Testing Beta-Pricing Models Using Large Cross-Sections *

Valentina Raponi Cesare Robotti Paolo Zaffaroni

December 18, 2018

Abstract

We propose a methodology for estimating and testing beta-pricing models when a large number of assets is available for investment but the number of time-series observations is fixed. We first consider the case of correctly specified models with constant risk premia, and then extend our framework to deal with time-varying risk premia, potentially misspecified models, firm characteristics, and unbalanced panels. We show that our large cross-sectional framework poses a serious challenge to common empirical findings regarding the validity of beta-pricing models. Firm characteristics are found to explain a much larger proportion of variation in estimated expected returns than betas.

Keywords: beta-pricing models; ex post risk premia; two-pass cross-sectional regressions; time-varying risk premia; model misspecification; firm characteristics; specification test; unbalanced panel; large-N asymptotics.

JEL classification: C12, C13, G12.

---

*Valentina Raponi, Imperial College Business School, e-mail: `v.raponi13@imperial.ac.uk`; Cesare Robotti, University of Warwick, e-mail: `Cesare.Robotti@wbs.ac.uk`; Paolo Zaffaroni (corresponding author), Imperial College Business School, e-mail: `p.zaffaroni@imperial.ac.uk`. We gratefully acknowledge comments from three anonimous referees, Adrian Buss, Fernando Chague, Victor DeMiguel, Francisco Gomes, Cam Harvey, Andrew Karolyi (Editor), Ralph Koijen, Ľuboš Pástor, Tarun Ramadorai, Krishna Ramaswamy, Olivier Scaillet, Jay Shanken, Pietro Veronesi, Grigory Vilkov, Guofu Zhou, and especially Raman Uppal, and seminar partecipants at CORE, Imperial College London, Luxembourg School of Finance, University of Georgia, University of Southampton, Toulouse School of Economics, Tinbergen Institute, University of Warwick, the 2015 Meetings of the Brazilian Finance Society, the CFE 2015, and the 2016 NBER/NSF Time Series Conference. 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".

Traditional econometric methodologies for estimating risk premia and testing beta-pricing models hinge on a large time-series sample size, T, and a small number of securities, N. At the same time, the thousands of stocks that are traded on a daily basis in financial markets provide a rich investment universe and an interesting laboratory for risk premia and cost of capital determination.[1] Moreover, although we have approximately a hundred years of US equity data, much shorter time series are typically used in empirical work to mitigate concerns of structural breaks and to bypass the difficult issue of modelling explicitly the time variation in risk premia. Finally, when considering non-US financial markets, only short time series are typically available.[2] Importantly, when N is large and T is small, the asymptotic distribution of any traditional risk premium estimator provides a poor approximation to its finite-sample distribution, thus rendering the statistical inference problematic.[3]

The main contribution of this paper is that it provides a methodology built on the large-N estimator of Shanken (1992), which allows us to perform valid inference on risk premia and assess the validity of the beta-pricing relation when N is large and T is fixed, possibly very small.[4] Our novel methods are first illustrated for correctly specified models with constant risk premia and then extended to deal with time variation in risk premia, potential model misspecification, firm characteristics in the risk-return relation, and unbalanced panels. We also demonstrate that methodologies specifically designed for a large T and fixed N environment are no longer applicable when a large number of assets is used. Proposition 3 below demonstrates the perils of inadvertently using the Fama and MacBeth (1973) t-ratios with the Shanken (1992) correction in our large N setting.

As emphasized by Shanken (1992), when T is fixed, one cannot reasonably hope for a consistent estimate of the traditional *ex ante* risk premium. For this reason, we focus on the *ex post* risk premia, which equal the ex ante risk premia plus the unexpected factor outcomes.[5]

[1] For example, one can download the returns on 18,474 US stocks for December 2013 from the Center for Research in Security Prices (CRSP), half of which are actively traded.

[2] For example, Table 1 in Hou et al. (2011) shows that, at most, only about thirty years of equity return data is available for emerging economies in Latin America, Europe-Middle East-Africa, and Asia-Pacific regions.

[3] The alternative approach of increasing the time-series frequency, although appealing, can lead to complications and is not always implementable. Potential problems with this approach include non-synchronous trading and market microstructure noise. Furthermore, for models that include non-traded (macroeconomic) risk factors, high-frequency data is not available.

[4] Our methodology offers an alternative to the common practice of employing a relatively small number of portfolios for the purpose of estimating and testing beta-pricing models. Although the use of portfolios is typically motivated by the attempt of reducing data noisiness, it can also cause loss of information and lead to misleading inference due to data aggregation. (See, for example, Brennan et al. (1998), Berk (2002), and Ang et al. (2018), among others.)

[5] The ex post risk premium is a parameter with several attractive properties. It is unbiased for the ex ante risk

We start by considering the baseline case of a correctly specified beta-pricing model with constant risk premia when a balanced panel of test asset returns is available. We show that the estimator of Shanken (1992) is free of any pre-testing biases and that no data has to be sacrificed for the preliminary estimation of the bias. (See Proposition 1 below). Next, we establish the asymptotic properties of the estimator, namely its $\sqrt{N}$-consistency and asymptotic normality. We derive an explicit expression for the estimator's asymptotic covariance matrix and show how this expression can be used to construct correctly sized confidence intervals for the risk premia. Our technical assumptions are relatively mild and easily verifiable. In particular, we allow for a substantial degree of cross-correlation among returns (conditional on the factors' realizations), and our assumptions are even weaker than the ones behind the Arbitrage Pricing Theory (APT) of Ross (1976).

In the first extension of the baseline methodology, we demonstrate that the estimator continues to exhibit attractive properties even when risk premia vary over time. In particular, it accurately describes the time-averages of the (time-varying) risk premia over a *fixed* time interval. We also derive a suitably modified version of the estimator that permits valid inference on risk premia at any given point in time. Noticeably, in our analysis we do not need to take a stand on the form of time variation in risk premia. Our time-varying risk premium estimator can accommodate non-traded as well as traded factors. For the latter, the traditional estimator based on the factors' rolling sample mean is asymptotically valid for the true risk premium at a given point in time only for specific sampling schemes, and it requires a very large T to work when time variation is allowed for. (See Internet Appendix IA.2 for details.)

Next, we allow for the possibility that the beta-pricing model is misspecified. We provide a new test of the validity of the beta-pricing relation and derive its large-N distribution under the null hypothesis that the model is correctly specified.[6] Moreover, we show that our test enjoys nice size and power properties. We then establish the statistical properties of the estimator when the beta-pricing model is misspecified. This extension is particularly relevant when we reject the model's

premium, and the beta-pricing model is still linear in the ex post risk premia under the assumptions of either correctly specified or misspecified models. Finally, the corresponding ex post pricing errors can be used to assess the validity of a given beta-pricing model when T is fixed. Naturally, when T becomes large, any discrepancy between the ex ante and ex post risk premia vanishes because the sample mean of the factors converges to its population mean.

[6]Since our test is specifically designed for scenarios in which N is large, it alleviates the concerns of Lewellen et al. (2010), Harvey et al. (2016), and Barillas and Shanken (2017) about a particular choice of test assets in the econometric analysis.

validity based on the outcome of the specification test, but we are still interested in estimating the risk premia of a model with a possibly incomplete set of factors. Finally, we study an important case of deviations from exact pricing, that is, the cross-sectional dependence of expected returns on firm characteristics. The asymptotic covariance matrix of the normally distributed characteristic premia estimator is derived in closed form, unlike most approaches in this literature that typically rely on simulation-based arguments for inference purposes. Our method can be used to determine whether the beta-pricing model is invalid and to quantify the economic importance of the characteristics when there are deviations from exact pricing. By employing a new measure, which is immune to the often-documented cross-correlation between estimated betas and characteristics, we are able to determine the relative contribution of betas and characteristics to the overall cross-sectional variation in expected returns.

In the last methodological extension of our baseline analysis, we consider the case of unbalanced panels. This is a useful extension because eliminating observations for the sole purpose of obtaining a balanced panel could result in unnecessarily large confidence intervals for the risk premia and loss of power of the specification test.

We demonstrate the usefulness of our methodology by means of several empirical analyses. The three prominent beta-pricing specifications that we consider are the Capital Asset Pricing Model (CAPM), the three-factor Fama and French (1993) model (FF3), and the recently proposed five-factor Fama and French (2015) model (FF5). We also consider variants of these models augmented with the non-traded liquidity factor of Pástor and Stambaugh (2003). Our proposed methods under potential model misspecification uncover a significant pricing ability for all the traded factors in each of the three models, even when using a relatively short time window of three years. In contrast, the risk premia estimates often appear to be statistically insignificant when using the traditional large-T approaches. Based on our methodology, the liquidity factor appears to be priced in only about one-fifth of the three-year rolling samples examined. We also document strong patterns of time variation in risk premia, for both traded and non-traded factors. In addition, our specification test rejects all beta-pricing models (with and without the liquidity factor), even when a short time window is used. Alternative methodologies, such as the finite-N approach of Gibbons et al. (1989) and the more recent test of Gungor and Luger (2016), seem to have substantially lower power in detecting model misspecification. Finally, our results indicate that five prominent firm characteristics (book-to-market ratio, asset growth, operating profitability, market capitalization, and six-month momentum) are important determinants of the cross-section of expected returns of individual assets. Although the characteristic premia estimates are not always found to be statistically significant, it seems that these characteristics jointly explain a fraction of the overall cross-sectional dispersion in expected returns that is about 30 times larger than the fraction explained by the estimated factors' betas, regardless of the beta-pricing model under consideration.

Our paper is related to a large number of studies in empirical asset pricing and financial econometrics. The traditional two-pass cross-sectional regression (CSR) methodology for estimating beta-pricing models, developed by Black et al. (1972) and Fama and MacBeth (1973), is valid when T is large and N is fixed. Shanken (1992) shows how the asymptotic standard errors of the second-pass CSR risk premia estimators are affected by the estimation error in the first-pass betas and provides standard errors that are robust to the errors-in-variables (EIV) problem.[7] Shanken and Zhou (2007) derive the large-T properties of the two-pass estimator in the presence of global model misspecification.[8] A different form of misspecification, not explored in this paper, can also occur when some of the factors have zero, or almost zero, betas, a situation that is referred to as the spurious or "useless" factors problem.[9] Lack of identification of the risk premia also arises when at least one of the betas is cross-sectionally quasi-constant, as documented by Ahn et al. (2013) with respect to the market factor empirical betas, a case also ruled out here.

Building on Litzenberger and Ramaswamy (1979), Shanken (1992) (Section 6) proposes a large-N estimator of the ex post risk premium and shows that it is asymptotically unbiased when N diverges and T is fixed. However, Shanken (1992) does not prove the consistency and asymptotic normality of this risk premium estimator.[10] Differently from Litzenberger and Ramaswamy (1979), Shanken (1992) demonstrates unbiasedness without imposing a rigid structure on the covariance

[7] Jagannathan and Wang (1998) relax the conditional homoskedasticity assumption of Shanken (1992). For a review of the large-T literature on beta-pricing models, see Shanken (1996), Jagannathan et al. (2010), and Kan and Robotti (2012).

[8] See also Hou and Kimmel (2006) and Kan et al. (2013).

[9] Several methods have been developed to deal with this particular form of model misspecification. See, for example, Jagannathan and Wang (1998), Kan and Zhang (1999a), Kan and Zhang (1999b), Kleibergen (2009), Ahn et al. (2013), Gospodinov et al. (2014), Burnside (2015), Bryzgalova (2016), Gospodinov et al. (2017), Ahn et al. (2018), Gospodinov et al. (2018), Kleibergen and Zhan (2018a), and Kleibergen and Zhan (2018b), among others.

[10] In the same paper, Shanken (1992) provides the well-known standard errors correction for ordinary least squares (OLS) and generalized least squares (GLS) estimators of the ex post risk premia, but his correction is only valid when T is large and N is fixed. (See his Section 3.2.)

matrix of the first-pass residuals.

Following these seminal contributions, other methods have been recently proposed to take advantage of the increasing availability of large cross-sections of individual securities. Our paper is close to Gagliardini et al. (2016) in the sense that both studies provide inferential methods for estimating and testing beta-pricing models. However, their work is developed in a joint-asymptotics setting, where both T and N need to diverge. Moreover, they focus on a slightly different parameter of interest (obtained as the difference between the ex ante risk premia and the factors' population mean), which can be derived from the ex post risk premium by netting out the sample mean of the factor. Like us, Gagliardini et al. (2016) need a bias adjustment because in their setting N is diverging at a much faster rate than T.[11] Moreover, while Gagliardini et al. (2016) assume random betas, as a consequence of their sampling framework with a continuum of assets, in our analysis we prefer to keep the betas nonrandom. This is for us mostly a convenience assumption since we show in the Internet Appendix that allowing for randomness of the betas in a large-N environment leaves our theoretical results unchanged. Gagliardini et al. (2016) characterize the time variation in risk premia by conditioning on observed state variables, whereas we leave the form of time variation unspecified. Like us, they show how to carry out inference when the beta-pricing model is globally misspecified. Finally, Gagliardini et al. (2016) allow for a substantial degree of cross-sectional dependence of the returns' residuals. Although our setup and assumptions differ from theirs (mainly because in our framework only N diverges), we also allow for a similar form of cross-sectional dependence in the residuals' covariance matrix.

Bai and Zhou (2015) investigate the joint asymptotics of the modified OLS and GLS CSR estimators of the ex ante risk premia. Although the CSR estimators are asymptotically unbiased when T diverges, they propose an adjustment to mitigate the finite-sample bias. Their bias adjustment differs from the one suggested by Litzenberger and Ramaswamy (1979) and Shanken (1992), and studied in this paper, because it relies on a large T for its validity. However, their simulation results suggest that their bias-adjusted estimator performs well for various values of N and T. Moreover, since T must be large in their setting, Bai and Zhou (2015) bias-adjustment is asymptotically negligible, implying that the asymptotic distribution of their CSR estimators is identical to the

[11] In contrast, recall that in the traditional analysis of the CSR estimator (where T diverges and N is fixed), no bias adjustment is required.

asymptotic distribution of the traditional OLS and GLS CSR estimators.[12] In contrast, we show that the asymptotic distribution of the risk premia estimator must necessarily change in the fixed-T case, where the traditional trade-off between bias and variance emerges. Moreover, consistent estimation of the asymptotic covariance matrix of our risk premia estimator requires a different analysis because only N is allowed to diverge. Bai and Zhou (2015) focus exclusively on the case of a balanced panel under the assumption of correctly specified models. Unlike us, they do not account for time variation in the risk premia and do not analyze model misspecification.

Giglio and Xiu (2017) propose a modification of the two-pass methodology based on principal components that is robust to omitted priced factors and mis-measured observed factors, and establish its validity under joint asymptotics.

Kim and Skoulakis (2018) employ the so-called regression calibration approach used in EIV models to derive a $\sqrt{N}$-consistent estimator of the ex post risk premia in a two-pass CSR setting.[13] Finally, Jegadeesh et al. (2018) propose instrumental-variable estimators of the ex post risk premia, exploiting the assumed independence over time of the return data.[14]

As for specification testing, Pesaran and Yamagata (2012) extend the classical test of Gibbons et al. (1989) to a large-N setting. Besides accommodating only traded factors, the feasible version of their tests requires joint asymptotics and N needs to diverge at a faster rate than T. Gungor and Luger (2016) propose a nonparametric testing procedure for mean-variance efficiency and spanning hypotheses (with tests of the beta-pricing restriction as a special case), and they derive (exact) bounds on the null distribution of the test statistics using resampling techniques. Their procedure, which is designed for traded factors only, is valid for any N and T, even though they show that the power of their test increases when both N and T diverge. Gagliardini et al. (2016) derive the asymptotic distribution of their specification test under joint asymptotics and, like us, they allow

[12] Gagliardini et al. (2016) show that the bias adjustment in their framework is not asymptotically negligible when N diverges at a much faster rate than T, a case not explicitly studied in Bai and Zhou (2015).

[13] Building on Jagannathan et al. (2010), the Kim and Skoulakis (2018) estimator can be seen as an alternative to the Shanken estimator, the only difference being that in Kim and Skoulakis (2018) the first- and second-pass regressions are evaluated on non-overlapping time periods.

[14] Besides the classical econometric challenges associated with the choice of potentially weak instruments, these instrumental-variable approaches require a relatively larger T in order to achieve the same statistical accuracy of the Shanken (1992) estimator. Moreover, the construction of the instruments in Jegadeesh et al. (2018) hinges upon the assumption of stochastic independence over time of the return data. The same assumption is also required in Kim and Skoulakis (2018). In contrast, it can be shown that the Shanken (1992) estimator retains its asymptotic properties even when the data is not independent over time. In fact, an arbitrary degree of serial dependence of the return data can be allowed for.

for general factors. Finally, Gagliardini et al. (2018) propose a diagnostic criterion for detecting the number of omitted factors from a given beta-pricing model and establish its statistical behavior under joint asymptotics.

Having detailed our contributions and related them to the existing literature, we now discuss when our methodology should be used, from three different angles. With respect to the sampling scheme, our methodology is theoretically justified when T is fixed and N diverges. In contrast, the limiting results for the traditional CSR estimators cited above are valid when T diverges with a fixed N as well as when both T and N diverge. Proposition 3 in the paper warns us about using these traditional methods under our reference sampling scheme. Moreover, based on numerous Monte Carlo experiments, previous studies have found that the large-T approximations of the CSR estimators are reliable only when five or more decades of data are used. (See Chen and Kan (2004) and Shanken and Zhou (2007), among others.) Therefore, our methodology could be useful also in scenarios where the time-series dimension is relatively large.

Starting from traded factors and assuming that the true risk premia are constant and the model is correctly specified, the sample means of the factors' excess returns or return spreads could be used as risk premia estimators of the true factors' means. However, a sufficiently large T is required for the sample means to converge to their population counterparts. For non-traded factors, for example, macroeconomic variables, a panel of test asset returns is required to pin down the factors' risk premia, as the time series of the factors do not suffice. Mimicking portfolio excess returns could also be used in place of the non-traded factors, with the population means of the mimicking portfolio excess returns serving as the true risk premia.[15] However, the mimicking portfolio projection requires $N < T$, which is violated under our reference sampling scheme.[16]

Finally, when the risk premia are time-varying, the argument for using our methodology appears even more compelling. Note that the considerations above regarding alternative estimation

[15]See Breeden et al. (1989), Chan et al. (1998), and Lamont (2001), among others, for empirical studies based on the mimicking portfolio methodology. Balduzzi and Robotti (2008) demonstrate by means of Monte Carlo simulations the greater accuracy of the mimicking portfolio risk premia estimates relative to the CSR risk premia estimates associated with the corresponding non-traded factors.

[16]When $N > T$, one could obtain the first $\tilde{N}$ principal components from a large panel of test assets returns, and then construct the mimicking portfolio for the non-traded factor using these $\tilde{N}$ assets (assuming that $\tilde{N} < T < N$). Although this approach is feasible and is used in our empirical application, the theoretical properties of this double-projection approach are difficult to derive; see Giglio and Xiu (2017) for a theoretical analysis of a similar approach. We are grateful to an anonymous referee for suggesting this approach to us.

procedures for the traded factors case hold for both constant and time-varying risk premia. In particular, the (rolling) sample mean of the excess return on the traded factor (or of the return spread) will capture, in general, the average, over T observations, of the true time-varying risk premium associated with the factor. Alternatively, one can adopt the sampling scheme typical of nonparametric methods, with the implication that now the (rolling) sample mean will capture the time-varying risk premium and not just its average. However, a very large T would be necessary to obtain accurate estimates and a certain degree of smoothness, over time, of the true time-varying risk premium would be required. (See the Internet Appendix IA.2 for further details.) Our method for time-varying risk premia works for any T and makes no smoothness assumption.

To summarize, compelling reasons for using our methodology arise when T is fairly small (and, in particular, smaller than N), when considering models with non-traded factors, and when interest lies in the time variation in risk premia on traded and non-traded factors. In addition, our methodology can handle potential model misspecification (due, for example, to omitted pervasive factors) and, in particular, it provides a natural framework to determine whether the rejection of the beta-pricing relation is due to priced firm characteristics. Finally, we can easily accommodate unbalanced panels in the analysis.

The rest of the paper is organized as follows. Section 1 surveys the two-pass OLS CSR methodology, introduces our main assumptions, and sets the notation. Section 2 presents the asymptotic results for constant and time-varying risk premia estimates under correctly specified models. Section 3 generalizes our theory to potentially misspecified beta-pricing models with and without firm characteristics. In Section 4, we investigate the empirical performance of FF5. Section 5 concludes. The technical proofs are in the Appendix.[17]

1. The Two-Pass Methodology

This section introduces the notation and summarizes the two-pass OLS CSR methodology. We assume that the asset returns $R_t = [R_{1t}, \ldots, R_{Nt}]'$ are governed by the following beta-pricing

[17] The Internet Appendix (IA) contains additional material: Section IA.1 provides a discussion of random betas; Section IA.2 describes the properties of nonparametric estimation methods for the risk premia on traded factors under various sampling schemes; Section IA.3 illustrates the finite-N sampling properties of the Shanken estimator and of the associated specification test using Monte Carlo simulations; Section IA.4 provides an extension of our baseline analysis to unbalanced panels; Section IA.5 contains empirical results for CAPM, FF3, and additional results for FF5.

model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it} = \alpha_i + \beta_i' f_t + \epsilon_{it}, \tag{1}$$

where i denotes the i-th asset, with $i = 1, \ldots, N$, t refers to time, with $t = 1, \ldots, T$, α_i is a scalar parameter 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 $f_t = [f_{1t}, \ldots, f_{Kt}]'$, and ϵ_{it} is the i-th return's idiosyncratic component. In matrix notation, we can write the model above as

$$R_t = \alpha + B f_t + \epsilon_t, \qquad t = 1, \ldots, T, \tag{2}$$

where $\alpha = [\alpha_1, \ldots, \alpha_N]'$, $B = [\beta_1, \ldots, \beta_N]'$, and $\epsilon_t = [\epsilon_{1t}, \ldots, \epsilon_{Nt}]'$. Let $\Gamma = [\gamma_0,\ \gamma_1']'$, where γ_0 the zero-beta rate and γ_1 is the K-vector of ex ante factor risk premia, and denote by $X = [1_N,\ B]$ the beta matrix augmented with 1_N, an N-vector of ones. The following assumption of exact pricing is used at various points in the analysis below.

Assumption 1

$$E[R_t] = X\Gamma. \tag{3}$$

Eq. (3) follows, for example, from no-arbitrage (see Condition A in Chamberlain (1983)) and a well-diversified mean-variance frontier (Definition 4 in Chamberlain (1983)).[18]

Averaging Eq. (2) over time, where we set $\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t = [\bar{R}_1, \ldots, \bar{R}_N]'$, $\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t$, and $\bar{f} = [\bar{f}_1, \ldots, \bar{f}_K]' = \frac{1}{T}\sum_{t=1}^{T} f_t$, imposing Assumption 1, and noting that $E[R_t] = \alpha + BE[f_t]$ from Eq. (2), yields

$$\bar{R} = X\Gamma^P + \bar{\epsilon}, \tag{4}$$

where $\Gamma^P = [\gamma_0,\ \gamma_1^{P\prime}]'$, and

$$\gamma_1^P = \gamma_1 + \bar{f} - E[f_t]. \tag{5}$$

From Eq. (4), average returns are linear in the asset betas *conditional* on the factor outcomes through the quantity γ_1^P, which, in turn, depends on the factors' sample mean innovations, $\bar{f} - E[f_t]$. The random coefficient vector γ_1^P in Eq. (5) is referred to as the vector of *ex post* risk premia.[19]

[18] It should be noted that the mere absence of arbitrage is not sufficient for exact pricing, that is, nonzero pricing errors can coexist with no-arbitrage, as in the case of the APT of Ross (1976).

[19] For traded factors, Eq. (5) reduces to $\gamma_1^P = \bar{f} - \gamma_0 1_K$, where 1_K is a K-vector of ones. (See Shanken (1992).)

Eq. (5) shows that Γ and Γ^P will coincide when $\bar{f} = E[f_t]$, which happens for $T \to \infty$. When T is small, ex ante and ex post risk premia can differ substantially, as emphasized in the empirical section of the paper, although γ_1^P remains an unbiased measure for the ex ante risk premia, γ_1.[20]

Note that Eq. (4) cannot be used to estimate the ex post risk premia Γ^P since X is not observed. For this reason, the popular two-pass OLS CSR method first obtains estimates of the betas by running the following multivariate regression for every i:

$$R_i = \alpha_i 1_T + F\beta_i + \epsilon_i, \tag{6}$$

where $R_i = [R_{i1}, \ldots, R_{iT}]'$, $\epsilon_i = [\epsilon_{i1}, \ldots, \epsilon_{iT}]'$, $F = [f_1, \ldots, f_T]'$ is the $T \times K$ matrix of factors, and 1_T is a T-vector of ones. Then, the OLS estimates of B are given by

$$\hat{B} = R'\tilde{F}(\tilde{F}'\tilde{F})^{-1} = B + \epsilon'\mathcal{P}, \tag{7}$$

where $\hat{B} = [\hat{\beta}_1, \ldots, \hat{\beta}_N]'$, $R = [R_1, \ldots, R_N]$, $\epsilon = [\epsilon_1, \ldots, \epsilon_N]$, and $\mathcal{P} = \tilde{F}(\tilde{F}'\tilde{F})^{-1}$ with $\tilde{F} = [\tilde{f}_1, \ldots, \tilde{f}_T]' = \left(I_T - \frac{1_T 1_T'}{T}\right) F = F - 1_T \bar{f}'$, where I_T is the identity matrix of order T. The corresponding matrix of OLS residuals is given by $\hat{\epsilon} = [\hat{\epsilon}_1, \ldots, \hat{\epsilon}_N] = R - 1_T\bar{R}' - \tilde{F}\hat{B}'$.

