkt}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.
Figure II.b

Fama and French (1993) three-factor model:
Estimates and confidence intervals of small-minus-big (smb) factor risk premium

The figure presents the estimates and the associated confidence intervals of the risk premium on the smb factor (the return difference between portfolios of stocks with small and large market capitalizations) obtained by estimating the Fama and French (1993) three-factor model. We report the time-series of the Shanken estimator $\hat{\gamma}_{smb}$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel), and 10-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{smb}$ based on the large-N standard errors of Theorem 2. We also report the time-series of the OLS estimator $\bar{\gamma}_{smb}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.
Figure II.c
Fama and French (1993) three-factor model:
Estimates and confidence intervals of high-minus-low ($hml$) factor risk premium

The figure presents the estimates and the associated confidence intervals of the risk premium on the $hml$ factor (the return difference between portfolios of stocks with high and low book-to-market) obtained by estimating the Fama and French (1993) three-factor model. We report the time-series of the Shanken estimator $\hat{\gamma}_{hml}$ (blue line), obtained with a rolling time window of three- (top panel), six- (central panel), and 10-year (bottom panel). The blue band represents the 95% confidence interval around $\hat{\gamma}_{hml}$ based on the large-N standard errors of Theorem 2. We also report the time-series of the OLS estimator $\tilde{\gamma}_{hml}$ (red line) with its associated 95% confidence interval (orange band) based on the large-T standard errors. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.
Figure IV
CAPM: Specification test $S^*$.

The figure reports the time-series of p-values (black line) corresponding to the asset pricing test $S^*$ corresponding to the null hypothesis that the CAPM holds, obtained with a rolling time window of three- (top panel), six- (central panel), and 10-year (bottom panel). The dotted red line indicates the 5% level. We use individual stock data from the CRSP, monthly observations from January 1966 until December 2013.
Figure V

This figure reports the time series of p-values (black line) corresponding to the asset pricing test of Gibbons et al. (1989) corresponding to the null hypothesis that the CAPM holds, obtained with a rolling time window of three- (top panel), six- (central panel), and 10-year (bottom panel). The dotted red line indicates the 5% level. We use individual stock data from the CRSP aggregated into 25 equally weighted portfolios, monthly observations from January 1966 until December 2013.