We then run a single CSR of the sample mean vector $\bar{R}$ on $\hat{X} = [1_N,\ \hat{B}]$ to estimate the risk premia. Note that we have two alternative feasible representations of Eq. (4), that is,

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

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

$$\bar{R} = \hat{X}\Gamma^P + \eta^P, \tag{9}$$

with residuals $\eta^P = \left[\bar{\epsilon} - (\hat{X} - X)\Gamma^P\right]$. The OLS CSR estimator applied to either Eq. (8) or Eq. (9) yields

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

However, when T is fixed, $\hat{\Gamma}$ cannot be used as a consistent estimator of the ex ante risk premia, Γ, in Eq. (8) and of the ex post risk premia, Γ^P, in Eq. (9). The reason is that neither $\hat{B}$ converges to B, nor $\bar{f}$ converges to $E[f_t]$ unless $T \to \infty$. Focusing on the representation in Eq. (9), the OLS CSR

[20] It should be noted that any valid estimator of γ_1^P provides, as a by-product, a valid estimator of the population parameter $\nu = \gamma_1 - E[f_t] = \gamma_1^P - \bar{f}$, namely the portion of the ex ante risk premia that is nonlinearly related to the factors. This is the quantity studied in Gagliardini et al. (2016).

estimator can be corrected as follows. Denote by $\text{tr}(\cdot)$ the trace operator and by 0_K a K-vector of zeros. In addition, let

$$\hat{\sigma}^2 = \frac{1}{N(T-K-1)}\text{tr}(\hat{\epsilon}'\hat{\epsilon}). \tag{11}$$

The bias-adjusted estimator of Shanken (1992) is then given by

$$\hat{\Gamma}^* = \begin{bmatrix} \hat{\gamma}_0^* \\ \hat{\gamma}_1^* \end{bmatrix} = \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}'\bar{R}}{N}, \tag{12}$$

where

$$\hat{\Sigma}_X = \frac{\hat{X}'\hat{X}}{N} \quad \text{and} \quad \hat{\Lambda} = \begin{bmatrix} 0 & 0_K' \\ 0_K & \hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix}. \tag{13}$$

The formula for the estimator $\hat{\Gamma}^*$ exhibits a *multiplicative* bias adjustment through the term $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}$.[21] This prompts us to explore the analogies of $\hat{\Gamma}^*$ with the more conventional class of *additive* bias-adjusted OLS CSR estimators. To this end, it is useful to consider the following expression for the OLS CSR estimator, $\hat{\Gamma}$, obtained from Bai and Zhou (2015) in their Theorem 1:

$$\begin{aligned} \hat{\Gamma} &= \Gamma^P + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \begin{bmatrix} 0 & 0_K' \\ 0_K & -\hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix} \Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= \Gamma^P - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda}\Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right). \end{aligned} \tag{14}$$

This formula suggests a simple way to construct an additive bias-adjusted estimator of Γ^P; that is,

$$\hat{\Gamma}^{bias-adj} = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda}\hat{\Gamma}^{prelim}, \tag{15}$$

where $\hat{\Gamma}^{prelim}$ is an arbitrary preliminary estimator of Γ^P.[22] The next proposition shows that, by imposing that the preliminary estimator, $\hat{\Gamma}^{prelim}$, and the bias-adjusted estimator, $\hat{\Gamma}^{bias-adj}$, coincide, the unique solution to Eq. (15) is the Shanken (1992) estimator $\hat{\Gamma}^*$ in Eq. (12).

Proposition 1 *Assume that $\hat{\Sigma}_X - \hat{\Lambda}$ is nonsingular. Then, the Shanken (1992) estimator $\hat{\Gamma}^*$ in Eq. (12) is the unique solution to the linear system of equations:*

$$\hat{\Gamma}^* = \hat{\Gamma} + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda}\hat{\Gamma}^*. \tag{16}$$

[21] Eq. (15) in Shanken (1992) differs slightly from our Eq. (12). The reason is that we do not impose the traded-factor restriction of Shanken (1992) in our setting.

[22] For example, Bai and Zhou (2015) propose using the OLS CSR $\hat{\Gamma}$ itself as the preliminary estimator, plugging it into the formula above in place of $\hat{\Gamma}^{prelim}$. However, this adjustment is justified only when $T \to \infty$. In general, the use of a preliminary estimator would decrease the precision of the bias-adjusted estimator and, in addition, it would make its properties harder to study.

Proof: See Appendix B.

Therefore, $\hat{\Gamma}^*$ is the *unique* additive bias-adjusted OLS CSR estimator that does not require the preliminary estimation of the risk premia. As a computational precaution, it is possible that the EIV correction in Eq. (12) overshoots, making the matrix $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)$ almost singular for a given N and potentially leading to extreme values for the estimator. To alleviate this risk, our suggestion is to multiply the matrix $\hat{\Lambda}$ by a scalar k $(0 \leq k \leq 1)$ and to substitute $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}$ with $\left(\hat{\Sigma}_X - k\hat{\Lambda}\right)^{-1}$ in Eq. (12), effectively yielding a shrinkage estimator.[23] If k is zero, we obtain the OLS CSR estimator $\hat{\Gamma}$, whereas if k is one, we obtain the Shanken (1992) estimator $\hat{\Gamma}^*$.[24] In our simulation experiments, we find that this shrinkage estimator is virtually unbiased, leading to $k = 1$. In contrast, in our empirical application in Section 4, shrinking is applied to roughly 75% of the cases (the average k is 0.58) when $T = 36$ and to 5% of the cases (the average k is 0.71) when $T = 120$. Our shrinkage adjustment can also alleviate the documented evidence of cross-sectional quasi-homogeneity for the loadings associated with certain risk factors, in particular for the market factor (see Ahn et al. (2013)).[25]

Before turning to the challenging task of deriving the large-N distribution of the Shanken (1992) estimator (and the associated standard errors), we discuss the perils of using the traditional t-ratios (specifically designed for a large-T environment) when N diverges. We first introduce the necessary assumptions and then present our results in Proposition 3 below.

[23] Our asymptotic theory would require $k = k_N$ to converge to unity at a suitably slow rate as N increases. We omit the details to simplify the exposition.

[24] The choice of the shrinkage parameter k can be based on the eigenvalues of the matrix $\left(\hat{\Sigma}_X - k\hat{\Lambda}\right)$ as follows. Starting from $k = 1$, if the minimum eigenvalue of this matrix is negative and/or the *condition number* of this matrix is larger than 20 (as suggested by Greene (2003), p. 60), then we lower k by an arbitrarily small amount. In our empirical application we set this amount equal to 0.05 and perform shrinkage whenever the absolute value of the relative change between the Shanken (1992) and the OLS CSR estimators is greater than 100%. We iterate this procedure until the minimum eigenvalue is positive and the condition number becomes less than 20. Gagliardini et al. (2016) rely on similar methods to implement their trimming conditions. Alternatively, one could use cross-validation to set the value of k.

[25] Ahn et al. (2013) propose the so-called invariance beta (IB) coefficient as a measure of cross-sectional homogeneity. Applying their measure to our data on FF5, we find that the IB coefficient corresponding to the market factor equals 0.74 and 0.81 for rolling samples of size $T = 36$ and $T = 120$, respectively (averages across rolling samples). The IB coefficient is equal to 0.93 when considering the whole sample. According to Ahn et al. (2013), these values signal a very moderate risk of multicollinearity due to cross-sectional homogeneity. Similar values of the IB coefficient associated with the loadings on the market factor are obtained when estimating CAPM and FF3.

Assumption 2 *As $N \to \infty$,*

$$\frac{1}{N}\sum_{i=1}^{N}\beta_i \to \mu_\beta \qquad \textit{and} \qquad \frac{1}{N}\sum_{i=1}^{N}\beta_i\beta_i' \to \Sigma_\beta, \tag{17}$$

$$\textit{such that the matrix } \begin{bmatrix} 1 & \mu_\beta' \\ \mu_\beta & \Sigma_\beta \end{bmatrix} \textit{ is positive-definite.} \tag{18}$$

Assumption 2 states that the limiting cross-sectional averages of the betas, and of the squared betas, exist. The second part of Assumption 2 rules out the possibility of spurious factors and situations in which at least one of the elements of β_i is cross-sectionally constant. (See Ahn et al. (2013).) It implies that X has full (column) rank for N sufficiently large. To simplify the exposition, we assume that the β_i are nonrandom.[26]

Assumption 3 *The vector ϵ_t is independently and identically distributed (i.i.d.) over time with*

$$E[\epsilon_t|F] = 0_N \tag{19}$$

and a positive-definite matrix,

$$Var[\epsilon_t|F] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{pmatrix} = \Sigma, \tag{20}$$

where 0_N is a N-vector of zeros, and σ_{ij} denotes the (i,j)-th element of Σ, for every $i, j = 1, \ldots, N$ with $\sigma_i^2 = \sigma_{ii}$.

The *i.i.d.* assumption over time is common to many studies, including Shanken (1992). However, our large N asymptotic theory, in principle, permits the ϵ_{it} to be arbitrarily correlated over time, but the expressions would be more complicated. Conditions (19) and (20) are verified if the factors f_t and the innovations ϵ_s are mutually independent for any s, t. Noticeably, Condition (20) is not imposing any specific structure on the elements of Σ. In particular, we are not assuming that the returns' innovations are uncorrelated across assets or exhibit the same variance. However, our large-N asymptotic theory needs to discipline the degree of cross-correlation among the residuals,

[26] See Gagliardini et al. (2016) for a treatment of the beta-pricing model with random betas. In Internet Appendix IA.1, we discuss the consequences of relaxing the nonrandomness of the β_i.

although still allowing for a substantial degree of heterogeneity in the cross-section of asset returns. (See Assumption 5 below.)

As for the factors, we impose minimal assumptions because our asymptotic analysis holds conditional on the factors' realizations.

Assumption 4 *$E[f_t]$ does not vary over time. Moreover, $\tilde{F}'\tilde{F}$ is a positive-definite matrix for every $T \geq K$.*

Assumption 5 *As $N \to \infty$,*

(i)

$$\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2 - \sigma^2\right) = o\left(\frac{1}{\sqrt{N}}\right), \tag{21}$$

for some $0 < \sigma^2 < \infty$.

(ii)

$$\sum_{i,j=1}^{N} \mid \sigma_{ij} \mid \mathbb{1}_{\{i \neq j\}} = o\left(N\right), \tag{22}$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

(iii)

$$\frac{1}{N}\sum_{i=1}^{N} \mu_{4i} \to \mu_4, \tag{23}$$

for some $0 < \mu_4 < \infty$ where $\mu_{4i} = E[\epsilon_{it}^4]$.

(iv)

$$\frac{1}{N}\sum_{i=1}^{N} \sigma_i^4 \to \sigma_4, \tag{24}$$

for some $0 < \sigma_4 < \infty$.

(v)

$$\sup_i \mu_{4i} \leq C < \infty, \tag{25}$$

for a generic constant C.

(vi)

$$E[\epsilon_{it}^3] = 0. \tag{26}$$

(vii)

$$\frac{1}{N}\sum_{i=1}^{N}\kappa_{4,iiii} \to \kappa_4, \tag{27}$$

for some $0 \leq |\kappa_4| < \infty$, where $\kappa_{4,iiii} = \kappa_4(\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it})$ denotes the fourth-order cumulant of the residuals $\{\epsilon_{it}, \epsilon_{it}, \epsilon_{it}, \epsilon_{it}\}$.

(viii) For every $3 \leq h \leq 8$, all the mixed cumulants of order h satisfy

$$\sup_{i_1}\sum_{i_2,\ldots,i_h=1}^{N}|\kappa_{h,i_1 i_2\ldots i_h}| = o\left(N\right), \tag{28}$$

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. In particular, Assumption 5(i) limits the cross-sectional heterogeneity of the return conditional variance. Assumption 5(ii) implies that the conditional correlation among asset returns is sufficiently weak. Assumptions 5(i) and 5(ii) allow for many forms of strong cross-sectional dependence, as emphasized by the following proposition, which considers the case in which the ϵ_{it} obey a factor structure.

Proposition 2 *Assume that*

$$\epsilon_{i,t} = \lambda_i u_t + \eta_{i,t}, \tag{29}$$

where

$$\sum_{i=1}^{N}|\lambda_i| = O(N^{\delta}),\ \ 0 \leq \delta < 1/2, \tag{30}$$

and (without loss of generality) for some fixed $q < N$ and some constant C,

$$\lambda_1 + \cdots + \lambda_q \sim CN^{\frac{\delta}{2}}, \tag{31}$$

with u_t i.i.d. $(0,1)$ and $\eta_{i,t}$ i.i.d. $(0,\sigma^2_\eta)$ over time and across units, where the u_t and the $\eta_{i,s}$ are mutually independent for every i, s, t. Then,

(i) Assumption 5(i) and 5(ii) are satisfied with $\sigma^2 = \sigma^2_\eta$.

(ii) The maximum eigenvalue of Σ diverges as $N \to \infty$.[27]

[27] The maximum eigenvalue of Σ is given by $\sup_{z s.t. \|z\|=1} z'\Sigma z$.

Proof: See Appendix B.

Note that the boundedness of the maximum eigenvalue is the most common assumption on the covariance matrix of the disturbances in beta-pricing models. (See, e.g., the generalization of the APT by Chamberlain and Rothschild (1983).) Our assumptions are weaker than the ones for the APT because the maximum eigenvalue can now diverge. This implies that the row-column norm of Σ, $\sup_{1\leq i\leq N}\sum_{j=1}^{N}|\sigma_{ij}|$, diverges.[28] Eq. (29) is adopted in our Monte Carlo experiments reported in the Internet Appendix. Other special cases nested by Assumption 5 for which the cross-covariances σ_{ij} are nonzero are network and spatial measures of cross-dependence and a suitably modified version of the block-dependence structure of Gagliardini et al. (2016).[29]

In Assumption 5(iii), we simply assume the existence of the limit of the conditional fourth-moment, averaged across assets. In Assumption 5(iv), the magnitude of σ_4 reflects the degree of cross-sectional heterogeneity of the conditional variance of the 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. Without 5(vi) the asymptotic distribution would be more involved, due to the presence of terms such as the third moment of the disturbance (averaged across assets). Assumption 5(vii) allows for non-Gaussianity of the 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 (1992) estimator, one needs to set $\kappa_4 = 0$ merely for identification purposes, as explained in Lemma 6 in Appendix A. However, higher-order cumulants are not constrained to be zero, implying that $\kappa_4 = 0$ is not equivalent to Gaussianity. We are now ready to state our Proposition 3.

Proposition 3 *Under Assumptions 1-5 and as $N \to \infty$, the Fama and MacBeth (1973) t-ratios for $\hat{\Gamma} = [\hat{\gamma}_0, \hat{\gamma}_{11}, \ldots, \hat{\gamma}_{1k}, \ldots, \hat{\gamma}_{1K}]'$ based on the correction of Shanken (1992) satisfy the following relations.*

[28] Assumption 5 allows for the maximum eigenvalue of Σ to diverge at rate $o\left(\sqrt{N}\right)$. (See the proof of Proposition 2 for details.) Gagliardini et al. (2016) can allow for a faster rate, $o(N)$, of divergence of the maximum eigenvalue of Σ because both T and N diverge in their double-asymptotics setting.

[29] Gagliardini et al. (2016) Assumption BD.2 on block sizes and block numbers requires that the largest block size shrinks with N and that there are not too many large blocks; that is, the partition in independent blocks is sufficiently fine-grained asymptotically. They show formally that such block-dependence structure is compatible with the unboundedness of the maximum eigenvalue of Σ.

(i) For the ex ante risk premia $\Gamma = [\gamma_0, \gamma_{11}, \ldots, \gamma_{1k}, \ldots, \gamma_{1K}]'$, we have

$$|t_{FM}(\hat{\gamma}_0)| = \frac{|\hat{\gamma}_0 - \gamma_0|}{SE_0^{FM}} \to_p \infty \tag{32}$$

and

$$|t_{FM}(\hat{\gamma}_{1k})| = \frac{|\hat{\gamma}_{1k} - \gamma_{1k}|}{SE_k^{FM}} \to_p \left| \frac{\bar{f}_k - E[f_{kt}]}{\hat{\sigma}_k/\sqrt{T}} - \frac{\imath_{k,K}' A^{-1} C \gamma_1^P}{\hat{\sigma}_k/\sqrt{T}} \right| \text{ for } k \geq 1. \tag{33}$$

(ii) For the ex post risk premia $\Gamma^P = [\gamma_0, \gamma_{11}^P, \ldots, \gamma_{1k}^P, \ldots, \gamma_{1K}^P]'$, we have

$$|t_{FM,P}(\hat{\gamma}_0)| = \frac{|\hat{\gamma}_0 - \gamma_0|}{SE_0^{FM,P}} \to_p \infty \tag{34}$$

and

$$|t_{FM,P}(\hat{\gamma}_{1k})| = \frac{|\hat{\gamma}_{1k} - \gamma_{1k}^P|}{SE_k^{FM,P}} \to_p \infty \text{ for } k \geq 1, \tag{35}$$

where SE_k^{FM} and $SE_k^{FM,P}$ are the Fama and MacBeth (1973) standard errors with the Shanken (1992) correction corresponding to the ex ante and ex post risk premia, respectively (see Appendix B for details), and where $\imath_{k,K}$ is k-th column of the identity matrix I_K, $\hat{\sigma}_k^2$ is the (k,k)-th element of $\tilde{F}'\tilde{F}/T$, $A = \Sigma_\beta - \mu_\beta \mu_\beta' + C$, and $C = \sigma^2(\tilde{F}'\tilde{F})^{-1}$.

Proof: See Appendix B.

In summary, Proposition 3 shows that a methodology designed for a fixed N and a large T, such as the one based on the Fama and MacBeth (1973) standard errors with the Shanken's correction, is likely to lead to severe over-rejections when N is large, thus rendering the inference on the beta-pricing model invalid.[30] Our Monte Carlo simulations corroborate this finding, as emphasized in the Internet Appendix. Moreover, Proposition 3 shows that when N and T are large, there is no need to apply the correction of Shanken (1992) to the Fama and MacBeth (1973) standard errors.

2. Asymptotic Analysis under Correctly Specified Models

In this section, we establish the limiting distribution of the Shanken (1992) bias-adjusted estimator, $\hat{\Gamma}^*$, and explain how its asymptotic covariance matrix can be consistently estimated.

[30]In particular, the t-ratio of the OLS CSR estimator for a particular element of the ex ante risk premium vector, γ_1, equals the standardized sample mean of the associated factor plus a bias term. When T is allowed to diverge, the convergence of this t-ratio to a standard normal is re-obtained, but, for any given T, the deviations from normality can be substantial.

2.1 Baseline case

Our baseline case assumes that the beta-pricing model is correctly specified, that the risk premia are constant, and that the panel is balanced. This corresponds to the setup of Shanken (1992).

Let $\Sigma_X = \begin{bmatrix} 1 & \mu_\beta' \\ \mu_\beta & \Sigma_\beta \end{bmatrix}$, $\sigma^2 = \lim \frac{1}{N}\sum_{i=1}^N \sigma_i^2$, $U_\epsilon = \lim \frac{1}{N}\sum_{i,j=1}^N E\Big[\text{vec}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\text{vec}(\epsilon_j\epsilon_j' - \sigma_j^2 I_T)'\Big]$, $M = I_T - D(D'D)^{-1}D'$, where μ_β, Σ_β, and σ_i^2 are defined in our assumptions above, U_ϵ is described in Appendix C, $D = [1_T,\ F]$, $Q = \frac{1_T}{T} - \mathcal{P}\gamma_1^P$, $Z = (Q \otimes \mathcal{P}) + \frac{\text{vec}(M)}{T-K-1}\gamma_1^{P'}\mathcal{P}'\mathcal{P}$, and $\otimes$ and $\text{vec}(\cdot)$ denote the Kronecker product operator and the vec operator, respectively.

We make the following further assumption to derive the large-N distribution of the Shanken (1992) estimator.

Assumption 6 *As $N \to \infty$, we have*

(i)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \epsilon_i \xrightarrow{d} \mathcal{N}\left(0_T, \sigma^2 I_T\right). \tag{36}$$

(ii)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \text{vec}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T) \xrightarrow{d} \mathcal{N}(0_{T^2}, U_\epsilon). \tag{37}$$

(iii) For a generic T-vector C_T,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(C_T' \otimes \begin{pmatrix} 1 \\ \beta_i \end{pmatrix}\right)\epsilon_i \xrightarrow{d} \mathcal{N}(0_{K+1}, V_c), \tag{38}$$

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 at the cost of raising the level of complexity of our proofs. For instance, when Eqs. (29)-(30) hold, then Eq. (36) follows by Theorem 2 of Kuersteiner and Prucha (2013) when the η_{it} satisfy their martingale difference assumptions. (See their Assumptions 1 and 2.) This result extends easily to Eqs. (37)-(38) under suitable additional assumptions. (Details are available upon request.) We are now ready to state our first theorem.

Theorem 1 *As $N \to \infty$, we have* *(i) Under Assumptions 1–5,*

$$\hat{\Gamma}^* - \Gamma^P = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{39}$$

(ii) Under Assumptions 1–6,

$$\sqrt{N}\left(\hat{\Gamma}^* - \Gamma^P\right) \to_d \mathcal{N}\left(0_{K+1}, V + \Sigma_X^{-1} W \Sigma_X^{-1}\right), \tag{40}$$

where

$$V = \frac{\sigma^2}{T}\left[1 + \gamma_1^{P\prime}\left(\frac{\tilde{F}'\tilde{F}}{T}\right)^{-1}\gamma_1^P\right]\Sigma_X^{-1} \tag{41}$$

and

$$W = \begin{bmatrix} 0 & 0_K' \\ 0_K & Z'U_\epsilon Z \end{bmatrix}. \tag{42}$$

Proof: See Appendix B.

The expression in Eq. (40) is remarkably simple and has a neat interpretation. The first term of this asymptotic covariance, V, accounts for the estimation error in the betas, and it is essentially identical to the large-T expression of the asymptotic covariance matrix associated with the OLS CSR estimator in Shanken (1992). (See his Theorem 1(ii).) The term $\frac{\sigma^2}{T}\Sigma_X^{-1}$ in Eq. (41) is the classical OLS CSR covariance matrix, which one would obtain if the betas were observed. The term $c = \gamma_1^{P\prime}\left(\tilde{F}'\tilde{F}/T\right)^{-1}\gamma_1^P$ is an asymptotic EIV adjustment, with $c\frac{\sigma^2}{T}\Sigma_X^{-1}$ being the corresponding overall EIV contribution to 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 the residual variance, σ^2, and inversely related to the factors' variability, $\left(\tilde{F}'\tilde{F}/T\right)^{-1}$. The last term of the asymptotic covariance, $\Sigma_X^{-1}W\Sigma_X^{-1}$ in Eq. (40), arises because of the bias adjustment that characterizes $\hat{\Gamma}^*$. The W matrix in Eq. (42) accounts for the cross-sectional variation in the residual variances of the asset returns through U_ϵ. This term will vanish when $T \to \infty$. In Appendix C, we provide an explicit expression for U_ϵ, and we show that U_ϵ only depends on the fourth-moment structure of the ϵ_{it}, that is, on κ_4 and σ_4.[31] The $\sqrt{N}$-rate of convergence obtained in Theorem 1-(i) coincides with the rate of convergence established by Gagliardini et al. (2016) with respect to their $\sqrt{NT}$-consistent estimator of $\nu = \gamma_1^P - \bar{f}$ when T is fixed.

[31] See Assumption 5 for the definition of κ_4 (the cross-sectional average of the fourth-order cumulants of the ϵ_{it}) and σ_4 (the cross-sectional average of the σ_i^4).

To conduct statistical inference, we need a consistent estimator of the asymptotic covariance matrix, which we present in the next theorem. Let $M^{(2)} = M \odot M$, where $\odot$ denotes the Hadamard product operator. In addition, define

$$\hat{Z} = (\hat{Q} \otimes \mathcal{P}) + \frac{\text{vec}(M)}{T-K-1} \hat{\gamma}_1^{*\prime} \mathcal{P}' \mathcal{P} \qquad \text{with} \qquad \hat{Q} = \frac{1_T}{T} - \mathcal{P}\hat{\gamma}_1^*. \tag{43}$$

Theorem 2 *Under Assumptions 1-5 and the identification condition $\kappa_4 = 0$, as $N \to \infty$, we have*

$$\hat{V} + \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \hat{W} \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \to_p V + \Sigma_X^{-1} W \Sigma_X^{-1}, \tag{44}$$

where

$$\hat{V} = \frac{\hat{\sigma}^2}{T} \left[1 + \hat{\gamma}_1^{*\prime} \left(\frac{\tilde{F}'\tilde{F}}{T} \right)^{-1} \hat{\gamma}_1^* \right] (\hat{\Sigma}_X - \hat{\Lambda})^{-1}, \tag{45}$$

$$\hat{W} = \begin{bmatrix} 0 & 0_K' \\ 0_K & \hat{Z}'\hat{U}_\epsilon \hat{Z} \end{bmatrix}, \tag{46}$$

and $\hat{U}_\epsilon$ is a consistent estimator of U_ϵ (see Appendix C), obtained replacing σ_4 with

$$\hat{\sigma}_4 = \frac{\frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \hat{\epsilon}_{it}^4}{3\text{tr}\left(M^{(2)}\right)}. \tag{47}$$

Proof: See Appendix B.

A remarkable feature of the result above is that a consistent estimate of the asymptotic covariance matrix of $\hat{\Gamma}^*$ can be obtained while leaving the residual covariance matrix Σ unspecified. In fact, with Σ having in general $N(N+1)/2$ distinct elements and our asymptotic theory being valid only for $N \to \infty$, consistent estimation of Σ would be infeasible. A convenient feature of the Shanken (1992) estimator is that it depends on Σ only through the *average* of the σ_i^2. Moreover, its asymptotic covariance matrix depends on the limits of $\sum_{i,j=1}^N \sigma_{ij}/N$ and $\sum_{i=1}^N \sigma_i^4/N$. Our large N asymptotic theory shows how these quantities can be estimated consistently. In contrast, the individual covariances σ_{ij} cannot be consistently estimated due to the fixed T. The condition $\kappa_4 = 0$ is required as a consequence of the small-T and large-N framework.[32] However, $\kappa_4 = 0$

[32] As we show in detail in Lemma 6 of Appendix A, the limit of $\hat{\sigma}_4$ in Eq. (47) converges to a linear combination of k_4 and σ_4. These two parameters could be identified and consistently estimated only under the stronger assumption of independence across assets, since, in this case, σ_4 would reduce to σ^4 (which could be easily estimated using the square of $\hat{\sigma}^2$). In contrast, allowing for some arbitrary degree of cross-correlation implies that k_4 and σ_4 cannot be separately identified. This is the reason for setting $k_4 = 0$.

is not as restrictive as it may seem. A sufficiently large level of heterogeneity in the σ_i^2 generates a substantial level of volatility in the conditional distribution of assets' returns by inducing a mixture-distribution effect.[33]

2.2 Time-varying case

In this section, we study the behavior of the estimator $\hat{\Gamma}^*$ when the risk premia are allowed to be time-varying, again under the assumption of correct model specification. It turns out that $\hat{\Gamma}^*$ is suitable for time-varying risk premia estimation because it estimates accurately *local* averages (over the, possibly very short, time window of size $T > K + 1$) of the true time-varying risk premia, regardless of their form and degree of time variation. Noticeably, we are also able to derive a consistent estimator of the true t-th period risk premia and to characterize its asymptotic distribution.[34]

Throughout this section, we substitute Assumption 1 with

$$E_{t-1}[R_{it}] = \gamma_{0,t-1} + \beta_i'\gamma_{1,t-1}, \tag{48}$$

where $E_{t-1}[\cdot]$ denotes the conditional expectation with respect to all the available information up to time $t-1$. Importantly, our theory does not need to restrict the type of time variation in $\Gamma_{t-1} = \left[\gamma_{0,t-1}, \gamma_{1,t-1}'\right]'$. To simplify the treatment of time variation in the premia, without altering the estimation procedure developed in this paper, we maintain the β_i in Eq. (48) constant over time.[35] Our results below easily extend to the case of $\beta_{i,t-1} = B_i z_{t-1}$, for some (vector of) predetermined state variables z_{t-1} and a suitable matrix of loadings B_i.

Under Eq. (48), asset returns are now given by $R_{it} = [1,\ \beta_i']\Gamma_{t-1}^P + \epsilon_{it}$, where Γ_{t-1}^P are the $(t-1)$-th ex post risk premia:

$$\Gamma_{t-1}^P = \Gamma_{t-1} + f_t - E_{t-1}[f_t], \text{ with a sample average } \bar{\Gamma}^P = \frac{1}{T}\sum_{t=1}^{T}\Gamma_{t-1}^P. \tag{49}$$

By construction, the ex post time-varying risk premia Γ_{t-1}^P have a conditional mean that equals Γ_{t-1}, the ex ante time-varying risk premia.

[33] In our empirical applications our estimate σ_4 is about 10 times the estimate for σ^4.

[34] Our new estimator for the time-varying risk premia appears useful also for traded factors, and not just for non-traded factors, particularly within our fixed-T environment (see the Internet Appendix IA.2 for further details), especially when T is assumed to be very small.

[35] See, e.g., Ferson and Harvey (1991) who argue that the time variation in expected returns is mainly due to time variation in the premia as opposed to time variation in the betas.

To estimate the (t–1)-th risk premia, for $t = 1, \ldots, T$, we introduce the following novel estimator:

$$\hat{\Gamma}^*_{t-1} = \left[\begin{array}{c} \hat{\gamma}^*_{0,t-1} \\ \hat{\gamma}^*_{1,t-1} \end{array}\right] = \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}' R_t}{N} - \hat{\sigma}^2 \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left(\begin{array}{c} 0 \\ (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\imath_{t,T} \end{array}\right), \tag{50}$$

where, as before, $\imath_{t,T}$ denotes the t-th column, for $t = 1, \ldots, T$, of the identity matrix I_T.[36] The next theorem derives the large-N behavior of both $\hat{\Gamma}^*$ and $\hat{\Gamma}^*_{t-1}$.

Theorem 3 *Under Eq. (48) and Assumptions 2-6, as $N \to \infty$, we have*

(i) $\hat{\Gamma}^*$ *and* $\sqrt{N}(\hat{\Gamma}^* - \bar{\Gamma}^P)$ *satisfy Theorem 1 with* Γ^P *replaced by* $\bar{\Gamma}^P$.

(ii) $\hat{\Gamma}^*_{t-1} - \Gamma^P_{t-1} = O_p\left(\frac{1}{\sqrt{N}}\right)$ *and*

$$\sqrt{N}\left(\hat{\Gamma}^*_{t-1} - \Gamma^P_{t-1}\right) \to_d \mathcal{N}\left(0_{K+1}, V_{t-1} + \Sigma_X^{-1} W_{t-1} \Sigma_X^{-1}\right), \tag{51}$$

where $V_{t-1} = \sigma^2 Q'_{t-1} Q_{t-1} \Sigma_X^{-1}$, $W_{t-1} = \left[\begin{array}{cc} 0 & 0'_K \\ 0_K & Z'_{t-1} U_\epsilon Z_{t-1} \end{array}\right]$, $Q_{t-1} = \imath_{t,T} - \mathcal{P}\gamma^P_{1,t-1}$, *and* $Z_{t-1} = (Q_{t-1} \otimes \mathcal{P}) - \frac{\mathrm{vec}(M)}{T-K-1} Q'_{t-1}\mathcal{P}$, *with* U_ϵ *as in Theorem 1.*

Proof: See Appendix B.

Theorem 3 states that, when Eq. (48) holds, $\hat{\Gamma}^*$ consistently estimates the local average of the ex post time-varying risk premia over T periods, the only requirement being that $T > K+1$. If one is interested in the ex post risk premia for a specific time period, Γ^P_{t-1}, then asymptotically correct inference can be carried out by using $\hat{\Gamma}^*_{t-1}$. Interestingly, $\hat{\Gamma}^*$ is numerically identical to the sample mean of $\hat{\Gamma}^*_{t-1}$, over $t = 1, \ldots, T$, because the additive bias adjustment, on the right-hand side of Eq. (50), vanishes due to the identity $\sum_{t=1}^T \tilde{F}'\imath_{t,T} = \tilde{F}'1_T = 0$.

To better understand the importance of our large-N results, it is useful to consider the behavior of the OLS CSR estimator $\hat{\Gamma}$ when Eq. (48) holds. In this case, we have

$$\hat{\Gamma} \to_p \Gamma_\infty \quad \text{as } T \to \infty, \tag{52}$$

where $\Gamma_\infty = \lim_{T\to\infty} \frac{1}{T}\int_0^T \Gamma_s ds$ denotes the *integrated* risk premia, namely the long-run average over the entire timeline.[37] Next, consider $\hat{\Gamma}_{t-1} = (\hat{X}'\hat{X})^{-1}\hat{X}'R_t$, which can be thought of as the

[36] Note that $\hat{\Gamma}^*_{t-1}$ is a new estimator that successfully tackles the problem of estimating time-varying risk premia in a large-N setting. It should not be confused with the Shanken (1992) formula in his Theorem 5.

[37] If one assumes, as in Ang and Kristensen (2012), that $\Gamma_t = \Gamma(t/T)$, $1 \le t \le T$, for a smooth function $\Gamma(\cdot)$, then the integrated risk premia Γ_∞ become $\int_0^1 \Gamma_s ds$.

OLS CSR estimator for the $(t-1)$-th risk premia.[38] It follows that

$$\hat{\Gamma}_{t-1} \to_p \Gamma_{t-1}^P + \begin{pmatrix} N & 1_N' B \\ B' 1_N & B'B \end{pmatrix}^{-1} \begin{pmatrix} 1_N' \\ B' \end{pmatrix} \epsilon_t \quad \text{as } T \to \infty. \tag{53}$$

Hence, the limit of $\hat{\Gamma}_{t-1}$ is the sum of two components, that is, the $(t-1)$-th ex post risk premia Γ_{t-1}^P and a random term that is a function of ϵ_t. This last term cannot be consistently estimated, thus making $\hat{\Gamma}_{t-1}$ an unreliable estimator of both Γ_{t-1} and Γ_{t-1}^P, even when $T \to \infty$. In contrast, in our large-N framework, $\left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}' R_t}{N} \to_p \Gamma_{t-1}^P + \sigma^2 \Sigma_X^{-1} \begin{pmatrix} 0 \\ (\tilde{F}'\tilde{F})^{-1} \tilde{F}' \imath_{t,T} \end{pmatrix}$ as $N \to \infty$, where the bias term $\sigma^2 \Sigma_X^{-1} \begin{pmatrix} 0 \\ (\tilde{F}'\tilde{F})^{-1} \tilde{F}' \imath_{t,T} \end{pmatrix}$ can now be consistently estimated, leading to the bias-adjusted estimator $\hat{\Gamma}_{t-1}^*$ in Eq. (50). Finally, a consistent estimator of the asymptotic covariance matrix of $\hat{\Gamma}_{t-1}^*$ in Eq. (51) can be easily obtained. (See Theorem 2 and its proof.)

3. Asymptotic Analysis under Potentially Misspecified Models

In this section, we explore the implications of model misspecification for model and parameter testing. Under the full rank assumption on the X matrix, the focus of the analysis is on the fixed (global) type of misspecification considered in Shanken and Zhou (2007) and several follow-up papers. A beta-pricing model is misspecified if there exists no value of the risk premia Γ for which the associated vector of pricing errors is zero. This misspecification might be due, for example, to the omission of some relevant risk factor, imperfect measurement of the factors, or failure to incorporate some relevant aspect of the economic environment – taxes, transaction costs, irrational investors, and the like. Thus, misspecification of some sort seems inevitable, given the inherent limitations of beta-pricing models.

This section is organized as follows. In Section 3.1, we propose a new specification test that is appropriately designed to detect model misspecification of unknown form. Section 3.2 deals with risk premia estimation and provides standard errors that are valid under potential model misspecification. Finally, Section 3.3 explores the situation in which the beta-pricing model is misspecified due to priced firm characteristics.

[38] The quantity $\hat{\Gamma}_{t-1}$ is well-known in empirical finance because its sample variance is routinely used to compute the Fama and MacBeth (1973) standard errors of $\hat{\Gamma}$.

3.1 Testing for model misspecification

When a beta-pricing model is correctly specified (see Assumption 1),

$$H_0: \; e_i = 0 \qquad \text{for every } i = 1, 2, \ldots, \tag{54}$$

where $e_i = E[R_{it}] - \gamma_0 - \beta_i' \gamma_1$ is the population (ex ante) pricing error associated with asset i. Denoting the vector of sample ex post pricing errors by

$$\hat{e}^P = (\hat{e}_1^P, \ldots, \hat{e}_N^P)' = \bar{R} - \hat{X}\hat{\Gamma}^*, \tag{55}$$

we have

$$\begin{aligned} \hat{e}_i^P &= \bar{R}_i - \hat{X}_i \hat{\Gamma}^* \\ &= e_i + Q'\epsilon_i - \hat{X}_i \left(\hat{\Gamma}^* - \Gamma^P \right). \end{aligned} \tag{56}$$

Theorem 1(i) implies that, for every i,

$$\hat{e}_i^P \to_p e_i + Q'\epsilon_i \equiv e_i^P. \tag{57}$$

Eq. (57) shows that even when the ex ante pricing errors, e_i, are zero, $\hat{e}_i^P$ will not converge in probability to zero because T is fixed. Nonetheless, a test of H_0 with correct size and good power can be developed. Define the sum of the sample squared ex post pricing errors as

$$\hat{\mathcal{Q}} = \frac{1}{N} \sum_{i=1}^{N} (\hat{e}_i^P)^2. \tag{58}$$

Consider the centered statistic

$$\mathcal{S} = \sqrt{N} \left(\hat{\mathcal{Q}} - \frac{\hat{\sigma}^2}{T} \left(1 + \hat{\gamma}_1^{*\prime} (\tilde{F}'\tilde{F}/T)^{-1} \hat{\gamma}_1^* \right) \right). \tag{59}$$

The centering is needed because of Eq. (57). To see this, from the population ex post pricing errors, e_i^P, we have

$$\frac{1}{N} \sum_{i=1}^{N} (e_i^P)^2 = \frac{1}{N} \sum_{i=1}^{N} e_i^2 + Q' \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' \right) Q + o_p(1) = \frac{1}{N} \sum_{i=1}^{N} e_i^2 + \sigma^2 Q'Q + o_p(1). \tag{60}$$

Therefore, even under $H_0 : e_i = 0$ for all i, the average of the population squared ex post pricing errors will not converge to zero but rather to $\sigma^2 Q'Q = \sigma^2 (1 + \gamma_1^{*\prime} (\tilde{F}'\tilde{F}/T)^{-1} \gamma_1^*)$. This is the quantity whose consistent estimate we need to demean our test statistic by in order to obtain its limiting distribution. The following theorem provides the limiting distribution of $\mathcal{S}$ under $H_0 : e_i = 0$ for every i.

Theorem 4 *Under Eq. (54) and Assumptions 1-6, as $N \to \infty$, we have*

$$\mathcal{S} \to_d \mathcal{N}(0, \mathcal{V}), \tag{61}$$

where $\mathcal{V} = Z_Q' U_\epsilon Z_Q$ *and* $Z_Q = (Q \otimes Q) - \frac{\mathrm{vec}(M)}{T-K-1} Q'Q$.

Proof: See Appendix B.

The asymptotic variance of the test in Eq. (61) can be consistently estimated by replacing Q with $\hat{Q}$ and U_ϵ with $\hat{U}_\epsilon$. Specifically, using Theorem 2 and Lemma 6 in Appendix A, we have

$$\hat{\mathcal{V}} = \hat{Z}_Q' \hat{U}_\epsilon \hat{Z}_Q \to_p Z_Q' U_\epsilon Z_Q, \tag{62}$$

where

$$\hat{Z}_Q = \left(\hat{Q} \otimes \hat{Q}\right) - \frac{\mathrm{vec}(M)}{T-K-1} \hat{Q}'\hat{Q}. \tag{63}$$

Then, under H_0, it follows that

$$\mathcal{S}^* = \frac{\mathcal{S}}{\hat{\mathcal{V}}^{\frac{1}{2}}} \to_d \mathcal{N}(0, 1). \tag{64}$$

It turns out that our test statistic $\mathcal{S}^*$ has power when e_i^2 is greater than zero for the majority of the test assets.[39] Moreover, it is straightforward to show that the distribution of our test under the null hypothesis is invariant to asset repackaging.

3.2 Estimation under potential model misspecification

If the null hypothesis of correct model specification, for the beta-pricing model under consideration, is rejected, one has two options. The first possibility is to conclude that the model is wrong, and to modify the model accordingly before proceeding with risk premia estimation. If one still wishes to conduct inference on risk premia with the same beta-pricing model, then the standard errors of the risk premia estimates need to be robustified against potential model misspecification. This is the approach we propose in this section. Suppose that Assumption 1 is violated and assume that

$$E[R_t] = 1_N \tilde{\gamma}_0 + B\tilde{\gamma}_1 + e, \tag{65}$$

[39] Specifically, our test will reject H_0 when the pricing errors e_i are zero for only a number N_0 of assets, such that $N_0/N \to 0$ as $N \to \infty$. This condition allows N_0 to diverge, although not too fast. A formal power analysis can be developed by using the notion of local alternatives as in Gagliardini et al. (2016). In the Internet Appendix, we present a Monte Carlo simulation experiment calibrated to real data that demonstrates the desirable size and power properties of our test.

where, following Shanken and Zhou (2007), the (pseudo)-true values $\tilde{\Gamma} = [\tilde{\gamma}_0, \tilde{\gamma}_1']'$ are given by

$$\tilde{\Gamma} = argmin_C \frac{(E[R_t] - XC)'(E[R_t] - XC)}{N}, \text{ for an arbitrary } (K+1)\text{-vector } C. \tag{66}$$

When the model is correctly specified, $\tilde{\Gamma} = \Gamma$, the vector of ex ante risk premia.[40]

We now introduce an additional assumption that governs the behavior of the population pricing errors in terms of cross-sectional moments with the returns' innovations.

Assumption 7 *As $N \to \infty$, we have*

(i)

$$\frac{1}{N} \sum_{i=1}^{N} \epsilon_i e_i \to_p 0. \tag{67}$$

(ii)

$$\frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' e_i^2 \to_p \tau_\Omega I_T. \tag{68}$$

(iii)

$$\frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' e_i \to_p \tau_\Phi I_T. \tag{69}$$

(iv)

$$\sum_{i,j=1}^{N} \mid \sigma_{ij} e_i e_j \mid \mathbb{1}_{\{i \neq j\}} = o\left(N\right), \tag{70}$$

for some constants $\tau_\Omega = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \epsilon_{it}^2 e_i^2$ and $\tau_\Phi = \text{plim} \frac{1}{N} \sum_{i=1}^{N} \epsilon_{it}^2 e_i$.

Assumption 7(i) implies that the ϵ_{it} and the pricing errors are cross-sectionally uncorrelated, although, by Assumption 7(ii) and 7(iii), they could be cross-sectionally dependent in terms of second moments of the ϵ_{it}. Assumption 7(iv) implies that the pricing errors are not altering the degree of cross-sectional dependence of the ϵ_{it}.

[40] Under the *i.i.d.* normality assumption and Eq. (65), Shanken and Zhou (2007) establish the asymptotic distribution of the OLS and GLS CSR estimators of $\tilde{\Gamma}$ as $T \to \infty$. (See also Hou and Kimmel (2006).) Kan et al. (2013) generalize their results to the case of temporally dependent and nonnormal test asset returns and factors, and derive the large-T distribution of the OLS and GLS CSR R^2.

Let $\tilde{\Gamma}^P = \tilde{\Gamma} + \bar{f} - E[f_t]$. The following theorem extends Theorems 1 and 2 to the case of globally misspecified beta-pricing models.

Theorem 5 *As $N \to \infty$, we have*

(i) Under Assumptions 2-5, Assumption 7, and Eq. (65),

$$\hat{\Gamma}^* - \tilde{\Gamma}^P = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{71}$$

(ii) Under Assumptions 2-7 and Eq. (65),

$$\sqrt{N}\left(\hat{\Gamma}^* - \tilde{\Gamma}^P\right) \to_d \mathcal{N}\left(0_{K+1}, V + \Sigma_X^{-1}\left(W + \Omega + \Phi + \Phi'\right)\Sigma_X^{-1}\right), \tag{72}$$

where V and W are defined in Theorem 1 by replacing γ_1^P with $\tilde{\gamma}_1^P$,

$$\Omega = \begin{bmatrix} 0 & 0_K' \\ 0_K & \tau_\Omega \mathcal{P}'\mathcal{P} \end{bmatrix} \quad and \quad \Phi = \begin{bmatrix} 0 & \tau_\Phi Q'\mathcal{P} \\ 0_K & \tau_\Phi\left(Q' \otimes \mu_\beta\right)\mathcal{P} \end{bmatrix}. \tag{73}$$

(iii) Under Assumptions 2-5, Assumption 7, Eq. (65), and $\kappa_4 = 0$,

$$\hat{V} + \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}(\hat{W} + \hat{\Omega} + \hat{\Phi} + \hat{\Phi}')\left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \to_p V + \Sigma_X^{-1}(W + \Omega + \Phi + \Phi')\Sigma_X^{-1}, \tag{74}$$

where $\hat{V}$ and $\hat{W}$ are defined in Theorem 2,

$$\hat{\Omega} = \begin{bmatrix} 0 & 0_K' \\ 0_K & \hat{\tau}_\Omega \mathcal{P}'\mathcal{P} \end{bmatrix} \; and \; \hat{\Phi} = \begin{bmatrix} 0 & \hat{\tau}_\Phi Q'\mathcal{P} \\ 0_K & \hat{\tau}_\Phi\left(Q' \otimes \frac{\hat{B}'1_N}{N}\right)\mathcal{P} \end{bmatrix}, \tag{75}$$

and $\hat{\tau}_\Phi$ and $\hat{\tau}_\Omega$ are defined in Lemmas 8 and 9 in Appendix A, respectively.

Proof: See Appendix B.

Similar to the expressions in Shanken and Zhou (2007) and Kan et al. (2013), the asymptotic covariance of $\hat{\Gamma}^*$ contains three additional terms, Ω, Φ, and Φ'. The contribution of the pricing errors to the overall asymptotic covariance increases when the variability of the residuals ϵ_{it} increases or, alternatively, when the variability of the pricing errors e_i increases, leading to a larger τ_Ω.

Notice that under model misspecification $\tilde{\Gamma}$ changes with N and, as a consequence, one can define the limit risk premia $\tilde{\Gamma}_\infty = \lim_{N\to\infty}\tilde{\Gamma}$. Theorem 3 of Ingersoll (1984) provides the conditions for the existence and the uniqueness of $\tilde{\Gamma}_\infty$.[41] It follows that, by Theorem 5, $\hat{\Gamma}^*$ also converges to $\tilde{\Gamma}^P_\infty = \left[\tilde{\gamma}^P_{0,\infty}, \tilde{\gamma}^{P\prime}_{1,\infty}\right]' = \tilde{\Gamma}_\infty + \bar{f} - E[f_t]$. Moreover, if $\tilde{\Gamma} - \tilde{\Gamma}_\infty$ is $o\left(1/\sqrt{N}\right)$, then the asymptotic distribution of $\hat{\Gamma}^*$ around $\tilde{\Gamma}^P_\infty$ is the same as the one in Eq. (72).[42] Interestingly, even under model misspecification, there is no loss of speed of convergence. This differs from Gagliardini et al. (2016), who obtain a slower rate of convergence, $O\left(\sqrt{N}\right)$ instead of $O\left(\sqrt{NT}\right)$, of their estimator to the true ex ante risk premia, $\tilde{\Gamma}_\infty$, when the model is misspecified.

3.3 Misspecification due to priced characteristics

We follow Section 3.3 of Shanken (1992) and allow for Assumption 1 to be potentially violated because the cross-section of expected returns now satisfies

$$E[R_{it}] = \gamma_0 + \gamma_1'\beta_i + \delta' c_i, \tag{76}$$

where c_i denotes a K_c-vector of time-invariant firm characteristics and δ denotes the corresponding vector of characteristic premia. Our theory requires characteristics and loadings to be sufficiently heterogenous across assets although we allow them to be (almost) arbitrarily cross-sectionally correlated.[43] Since characteristics exhibit only modest changes over short time windows, Eq. (76) would be a good approximation to the true data generating process also in a time-varying setting with a small T.[44]

Imposing Eq. (76), averaging (2) over time, and replacing X with $\hat{X}$, we obtain

$$\bar{R} = \hat{X}\Gamma^P + C\delta + \eta^P, \tag{77}$$

[41] In particular, asymptotic no-arbitrage (see Ingersoll (1984), Eq. (7)), our Assumption 2, and boundedness of the maximum eigenvalue of Σ imply Ingersoll's result.

[42] It can be shown that (deterministic) convergence of $\tilde{\Gamma}$ to $\tilde{\Gamma}_\infty$ occurs at most at rate $O\left(1/\sqrt{\sum_{i=1}^N \beta_i'\beta_i}\right)$, which equals $O\left(1/\sqrt{N}\right)$ by Assumption 2, although any faster rate is allowed for in principle. Notice that if $\tilde{\Gamma} - \tilde{\Gamma}_\infty$ is exactly $O\left(1/\sqrt{N}\right)$, then we need to modify our sampling scheme and select an arbitrary, slightly smaller, set of assets n such that $n/N \to 0$ as N diverges. When evaluating $\hat{\Gamma}^*$ using these n assets, then the slower $O\left(\sqrt{n}\right)$ rate of convergence to $\tilde{\Gamma}^P_\infty$ is obtained.

[43] The case for (linear or nonlinear) dependence, whereby $\beta_i = \beta(c_i)$, has been forcefully made by both the empirical (see Connor et al. (2012), Chordia et al. (2015), and Kelly et al. (2018), among others) and theoretical literature (see the survey in Kogan and Papanikolaou (2013)) in order to resolve the debate on systematic risk- versus characteristic-based stories of expected returns that was spurred from the influential empirical findings of Daniel and Titman (1997).

[44] Chordia et al. (2015) highlight the challenges that arise when estimating time-varying characteristic premia and propose a bootstrap procedure to perform correct inference in this setting.

where $C = [c_1, \ldots, c_N]'$ and $\eta^P = \left(\bar{\epsilon} - (\hat{X} - X)\Gamma^P\right)$. The estimates of Γ^P and δ are given by

$$\begin{bmatrix} \hat{\Gamma}^* \\ \hat{\delta}^* \end{bmatrix} = \begin{bmatrix} \hat{X}'\hat{X} - N\hat{\Lambda} & \hat{X}'C \\ C'\hat{X} & C'C \end{bmatrix}^{-1} \begin{bmatrix} \hat{X}'\bar{R} \\ C'\bar{R} \end{bmatrix}, \tag{78}$$

where $\hat{\Lambda}$ is the bias adjustment from Theorem 1. In line with the discussion around Theorem 3, $\hat{\Gamma}^*$ and $\hat{\delta}^*$ will also estimate (consistently) the local averages of the risk and characteristic premia if these are allowed to be time-varying.

In this setting with characteristics, we need to make the following additional assumption. Let $z_i = \epsilon_i \otimes c_i$ and $\Sigma_{zz,ij} = \mathrm{Cov}(z_i, z_j') = \sigma_{ij}\left[I_T \otimes c_i c_j'\right]$.

Assumption 8 *As $N \to \infty$,*

(i)

$$\hat{\mu}_C = \frac{C'1_N}{N} \qquad \to_p \mu_C = [\mu_{c1}, \ldots, \mu_{cK}]', \ \textit{a finite } K_c\textit{-vector}, \tag{79}$$

$$\hat{\Sigma}_{CC} = \frac{C'C}{N} \qquad \to_p \Sigma_{CC}, \ \textit{a finite positive-definite } (K_c \times K_c) \ \textit{matrix}, \tag{80}$$

$$\hat{\Sigma}_{CB} = \frac{C'B}{N} \qquad \to_p \Sigma_{CB}, \ \textit{a finite } (K_c \times K) \ \textit{matrix}, \tag{81}$$

with positive-definite matrices

$$\Sigma_{CC} - \mu_C\mu_C' \ \textit{and} \ \begin{bmatrix} \Sigma_{CC} & \Sigma_{CB} \\ \Sigma_{CB}' & \Sigma_\beta \end{bmatrix} - \begin{bmatrix} \mu_C \\ \mu_\beta \end{bmatrix}\begin{bmatrix} \mu_C \\ \mu_\beta \end{bmatrix}'. \tag{82}$$

(ii)

$$\frac{C'\epsilon'}{N} \to_p 0_{(K_c \times T)}. \tag{83}$$

(iii)

$$\frac{1}{N}\sum_{i=1}^{N} \Sigma_{zz,ii} \to \sigma^2(I_T \otimes \Sigma_{CC}) \ \textit{and} \ \sum_{i,j=1}^{N} \Sigma_{zz,ij} 1_{\{i \neq j\}} = o(N). \tag{84}$$

(iv)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} z_i \to_d \mathcal{N}\left(0_{K_cT}, \sigma^2(I_T \otimes \Sigma_{CC})\right). \tag{85}$$

Since $\begin{bmatrix} \Sigma_{CC} & \Sigma_{CB} \\ \Sigma_{CB}' & \Sigma_\beta \end{bmatrix} - \begin{bmatrix} \mu_C \\ \mu_\beta \end{bmatrix}\begin{bmatrix} \mu_C \\ \mu_\beta \end{bmatrix}$ in Assumption 8(i) is positive-definite, then $\begin{bmatrix} \Sigma_{CC} & \Sigma_{CB} \\ \Sigma_{CB}' & \Sigma_\beta \end{bmatrix}$ is also positive-definite, and this implies that the β_i and the c_i cannot be proportional.

In the next two theorems, we characterize the asymptotic properties of the estimators $\hat{\Gamma}^*$ and $\hat{\delta}^*$.

Theorem 6 *As $N \to \infty$, we have*

(i) Under Assumptions 2-5 and 8, and Eq. (76),

$$\hat{\Gamma}^* - \Gamma^P = O_p\left(\frac{1}{\sqrt{N}}\right), \hat{\delta}^* - \delta = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{86}$$

(ii) Under Assumptions 2-6 and 8, and Eq. (76),

$$\sqrt{N}\begin{bmatrix}\hat{\Gamma}^* - \Gamma^P \\ \hat{\delta}^* - \delta\end{bmatrix} \to_d \mathcal{N}\left(0_{K+K_c+1}, \sigma^2(Q'Q)L^{-1} + L^{-1}OL^{-1}\right), \tag{87}$$

with

$$L = \begin{bmatrix} \Sigma_X & \begin{bmatrix}\mu_C' \\ \Sigma_{CB}'\end{bmatrix} \\ \begin{bmatrix}\mu_C & \Sigma_{CB}\end{bmatrix} & \Sigma_{CC}\end{bmatrix}, \quad O = \begin{bmatrix} \begin{bmatrix} 0 & 0_K' \\ 0_K & Z'U_\epsilon Z\end{bmatrix} & 0_{(K+1)\times K_c} \\ 0_{K_c\times(K+1)} & 0_{K_c\times K_c}\end{bmatrix}, \tag{88}$$

where Q, Z, and U_ϵ are defined in Theorem 1.

Proof: See Appendix B.

A consistent estimator of the asymptotic covariance matrix of $\hat{\Gamma}^*$ and $\hat{\delta}^*$ is provided in the next theorem.[45]

Theorem 7 *Under Assumptions 2-5 and 8, Eq. (76), and the identification condition $\kappa_4 = 0$, as $N \to \infty$, we have*

$$\hat{\sigma}^2(\hat{Q}'\hat{Q})\hat{L}^{-1} + \hat{L}^{-1}\hat{O}\hat{L}^{-1} \to_p \sigma^2(Q'Q)L^{-1} + L^{-1}OL^{-1}, \tag{89}$$

with

$$\hat{L} = \begin{bmatrix} \hat{\Sigma}_X - \hat{\Lambda} & \begin{bmatrix}\hat{\mu}_C' \\ \hat{\Sigma}_{CB}'\end{bmatrix} \\ \begin{bmatrix}\hat{\mu}_C & \hat{\Sigma}_{CB}\end{bmatrix} & \hat{\Sigma}_{CC}\end{bmatrix}, \quad \hat{O} = \begin{bmatrix} \begin{bmatrix} 0 & 0_K' \\ 0_K & \hat{Z}'\hat{U}_\epsilon \hat{Z}\end{bmatrix} & 0_{(K+1)\times K_c} \\ 0_{K_c\times(K+1)} & 0_{K_c\times K_c}\end{bmatrix}, \tag{90}$$

where $\hat{\sigma}^2$ is defined in Eq. (11), and $\hat{Q}$, $\hat{Z}$, and $\hat{U}_\epsilon$ are defined in Theorem 2.

[45] The proof of Theorem 7 follows the same steps of the proof of Theorem 2 and is therefore omitted.

4. Empirical Analysis

In this section, we show empirically that the results obtained with our fixed-T and large-N methodology can differ substantially from the results obtained with traditional large-T and fixed-N methods. Using a large number of individual equity returns from CRSP, we estimate and test FF5 and an extension of this model that includes the non-traded liquidity factor of Pástor and Stambaugh (2003).[46] The demonstrated empirical success of FF5 in explaining the cross-sectional variation in expected equity returns is what motivates our interest in this model.[47] In the second part of this section, we analyze the extent to which firm characteristics contribute to explaining the cross-section of expected equity returns.

The risk and characteristic premia estimators, their confidence intervals, and the various test statistics employed are based on our theoretical analysis in Sections 2 and 3.

4.1 Data

The monthly data on the traded factors of FF5 is available from Kenneth French's website and the non-traded liquidity factor of Pástor and Stambaugh (2003) is taken from Lubos Pástor's website.[48] As for the test assets, we download monthly stock returns (from January 1966 to December 2013) from CRSP 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. We perform the empirical analysis using balanced panels over fixed-time windows of three and 10 years (that is, $T = 36$ and 120), respectively. We obtain time series of estimated risk premia and test statistics by shifting the time window month by month over the 1966-2013 period. After filtering the data, we obtain an average number (over the overlapping time windows) of approximately $2,800$ stocks

[46] The Internet Appendix reports further empirical results for FF5, as well as results for CAPM and FF3.

[47] Several studies (see Kozak et al. (2018), Kelly et al. (2018), and Huang et al. (2018), among others) have shown that these five factors are highly correlated with appropriately constructed latent factors such as the first five principal components, and variations of, from the data.

[48] The five traded factors of FF5 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).

when $T = 36$ and $1,200$ stocks when $T = 120$.

4.2 Specification testing

For the analysis with traded factors only, we report the p-values of our specification test, $\mathcal{S}^*$, as well as the p-values of two alternative tests, the Gibbons et al. (1989) (GRS) and Gungor and Luger (2016) (GL) tests. It should be noted that GRS requires N to be fixed, while the Gungor and Luger (2016) test is valid for any N and T. All three tests are tests of the same null hypothesis; that is, $H_0 : e_i = 0$, for every $i = 1, 2, \ldots$.

(i) $\mathcal{S}^*$ test

We first assess the performance of FF5 using $\mathcal{S}^*$.

Figure 1 about here

The black line in Figure 1 denotes the time series of p-values associated with our test statistic $\mathcal{S}^*$ for time windows of three years (top panel) and 10 years (bottom panel), respectively. When the black line is below the 5% significance level (dotted red line), we reject FF5. Figure 1 shows that based on our test, we reject the validity of FF5 about 60% of the times when $T = 36$. As expected, the rejection of FF5 happens more frequently when we increase the time window from $T = 36$ to $T = 120$. The rejection of FF5 occurs in about 95% of the cases when the latter scenario is considered. Given the availability of a time series of p-values, one could cast the analysis in a multiple testing framework, as suggested by Barras et al. (2010). Applying their methodology to $\mathcal{S}^*$, we reject the null of correct model specification in 61% and 95% of the cases for $T = 36$ and $T = 120$, respectively. In Figure 2, we perform the same analysis for the liquidity-augmented FF5.

Figure 2 about here

This variant of FF5 turns out to be strongly rejected, even when $T = 36$. The rejection frequencies are approximately equal to 82% and 92% for $T = 36$ and 120, respectively. Overall, the frequent and strong rejections of FF5 justify our use of confidence intervals that are robust to model misspecification in the subsequent analysis.

(ii) GRS and GL tests

Figure 3 reports the GRS p-values (blue line) as well as the GL p-values (green line).

Figure 3 about here

Unlike ours, these two tests are only applicable to beta-pricing models with traded factors. As a consequence, we consider only FF5 here. Since GRS is a GLS-based test, effectively, it is implementable only when N is substantially smaller than T. Therefore, we construct 25 equally weighted portfolio returns from our individual stock returns and analyze the performance of these two tests, using this smaller asset set.[49] Differently from our large-N test, we are much less likely to reject FF5 based on the GRS test. When considering time windows of $T = 36$, the average rejection rate for FF5 is only about 30%. In addition, FF5 is rejected almost always when $T = 120$. We obtain similar results when using the GL test, although it is harder to quantify the rejection rates in this case because the GL test often leads to an inconclusive outcome. Based on the GL test, FF5 is not rejected in about 70% of the cases when $T = 36$, but the test is inconclusive about 29% of the time. Moreover, FF5 is not rejected in only about 18% of the cases when $T = 120$, but the test is inconclusive about 76% of the time. The main message here is that using our test can lead to qualitatively different conclusions relative to existing methods.

4.3 Risk premia estimates

Since our test, $\mathcal{S}^*$, points to serious misspecification of the risk-return relation, in this section we perform parameter testing by means of standard errors that are robust to model misspecification. Specifically, we use the large-N standard errors derived in Theorems 5. To highlight the differences between our approach and standard large-T methods, we also consider the OLS CSR estimator and the corresponding large-T standard errors from Theorem 1(ii) in Shanken (1992). For traded factors, we also report the rolling sample mean of the factor returns, which is a valid risk premium estimator when T is large. In contrast, when considering non-traded factors such as liquidity, we consider the rolling sample mean of the corresponding mimicking portfolio return. (See footnote 16 above.)

(i) FF5

[49] The results in Figure 3 are obtained by randomly assigning the various stocks to 25 portfolios. For instance, when $T = 36$, each of the 25 portfolios contains approximately 110 randomly selected stocks. We also experimented with 25 portfolios formed on CAPM betas. The results of the analysis are qualitatively similar to those in Figure 3.

Based on a time window of three years, the top panel of Figure 4 presents the rolling-window estimates of the risk premium on the market factor and the corresponding 95% confidence intervals. (The results for the other four factors are in the Internet Appendix.)

Figure 4 about here

In the figure, the bold black line and the dotted red line refer to the Shanken (1992) and OLS CSR estimators, respectively. The grey band represents the large-N 95% confidence intervals that are robust to model misspecification, whereas the striped orange band is for the large-T confidence intervals. Finally, the dashed black line displays the rolling factor sample mean. Noticeably, the large-T confidence intervals include the zero value in about 60% of the cases. In contrast, our large-N confidence intervals include the zero value only about 30% of the time. Not surprisingly, the bottom panel of Figure 4 ($T = 120$ case) shows that the risk premia estimates are smoother than in the $T = 36$ scenario. However, the large-T confidence intervals are still larger than the corresponding large-N confidence intervals, and they indicate that the OLS CSR and the Shanken (1992) estimates are statistically significant 30% and 80% of the time, respectively. The large-N estimates appear to be systematically larger than the corresponding large-T estimates for most dates, especially for the longer time window. This is the result of the systematic (negative) bias that affects the OLS CSR estimator when N is large. The relationship between the large-N and the rolling sample mean estimates (the latter are based on windows of $T = 36$ and $T = 120$ monthly data, respectively) is less stable. The two sets of estimates exhibit a correlation of about 0.5 when $T = 36$ and 0.7 when $T = 120$. Figure 4 shows that the large-T approach supports the hypothesis of constant risk premia, whereas our large-N results point toward a significant time variation in risk premia. Therefore, it seems plausible to interpret $\hat{\Gamma}^*$ as the estimator of the local average, over T periods, of the (time-varying) risk premia, $\bar{\Gamma}$, as explained in Section 2.2.

The top panel of Figure 5 reports the Shanken (1992) large-N estimates, expressed in terms of a single line (black line) and in terms of local averages (horizontal bars of length $T = 36$, blue lines), with the corresponding 95% confidence intervals for these local averages based on the large-N standard errors of Theorem 5 (grey band).

Figure 5 about here

The local average estimates appear to be significantly different from each other in most cases, which is a clear symptom of time variation in risk premia. In the same panel, we also report the rolling sample mean (over fixed windows of six months of daily data) of the market excess return (dashed dotted red line) and the corresponding 95% confidence interval (orange band). As our results indicate, although the latter is a suitable (nonparametric) estimator of the time-varying risk premium, it requires a large number of observations (over a short time window) to produce sufficiently narrow confidence intervals. The correlation between the Shanken (1992) large-N estimator and the six-month rolling sample mean based on daily data is positive but small (the sample correlation coefficient is 0.14). In addition, differently from the Shanken (1992) large-N estimator, the six-month rolling sample mean based on daily data appears to be very noisy.

Given the pronounced time variation in risk premia, the bottom panel of Figure 5 reports our novel estimator $\hat{\gamma}^*_{1,t-1}$ (black line), formally defined in Eq. (50), and the corresponding 95% confidence interval (grey band). Although noisier than $\hat{\gamma}^*_1$, the $\hat{\gamma}^*_{1,t-1}$ estimates are still statistically significant about 50% of the time. As the figure indicates, there is a sharp increase in risk premia volatility in correspondence and in the aftermath of major economic and financial crises and episodes such as the Black Monday of October 1987 and the US savings and loan crisis of the 80s and 90s. Our empirical findings on risk premia counter-cyclicality confirm the results in Gagliardini et al. (2016) and corroborate the predictions of many theoretical models. (See the discussion in Section 4.3 of Gagliardini et al. (2016).)

(ii) Liquidity-augmented FF5

As for the liquidity-augmented FF5, Figure 6 presents the estimated liquidity risk premium in the time-invariant setting.

Figure 6 about here

The estimated liquidity risk premia in Figure 6 are positive 55% and 37% of the time for $T = 36$ and $T = 120$, respectively. However, the risk premia estimates are statistically significant at the 5% level only in the 21% and 32% of the cases, for $T = 36$ and $T = 120$, respectively. In the same figure, we also report the OLS CSR estimator and the corresponding mimicking portfolio rolling sample mean (based on windows of $T = 36$ and $T = 120$ monthly data). The OLS CSR estimates in this case are not too far from the Shanken (1992) estimates. In contrast, the rolling mimicking portfolio sample means are now only mildly positively correlated with the $\hat{\Gamma}^*$ estimates. (The correlation coefficients are 0.15 and 0.27 for $T = 36$ and $T = 120$, respectively.)

As in the traded factor case, Figure 7 indicates that the time variation in risk premia is pronounced.

Figure 7 about here

Based on the top panel of Figure 7, the correlation between the mimicking portfolio six-month rolling sample mean and the Shanken (1992) large-N estimates is about 0.19. Similar to the FF5 case, the large-N estimator seems to exhibit a higher precision. Looking at the bottom panel of Figure 7, the risk premia counter-cyclicality emerges again, especially around major economic and financial downturns.

Finally, Table 1 reports the percentage difference (averaged over rolling time windows of size $T = 36$ and $T = 120$, respectively) between the Shanken (1992) estimator, $\hat{\Gamma}^*$, and the OLS CSR estimator, $\hat{\Gamma}$, for the various risk premia in CAPM, FF3, and FF5.

Table 1 about here

Panel A shows that the percentage difference between estimators is quite large (about 64% when $T = 36$ and 27% when $T = 120$). As for FF3 in Panel B, the discrepancy between the two estimators is sizeable for *hml*, ranging from 31% to 52%, and less pronounced for *mkt* and *smb*. Moreover, relative to FF5, Panel C indicates that the percentage difference between the two estimators is relatively large for *cma*, ranging from 33% to about 43%. Finally, sizeable differences between the two estimators exist for *liq*, especially in Panel A.

In summary, we often find significant differences between the results based on our large-N approach and the results based on conventional large-T methods. The difference mainly stems from the smaller standard errors of the Shanken (1992) estimator relative to the OLS CSR estimator and the nontrivial bias correction induced by the Shanken (1992) estimator when N is large. These differences are even more pronounced when comparing the results based on the Shanken (1992) estimator with those based on the rolling sample mean estimator. Finally, the estimated risk premium on the (non-traded) liquidity factor of Pástor and Stambaugh (2003) is often found to be statistically insignificant.

4.4 Characteristics

In this section, for ease of comparison with Chordia et al. (2015), we use balanced panel data from January 1980 to December 2015.[50] In the dataset we use, an average of 3,071 firms have return data in a particular month. Consistent with Daniel and Titman (1997) and Chordia et al. (2015), among others, we focus on five firm characteristics that have often been found to be related to the cross-section of expected returns: book-to-market ratio (B/M), asset growth ($ASSGR$), operating profitability ($OPERPROF$), market capitalization ($MCAPIT$), and six-month momentum ($MOM6$). As it is common in this literature, we cross-sectionally standardize the characteristics.

In the interest of space, we focus only on the $T = 36$ case. For each time window, we compute the average of the characteristics. In the first pass, we obtain beta estimates for CAPM, FF3, and FF5. We then estimate the ex post risk and characteristic premia using our second-pass CSR estimator in Eq. (78). Figure 8 reports the time series of the characteristic premia estimates, $\hat{\delta}^*$, and the 95% confidence intervals for each model.

Figure 8 about here

Although the confidence intervals tend to widen when moving from CAPM to FF5, averaging across the three models, the estimated B/M premium is positive about 59% of the time, but it is only statistically significant at the 5% level in about 3% of the cases. The estimated $ASSGR$ premium is almost always negative (in 81% of the cases) and significantly so about 16% of the time, whereas the estimated $OPERPROF$ premium is positive in about 32% of the cases and statistically significant only about 19% of the time. For $MCAPIT$, the estimated premium is positive 32% of the time and statistically significant in about 12% of the cases, while the $MOM6$ estimate is almost always positive (99.6% of the time) and significant in 86% of the cases.

We now analyze the joint importance of the five characteristics in explaining deviations from correct model specification; that is, we assess whether the expected returns on individual stocks represent a compensation for risk or firm characteristics. We consider two alternative approaches. First, we conduct formal tests of the two hypotheses, $H_0 : \gamma_1^P = 0_K$ and $H_0 : \delta = 0_{K_c}$ using the asymptotic distribution theory in Theorems 6 and 7. The results are in Panel A of Table 2. The F-tests indicate that the characteristic premia estimates are statistically significant at any

[50] We thank Alberto Martín-Utrera for sharing his data with us and refer to DeMiguel et al. (2018) for data details.

conventional level, with the average F-test (over rolling windows of size $T = 36$) for the null hypothesis $H_0 : \delta = 0_{K_c}$ being equal to 1278.60, 1108.41, and 927.04 for CAPM, FF3, and FF5, respectively. In contrast, the average F-test for the null hypothesis $H_0 : \gamma_1^P = 0_K$ equals 12.45, 17.19, and 57.18 for CAPM, FF3, and FF5, respectively, with rejections rates, in the order, of 25.70%, 25.90%, and 37.90%.

Next, Panel B of Table 2 presents the cross-sectional variance contribution of betas and characteristics to the overall cross-sectional dispersion in the (sample) average returns, $\bar{R}_i$. Chordia et al. (2015) suggest to consider the ratios of the (cross-sectional) variance of the beta component (betas times the factor risk premia) and of the characteristics component (characteristics times the characteristic premia), with respect to the overall (cross-sectional) variance of average returns. However, since the beta and characteristics components are not orthogonal cross-sectionally, this can lead to a percentage of the cross-sectional variance explained by the betas and by the characteristics that is jointly greater than 100%.[51] In addition, the estimated pricing errors based on our bias-adjusted estimator are not necessarily orthogonal to the regressors of the CSR, thus complicating the interpretation even further.

We modify the approach of Chordia et al. (2015) as follows. From the estimated CSR, we have $\bar{R} = \hat{X}\hat{\Gamma}^* + C\hat{\delta}^* + \hat{\eta}^P$, where $\hat{\eta}^P$ are the sample counterparts of η^P in Eq. (77). Consider the orthogonalization of the estimated pricing errors, $\hat{\eta}^P$,

$$
\begin{aligned}
\bar{R} \quad &= \hat{X}\hat{\Gamma}^* + C\hat{\delta}^* + P_{\hat{Z}}\hat{\eta}^P + (I_N - P_{\hat{Z}})\hat{\eta}^P \\
&\equiv \hat{X}\hat{\Gamma}^* + C\hat{\delta}^* + P_{\hat{Z}}\hat{\eta}^P + \hat{\eta}^{*P},
\end{aligned} \tag{91}
$$

where $P_{\hat{Z}} = \hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'$ with $\hat{Z} = [\hat{X},\ C]$, and I_N denotes the identity matrix of order N. By construction, the orthogonalized estimated pricing errors, $\hat{\eta}^{*P} = (I_N - P_{\hat{Z}})\hat{\eta}^P$, satisfy $\hat{Z}'\hat{\eta}^{*P} = 0_{K+K_c+1}$. Setting $P_C = C(C'C)^{-1}C'$, rewrite the estimated CSR as

$$
\begin{aligned}
\bar{R} \quad &= \left(\hat{X}\hat{\Gamma}^* + P_{\hat{Z}}\hat{\eta}^P\right) + C\hat{\delta}^* + \hat{\eta}^{*P} \\
&= \left[(I_N - P_C)\left(\hat{X}\hat{\Gamma}^* + P_{\hat{Z}}\hat{\eta}^P\right)\right] + \left[P_C\left(\hat{X}\hat{\Gamma}^* + P_{\hat{Z}}\hat{\eta}^P\right) + C\hat{\delta}^*\right] + \hat{\eta}^{*P} \\
&\equiv \bar{R}_{\perp C} + \bar{R}_C + \hat{\eta}^{*P},
\end{aligned} \tag{92}
$$

where $\bar{R}_{\perp C} \equiv (I_N - P_C)\left(\hat{X}\hat{\Gamma}^* + P_{\hat{Z}}\hat{\eta}^P\right)$ is the component of the average returns that is explained only by the estimated betas, and thus (perfectly) uncorrelated with C in sample, and $\bar{R}_C \equiv$

[51] This problem is acknowledged, although not solved, in Chordia et al. (2015).

$P_C\left(\hat{X}\hat{\Gamma}^* + P_{\hat{Z}}\hat{\eta}^P\right) + C\hat{\delta}^*$ is the component of the average returns due to C only. Since $\bar{R}_{\perp C}$ and $\bar{R}_C$ are orthogonal to each other and to $\hat{\eta}^{*P}$, the sample variance of the average returns equals the sum of the sample variances of the beta component, of the characteristics component, and of the orthogonalized pricing errors, that is,

$$S^2_{\bar{R}} = \frac{\bar{R}'\bar{R}}{N} - \left(\frac{1_N'\bar{R}}{N}\right)^2 = \left(\frac{\bar{R}'_{\perp C}\bar{R}_{\perp C}}{N} - (\frac{1_N'\bar{R}_{\perp C}}{N})^2\right) + \left(\frac{\bar{R}'_C\bar{R}_C}{N} - (\frac{1_N'\bar{R}_C}{N})^2\right) + \frac{\hat{\eta}^{*P\prime}\hat{\eta}^{*P}}{N}$$
$$\equiv S^2_{\bar{R}_{\perp C}} + S^2_{\bar{R}_C} + S^2_{\hat{\eta}^{*P}}. \tag{93}$$

Panel B of Table 2 reports the average, over rolling windows of size $T = 36$, of the variance ratios $100 \times S^2_{\bar{R}_C}/S^2_{\bar{R}}$ and $100 \times S^2_{\bar{R}_{\perp C}}/S^2_{\bar{R}}$.

Table 2 about here

The results are largely supportive of our findings based on the F-tests; that is, characteristics overwhelmingly dominate the cross-sectional variation in average individual stock returns. Averaging across the three beta-pricing models, the characteristic variance ratio, $100 \times S^2_{\bar{R}_C}/S^2_{\bar{R}}$, is about 76%, whereas the beta variance ratio, $100 \times S^2_{\bar{R}_{\perp C}}/S^2_{\bar{R}}$, is about 2.8%. The rest (about 21.5%) represents the unexplained portion of the average return cross-sectional variance.[52] Overall, our empirical findings support the conclusions of Chordia et al. (2015), who argue that regardless of the beta-pricing model and whether the premia are allowed to be time-varying, it is mainly the characteristics that contribute to the cross-sectional variation in expected stock returns.

5. Conclusion

This paper is concerned with estimation of risk premia and testing of beta-pricing models when the data is available for a large cross-section of securities, N, but only for a fixed number of time periods, T. Since in this context the traditional OLS CSR estimator of the risk premia is asymptotically biased and inconsistent, we provide a new methodology built on the appealing bias-adjusted estimator of the ex post risk premia proposed by Shanken (1992). We establish its consistency and asymptotic normality for the baseline case of correctly specified beta-pricing models with constant risk premia, and then extend our setting to deal with time-varying risk premia. We

[52] Confidence intervals for these variance ratios could be computed based on our asymptotic results. The details are available upon request.

also explore in detail the case of misspecified beta-pricing models. We derive a new specification test and its large-N properties, and we then show how to robustify the asymptotic standard errors of the risk premia estimator when the beta-pricing relation is violated. The important case of misspecification due to priced firm characteristics is considered. Finally, we analyze the case of unbalanced panels.

We apply our large-N methodology to empirically investigate the performance of some prominent beta-pricing specifications using individual stock return data, that is, the monthly returns (from CRSP) on about 3,500 individual stocks for the January 1966 – December 2013 period. We consider three beta-pricing models: the CAPM, the three-factor model of Fama and French (1993), and the five-factor model of Fama and French (2015). We also augment these models with the (non-traded) liquidity factor of Pástor and Stambaugh (2003).

Our large-N test often rejects the Fama and French (2015) model, with and without the liquidity factor, at conventional significance levels even for short time windows of three years. In contrast, when using a suitable aggregation of the same data, in most cases we are unable to reject the Fama and French (2015) model using the traditional large-T methodologies. Similar conclusions hold when testing the validity of the CAPM and the Fama and French (1993) three-factor model, with and without the liquidity factor. The empirical rejection of these models suggests that the misspecification-robust standard errors derived in this paper should be employed when performing inference on risk premia.

Turning to estimation, our results indicate that all the traded-factor risk premia estimates are statistically significant most of the time, even over short time windows of three years. In contrast, the (non-traded) liquidity factor is often not priced. We also provide evidence of significant time variation in risk premia for both traded and non-traded factors. Our overall evidence of pricing is at odds with the results obtained using the traditional approach based on the large-T Shanken (1992) standard errors.

Finally, allowing for characteristics in the risk-return relation, we find that the book-to-market ratio, asset growth, operating profitability, market capitalization, and six-month momentum explain most of the cross-sectional variation in estimated expected stock returns. Monte Carlo simulations (in the Internet Appendix) corroborate our theoretical findings, both in terms of estimation and in terms of testing of the beta-pricing restriction.

Appendix A: Lemmas

Lemma 1 *Under Assumptions 3-5,*

$$\hat{\sigma}^2 - \sigma^2 = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{A.1}$$

Proof. Rewrite $\hat{\sigma}^2 - \sigma^2$ as

$$\begin{aligned}
\hat{\sigma}^2 - \sigma^2 &= \left(\hat{\sigma}^2 - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right) + \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 - \sigma^2\right) \\
&= \left(\hat{\sigma}^2 - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right) + o\left(\frac{1}{\sqrt{N}}\right) \qquad \text{(A.2)}
\end{aligned}$$

by Assumption 5(i). Moreover,

$$\begin{aligned}
\hat{\sigma}^2 - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 &= \frac{\operatorname{tr}(M\epsilon\epsilon')}{N(T-K-1)} - \frac{\operatorname{tr}(M)}{T-K-1}\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \\
&= \frac{\operatorname{tr}\left(P\left(\sum_{i=1}^{N}\sigma_i^2 I_T - \epsilon\epsilon'\right)\right)}{N(T-K-1)} + \frac{\operatorname{tr}(\epsilon\epsilon') - T\sum_{i=1}^{N}\sigma_i^2}{N(T-K-1)}. \qquad \text{(A.3)}
\end{aligned}$$

As for the second term on the right-hand side of Eq. (A.3), we have

$$\begin{aligned}
\frac{\operatorname{tr}(\epsilon\epsilon') - T\sum_{i=1}^{N}\sigma_i^2}{N(T-K-1)} &= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\epsilon_{it}^2 - \sigma_i^2\right)}{N(T-K-1)} \\
&= O_p\left(\frac{1}{\sqrt{N}}\frac{\sqrt{T}}{(T-K-1)}\right) = O_p\left(\frac{1}{\sqrt{N}}\right). \qquad \text{(A.4)}
\end{aligned}$$

As for the first term on the right-hand side of Eq. (A.3), we have

$$\begin{aligned}
\frac{\operatorname{tr}\left(P\left(\sum_{i=1}^{N}\sigma_i^2 I_T - \epsilon\epsilon'\right)\right)}{N(T-K-1)} &= \frac{\sum_{t=1}^{T} d_t\left(D'D\right)^{-1}D'\left(\sum_{i=1}^{N}\sigma_i^2\imath_{t,T} - \sum_{i=1}^{N}\epsilon_i\epsilon_{it}\right)}{N(T-K-1)} \\
&= \frac{\sum_{t=1}^{T} p_t\left(\sum_{i=1}^{N}\sigma_i^2\imath_{t,T} - \sum_{i=1}^{N}\epsilon_i\epsilon_{it}\right)}{N(T-K-1)}, \qquad \text{(A.5)}
\end{aligned}$$

where $\imath_{t,T}$ is a T-vector with one in the t-th position and zeros elsewhere, d_t is the t-th row of $D = [1_T,\ F]$, and $p_t = d_t\left(D'D\right)^{-1}D'$. Since Eq. (A.5) has a zero mean, we only need to consider its variance to determine the rate of convergence. We have

$$
\begin{aligned}
& \operatorname{Var}\left(\frac{\sum_{t=1}^{T} p_t\left(\sum_{i=1}^{N} \sigma_i^2 \imath_{t,T}-\sum_{i=1}^{N} \epsilon_i \epsilon_{it}\right)}{N(T-K-1)}\right) \\
= & \frac{1}{N^2(T-K-1)^2} E\left[\sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t\left(\sigma_i^2 \imath_{t,T}-\epsilon_i \epsilon_{it}\right)\left(\sigma_j^2 \imath_{s,T}-\epsilon_j \epsilon_{js}\right)' p_s'\right] \\
= & \frac{1}{N^2(T-K-1)^2} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} p_t E\left[\left(\sigma_i^2 \imath_{t,T}-\epsilon_i \epsilon_{it}\right)\left(\sigma_j^2 \imath_{s,T}-\epsilon_j \epsilon_{js}\right)'\right] p_s'. \qquad \text{(A.6)}
\end{aligned}
$$

Moreover, we have

$$
E\left[\left(\sigma_i^2 \imath_{t,T}-\epsilon_i \epsilon_{it}\right)\left(\sigma_j^2 \imath_{s,T}-\epsilon_j \epsilon_{js}\right)'\right]=E\left[\sigma_i^2 \sigma_j^2 \imath_{t,T} \imath_{s,T}'+\epsilon_i \epsilon_j' \epsilon_{it} \epsilon_{js}-\sigma_i^2 \imath_{t,T} \epsilon_j' \epsilon_{js}-\sigma_j^2 \epsilon_{it} \epsilon_i \imath_{s,T}'\right]
$$

$$
=\begin{cases}
\mu_{4i} \imath_{t,T} \imath_{t,T}'+\sigma_i^4\left(I_T-2 \imath_{t,T} \imath_{t,T}'\right) & \text{if} \quad i=j,\ t=s \\
\kappa_{4,iijj} \imath_{t,T} \imath_{t,T}'+\sigma_{ij}^2\left(I_T+\imath_{t,T} \imath_{t,T}'\right) & \text{if} \quad i \neq j,\ t=s \\
\sigma_i^4 \imath_{s,T} \imath_{t,T}' & \text{if} \quad i=j,\ t \neq s \\
\sigma_{ij}^2 \imath_{s,T} \imath_{t,T}' & \text{if} \quad i \neq j,\ t \neq s.
\end{cases} \qquad \text{(A.7)}
$$

It follows that

$$
\begin{aligned}
& \operatorname{Var}\left(\frac{\sum_{t=1}^{T} p_t\left(\sum_{i=1}^{N} \sigma_i^2 \imath_{t,T}-\sum_{i=1}^{N} \epsilon_i \epsilon_{it}\right)}{N(T-K-1)}\right) \\
= & \frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} p_t\left(\mu_{4i} \imath_{t,T} \imath_{t,T}'+\sigma_i^4\left(I_T-2 \imath_{t,T} \imath_{t,T}\right)\right) p_t' \\
& +\frac{1}{N^2(T-K-1)^2} \sum_{t=1}^{T} \sum_{i \neq j} p_t\left(\kappa_{4,iijj} \imath_{t,T} \imath_{t,T}'+\sigma_{ij}^2\left(I_T+\imath_{t,T} \imath_{t,T}'\right)\right) p_t' \\
& +\frac{1}{N^2(T-K-1)^2} \sum_{i=1}^{N} \sigma_i^4 \sum_{t \neq s} p_t \imath_{s,T} \imath_{t,T}' p_s' \\
& +\frac{1}{N^2(T-K-1)^2} \sum_{i \neq j} \sigma_{ij}^2 \sum_{t \neq s} p_t \imath_{s,T} \imath_{t,T}' p_s' \\
= & O\left(\frac{1}{N}\right) \qquad \text{(A.8)}
\end{aligned}
$$

by Assumptions 5(ii), 5(iii), 5(iv), and 5(viii), which implies that the first term on the right-hand side of Eq. (A.3) is $O_p\left(\frac{1}{\sqrt{N}}\right)$. Putting the pieces together concludes the proof. ■

Lemma 2 *Let*

$$\Lambda = \left[\begin{array}{cc} 0 & 0_K' \\ 0_K & \sigma^2(\tilde{F}'\tilde{F})^{-1} \end{array} \right]. \tag{A.9}$$

(i) Under Assumptions 2-5,

$$\hat{X}'\hat{X} = O_p(N). \tag{A.10}$$

In addition, under Assumption 6,

(ii)

$$\hat{\Sigma}_X \rightarrow_p \Sigma_X + \Lambda, \tag{A.11}$$

and

(iii)

$$\frac{(\hat{X}-X)'(\hat{X}-X)}{N} \rightarrow_p \Lambda. \tag{A.12}$$

Proof.

(i) Consider

$$\hat{X}'\hat{X} = \left[\begin{array}{cc} N & 1_N'\hat{B} \\ \hat{B}'1_N & \hat{B}'\hat{B} \end{array} \right]. \tag{A.13}$$

Then,

$$\hat{B}'1_N = \sum_{i=1}^{N} \hat{\beta}_i = \sum_{i=1}^{N} \beta_i + \mathcal{P}' \sum_{i=1}^{N} \epsilon_i. \tag{A.14}$$

Under Assumptions 4-5,

$$\begin{aligned} \mathrm{Var}\left(\sum_{t=1}^{T}\sum_{i=1}^{N} \epsilon_{it}(f_t - \bar{f})\right) &= \sum_{t,s=1}^{T}\sum_{i,j=1}^{N} (f_t-\bar{f})(f_s-\bar{f})'E[\epsilon_{it}\epsilon_{js}] \\ &\leq \sum_{t=1}^{T}\sum_{i,j=1}^{N} (f_t-\bar{f})(f_t-\bar{f})'|\sigma_{ij}| \\ &= O\left(N\sigma^2 \sum_{t=1}^{T}(f_t-\bar{f})(f_t-\bar{f})'\right) = O\left(NT\right). \end{aligned} \tag{A.15}$$

Using Assumption 2, we have

$$\hat{B}'1_N = O_p\left(N + \left(\tfrac{N}{T}\right)^{\frac{1}{2}}\right) = O_p(N). \tag{A.16}$$

Next, consider

$$\begin{aligned}\hat{B}'\hat{B} &= \sum_{i=1}^{N} \hat{\beta}_i \hat{\beta}_i' \\ &= \sum_{i=1}^{N} \left(\beta_i + \mathcal{P}'\epsilon_i\right)\left(\beta_i{}' + \epsilon_i{}'\mathcal{P}\right) \\ &= \sum_{i=1}^{N} \beta_i \beta_i{}' + \mathcal{P}'\left(\sum_{i=1}^{N} \epsilon_i \epsilon_i{}'\right)\mathcal{P} \\ &\quad +\mathcal{P}'\left(\sum_{i=1}^{N} \epsilon_i \beta_i{}'\right) + \left(\sum_{i=1}^{N} \beta_i \epsilon_i{}'\right)\mathcal{P}. \end{aligned} \tag{A.17}$$

By Assumption 2,

$$\sum_{i=1}^{N} \beta_i \beta_i{}' = O(N). \tag{A.18}$$

Using similar arguments as for Eq. (A.15),

$$\mathcal{P}'\left(\sum_{i=1}^{N} \epsilon_i \beta_i{}'\right) = O_p\left(\left(\frac{N}{T}\right)^{\frac{1}{2}}\right) \tag{A.19}$$

and

$$\left(\sum_{i=1}^{N} \beta_i \epsilon_i{}'\right)\mathcal{P} = O_p\left(\left(\frac{N}{T}\right)^{\frac{1}{2}}\right). \tag{A.20}$$

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 4-5,

$$
\begin{aligned}
& \left\| E\left[\tilde{F}'\left(\sum_{i=1}^{N}\epsilon_i{\epsilon_i}'\right)\tilde{F}\right]\right\| \\
= & \left\| E\left[\sum_{t,s=1}^{T}\sum_{i=1}^{N}(f_t-\bar{f})(f_s-\bar{f})'\epsilon_{it}\epsilon_{is}\right]\right\| \\
\leq & \sum_{t,s=1}^{T}\sum_{i=1}^{N}\|(f_t-\bar{f})(f_s-\bar{f})'\||E\left[\epsilon_{it}\epsilon_{is}\right]| \\
= & \sum_{t=1}^{T}\sum_{i=1}^{N}\left\|\left(f_t-\bar{f}\right)\left(f_t-\bar{f}\right)'\right\|\sigma_i^2 \\
= & O\left(N\sigma^2\sum_{t=1}^{T}\|(f_t-\bar{f})(f_t-\bar{f})'\|\right) = O(NT). \qquad \text{(A.21)}
\end{aligned}
$$

Then, we have

$$
\mathcal{P}'\left(\sum_{i=1}^{N}\epsilon_i{\epsilon_i}'\right)\mathcal{P} = O_p\left(\frac{N}{T}\right) \qquad \text{(A.22)}
$$

and

$$
\hat{B}'\hat{B} = O_p\left(N + \left(\frac{N}{T}\right)^{\frac{1}{2}} + \frac{N}{T}\right) = O_p(N). \qquad \text{(A.23)}
$$

This concludes the proof of part (i).

(ii) Using part (i) and under Assumptions 3-6, we have

$$
N^{-1}\hat{B}'1_N = \frac{1}{N}\sum_{i=1}^{N}\beta_i + O_p\left(\frac{1}{\sqrt{N}}\right) \qquad \text{(A.24)}
$$

and

$$
\begin{aligned}
N^{-1}\hat{B}'\hat{B} &= \frac{1}{N}\sum_{i=1}^{N}\beta_i\beta_i' + \mathcal{P}'\left(\frac{1}{N}\sum_{i=1}^{N}\epsilon_i\epsilon_i'\right)\mathcal{P} + \mathcal{P}'\left(\frac{1}{N}\sum_{i=1}^{N}\epsilon_i\beta_i'\right) + \left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\epsilon_i'\right)\mathcal{P} \\
&= \frac{1}{N}\sum_{i=1}^{N}\beta_i\beta_i' + \mathcal{P}'\left(\frac{1}{N}\sum_{i=1}^{N}\epsilon_i\epsilon_i' - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 I_T + \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 I_T - \sigma^2 I_T + \sigma^2 I_T\right)\mathcal{P} \\
&\quad + \mathcal{P}'\left(\frac{1}{N}\sum_{i=1}^{N}\epsilon_i\beta_i'\right) + \left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\epsilon_i'\right)\mathcal{P} \\
&= \frac{1}{N}\sum_{i=1}^{N}\beta_i\beta_i' + \mathcal{P}'\left(\frac{1}{N}\sum_{i=1}^{N}\left(\epsilon_i\epsilon_i' - \sigma_i^2 I_T\right)\right)\mathcal{P} + \frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2 - \sigma^2\right)\mathcal{P}'\mathcal{P} + \sigma^2\mathcal{P}'\mathcal{P} \\
&\quad + \mathcal{P}'\left(\frac{1}{N}\sum_{i=1}^{N}\epsilon_i\beta_i'\right) + \left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\epsilon_i'\right)\mathcal{P} \\
&= \frac{1}{N}\sum_{i=1}^{N}\beta_i\beta_i' + \sigma^2\mathcal{P}'\mathcal{P} + O_p\left(\frac{1}{\sqrt{N}}\right) + o\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}
\tag{A.25}
$$

Assumption 2 concludes the proof of part (ii).

(iii) Note that

$$
\begin{aligned}
\frac{(\hat{X}-X)'(\hat{X}-X)}{N} &= \frac{1}{N}\begin{bmatrix} 0_N' \\ (\hat{B}-B)' \end{bmatrix}\left[0_N,\ (\hat{B}-B)\right] \\
&= \begin{bmatrix} 0 & 0_K' \\ 0_K & \mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P} \end{bmatrix}.
\end{aligned}
\tag{A.26}
$$

As in part (ii) we can write

$$
\frac{\epsilon\epsilon'}{N} = \frac{1}{N}\sum_{i=1}^{N}\left(\epsilon_i\epsilon_i' - \sigma_i^2 I_T\right) + \left(\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2 - \sigma^2\right)\right) I_T + \sigma^2 I_T. \tag{A.27}
$$

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

$$
\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P} = \sigma^2\mathcal{P}'\mathcal{P} + O_p\left(\frac{1}{\sqrt{N}}\right) + o\left(\frac{1}{\sqrt{N}}\right).\blacksquare \tag{A.28}
$$

Lemma 3

Under Assumptions 2-5,

$$
X'\bar{\epsilon} = O_p\left(\sqrt{N}\right). \tag{A.29}
$$

Proof. We have

$$X'\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^{T}\left[\begin{array}{c} 1_N' \\ B' \end{array}\right]\epsilon_t \tag{A.30}$$

and

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}1_N'\epsilon_t\right) &= \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{i,j=1}^{N}E[\epsilon_{it}\epsilon_{js}] \\
&\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i,j=1}^{N}|\sigma_{ij}| \\
&= O\left(\frac{NT}{T^2}\sigma^2\right) = O\left(N\right).
\end{aligned} \tag{A.31}$$

Moreover, using Assumptions 2 and 5(ii),

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}B'\epsilon_t\right) &= \frac{1}{T^2}\sum_{t,s=1}^{T}\sum_{i,j=1}^{N}E[\epsilon_{it}\epsilon_{js}]\beta_i\beta_j' \\
&\leq \frac{1}{T^2}\sum_{t=1}^{T}\sum_{i,j=1}^{N}|\beta_i\beta_j'||\sigma_{ij}| \\
&= O\left(\frac{NT}{T^2}\sigma^2\right) = O\left(N\right).
\end{aligned} \tag{A.32}$$

Putting the pieces together, $X'\bar{\epsilon} = O_p\left(\sqrt{N}\right)$. ■

Lemma 4

Under Assumptions 3-5,

$$(\hat{X}-X)'X\Gamma^P = O_p\left(\sqrt{N}\right). \tag{A.33}$$

Proof. We have

$$(\hat{X}-X)'X\Gamma^P = \left[\begin{array}{c} 0_N' \\ \mathcal{P}'\epsilon \end{array}\right]X\Gamma^P. \tag{A.34}$$

Using similar arguments to Eq. (A.15) concludes the proof. ■

Lemma 5

Under Assumptions 3-5,

$$(\hat{X}-X)'\bar{\epsilon} = O_p\left(\sqrt{N}\right). \tag{A.35}$$

Proof.

$$
\begin{aligned}
(\hat{X}-X)'\bar{\epsilon} &= \begin{bmatrix} 0 \\ \mathcal{P}'\epsilon\bar{\epsilon} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathcal{P}'\epsilon\epsilon'\frac{1_T}{T} \end{bmatrix} \\
&= \begin{bmatrix} 0 \\ \mathcal{P}'\left[\left(\epsilon\epsilon' - \sum_{i=1}^N \sigma_i^2 I_T\right) + \left(\sum_{i=1}^N \sigma_i^2 - N\sigma^2\right) I_T\right]\frac{1_T}{T} \end{bmatrix} = O_p(\sqrt{N})
\end{aligned} \tag{A.36}
$$

by Assumption 5. ■

Lemma 6 *Under Assumption 5 and the identification assumption $\kappa_4 = 0$, we have*

$$
\hat{\sigma}_4 \to_p \sigma_4. \tag{A.37}
$$

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

(i) By Assumptions 5(iv), 5(vi), and 5(vii), we have

$$
\begin{aligned}
E\left[\frac{1}{N}\sum_{t=1}^T\sum_{i=1}^N \hat{\epsilon}_{it}^4\right] &= \frac{1}{N}\sum_{t=1}^T\sum_{i=1}^N E\left[\hat{\epsilon}_{it}^4\right] \\
&= \frac{1}{N}\sum_{t=1}^T\sum_{i=1}^N \sum_{s_1,s_2,s_3,s_4=1}^T m_{ts_1}m_{ts_2}m_{ts_3}m_{ts_4}E\left[\epsilon_{is_1}\epsilon_{is_2}\epsilon_{is_3}\epsilon_{is_4}\right] \\
&= \frac{1}{N}\sum_{t=1}^T\sum_{i=1}^N \kappa_{4,iiii}\sum_{s=1}^T m_{ts}^4 + 3\frac{1}{N}\sum_{t=1}^T\sum_{i=1}^N \sigma_i^4\left(\sum_{s=1}^T m_{ts}^2\right)^2 \\
&\to \kappa_4\sum_{t=1}^T\sum_{s=1}^T m_{ts}^4 + 3\sigma_4\sum_{t=1}^T\left(\sum_{s=1}^T m_{ts}^2\right)^2,
\end{aligned} \tag{A.38}
$$

where $\hat{\epsilon}_{it} = \imath_{t,T}' M\epsilon_i$ and $M = [m_{ts}]$ for $t, s = 1, \ldots, T$. Note that

$$
\begin{aligned}
\sum_{s=1}^T m_{ts}^2 &= ||m_t||^2 \\
&= \imath_t' M \imath_t \\
&= \imath_t'\left(I_T - D(D'D)^{-1}D'\right)\imath_t \\
&= 1 - \mathrm{tr}\left(D(D'D)^{-1}D'\imath_t\imath_t'\right) \\
&= 1 - \mathrm{tr}\left(P\imath_t\imath_t'\right) \\
&= 1 - p_{tt} \\
&= m_{tt},
\end{aligned} \tag{A.39}
$$

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

$$\sum_{t=1}^{T}\left(\sum_{s=1}^{T} m_{ts}^2\right)^2 = \sum_{t=1}^{T} m_{tt}^2 = \mathrm{tr}\left(M^{(2)}\right). \tag{A.40}$$

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

$$E\left[\hat{\sigma}_4\right] \quad \to \quad \sigma_4. \tag{A.41}$$

This concludes the proof of part (i).

(ii) As for the variance of $\hat{\sigma}_4$, we have

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\epsilon}_{it}^4\right) &= \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathrm{Cov}\left(\hat{\epsilon}_{it}^4, \hat{\epsilon}_{js}^4\right) \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sum_{\substack{u_1,u_2,\\ u_3,u_4=1}}^{T}\sum_{\substack{v_1,v_2,\\ v_3,v_4=1}}^{T} m_{tu_1}m_{tu_2}m_{tu_3}m_{tu_4}m_{sv_1}m_{sv_2}m_{sv_3}m_{sv_4} \\
&\quad \times \mathrm{Cov}\left(\epsilon_{iu_1}\epsilon_{iu_2}\epsilon_{iu_3}\epsilon_{iu_4}, \epsilon_{jv_1}\epsilon_{jv_2}\epsilon_{jv_3}\epsilon_{jv_4}\right) \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\sum_{\substack{u_1,u_2,\\ u_3,u_4=1}}^{T}\sum_{\substack{v_1,v_2,\\ v_3,v_4=1}}^{T} m_{tu_1}m_{tu_2}m_{tu_3}m_{tu_4}m_{sv_1}m_{sv_2}m_{sv_3}m_{sv_4} \\
&\quad \times \Bigg(\kappa_8\left(\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4}\right) \\
&\quad + \sum^{(6,2)}\kappa_6\left(\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_1}, \epsilon_{jv_2}\right)\mathrm{Cov}\left(\epsilon_{jv_3}, \epsilon_{jv_4}\right) \\
&\quad + \sum^{(4,4)}\kappa_4\left(\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{jv_1}, \epsilon_{jv_2}\right)\kappa_4\left(\epsilon_{iu_3}, \epsilon_{iu_4}, \epsilon_{jv_3}, \epsilon_{jv_4}\right) \\
&\quad + \sum^{(4,2,2)}\kappa_4\left(\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{jv_1}, \epsilon_{jv_2}\right)\mathrm{Cov}\left(\epsilon_{iu_3}, \epsilon_{iu_4}\right)\mathrm{Cov}\left(\epsilon_{jv_3}, \epsilon_{jv_4}\right) \\
&\quad + \sum^{(2,2,2,2)}\mathrm{Cov}\left(\epsilon_{iu_1}, \epsilon_{iu_2}\right)\mathrm{Cov}\left(\epsilon_{iu_3}, \epsilon_{jv_1}\right)\mathrm{Cov}\left(\epsilon_{iu_4}, \epsilon_{jv_2}\right)\mathrm{Cov}\left(\epsilon_{jv_3}, \epsilon_{jv_4}\right)\Bigg),
\end{aligned} \tag{A.42}$$

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 $E[\epsilon_{it}] = E\left[\epsilon_{it}^3\right] = 0$, we do not need to consider further partitions in the relation above.[53] Then, under Assumptions 5(i), 5(ii), 5(v), and 5(viii), it follows that

$$\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\epsilon}_{it}^4\right) = O\left(\frac{1}{N}\right) \tag{A.43}$$

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

Lemma 7 *Let $w = [w_1, \ldots, w_T]'$ and $s = [s_1, \ldots, s_T]'$ be two arbitrary T-vectors. Then, under Eq. (65) and Assumptions 2-7,*

$$\frac{1}{N(T-K)}\sum_{i=1}^{N}\hat{\epsilon}_i'\hat{\epsilon}_i\sum_{k=1}^{T}w_k\epsilon_{ki}\sum_{r=1}^{T}s_r\epsilon_{ri} \to_p \frac{\mathrm{tr}\,(M\,(S_1+S_2))}{(T-K)}, \tag{A.44}$$

where $S_1 = diag\left[(s_1w_1\mu_4 + \sigma^4\sum_{k\neq 1}^{T}w_ks_k), \ldots, (s_Tw_T\mu_4 + \sigma^4\sum_{k\neq T}^{T}w_ks_k)\right]$ and
$S_2 = \sigma^4\left(ws' + sw' - 2diag(w_1s_1, \ldots, w_Ts_T)\right)$.

Proof. Note that

$$\begin{aligned}
&\frac{1}{N(T-K)}\sum_{i=1}^{N}\hat{\epsilon}_i'\hat{\epsilon}_i\sum_{k=1}^{T}\epsilon_{ki}\sum_{r=1}^{T}s_r\epsilon_{ri} = \\
&= \frac{1}{N(T-K)}\mathrm{tr}\left(M\left(\sum_{i=1}^{N}\epsilon_i\epsilon_i'\left(\sum_{k=r=1}^{T}w_ks_k\epsilon_{ki}^2 + \sum_{r>k}^{T}w_ks_r\epsilon_{ik}\epsilon_{ir} + \sum_{r<k}^{T}w_ks_r\epsilon_{ik}\epsilon_{ir}\right)\right)\right).
\end{aligned} \tag{A.45}$$

For the first term of Eq. (A.45),

$$\begin{aligned}
\frac{1}{N(T-K)}\mathrm{tr}\left(M\left(\sum_{i=1}^{N}\epsilon_i\epsilon_i'\sum_{k=r=1}^{T}w_ks_k\epsilon_{ki}^2\right)\right) &= \frac{1}{(T-K)}\mathrm{tr}\left(M\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{k=r=1}^{T}\epsilon_i\epsilon_i'w_ks_k\epsilon_{ki}^2\right)\right) \\
&\to_p \frac{1}{(T-K)}\mathrm{tr}\,(MS_1)\,,
\end{aligned} \tag{A.46}$$

where

$$S_1 = \mathrm{plim}\frac{1}{N}\sum_{i=1}^{N}\sum_{k=r=1}^{T}\epsilon_i\epsilon_i'w_ks_k\epsilon_{ki}^2 = \mathrm{diag}\left[(s_1w_1\mu_4 + \sigma^4\sum_{k\neq 1}^{T}w_ks_k), \ldots, (s_Tw_T\mu_4 + \sigma^4\sum_{k\neq T}^{T}w_ks_k)\right]. \tag{A.47}$$

[53] According to the theory on cumulants (Brillinger (2001)), evaluation of $\mathrm{Cov}\left(\epsilon_{iu_1}\epsilon_{iu_2}\epsilon_{iu_3}\epsilon_{iu_4}, \epsilon_{jv_1}\epsilon_{jv_2}\epsilon_{jv_3}\epsilon_{jv_4}\right)$ requires considering the indecomposable partitions of the two sets, $\{\epsilon_{iu_1}, \epsilon_{iu_2}, \epsilon_{iu_3}, \epsilon_{iu_4}\}$ and $\{\epsilon_{jv_1}, \epsilon_{jv_2}, \epsilon_{jv_3}, \epsilon_{jv_4}\}$, meaning that there must be at least one subset that includes an element of both sets.

For the second and third terms of Eq. (A.45), we obtain

$$\begin{aligned}
& \frac{1}{N(T-K)}\mathrm{tr}\left(M\left(\sum_{i=1}^{N}\epsilon_i\epsilon_i'\left(\sum_{r>k}^{T}w_k s_r\epsilon_{ik}\epsilon_{ir}+\sum_{r<k}^{T}w_k s_r\epsilon_{ik}\epsilon_{ir}\right)\right)\right) \\
= & \frac{1}{(T-K)}\mathrm{tr}\left(M\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{r>k}^{T}\epsilon_i\epsilon_i' w_k s_r\epsilon_{ik}\epsilon_{ir}+\frac{1}{N}\sum_{i=1}^{N}\sum_{r<k}^{T}\epsilon_i\epsilon_i' w_k s_r\epsilon_{ik}\epsilon_{ir}\right)\right) \\
\to_p & \frac{1}{(T-K)}\mathrm{tr}(MS_2),
\end{aligned} \tag{A.48}$$

where

$$\begin{aligned}
S_2 &= \mathrm{plim}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{r>k}^{T}\epsilon_i\epsilon_i' w_k s_r\epsilon_{ik}\epsilon_{ir}+\frac{1}{N}\sum_{i=1}^{N}\sum_{r<k}^{T}\epsilon_i\epsilon_i' w_k s_r\epsilon_{ik}\epsilon_{ir}\right) \\
&= \sigma^4\left(ws'+sw'-2\mathrm{diag}(w_1 s_1,\ldots,w_T s_T)\right).\blacksquare
\end{aligned} \tag{A.49}$$

Lemma 8 *Let* $\hat{\tau}_\Phi = \frac{1}{N(T-K)}\sum_{i=1}^{N}\hat{\epsilon}_i'\hat{\epsilon}_i\hat{e}_i^P$*. Then, under Eq. (65) and Assumptions 2-7,*

$$\hat{\tau}_\Phi \to_p \tau_\Phi. \tag{A.50}$$

Proof. Given

$$\begin{aligned}
\hat{e}_i^P &= \bar{R}_i - \hat{X}_i'\hat{\Gamma}^* \\
&= X_i'\tilde{\Gamma}^P + e_i + \bar{\epsilon}_i - \hat{X}_i'\hat{\Gamma}^* \\
&= e_i + \bar{\epsilon}_i - (\hat{X}_i - X_i)'\tilde{\Gamma}^P - \hat{X}_i'(\hat{\Gamma}^* - \tilde{\Gamma}^P),
\end{aligned} \tag{A.51}$$

using the fact that $\hat{\epsilon}_i = M\epsilon_i$ and Eq. (A.51), we can write

$$\begin{aligned}
\hat{\tau}_\Phi &= \frac{1}{N(T-K)}\sum_{i=1}^{N}\hat{\epsilon}_i'\hat{\epsilon}_i\hat{e}_i^P = \frac{1}{N(T-K)}\sum_{i=1}^{N}\epsilon_i' MM\epsilon_i\hat{e}_i \\
&= \frac{1}{N(T-K)}\sum_{i=1}^{N}\mathrm{tr}(M\epsilon_i'\epsilon_i)(e_i+\bar{\epsilon}_i-(\hat{X}_i-X_i)'\tilde{\Gamma}^P-\hat{X}_i'(\hat{\Gamma}^*-\tilde{\Gamma}^P)) \\
&= \frac{1}{N(T-K)}\sum_{i=1}^{N}\mathrm{tr}(M\epsilon_i'\epsilon_i e_i)+o_p(1)\to_p\frac{1}{(T-K)}\mathrm{tr}\left(M\tau_\Phi\right)=\tau_\Phi.\blacksquare
\end{aligned} \tag{A.52}$$

Lemma 9 *Let*

$$\hat{\tau}_\Omega = \frac{1}{N(T-K)}\sum_{i=1}^{N}\hat{\epsilon}_i'\hat{\epsilon}_i(\hat{e}_i^P)^2 - \frac{\sigma^4}{T}\left(1+\frac{2\mathrm{tr}(M1_T1_T')}{T(T-K)}\right)-\frac{\mathrm{tr}(MS_F)}{(T-K)}+2\frac{\mathrm{tr}(MC_F)}{T(T-K)}, \tag{A.53}$$

where

$$S_F = \sigma^4 \begin{bmatrix} A'\left(3\tilde{f}_1\tilde{f}_1' + \sum_{t\neq 1}^T \tilde{f}_t\tilde{f}_t'\right)A & 2A'\tilde{f}_1\tilde{f}_2'A & \cdots & 2A'\tilde{f}_1\tilde{f}_T'A \\ 2A'\tilde{f}_2\tilde{f}_1'A & A'\left(3\tilde{f}_2\tilde{f}_2' + \sum_{t\neq 2}^T \tilde{f}_t\tilde{f}_t'\right) & \cdots & 2A'\tilde{f}_2\tilde{f}_T'A \\ \vdots & \vdots & \ddots & \vdots \\ 2A'\tilde{f}_T\tilde{f}_1'A & 2A'\tilde{f}_T\tilde{f}_2'A & \cdots & A'\left(3\tilde{f}_1\tilde{f}_1' + \sum_{t\neq T}^T \tilde{f}_T\tilde{f}_T'\right)A \end{bmatrix} \tag{A.54}$$

and

$$C_F = \sigma^4 \begin{bmatrix} 3\tilde{f}_1'A + \sum_{t\neq 1}^T \tilde{f}_t'A & (\tilde{f}_1+\tilde{f}_2)'A & \cdots & (\tilde{f}_1+\tilde{f}_T)'A \\ (\tilde{f}_2+\tilde{f}_1)'A & 3\tilde{f}_2'A + \sum_{t\neq 2}^T \tilde{f}_t'A & \cdots & (\tilde{f}_2+\tilde{f}_T)'A \\ \vdots & \vdots & \ddots & \vdots \\ (\tilde{f}_T+\tilde{f}_1)'A & (\tilde{f}_T+\tilde{f}_2)'A & \cdots & 3\tilde{f}_T'A + \sum_{t\neq T}^T \tilde{f}_t'A \end{bmatrix}, \tag{A.55}$$

with $A = (\tilde{F}'\tilde{F})^{-1}\tilde{\gamma}_1^P$. *Then, under Eq. (65) and Assumptions 2-7,*

$$\hat{\tau}_\Omega \to_p \tau_\Omega. \tag{A.56}$$

Proof: By Eq. (A.51), we have

$$\begin{aligned} (\hat{e}_i^P)^2 &= e_i^2 + \bar{\epsilon}_i^2 + \left((\hat{\beta}_i - \beta_i)'\tilde{\gamma}_1^P\right)^2 + \left([1, \hat{\beta}_i'](\hat{\Gamma}^* - \tilde{\Gamma}^P)\right)^2 \\ &+ 2e_i\left(\bar{\epsilon}_i - (\hat{\beta}_i - \beta_i)'\tilde{\gamma}_1^P - [1, \hat{\beta}_i'](\hat{\Gamma}^* - \tilde{\Gamma}^P)\right) \\ &+ 2\bar{\epsilon}_i\left(-(\hat{\beta}_i - \beta_i)'\tilde{\gamma}_1^P - [1, \hat{\beta}_i'](\hat{\Gamma}^* - \tilde{\Gamma}^P)\right) \\ &+ 2(\hat{\beta}_i - \beta_i)'\tilde{\gamma}_1^P[1, \hat{\beta}_i'](\hat{\Gamma}^* - \tilde{\Gamma}^P). \end{aligned} \tag{A.57}$$

Then,

$$
\begin{aligned}
\hat{\tau}_\Omega &= \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i (\hat{e}_i^P)^2 \\
&= \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i e_i^2 + \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \bar{\epsilon}_i^2 + \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \left((\hat{\beta}_i - \beta_i)' \tilde{\gamma}_1^P \right)^2 \\
&- 2 \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \bar{\epsilon}_i (\hat{\beta}_i - \beta_i)' \tilde{\gamma}_1^P + o_p(1), \qquad \text{(A.58)}
\end{aligned}
$$

where all terms involving $(\hat{\Gamma}^* - \tilde{\Gamma}^P)$ are condensed into the $o_p(1)$ term. By Assumption 7, the first term in Eq. (A.58) satisfies

$$
\frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i e_i^2 \quad = \quad \frac{1}{(T-K)} \mathrm{tr} \left(M \frac{1}{N} \sum_{i=1}^{N} \epsilon_i \epsilon_i' . e_i^2 \right) \to_p \frac{1}{(T-K)} \mathrm{tr}(M \tau_\Omega) = \tau_\Omega. \quad \text{(A.59)}
$$

For the second term in Eq. (A.58), we have

$$
\frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \bar{\epsilon}_i^2 \quad = \quad \frac{1}{T^2} \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \sum_{t=1}^{T} \epsilon_{it} \sum_{s=1}^{T} \epsilon_{is}. \qquad \text{(A.60)}
$$

Then, applying Lemma 7 with $w = s = [1, \ldots, 1]'$, we have

$$
\frac{1}{T^2} \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \sum_{t=1}^{T} \epsilon_{it} \sum_{s=1}^{T} \epsilon_{is} \to_p \frac{\sigma^4}{T} \left(1 + \frac{2\mathrm{tr}(M 1_T 1_T')}{T(T-K)} \right). \qquad \text{(A.61)}
$$

For the third term in Eq. (A.58), we have

$$
\frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \left((\hat{\beta}_i - \beta_i)' \tilde{\gamma}_1^P \right)^2 \quad = \quad \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \sum_{t=1}^{T} \tilde{\gamma}_1^{P\prime} (\tilde{F}' \tilde{F})^{-1} f_t \epsilon_{it} \sum_{s=1}^{T} \tilde{\gamma}_1^{P\prime} (\tilde{F}' \tilde{F})^{-1} f_s \epsilon_{is}, \qquad \text{(A.62)}
$$

and by Lemma 7 with $w = s = \left[\tilde{\gamma}_1^{P\prime} (\tilde{F}' \tilde{F})^{-1} f_1, \ldots, \tilde{\gamma}_1^{P\prime} (\tilde{F}' \tilde{F})^{-1} f_T \right]'$, one obtains

$$
\frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \left((\hat{\beta}_i - \beta_i)' \tilde{\gamma}_1^P \right)^2 \to_p \frac{\mathrm{tr}(M S_F)}{(T-K)}. \qquad \text{(A.63)}
$$

Finally, for the fourth term in Eq. (A.58), rewriting it as

$$
-2 \frac{1}{N(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \bar{\epsilon}_i (\hat{\beta}_i - \beta_i)' \tilde{\gamma}_1^P \quad = \quad -2 \frac{1}{NT(T-K)} \sum_{i=1}^{N} \hat{\epsilon}_i' \hat{\epsilon}_i \sum_{t=1}^{T} \epsilon_{it} \sum_{s=1}^{T} \epsilon_{is} \tilde{f}_s' (\tilde{F}' \tilde{F})^{-1} \tilde{\gamma}_1^P, \qquad \text{(A.64)}
$$

and applying again Lemma 7 with $w = [1, \ldots, 1]'$ and $s = [A'\tilde{f}_1, \ldots, A'\tilde{f}_T]'$, we obtain

$$-2\frac{1}{N(T-K)}\sum_{i=1}^{N}\hat{\epsilon}_i'\hat{\epsilon}_i\bar{\epsilon}_i(\hat{\beta}_i - \beta_i)'\tilde{\gamma}_1^P \quad \to_p \quad -2\frac{\mathrm{tr}(MC_F)}{T(T-K)}.\blacksquare \tag{A.65}$$

Appendix B: Proofs of Propositions and Theorems

Proof of Proposition 1. Consider the class of additive bias-adjusted estimators $\hat{\Gamma}^{bias-adj}$ for Γ^P:

$$\hat{\Gamma}^{bias-adj} = \hat{\Gamma} + (\frac{\hat{X}'\hat{X}}{N})^{-1}\hat{\Lambda}\hat{\Gamma}^{prelim} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R} + (\frac{\hat{X}'\hat{X}}{N})^{-1}\hat{\Lambda}\hat{\Gamma}^{prelim}, \tag{B.1}$$

where $\hat{\Gamma}^{prelim}$ denotes any preliminary $\sqrt{N}$-consistent estimator of Γ^P. Setting $\hat{\Gamma}^{bias-adj} = \hat{\Gamma}^{prelim}$ and rearranging terms, we obtain

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

which implies that

$$\hat{\Gamma}^{bias-adj} = \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}\frac{\hat{X}'\bar{R}}{N} = \hat{\Gamma}^*.\blacksquare \tag{B.3}$$

Proof of Proposition 2. By means of simple calculations, $\Sigma = \lambda\lambda' + \sigma_\eta^2 I_N$. Thus, $\sum_{i=1}^N \sigma_i^2/N = \sum_{i=1}^N(\lambda_i^2 + \sigma_\eta^2)/N \to \sigma_\eta^2$ because $\sum_{i=1}^N \lambda_i^2 \leq (\sum_{i=1}^N |\lambda_i|)^2 = O(N^{2\delta}) = o(N)$. Therefore, setting $\sigma^2 = \sigma_\eta^2$, one obtains $\sum_{i=1}^N(\sigma_i^2 - \sigma^2)/N = \sum_{i=1}^N \lambda_i^2/N = (\lambda_1^2 + \cdots + \lambda_q^2)/N + \sum_{i=q+1}^N \lambda_i^2/N = O(N^{\delta-1} + N^{2\delta-1}) = o(\sqrt{N})$ since $\delta < 1/2$. It follows that Assumption 5(i) is satisfied.

Next, given that $\sigma_{ij} = \lambda_i\lambda_j$ for $i \neq j$, we obtain $\sum_{i\neq j=1}^N |\sigma_{ij}| \leq (\sum_{i=1}^N |\lambda_i|)^2 = O(N^{2\delta}) = o(N)$, thus satisfying Assumption 5(ii).

The maximum eigenvalue of Σ is bounded from below by the maximum eigenvalue of $\lambda\lambda'$, which equals $\lambda'\lambda$ (all the other $N-1$ eigenvalues of $\lambda\lambda'$ are zero), where $\lambda_1^2 + \cdots + \lambda_q^2 \leq \lambda'\lambda = O(N^{2\delta})$. Therefore, the maximum eigenvalue diverges at least at rate $o(\sqrt{N})$. $\blacksquare$

Proof of Proposition 3. The Fama and MacBeth (1973) standard errors with the Shanken (1992) correction are given by

$$SE_k^{FM} = \left((1+\hat{c})(\hat{W}_k - \mathbb{1}_{\{k>0\}}\hat{\sigma}_k^2) + \mathbb{1}_{\{k>0\}}\hat{\sigma}_k^2/T\right)^{\frac{1}{2}} \text{ and } SE_k^{FM,P} = \left((1+\hat{c})(W_k - \mathbb{1}_{\{k>0\}}\hat{\sigma}_k^2)\right)^{\frac{1}{2}}, \tag{B.4}$$

for $k = 0, \ldots, K$, where $\hat{W}_k = \imath'_{k+1,K+1}\sum_{t=1}^{T}(\hat{\Gamma}_t - \bar{\hat{\Gamma}})(\hat{\Gamma}_t - \bar{\hat{\Gamma}})'\imath_{k+1,K+1}/(T-1)$, $\hat{\Gamma}_t = (\hat{X}'\hat{X})^{-1}\hat{X}'R_t$ with sample mean $\bar{\hat{\Gamma}}$, $\imath_{j,J}$ denotes the j-th column, for $j = 1, \ldots, J$, of the identity matrix I_J, $\hat{c} = \hat{\gamma}'_1\left(\tilde{F}'\tilde{F}/T\right)^{-1}\hat{\gamma}_1$, $\mathbb{1}_{\{\}}$ is the indicator function, and $\hat{\sigma}^2_k$ denotes the (k,k)-th element of $\tilde{F}'\tilde{F}/T$.

Consider the numerator of the t-ratios first. By Lemma 2(ii) and Lemmas 4 and 5, we obtain $\hat{\Gamma} = [\hat{\gamma}_0,\ \hat{\gamma}'_1]' = (\Sigma_X + \Lambda)^{-1}\Sigma_X\Gamma^P + O_p\left(\frac{1}{\sqrt{N}}\right)$. By the blockwise formula of the inverse of a matrix (Magnus and Neudecker (2007), Section 1-11),

$$
\begin{aligned}
(\Sigma_X + \Lambda)^{-1}\Sigma_X\Gamma^P &= \begin{bmatrix} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta + C \end{bmatrix}^{-1}\begin{bmatrix} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta \end{bmatrix}\Gamma^P \\
&= \begin{bmatrix} 1 + \mu'_\beta A^{-1}\mu_\beta & -\mu'_\beta A^{-1} \\ -A^{-1}\mu_\beta & A^{-1} \end{bmatrix}\begin{bmatrix} 1 & \mu'_\beta \\ \mu_\beta & \Sigma_\beta \end{bmatrix}\Gamma^P \\
&= \begin{bmatrix} 1 & \mu'_\beta - \mu'_\beta A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta) \\ 0 & A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta) \end{bmatrix}\Gamma^P. \qquad \text{(B.5)}
\end{aligned}
$$

Then,

$$
\begin{aligned}
(\Sigma_X + \Lambda)^{-1}\Sigma_X\Gamma^P - \Gamma &= \begin{bmatrix} 1 & \mu'_\beta - \mu'_\beta A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta) \\ 0 & A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta) \end{bmatrix}\Gamma^P - \Gamma \\
&= \begin{bmatrix} 0 & \mu'_\beta(I_K - A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta)) \\ 0 & -(I_K - A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta)) \end{bmatrix}\Gamma \\
&+ \begin{bmatrix} 1 & \mu'_\beta(I_K - A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta)) \\ 0 & A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta) \end{bmatrix}\begin{bmatrix} 0 \\ \bar{f} - E[f_t] \end{bmatrix}. \qquad \text{(B.6)}
\end{aligned}
$$

Hence, $\text{plim } \hat{\gamma}_0 - \gamma_0 = \mu'_\beta(I_K - A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta))\gamma_1^P = \mu'_\beta A^{-1}C\gamma_1^P$ and, for every $j = 1, \ldots, K$, $\text{plim } \hat{\gamma}_{1j} - \gamma_{1j} = -\imath'_{j,K}(I_K - A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta))\gamma_1 + \imath'_{j,K}A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta)(\bar{f} - E[f_t])$ and $\text{plim } \hat{\gamma}_{1j} - \gamma^P_{1j} = -\imath'_{j,K}(I_K - A^{-1}(\Sigma_\beta - \mu_\beta\mu'_\beta))\gamma_1^P$. Consider now the behavior of the denominator of the t-ratios. It is easy to see that $\hat{W} = \frac{1}{T-1}\sum_{t=1}^{T}(\Gamma_t - \bar{\Gamma})(\Gamma_t - \bar{\Gamma})' = \hat{W}_a + \hat{W}_b + \hat{W}_c$, where

$$
\hat{W}_a = (\hat{X}'\hat{X})^{-1}\hat{X}'\left[\frac{1}{T-1}\sum_{t=1}^{T}(\epsilon_t - \bar{\epsilon})(\epsilon_t - \bar{\epsilon})'\right]\hat{X}(\hat{X}'\hat{X})^{-1}, \qquad \text{(B.7)}
$$

$$
\hat{W}_b = (\hat{X}'\hat{X})^{-1}\hat{X}'B\left[(\frac{1}{T-1}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})'\right]B'\hat{X}(\hat{X}'\hat{X})^{-1} \text{ and} \qquad \text{(B.8)}
$$

$$
\begin{aligned}
\hat{W}_c &= (\hat{X}'\hat{X})^{-1}\hat{X}'\left[\frac{\sum_{t=1}^{T}(\epsilon_t - \bar{\epsilon})(f_t - \bar{f})'}{T-1}\right]B'\hat{X}(\hat{X}'\hat{X})^{-1} \\
&+(\hat{X}'\hat{X})^{-1}\hat{X}'B\left[\frac{\sum_{t=1}^{T}(f_t - \bar{f})(\epsilon_t - \bar{\epsilon})'}{T-1}\right]\hat{X}(\hat{X}'\hat{X})^{-1}. \qquad \text{(B.9)}
\end{aligned}
$$

Based on Lemmas 2-4 (details are available upon request), we obtain

$$\hat{W} \to_p W = W_a + W_b + W_c \equiv (\Sigma_X + \Lambda)^{-1} \begin{bmatrix} 0 & 0_K' \\ 0_K & \frac{\sigma^4}{(T-1)}(\tilde{F}'\tilde{F})^{-1} \end{bmatrix} (\Sigma_X + \Lambda)^{-1}$$

$$+ (\Sigma_X + \Lambda)^{-1} \begin{bmatrix} \mu_\beta' \\ \Sigma_\beta \end{bmatrix} \left[\frac{\tilde{F}'\tilde{F}}{T-1} \right] [\mu_\beta, \Sigma_\beta] (\Sigma_X + \Lambda)^{-1}$$

$$+ (\Sigma_X + \Lambda)^{-1} \frac{\sigma^2}{T-1} \begin{bmatrix} 0 & \mu_\beta' \\ \mu_\beta & 2\Sigma_\beta \end{bmatrix} (\Sigma_X + \Lambda)^{-1}. \quad \text{(B.10)}$$

It follows that

$$W = \begin{bmatrix} 0 & 0_K' \\ 0_K & \frac{(\tilde{F}'\tilde{F})}{T-1} \end{bmatrix}. \quad \text{(B.11)}$$

Therefore, since $\hat{W}_k = \imath_{k+1,K+1}' \hat{W} \imath_{k+1,K+1}$ for $k = 0, \ldots, K$, we have $(1+\hat{c})(\hat{W}_k - \mathbb{1}_{\{k>0\}}\hat{\sigma}_k^2) \to_p 0$ for any value of $\hat{c}$. It follows that $SE_k^{FM} \to_p \hat{\sigma}_k/\sqrt{T}$ and $SE_k^{FM,P} \to_p 0$. The proof of parts (i) and (ii) follows from dividing $\hat{\gamma}_0 - \gamma_0$, $\hat{\gamma}_{1k} - \gamma_{1k}$, and $\hat{\gamma}_{1k} - \gamma_{1k}^P$ by SE_k^{FM} and $SE_k^{FM,P}$, for the ex ante and ex post risk premia, respectively, and then taking the limit as $N \to \infty$. ■

Proof of Theorem 1. For part (i), starting from Eq. (12), we have

$$\begin{aligned}
\hat{\Gamma}^* &= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}'\bar{R}}{N} \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \frac{\hat{X}'}{N} \left[\hat{X}\Gamma^P + \bar{\epsilon} - (\hat{X} - X)\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'\hat{X}}{N}\Gamma^P + \frac{\hat{X}'}{N}\bar{\epsilon} - \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left(\frac{\hat{X}'\hat{X}}{N}\right) \left[\Gamma^P + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}\bar{\epsilon} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P\right] \\
&= \left[I_{K+1} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \hat{\Lambda}\right]^{-1} \left[\Gamma^P + \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}\bar{\epsilon} - \left(\frac{\hat{X}'\hat{X}}{N}\right)^{-1} \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P\right].
\end{aligned} \quad \text{(B.12)}$$

Hence,

$$\begin{aligned}
\hat{\Gamma}^* - \Gamma^P &= \left(\frac{\hat{X}'\hat{X}}{N} - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'}{N}\bar{\epsilon} - \frac{\hat{X}'}{N}(\hat{X} - X)\Gamma^P + \hat{\Lambda}\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'}{N}\bar{\epsilon} - \left(\frac{\hat{X}'}{N}(\hat{X} - X) - \hat{\Lambda}\right)\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'}{N}\bar{\epsilon} - \begin{bmatrix} 1_N' \frac{\epsilon'}{N} \mathcal{P}\gamma_1^P \\ \frac{B'\epsilon'}{N}\mathcal{P}\gamma_1^P + \mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^P - \hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \end{bmatrix}\right].
\end{aligned} \quad \text{(B.13)}$$

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

$$\begin{aligned} \frac{\hat{X}'\bar{\epsilon}}{N} &= \frac{1}{N}(\hat{X} - X)'\bar{\epsilon} + \frac{1}{N}X'\bar{\epsilon} \\ &= O_p\left(\frac{1}{\sqrt{N}}\right), \end{aligned} \tag{B.14}$$

and Assumption 6(i) implies that

$$\mathcal{P}'\sum_{i=1}^{N}\epsilon_i = O_p\left(\sqrt{N}\right). \tag{B.15}$$

Note that

$$\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^P - \hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1}\gamma_1^P \tag{B.16}$$

can be rewritten as

$$\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N} - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 I_T\right)\mathcal{P}\gamma_1^P - \left[(\hat{\sigma}^2 - \sigma^2) - \left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 - \sigma^2\right)\right](\tilde{F}'\tilde{F})^{-1}\gamma_1^P. \tag{B.17}$$

Assumption 6(ii) implies that

$$\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N} - \frac{\sum_{i=1}^{N}\sigma_i^2}{N}I_T\right)\mathcal{P}\gamma_1^P = O_p\left(\frac{1}{\sqrt{N}}\right). \tag{B.18}$$

Using Lemma 1 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 $\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 - \sigma^2 = o\left(\frac{1}{\sqrt{N}}\right)$.

For part (ii), starting from (B.13), we have

$$
\begin{aligned}
\sqrt{N}(\hat{\Gamma}^* - \Gamma^P) &= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{\sqrt{N}} - \left(\frac{\hat{X}'}{\sqrt{N}}(\hat{X} - X)\Gamma^P\right) + \sqrt{N}\hat{\Lambda}\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{\hat{X}'\bar{\epsilon}}{\sqrt{N}} - \begin{bmatrix} 1_N' \\ \hat{B}' \end{bmatrix} \left[0_N,\ \frac{\epsilon'\mathcal{P}}{\sqrt{N}}\right]\Gamma^P + \sqrt{N}\hat{\Lambda}\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\frac{X'\bar{\epsilon}}{\sqrt{N}} + \frac{1}{\sqrt{N}} \begin{bmatrix} 0_N' \\ \mathcal{P}'\epsilon \end{bmatrix} \frac{\epsilon' 1_T}{T} - \frac{1}{\sqrt{N}} \begin{bmatrix} 1_N'\epsilon'\mathcal{P} \\ \hat{B}'\epsilon'\mathcal{P} \end{bmatrix} \gamma_1^P + \sqrt{N}\hat{\Lambda}\Gamma^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\begin{bmatrix} 1_N' \\ B' \end{bmatrix} \frac{\epsilon' 1_T}{T\sqrt{N}} + \begin{bmatrix} -1_N' \frac{\epsilon'\mathcal{P}}{\sqrt{N}}\gamma_1^P \\ \mathcal{P}' \frac{\epsilon\epsilon'}{\sqrt{N}} \frac{1_T}{T} - B' \frac{\epsilon'\mathcal{P}}{\sqrt{N}}\gamma_1^P - \mathcal{P}' \frac{\epsilon\epsilon'}{\sqrt{N}}\mathcal{P}\gamma_1^P \end{bmatrix}\right. \\
&\quad \left. + \sqrt{N}\hat{\sigma}^2(\tilde{F}'\tilde{F})^{-1}\gamma_1^P\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \begin{bmatrix} \frac{1_N'}{\sqrt{N}}\epsilon' \left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P\right) \\ \frac{B'\epsilon'}{\sqrt{N}}\left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P\right) + \mathcal{P}' \frac{\epsilon\epsilon'}{\sqrt{N}} \left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P\right) + \frac{\mathrm{tr}(M\epsilon\epsilon')}{\sqrt{N}(T-K-1)}\mathcal{P}'\mathcal{P}\gamma_1^P \end{bmatrix} \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} \left[\begin{bmatrix} \frac{1_N'\epsilon'}{\sqrt{N}}Q \\ \frac{B'\epsilon'}{\sqrt{N}}Q \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{P}' \frac{\epsilon\epsilon'}{\sqrt{N}}Q + \frac{\mathrm{tr}(M\epsilon\epsilon')}{\sqrt{N}(T-K-1)}\mathcal{P}'\mathcal{P}\gamma_1^P \end{bmatrix}\right] \\
&= \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1} (I_1 + I_2)\,. \qquad \text{(B.19)}
\end{aligned}
$$

Using Lemmas 1 and 2(ii), we have

$$
\left(\hat{\Sigma}_X - \hat{\Lambda}\right) \xrightarrow{p} \left(\begin{bmatrix} 1 & \mu_\beta' \\ \mu_\beta & \Sigma_\beta + \sigma^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix} - \begin{bmatrix} 0 & 0_K' \\ 0_K & \sigma^2(\tilde{F}'\tilde{F})^{-1} \end{bmatrix}\right) = \Sigma_X. \qquad \text{(B.20)}
$$

Consider now the terms I_1 and I_2. Both terms have a zero mean and, under Assumption 5(vi), they are asymptotically uncorrelated. Assumptions 2, 5(i), 6(i), and 6(iii) imply that

$$
\begin{aligned}
\mathrm{Var}(I_1) &= E\begin{bmatrix} Q'\frac{1}{\sqrt{N}}\sum_{i=1}^N \epsilon_i \frac{1}{\sqrt{N}}\sum_{j=1}^N \epsilon_j' Q & Q'\frac{1}{\sqrt{N}}\sum_{i=1}^N \epsilon_i \frac{1}{\sqrt{N}}\sum_{j=1}^N \epsilon_j'(Q\otimes\beta_j') \\ \frac{1}{\sqrt{N}}\sum_{i=1}^N (Q'\otimes\beta_i)\epsilon_i \frac{1}{\sqrt{N}}\sum_{j=1}^N \epsilon_j' Q & \frac{1}{\sqrt{N}}\sum_{i=1}^N (Q'\otimes\beta_i)\epsilon_i \frac{1}{\sqrt{N}}\sum_{j=1}^N \epsilon_j'(Q\otimes\beta_j') \end{bmatrix} \\
&= \begin{bmatrix} Q'\frac{1}{N}\sum_{i=1}^N E[\epsilon_i\epsilon_i']Q & Q'\frac{1}{N}\sum_{i=1}^N E[\epsilon_i\epsilon_i'](Q\otimes\beta_i') \\ \frac{1}{N}\sum_{i=1}^N (Q'\otimes\beta_i)E[\epsilon_i\epsilon_i']Q & \frac{1}{N}\sum_{i=1}^N (Q'\otimes\beta_i)E[\epsilon_i\epsilon_i'](Q\otimes\beta_i') \end{bmatrix} + o(1) \\
&\to \begin{bmatrix} \sigma^2 Q'Q & \sigma^2 Q'(Q\otimes\mu_\beta') \\ \sigma^2(Q'\otimes\mu_\beta)Q & \sigma^2(Q'Q\otimes\Sigma_\beta) \end{bmatrix} \\
&= \sigma^2 Q'Q\Sigma_X = \frac{\sigma^2}{T}\left[1 + \gamma_1^{P\prime}\left(\tilde{F}'\tilde{F}/T\right)^{-1}\gamma_1^P\right]\Sigma_X. \qquad \text{(B.21)}
\end{aligned}
$$

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

$$\begin{aligned} I_2 &= \left[\begin{array}{c} 0 \\ (Q'\otimes P')\text{vec}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\right) + \frac{1}{T-K-1}\text{tr}\left(M\frac{1}{\sqrt{N}}\sum_{i=1}^N(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\right)\mathcal{P}'\mathcal{P}\gamma_1^P \end{array}\right] \\ &= \left[\begin{array}{c} 0 \\ I_{22} \end{array}\right]. \end{aligned} \tag{B.22}$$

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

$$\text{Var}(I_2) = \left[\begin{array}{cc} 0 & 0_K' \\ 0_K & E\left[I_{22}I_{22}'\right] \end{array}\right]. \tag{B.23}$$

Under Assumptions 5(i) and 6(ii), we have

$$\begin{aligned} E\left[I_{22}I_{22}'\right] &= E\Bigg[(Q'\otimes\mathcal{P}')\frac{1}{\sqrt{N}}\sum_{i=1}^N \text{vec}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\frac{1}{\sqrt{N}}\sum_{j=1}^N \text{vec}(\epsilon_j\epsilon_j' - \sigma_j^2 I_T)'(Q\otimes\mathcal{P})\Bigg] \\ &\quad + E\Bigg[(Q'\otimes\mathcal{P}')\frac{1}{\sqrt{N}}\sum_{i=1}^N \text{vec}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\frac{1}{\sqrt{N}}\sum_{j=1}^N \text{vec}(\epsilon_j\epsilon_j' - \sigma_j^2 I_T)'\frac{\text{vec}(M)}{T-K-1}\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\Bigg] \\ &\quad + E\Bigg[\mathcal{P}'\mathcal{P}\gamma_1^P\frac{\text{vec}(M)'}{T-K-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \text{vec}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\frac{1}{\sqrt{N}}\sum_{j=1}^N \text{vec}(\epsilon_j\epsilon_j' - \sigma_j^2 I_T)'(Q\otimes\mathcal{P})\Bigg] \\ &\quad + E\Bigg[\mathcal{P}'\mathcal{P}\gamma_1^P\frac{\text{vec}(M)'}{T-K-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \text{vec}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\frac{1}{\sqrt{N}}\sum_{j=1}^N \text{vec}(\epsilon_j\epsilon_j' - \sigma_j^2 I_T)'\frac{\text{vec}(M)}{T-K-1} \\ &\quad \times\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\Bigg] \\ &\to \left[(Q'\otimes\mathcal{P}') + \mathcal{P}'\mathcal{P}\gamma_1^P\frac{\text{vec}(M)'}{T-K-1}\right]U_\epsilon\left[(Q\otimes\mathcal{P}) + \frac{\text{vec}(M)}{T-K-1}\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\right]. \end{aligned} \tag{B.24}$$

Defining $Z = \left[(Q\otimes\mathcal{P}) + \frac{\text{vec}(M)}{T-K-1}\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\right]$ concludes the proof of part (ii). ■

Proof of Theorem 2. By Theorem 1(i), $\hat{\gamma}_1^* \to_p \gamma_1^P$. Lemma 1 implies that $\hat{\Lambda}$ is a consistent estimator of Λ. Hence, using Lemma 2(ii), we have $\left(\hat{\Sigma}_X - \hat{\Lambda}\right) \to_p \Sigma_X$, which implies that $\hat{V} \to_p V$. A consistent estimator of W requires a consistent estimate of the matrix U_ϵ, which can be obtained using Lemma 6. This concludes the proof of Theorem 2. ■

Proof of Theorem 3. Writing

$$
\begin{aligned}
(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\frac{\hat{X}'R_t}{N} &= (\hat{\Sigma}_X-\hat{\Lambda})^{-1}\hat{\Sigma}_X\Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\hat{X}'\epsilon'\imath_{t,T}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\hat{X}'(X-\hat{X})\Gamma^P_{t-1}\\
&= (\hat{\Sigma}_X-\hat{\Lambda})^{-1}(\hat{\Sigma}_X-\hat{\Lambda}+\hat{\Lambda})\Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\hat{X}'\epsilon'\imath_{t,T}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\hat{X}'(X-\hat{X})\Gamma^P_{t-1}\\
&= \Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left(\frac{\hat{X}'\epsilon'\imath_{t,T}}{N}+\frac{\hat{X}'(X-\hat{X})}{N}\Gamma^P_{t-1}+\hat{\Lambda}\Gamma^P_{t-1}\right)\\
&= \Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left(\frac{X'\epsilon'\imath_{t,T}}{N}+\frac{(\hat{X}-X)'\epsilon'\imath_{t,T}}{N}+\frac{\hat{X}'(X-\hat{X})}{N}\Gamma^P_{t-1}+\hat{\Lambda}\Gamma^P_{t-1}\right)\\
&= \Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left(\left[\begin{array}{c}1_N'\\ B'\end{array}\right]\frac{\epsilon'\imath_{t,T}}{N}+\frac{1}{N}\left[\begin{array}{c}0_N'\\ \mathcal{P}'\epsilon\end{array}\right]\epsilon'\imath_{t,T}\right.\\
&\left.+\frac{1}{N}\left[\begin{array}{c}-1_N'\epsilon'\mathcal{P}\gamma^P_{1,t-1}\\ -B'\epsilon'\mathcal{P}\gamma^P_{1,t-1}-\mathcal{P}'\epsilon\epsilon'\mathcal{P}\gamma^P_{1,t-1}\end{array}\right]+\hat{\Lambda}\Gamma^P_{t-1}\right)\\
&= \Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left(\left[\begin{array}{c}\frac{1_N'\epsilon'}{N}Q_{t-1}\\ \frac{B'\epsilon'}{N}Q_{t-1}\end{array}\right]+\left[\begin{array}{c}0\\ \frac{P'\epsilon\epsilon'}{N}Q_{t-1}\end{array}\right]+\hat{\Lambda}\Gamma^P_{t-1}\right)\\
&= \Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left(\left[\begin{array}{c}\frac{1_N'\epsilon'}{N}Q_{t-1}\\ \frac{B'\epsilon'}{N}Q_{t-1}\end{array}\right]+\left[\begin{array}{c}0\\ -\frac{P'\epsilon\epsilon'}{N}P\gamma^P_{1,t-1}\end{array}\right]+\hat{\Lambda}\Gamma^P_{t-1}+\left[\begin{array}{c}0\\ \frac{P'\epsilon\epsilon'}{N}\imath_{t,T}\end{array}\right]\right)
\end{aligned}
\tag{B.25}
$$

with

$$
E\left(\left[\begin{array}{c}0\\ -\frac{\mathcal{P}'\epsilon\epsilon'}{N}\mathcal{P}\gamma^P_{1,t-1}\end{array}\right]+\hat{\Lambda}\Gamma^P_{t-1}\right)=E\left(\left[\begin{array}{c}0\\ -\frac{\mathcal{P}'\epsilon\epsilon'}{N}\mathcal{P}\gamma^P_{1,t-1}\end{array}\right]+\left[\begin{array}{c}0\\ \frac{\mathrm{tr}(M\epsilon\epsilon')}{N(T-K-1)}\mathcal{P}'\mathcal{P}\gamma^P_{1,t-1}\end{array}\right]\right)=0_{K+1}
\tag{B.26}
$$

and

$$
\left[\begin{array}{c}0\\ \frac{P'\epsilon\epsilon'}{N}\imath_{t,T}\end{array}\right]\to_p\left[\begin{array}{c}0\\ \sigma^2P'\imath_{t,T}\end{array}\right]=\left[\begin{array}{c}0\\ \sigma^2(\tilde{F}'\tilde{F})^{-1}\tilde{f}_t\end{array}\right]
\tag{B.27}
$$

yields part (i).

Next,

$$
\begin{aligned}
\hat{\Gamma}^*_{t-1} &= (\hat{\Sigma}_X-\hat{\Lambda})^{-1}\frac{\hat{X}'R_t}{N}-(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left[\begin{array}{c}0\\ \hat{\sigma}^2P'\imath_{t,T}\end{array}\right]\\
&= \Gamma^P_{t-1}+(\hat{\Sigma}_X-\hat{\Lambda})^{-1}\left(\left[\begin{array}{c}\frac{1_N'\epsilon'}{N}Q_{t-1}\\ \frac{B'\epsilon'}{N}Q_{t-1}\end{array}\right]+\left[\begin{array}{c}0\\ \frac{P'\epsilon\epsilon'}{N}Q_{t-1}\end{array}\right]-\left[\begin{array}{c}0\\ \hat{\sigma}^2P'Q_{t-1}\end{array}\right]\right).
\end{aligned}
\tag{B.28}
$$

The part of $\sqrt{N}(\hat{\Gamma}^*_{t-1} - \Gamma^P_{t-1})$ that depends on $\epsilon\epsilon'$ can be written as

$$\begin{aligned}
&(\hat{\Sigma}_X - \hat{\Lambda})^{-1}[(Q'_{t-1} \otimes P') - P'Q_{t-1}\text{vec}(M)']\text{vec}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\right) \\
&= (\hat{\Sigma}_X - \hat{\Lambda})^{-1}Z'_{t-1}\text{vec}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\epsilon_i\epsilon_i' - \sigma_i^2 I_T)\right), \qquad \text{(B.29)}
\end{aligned}$$

and the result follows along the proof of Theorem 1(ii). ■

Proof of Theorem 4. We first establish a simpler, asymptotically equivalent, expression for $\sqrt{N}\left(\frac{\hat{e}^{P\prime}\hat{e}^P}{N} - \hat{\sigma}^2\hat{Q}'\hat{Q}\right)$. Then, we derive the asymptotic distribution of this approximation. Consider the sample ex post pricing errors,

$$\hat{e}^P = \bar{R} - \hat{X}\hat{\Gamma}^*. \qquad \text{(B.30)}$$

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

$$\begin{aligned}
\hat{e}^P &= \hat{X}\Gamma^P + \bar{\epsilon} - (\hat{X} - X)\Gamma^P - \hat{X}\hat{\Gamma}^* \\
&= \bar{\epsilon} - \hat{X}(\hat{\Gamma}^* - \Gamma^P) - (\hat{X} - X)\Gamma^P. \qquad \text{(B.31)}
\end{aligned}$$

Then,

$$\begin{aligned}
\hat{e}^{P\prime}\hat{e}^P &= \bar{\epsilon}'\bar{\epsilon} + \Gamma^{P\prime}(\hat{X} - X)'(\hat{X} - X)\Gamma^P - 2(\hat{\Gamma}^* - \Gamma^P)'\hat{X}'\bar{\epsilon} - 2\Gamma^{P\prime}(\hat{X} - X)'\bar{\epsilon} \\
&\quad + 2\Gamma^{P\prime}(\hat{X} - X)'\hat{X}(\hat{\Gamma}^* - \Gamma^P) + (\hat{\Gamma}^* - \Gamma^P)'\hat{X}'\hat{X}(\hat{\Gamma}^* - \Gamma^P).
\end{aligned}$$

Note that

$$\frac{\bar{\epsilon}'\bar{\epsilon}}{N} = \frac{1}{T^2}1_T'\frac{\epsilon\epsilon'}{N}1_T \to_p \frac{\sigma^2}{T}, \qquad \text{(B.32)}$$

and, by Lemma 2(iii),

$$\Gamma^{P\prime}\frac{(\hat{X} - X)'(\hat{X} - X)}{N}\Gamma^P = \gamma_1^{P\prime}\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^P \to_p \sigma^2\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^P. \qquad \text{(B.33)}$$

Using Lemmas 3 and 5 and Theorem 1, we have

$$\frac{(\hat{\Gamma}^* - \Gamma^P)'\hat{X}'\bar{\epsilon}}{N} = \frac{(\hat{\Gamma}^* - \Gamma^P)'(\hat{X} - X)'\bar{\epsilon}}{N} + \frac{(\hat{\Gamma}^* - \Gamma^P)'X'\bar{\epsilon}}{N} = O_p\left(\frac{1}{N}\right) \qquad \text{(B.34)}$$

and

$$\frac{\Gamma^{P\prime}(\hat{X} - X)'\bar{\epsilon}}{N} = O_p\left(\frac{1}{\sqrt{N}}\right). \qquad \text{(B.35)}$$

In addition, using Lemmas 2(i), 2(iii), 4 and Theorem 1, we have

$$\begin{aligned}
\frac{\Gamma^{P\prime}(\hat{X}-X)'\hat{X}(\hat{\Gamma}^{*}-\Gamma^{P})}{N} &= \frac{\Gamma^{P\prime}(\hat{X}-X)'(\hat{X}-X)(\hat{\Gamma}^{*}-\Gamma^{P})}{N}+\frac{\Gamma^{P\prime}(\hat{X}-X)'X(\hat{\Gamma}^{*}-\Gamma^{P})}{N} \\
&= O_p\Big(\frac{1}{\sqrt{N}}\Big)+O_p\left(\frac{1}{N}\right) \qquad \text{(B.36)}
\end{aligned}$$

and

$$\frac{(\hat{\Gamma}^{*}-\Gamma^{P})'\hat{X}'\hat{X}(\hat{\Gamma}^{*}-\Gamma^{P})}{N} = O_p\left(\frac{1}{N}\right). \qquad \text{(B.37)}$$

It follows that

$$\frac{\hat{e}^{P\prime}\hat{e}^{P}}{N}\rightarrow_p\frac{\sigma^2}{T}+\sigma^2\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}=\sigma^2Q'Q. \qquad \text{(B.38)}$$

Collecting terms and rewriting explicitly only the ones that are $O_p\left(\frac{1}{\sqrt{N}}\right)$, we have

$$\begin{aligned}
\frac{\hat{e}^{P\prime}\hat{e}^{P}}{N} &= \frac{\bar{\epsilon}'\bar{\epsilon}}{N} && \text{(B.39)}\\
&\quad+\frac{\Gamma^{P\prime}(\hat{X}-X)'(\hat{X}-X)\Gamma^{P}}{N} && \text{(B.40)}\\
&\quad-2\,\frac{\Gamma^{P\prime}(\hat{X}-X)'\bar{\epsilon}}{N} && \text{(B.41)}\\
&\quad+2\,\frac{\Gamma^{P\prime}(\hat{X}-X)'(\hat{X}-X)(\hat{\Gamma}^{*}-\Gamma^{P})}{N} && \text{(B.42)}\\
&\quad+O_p\Big(\frac{1}{N}\Big). && \text{(B.43)}
\end{aligned}$$

Consider the sum of the three terms in Eqs. (B.39)–(B.41). Under Assumption 5(i), we have

$$\begin{aligned}
&\frac{\bar{\epsilon}'\bar{\epsilon}}{N}+\frac{\Gamma^{P\prime}(\hat{X}-X)'(\hat{X}-X)\Gamma^{P}}{N}-2\frac{\Gamma^{P\prime}(\hat{X}-X)'\bar{\epsilon}}{N}\\
=\;& \frac{1_T'}{T}\frac{\epsilon\epsilon'}{N}\frac{1_T}{T}+\gamma_1^{P\prime}\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^{P}-2\frac{1_T'}{T}\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^{P}\\
=\;& \frac{1_T'}{T}\frac{\epsilon\epsilon'}{N}\Big(\frac{1_T}{T}-\mathcal{P}\gamma_1^{P}\Big)-\frac{1_T'}{T}\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^{P}+\gamma_1^{P\prime}\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^{P}\\
=\;& \frac{1_T'}{T}\frac{\epsilon\epsilon'}{N}Q-Q'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^{P}\\
=\;& Q'\frac{\epsilon\epsilon'}{N}\frac{1_T}{T}-Q'\frac{\epsilon\epsilon'}{N}\mathcal{P}\gamma_1^{P}\\
=\;& Q'\frac{\epsilon\epsilon'}{N}Q=Q'\Big(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2I_T\Big)Q+\sigma^2Q'Q+o\left(\frac{1}{\sqrt{N}}\right), \qquad \text{(B.44)}
\end{aligned}$$

where the $o\left(\frac{1}{\sqrt{N}}\right)$ term comes from $(\bar{\sigma}^2-\sigma^2)Q'Q$. As for the term in Eq. (B.42), define

$$\Big(\hat{\Sigma}_X-\hat{\Lambda}\Big)^{-1}=\begin{bmatrix}\hat{\Sigma}_{11} & \hat{\Sigma}_{12}\\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22}\end{bmatrix}, \qquad \text{(B.45)}$$

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

$$
\begin{aligned}
& 2\frac{\Gamma^{P\prime}(\hat{X}-X)'(\hat{X}-X)(\hat{\Gamma}^{*}-\Gamma^{P})}{N} \\
= \;& 2\left[\gamma_1^{P\prime}\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\hat{\Sigma}_{21},\ \gamma_1^{P\prime}\mathcal{P}'\frac{\epsilon\epsilon'}{N}\mathcal{P}\hat{\Sigma}_{22}\right]\left[\begin{array}{c}\frac{1_N'\epsilon'Q}{N}\\ \frac{B'\epsilon'Q}{N}+Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\end{array}\right] \\
= \;& 2\gamma_1^{P\prime}\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon'Q}{N}+2\gamma_1^{P\prime}\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon'Q}{N} \\
& +2\gamma_1^{P\prime}\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right) \\
& +2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon'Q}{N}+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon'Q}{N} \\
& +2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)+o_p\left(\frac{1}{N}\right) \\
= \;& 2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon'Q}{N}+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon'Q}{N} \\
& +2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)+o_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{N}\right),
\end{aligned} \tag{B.46}
$$

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

$$
\begin{aligned}
& 2(\bar{\sigma}^2-\sigma^2)\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon'Q}{N}+2(\bar{\sigma}^2-\sigma^2)\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon'Q}{N} \\
& +2(\bar{\sigma}^2-\sigma^2)\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)=o_p\left(\frac{1}{N}\right)
\end{aligned} \tag{B.47}
$$

and

$$
\begin{aligned}
& 2\gamma_1^{P\prime}\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon'Q}{N}+2\gamma_1^{P\prime}\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon'Q}{N} \\
& +2\gamma_1^{P\prime}\mathcal{P}'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)=O_p\left(\frac{1}{N}\right),
\end{aligned} \tag{B.48}
$$

respectively. Therefore, we have

$$
\begin{aligned}
\frac{\hat{e}^{P\prime}\hat{e}^{P}}{N} = \;& Q'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)Q+\sigma^2 Q'Q \\
& +2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon'Q}{N}+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon'Q}{N} \\
& +2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)+O_p\left(\frac{1}{N}\right)+o_p\left(\frac{1}{N}\right)+o\left(\frac{1}{\sqrt{N}}\right).
\end{aligned} \tag{B.49}
$$

It follows that

$$\frac{\hat{e}^{P\prime}\hat{e}^{P}}{N}-\hat{\sigma}^2\hat{Q}'\hat{Q} = Q'\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)Q-\left(\hat{\sigma}^2\hat{Q}'\hat{Q}-\sigma^2 Q'Q\right)$$
$$+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon' Q}{N}+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon' Q}{N}$$
$$+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)+O_p\left(\frac{1}{N}\right)+o_p\left(\frac{1}{N}\right)+o\left(\frac{1}{\sqrt{N}}\right). \tag{B.50}$$

Note that

$$\begin{aligned}
&\hat{\sigma}^2\hat{Q}'\hat{Q}-\sigma^2 Q'Q \\
=\;& \frac{1}{T}(\hat{\sigma}^2-\sigma^2)+\hat{\sigma}^2\hat{\gamma}_1^{*\prime}(\tilde{F}'\tilde{F})^{-1}\hat{\gamma}_1^{*}-\sigma^2\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P} \\
=\;& \frac{1}{T}(\hat{\sigma}^2-\sigma^2)+(\hat{\sigma}^2-\sigma^2)\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}+2\sigma^2(\hat{\gamma}_1^{*}-\gamma_1^{P})'(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}+O_p\left(\frac{1}{N}\right) \\
=\;& (\hat{\sigma}^2-\sigma^2)\left(\frac{1}{T}+\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}\right)+2\sigma^2(\hat{\gamma}_1^{*}-\gamma_1^{P})'(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}+O_p\left(\frac{1}{N}\right) \\
=\;& (\hat{\sigma}^2-\sigma^2)\left(\frac{1}{T}+\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}\right)+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{21}\frac{1_N'\epsilon' Q}{N}+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}\frac{B'\epsilon' Q}{N} \\
&+2\sigma^2\gamma_1^{P\prime}\mathcal{P}'\mathcal{P}\hat{\Sigma}_{22}Z'\mathrm{vec}\left(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\right)+O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{N\sqrt{N}}\right),
\end{aligned} \tag{B.51}$$

where $\sigma^2(\hat{\gamma}_1^{*}-\gamma_1^{P})'(\tilde{F}'\tilde{F})^{-1}(\hat{\gamma}_1^{*}-\gamma_1^{P})+2\,(\hat{\sigma}^2-\sigma^2)(\hat{\gamma}_1^{*}-\gamma_1^{P})'(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}=O_p\left(\frac{1}{N}\right)$ and $(\hat{\sigma}^2-\sigma^2)(\hat{\gamma}_1^{*}-\gamma_1^{P})'(\tilde{F}'\tilde{F})^{-1}(\hat{\gamma}_1^{*}-\gamma_1^{P})=O_p\left(\frac{1}{N\sqrt{N}}\right)$. It follows that

$$\begin{aligned}
&\frac{\hat{e}'\hat{e}}{N}-\hat{\sigma}^2\hat{Q}'\hat{Q} \\
=\;& Q'\Big(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\Big)Q-(\hat{\sigma}^2-\sigma^2)\Big(\frac{1}{T}+\gamma_1^{P\prime}(\tilde{F}'\tilde{F})^{-1}\gamma_1^{P}\Big)+O_p\Big(\frac{1}{N\sqrt{N}}\Big)+O_p\Big(\frac{1}{N}\Big)+o\Big(\frac{1}{\sqrt{N}}\Big)+o_p\Big(\frac{1}{\sqrt{N}}\Big) \\
=\;& \Big[\Big(Q'\otimes Q'\Big)-\frac{Q'Q}{T-K-1}\mathrm{vec}(M)'\Big]\mathrm{vec}\Big(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\Big)+o_p\Big(\frac{1}{\sqrt{N}}\Big) \\
=\;& Z_Q'\mathrm{vec}\Big(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\Big)+o_p\Big(\frac{1}{\sqrt{N}}\Big),
\end{aligned} \tag{B.52}$$

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

$$\sqrt{N}\Big(\frac{\hat{e}'\hat{e}}{N}-\hat{\sigma}^2\hat{Q}'\hat{Q}\Big) = \sqrt{N}Z_Q'\mathrm{vec}\Big(\frac{\epsilon\epsilon'}{N}-\bar{\sigma}^2 I_T\Big)+o_p(1), \tag{B.53}$$

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

$$\sqrt{N}Z_Q'\text{vec}\Big(\frac{\epsilon\epsilon'}{N} - \bar{\sigma}^2 I_T\Big) \to_d \mathcal{N}\Big(0, Z_Q' U_\epsilon Z_Q\Big).\blacksquare \tag{B.54}$$

Proof of Theorem 5. For part (i), in view of Eq. (65), we obtain $\bar{R} = X\tilde{\Gamma}^P + e + \bar{\epsilon}$, where $\tilde{\Gamma}^P = \tilde{\Gamma} + \bar{f} - E[f_t]$. Using the same arguments as for Theorem 1,

$$\hat{\Gamma}^* - \tilde{\Gamma}^P \quad = \quad \left(\frac{\hat{X}'\hat{X}}{N} - \hat{\Lambda}\right)^{-1}\left[\frac{\hat{X}'\bar{\epsilon}}{N} - \left(\frac{\hat{X}'}{N}\left(\hat{X} - X\right) - \hat{\Lambda}\right)\tilde{\Gamma}^P + \frac{\hat{X}'e}{N}\right] \tag{B.55}$$

with $\left(\frac{\hat{X}'\hat{X}}{N} - \hat{\Lambda}\right) = O_p(1)$, $\frac{\hat{X}'\bar{\epsilon}}{N} = O_p\left(\frac{1}{\sqrt{N}}\right)$, and $\left(\frac{\hat{X}'}{N}\left(\hat{X} - X\right) - \hat{\Lambda}\right) = O_p\left(\frac{1}{\sqrt{N}}\right)$. As for the term $\frac{\hat{X}'e}{N}$,

$$\begin{aligned}\frac{\hat{X}'e}{N} = \frac{X'e}{N} + \frac{(\hat{X}-X)'e}{N} &= 0_{K+1} + \frac{1}{N}\begin{bmatrix}0\\ \mathcal{P}'\epsilon\, e\end{bmatrix}\\ &= 0_{K+1} + O_p\left(\frac{1}{\sqrt{N}}\right)\end{aligned} \tag{B.56}$$

since $\mathcal{P}'\epsilon\, e = O_p((\mathcal{P}'\sum_{i,j=1}^N \sigma_{ij}e_ie_j\mathcal{P})^{\frac{1}{2}}) = O_p(\sqrt{N})$ by Assumption 7(i)-(ii). Next,

$$\begin{aligned}\sqrt{N}\left(\hat{\Gamma}^* - \tilde{\Gamma}^P\right) \quad &= \quad \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}\left(\begin{bmatrix}\frac{1_N'\epsilon' Q}{\sqrt{N}}\\ \frac{B'\epsilon' Q}{\sqrt{N}}\end{bmatrix} + \begin{bmatrix}0\\ \frac{\mathcal{P}'\epsilon\epsilon' Q}{\sqrt{N}} + \frac{\text{tr}(M\epsilon\epsilon')}{\sqrt{N}(T-K-1)}\mathcal{P}'\mathcal{P}\tilde{\gamma}_1^P\end{bmatrix} + \begin{bmatrix}0\\ \mathcal{P}'\frac{1}{\sqrt{N}}\sum_{i=1}^N \epsilon_i e_i\end{bmatrix}\right)\\ &\equiv \quad \left(\hat{\Sigma}_X - \hat{\Lambda}\right)^{-1}(I_1 + I_2 + I_3)\,.\end{aligned} \tag{B.57}$$

As for terms I_1 and I_2, Theorem 1 applies, that is, $\left(\hat{\Sigma}_X - \hat{\Lambda}\right) \to_p \Sigma_X$, $\text{Var}(I_1) = \frac{\sigma^2}{T}\left[1 + \tilde{\gamma}_1^{P'}\left(\frac{\tilde{F}'\tilde{F}}{T}\right)^{-1}\tilde{\gamma}_1^P\right]\Sigma_X$ and $\text{Var}(I_2) = \begin{bmatrix}0 & 0_K'\\ 0_K & E[I_{22}I_{22}']\end{bmatrix}$, with $E[I_{22}I_{22}'] = \left[(Q'\otimes\mathcal{P}') + \mathcal{P}'\mathcal{P}\tilde{\gamma}_1^P\frac{\text{vec}(M)'}{T-K-1}\right]U_\epsilon\left[(Q\otimes\mathcal{P}) + \frac{\text{vec}M}{T-K-1}\tilde{\gamma}_1^{P'}\mathcal{P}'\mathcal{P}\right]$, where $\text{Cov}(I_1, I_2) = 0_{(K+1)\times(K+1)}$. Consider now the term I_3 and note that it has a zero mean. Its variance is equal to

$$\text{Var}(I_3) = E\begin{bmatrix}0 & 0_K'\\ 0_K & \mathcal{P}'\frac{1}{N}\sum_{i,j=1}^N \epsilon_i\epsilon_j' e_ie_j\mathcal{P}\end{bmatrix} \to_p \begin{bmatrix}0 & 0_K'\\ 0_K & \tau_\Omega\mathcal{P}'\mathcal{P}\end{bmatrix} \equiv \Omega, \tag{B.58}$$

and the covariance term satisfies

$$\text{Cov}(I_1, I_3) \quad = \quad E\begin{bmatrix}\frac{1_N'\epsilon' Q}{\sqrt{N}}\\ \frac{B'\epsilon' Q}{\sqrt{N}}\end{bmatrix}\left[0, \frac{1}{\sqrt{N}}\sum_{i=1}^N e_i\epsilon_i'\mathcal{P}\right] \to_p \begin{bmatrix}0 & \tau_\Phi Q'\mathcal{P}\\ 0_K & \tau_\Phi\left(Q'\otimes\mu_\beta\right)\mathcal{P}\end{bmatrix} \equiv \Phi, \tag{B.59}$$

while $\text{Cov}(I_2, I_3) = 0_{(K+1)\times(K+1)}$ by the assumption of zero third moment of the error term. Using Lemmas 8 and 9, the proof of part (ii) becomes very similar to the proof of Theorem 2 and is omitted. ■

Proof of Theorem 6. For part (i), rewrite

$$\begin{bmatrix}\hat{\Gamma}^* \\ \hat{\delta}^*\end{bmatrix} = \begin{bmatrix}\Gamma^P \\ \delta\end{bmatrix} + \begin{bmatrix}\hat{X}'\hat{X} - \hat{\Lambda} & \hat{X}'C \\ C'\hat{X} & C'C\end{bmatrix}^{-1} \left[\begin{bmatrix}\hat{\Lambda}\Gamma^P \\ 0_{K_c}\end{bmatrix} + \begin{bmatrix}\hat{X}' \\ C'\end{bmatrix}(\bar{\epsilon} + (X - \hat{X})\Gamma^P)\right].$$

As for the bias associated with $\hat{\Gamma}^*$ (see the proof of Theorem 1), we have

$$\hat{\Lambda}\Gamma^P + \frac{1}{N}\hat{X}'(\bar{\epsilon} + (X - \hat{X})\Gamma^P) = O_p(N^{-1/2}). \tag{B.60}$$

As for the bias associated with $\hat{\delta}^*$, we have

$$\frac{1}{N}C'(\bar{\epsilon} + (X - \hat{X})\Gamma^P) = \frac{1}{N}C'\epsilon'\left(\frac{1_T}{T} - \mathcal{P}\gamma_1^P\right) = \frac{1}{N}C'\epsilon'Q = O_p(N^{-1/2}) \tag{B.61}$$

since $N^{-1}C'\epsilon' \to_p 0_{K_c\times T}$ and

$$\begin{aligned}\text{Var}\left(\frac{1}{N}C'\epsilon'Q\right) &= (Q' \otimes I_{K_c})\frac{1}{N^2}\sum_{i,j=1}^{N}\Sigma_{zz,ij}(Q \otimes I_{K_c}) = \frac{1}{N^2}(Q' \otimes I_{K_c})\sum_{i,j=1}^{N}\sigma_{ij}(I_T \otimes c_i c_j')(Q \otimes I_{K_c}) \\ &= \frac{1}{N^2}\sum_{i,j=1}^{N}\sigma_{ij}(Q'Qc_i c_j') = \frac{1}{N}\sigma^2(Q'Q\Sigma_{CC}) + o\left(\frac{1}{N}\right)\end{aligned} \tag{B.62}$$

by Assumption 8.

For part (ii), by straightforward generalizations of Lemmas 1 and 2(ii), we have

$$\frac{1}{N}\begin{bmatrix}\hat{X}'\hat{X} - N\hat{\Lambda} & \hat{X}'C \\ C'\hat{X} & C'C\end{bmatrix} \to_p \begin{bmatrix}\Sigma_X & \begin{bmatrix}\mu_C' \\ \Sigma_{CB}'\end{bmatrix} \\ \begin{bmatrix}\mu_C & \Sigma_{CB}\end{bmatrix} & \Sigma_{CC}\end{bmatrix} = L. \tag{B.63}$$

We now prove that L is positive-definite. Using the blockwise formula for the inverse of a matrix, the invertibility of L follows from Σ_{CC} being positive-definite (see Assumption 8(i)) and the invertibility of $\begin{bmatrix}1 & \mu_\beta' \\ \mu_\beta & \Sigma_\beta\end{bmatrix} - \begin{bmatrix}\mu_C' \\ \Sigma_{CB}'\end{bmatrix}\Sigma_{CC}^{-1}\begin{bmatrix}\mu_C & \Sigma_{CB}\end{bmatrix}$. In turn, this holds if

$$D = \Sigma_\beta - \Sigma_{CB}'\Sigma_{CC}^{-1}\Sigma_{CB} \tag{B.64}$$

is positive-definite and

$$1 - \mu_C'\Sigma_{CC}^{-1}\mu_C - (\mu_\beta' - \mu_C'\Sigma_{CC}^{-1}\Sigma_{CB})D^{-1}(\mu_\beta - \Sigma_{CB}'\Sigma_{CC}^{-1}\mu_C) \tag{B.65}$$

is nonzero. The last equation can be rewritten as

$$1 - [\mu_C' \, \mu_\beta'] \begin{bmatrix} \Sigma_{CC} & \Sigma_{CB} \\ \Sigma_{CB}' & \Sigma_\beta \end{bmatrix}^{-1} \begin{bmatrix} \mu_C \\ \mu_\beta \end{bmatrix}. \tag{B.66}$$

The positiveness of Eq. (B.66) and the positive-definiteness of D follow from Assumption 8(i). Next, following the proof of Theorem 1,

$$\begin{aligned} \sqrt{N} \begin{bmatrix} \hat{\Gamma}^* - \Gamma^P \\ \hat{\delta}^* - \delta \end{bmatrix} &= \begin{bmatrix} \frac{\hat{X}'\hat{X}}{N} - \hat{\Lambda} & \frac{\hat{X}'C}{N} \\ \frac{C'\hat{X}}{N} & \frac{C'C}{N} \end{bmatrix}^{-1} \\ &\times \left(\begin{bmatrix} \frac{1_N'\epsilon'}{\sqrt{N}} Q \\ \frac{B'\epsilon'}{\sqrt{N}} Q \\ 0_{K_c} \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{P}' \frac{\epsilon\epsilon'}{\sqrt{N}} Q + \frac{\mathrm{tr}(M\epsilon\epsilon')}{\sqrt{N}(T-K-1)} \mathcal{P}'\mathcal{P}\gamma_1^P \\ 0_{K_c} \end{bmatrix} + \begin{bmatrix} 0 \\ 0_K \\ \frac{C'\epsilon'}{\sqrt{N}} Q \end{bmatrix} \right) \\ \equiv & \begin{bmatrix} \frac{\hat{X}'\hat{X}}{N} - \hat{\Lambda} & \frac{\hat{X}'C}{N} \\ \frac{C'\hat{X}}{N} & \frac{C'C}{N} \end{bmatrix}^{-1} (I_1 + I_2 + I_3). \end{aligned} \tag{B.67}$$

We now derive $\mathrm{Var}(I_3)$ and $\mathrm{Cov}(I_1, I_3')$ because the other terms can be directly obtained from Theorem 1 and $\mathrm{Cov}(I_2, I_3') = 0_{(K+K_c+1)\times(K+K_c+1)}$. We have

$$\mathrm{Var}(I_3) = \begin{bmatrix} 0_{(K+1)\times(K+1)}' & 0_{(K+1)\times K_c}' \\ 0_{K_c\times(K+1)} & \frac{Q'Q}{N}\sum_{i=1}^N \sigma_{ij}(c_i c_j') \end{bmatrix} \to \begin{bmatrix} 0_{(K+1)\times(K+1)}' & 0_{(K+1)\times K_c}' \\ 0_{K_c\times(K+1)} & \sigma^2 Q'Q\Sigma_{CC} \end{bmatrix}$$

and, by Theorem 1,

$$\mathrm{Cov}(I_1, I_3') = \begin{bmatrix} 0_{(K+1)\times(K+1)} & \frac{Q'Q}{N}\sum_{i=1}^N \sigma_{ij}\left(\begin{bmatrix} 1 \\ \beta_i \end{bmatrix} c_j'\right) \\ 0_{K_c\times(K+1)} & 0_{K_c\times K_c} \end{bmatrix} \to \begin{bmatrix} 0_{(K+1)\times(K+1)} & \sigma^2 Q'Q \begin{bmatrix} \mu_C' \\ \Sigma_{CB}' \end{bmatrix} \\ 0_{K_c\times(K+1)} & 0_{K_c\times K_c} \end{bmatrix}. \blacksquare$$

Appendix C: Explicit Form of U_ϵ

Denote by U_ϵ the $T^2 \times T^2$ matrix

$$U_\epsilon = \left[\begin{array}{c:c:c:c:c} U_{11} & \cdots & U_{1t} & \cdots & U_{1T} \\ \hdashline \vdots & \ddots & \vdots & \vdots & \vdots \\ \hdashline U_{t1} & \cdots & U_{tt} & \cdots & U_{tT} \\ \hdashline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hdashline U_{T1} & \cdots & U_{Tt} & \cdots & U_{TT} \end{array}\right]. \tag{C.1}$$

Each block of U_ϵ 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 σ_4 in the (s,s) position for every $s \neq t$; that is,

$$
U_{tt} \quad = \quad \begin{array}{r c}
 & \begin{array}{c} \downarrow \\ t\text{-th column} \end{array} \\
\begin{array}{c} \\ \\ \\ \rightarrow \\ t\text{-th row} \\ \\ \\ \end{array} &
\begin{bmatrix}
\sigma_4 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \sigma_4 & 0 & \cdots & \cdots & 0 \\
0 & \cdots & 0 & (\kappa_4 + 2\sigma_4) & 0 & \cdots & 0 \\
0 & \cdots & \cdots & 0 & \sigma_4 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & \sigma_4
\end{bmatrix}
\end{array}. \tag{C.2}
$$

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 σ_4, that is,

$$
U_{ts} \quad = \quad \begin{array}{r c}
 & \begin{array}{c} \downarrow \\ t\text{-th column} \end{array} \\
\begin{array}{c} \\ \\ \\ \rightarrow \\ s\text{-th row} \\ \\ \\ \end{array} &
\begin{bmatrix}
0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
0 & \cdots & 0 & \sigma_4 & 0 & \cdots & 0 \\
0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & 0
\end{bmatrix}
\end{array}. \tag{C.3}
$$

Under Assumption 5, by Lemma 6, it is easy to show that $\hat{U}_\epsilon$ in Theorem 2 is a consistent plug-in estimator of U_ϵ that only depends on $\hat{\sigma}_4$.

References

Ahn, S. C., A. R. Horenstein, and N. Wang. 2018. Beta matrix and common factors in stock returns. *Journal of Financial and Quantitative Analysis* 53:1417–1440.

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

Ang, A., and D. Kristensen. 2012. Testing conditional factor models. *Journal of Financial Economics* 106:132–156.

Ang, A., J. Liu, and K. Schwarz. 2018. Using stocks or portfolios in tests of factor models. *Journal of Financial and Quantitative Analysis*, forthcoming.

Bai, J., and G. Zhou. 2015. Fama-MacBeth two-pass regressions: Improving risk premia estimates. *Finance Research Letters* 15:31–40.

Balduzzi, P., and C. Robotti. 2008. Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business & Economic Statistics* 26:354–368.

Barillas, F., and J. Shanken. 2017. Which Alpha? *Review of Financial Studies* 30:1316–1338.

Barras, L., O. Scaillet, and R. Wermers. 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance* 65:179–216.

Berk, J. B. 2002. Sorting out sorts. *Journal of Finance* 55:407–427.

Black, F., M. C. Jensen, and M. Scholes. 1972. The Capital Asset Pricing Model: Some empirical tests. In *Studies in the Theory of Capital Markets*. New York: Praeger.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger. 1989. Empirical tests of the consumption-oriented CAPM. *Journal of Finance* 44:231–262.

Brennan, M., T. Chordia, and A. Subrahmanyam. 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. *Journal of Financial Economics* 49:345–373.

Brillinger, D. R. 2001. *Time Series: Data Analysis and Theory.* Philadelphia: Society for Industrial and Applied Mathematics.

Bryzgalova, S. 2016. Spurious factors in linear asset pricing models. Working paper, Stanford University.

Burnside, C. 2015. Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics* 14:295–330.

Chamberlain, G. 1983. Funds, factors, and diversification in arbitrage pricing models. *Econometrica* 51:1305–1323.

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

Chan, L. K. C., J. Karceski, and J. Lakonishok. 1998. The risk and return from factors. *Journal of Financial and Quantitative Analysis* 33:159–188.

Chen, R., and R. Kan. 2004. Finite sample analysis of two-pass cross-sectional regressions. Working paper, University of Toronto.

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

Connor, G., M. Hagmann, and O. Linton. 2012. Efficient semiparametric estimation of the Fama-French model and extensions. *Econometrica* 80:713–754.

Daniel, K., and S. Titman. 1997. Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance* 52:1–33.

DeMiguel, V., A. Martín-Utrera, F. J. Nogales, and R. Uppal. 2018. A transaction-cost perspective on the multitude of firm characteristics. Working paper, London Business School.

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

Fama, E. F., and K. R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–636.

Ferson, W. E., and C. R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* 99:385–415.

Gagliardini, P., E. Ossola, and O. Scaillet. 2016. Time-varying risk premium in large cross-sectional equity data sets. *Econometrica* 84:985–1046.

Gagliardini, P., E. Ossola, and O. Scaillet. 2018. A diagnostic criterion for approximate factor structure. Working paper, Swiss Finance Institute.

Gibbons, M. R., S. A. Ross, and J. Shanken. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57:1121–1152.

Giglio, S., and D. Xiu. 2017. Inference on risk premia in the presence of omitted factors. Working paper 23527, National Bureau of Economic Research.

Gospodinov, N., R. Kan, and C. Robotti. 2014. Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies* 27:2139–2170.

Gospodinov, N., R. Kan, and C. Robotti. 2017. Spurious inference in reduced-rank asset-pricing models. *Econometrica* 85:1613–1628.

Gospodinov, N., R. Kan, and C. Robotti. 2018. Too good to be true? Fallacies in evaluating risk factor models. *Journal of Financial Economics*, forthcoming.

Greene, W. H. 2003. *Econometric Analysis.* New Jersey: Pearson Education.

Gungor, S., and R. Luger. 2016. Multivariate tests of mean-variance efficiency and spanning with a large number of assets and time-varying covariances. *Journal of Business & Economic Statistics* 34:161–175.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. . . . and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.

Hou, K., G. A. Karolyi, and B. C. Kho. 2011. What factors drive global stock returns? *Review of Financial Studies* 24:2527–2574.

Hou, K., and R. Kimmel. 2006. On the estimation of risk premia in linear factor models. Working paper, Ohio State University.

Huang, D., J. Li, and G. Zhou. 2018. Shrinking factor dimension: A reduced-rank approach. Working paper, Washington University in St. Louis.

Ingersoll, J. E. 1984. Some results in the theory of arbitrage pricing. *Journal of Finance* 39:1021–1039.

Jagannathan, R., G. Skoulakis, and Z. Wang. 2010. The analysis of the cross-section of security returns. In Y. Aït-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics: Applications*, vol. 2 of *Handbooks in Finance*, pp. 73–134. San Diego: Elsevier.

Jagannathan, R., and Z. Wang. 1998. An asymptotic theory for estimating beta–pricing models using cross–sectional regression. *Journal of Finance* 53:1285–1309.

Jegadeesh, N., J. Noh, K. Pukthuanthong, R. Roll, and J. L. Wang. 2018. Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics*, forthcoming.

Kan, R., and C. Robotti. 2012. Evaluation of asset pricing models using two-pass cross-sectional regressions. In J. C. Duan, J. C. Gentle, and W. Hardle (eds.), *Handbook of Computational Finance*, Chapter 9, pp. 223–251. Berlin, Heidelberg: Springer.

Kan, R., C. Robotti, and J. Shanken. 2013. Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance* 68:2617–2649.

Kan, R., and C. Zhang. 1999a. GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics* 54:103–127.

Kan, R., and C. Zhang. 1999b. Two-pass tests of asset pricing models with useless factors. *Journal of Finance* 54:203–235.

Kelly, B. T., S. Pruitt, and Y. Su. 2018. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, forthcoming.

Kim, S., and G. Skoulakis. 2018. Ex-post risk premia estimation and asset pricing tests using large cross sections: The regression-calibration approach. *Journal of Econometrics* 204:159–188.

Kleibergen, F. 2009. Tests of risk premia in linear factor models. *Journal of Econometrics* 149:149–173.

Kleibergen, F., and Z. Zhan. 2018a. Asset pricing with consumption and robust inference. Working paper, University of Amsterdam.

Kleibergen, F., and Z. Zhan. 2018b. Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics* 16:155–190.

Kogan, L., and D. Papanikolaou. 2013. Firm characteristics and stock returns: The role of investment-specific shocks. *Review of Financial Studies* 26:2718–2759.

Kozak, S., S. Nagel, and S. Santosh. 2018. Shrinking the cross-section. *Journal of Financial Economics*, forthcoming.

Kuersteiner, G. M., and I. R. Prucha. 2013. Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics* 174:107–126.

Lamont, O. A. 2001. Economic tracking portfolios. *Journal of Econometrics* 105:161–184.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96:175–194.

Litzenberger, R. H., and K. Ramaswamy. 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics* 7:163–195.

Magnus, J., and R. H. Neudecker. 2007. *Matrix Differential Calculus with Applications in Statistics and Econometrics. Revised Edition.* Chicester (UK): J. Wiley & Sons.

Pástor, L., and R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111:642–685.

Pesaran, M. H., and T. Yamagata. 2012. Testing CAPM with a large number of assets. Working paper, Cambridge University.

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

Shanken, J. 1992. On the estimation of beta-pricing models. *Review of Financial Studies* 5:1–33.

Shanken, J. 1996. 23 Statistical methods in tests of portfolio efficiency: A synthesis. In G. Maddala and C. Rao (eds.), *Handbook of Statistics*, vol. 14, pp. 693–711. New York: Elsevier.

Shanken, J., and G. Zhou. 2007. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. *Journal of Financial Economics* 84:40–86.

Table 1
Percentage difference between estimated risk premia

Factor	$T = 36$	$T = 120$
	Panel A: CAPM (with liquidity)	
mkt	64.3%	27.2%
liq	41.3%	54.2%
	Panel B: FF3 (with liquidity)	
mkt	13.9%	7.3%
smb	14.7%	12.3%
hml	51.6%	31.2%
liq	22.9%	46.1%
	Panel C: FF5 (with liquidity)	
mkt	15.3%	11.1%
smb	13.2%	9.7%
hml	14.1%	15.2%
rmw	13.3%	15.2%
cma	43.3%	33.0%
liq	13.9%	38.7%

The table reports the percentage difference between the Shanken (1992) estimator, $\hat{\gamma}_1^*$, and the OLS CSR estimator, $\hat{\gamma}_1$, averaged over rolling windows of size $T = 36$ and $T = 120$, respectively. The three panels refer to the CAPM, Fama and French (1993) three-factor model (FF3), and Fama and French (2015) five-factor model (FF5). Each of these models has been augmented with the non-traded liquidity factor of Pástor and Stambaugh (2003). We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's and Lubos Pástor's websites from January 1966 to December 2013.

Table 2
Betas versus Characteristics

	CAPM	FF3	FF5
Panel A: F-tests and rejection frequencies			
	$H_0: \gamma_1^P = 0_K$		
F-tests	14.54	17.33	21.14
Rejection frequencies	25.84%	28.72%	29.91%
	$H_0: \delta = 0_{K_c}$		
F-tests	888.27	960.01	927.04
Rejection frequencies	100%	100%	100%
Panel B: Variance ratios			
$100 \times \frac{S^2_{\bar{R}_C}}{S^2_{\bar{R}}}$	73.84%	76.36%	76.70%
$100 \times \frac{S^2_{\bar{R}_{\perp C}}}{S^2_{\bar{R}}}$	2.21%	3.11%	3.19%

The top panel of the table reports the F-tests (average over rolling windows of size $T = 36$) for the null hypotheses $H_0: \gamma_1^P = 0_K$ and $H_0: \delta = 0_{K_c}$, respectively, and the rejection frequencies at the 95% confidence level (average over rolling windows of size $T = 36$). Each column refers to a different beta-pricing model, that is, the CAPM (first column), the Fama and French (1993) three-factor model (FF3, second column), and the Fama and French (2015) five-factor model (FF5, third column). The bottom panel reports the variance ratios $100 \times S^2_{\bar{R}_C}/S^2_{\bar{R}}$ and $100 \times S^2_{\bar{R}_{\perp C}}/S^2_{\bar{R}}$ defined in Section 4.4 (average over rolling windows of size $T = 36$). The data is from DeMiguel et al. (2018) and Kenneth French's website (from January 1980 to December 2015).

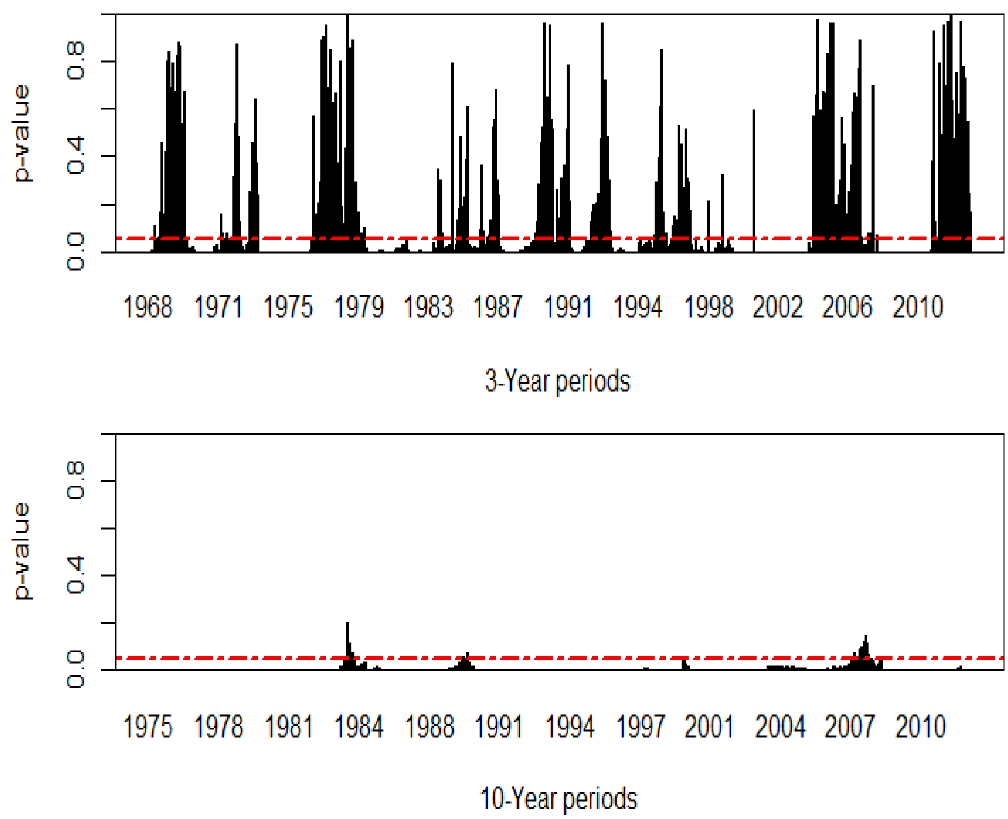


Figure 1
Specification testing for the Fama and French (2015) five-factor model
The figure presents the time series of p-values (black line) of $\mathcal{S}^*$ for FF5. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's website from January 1966 to December 2013.

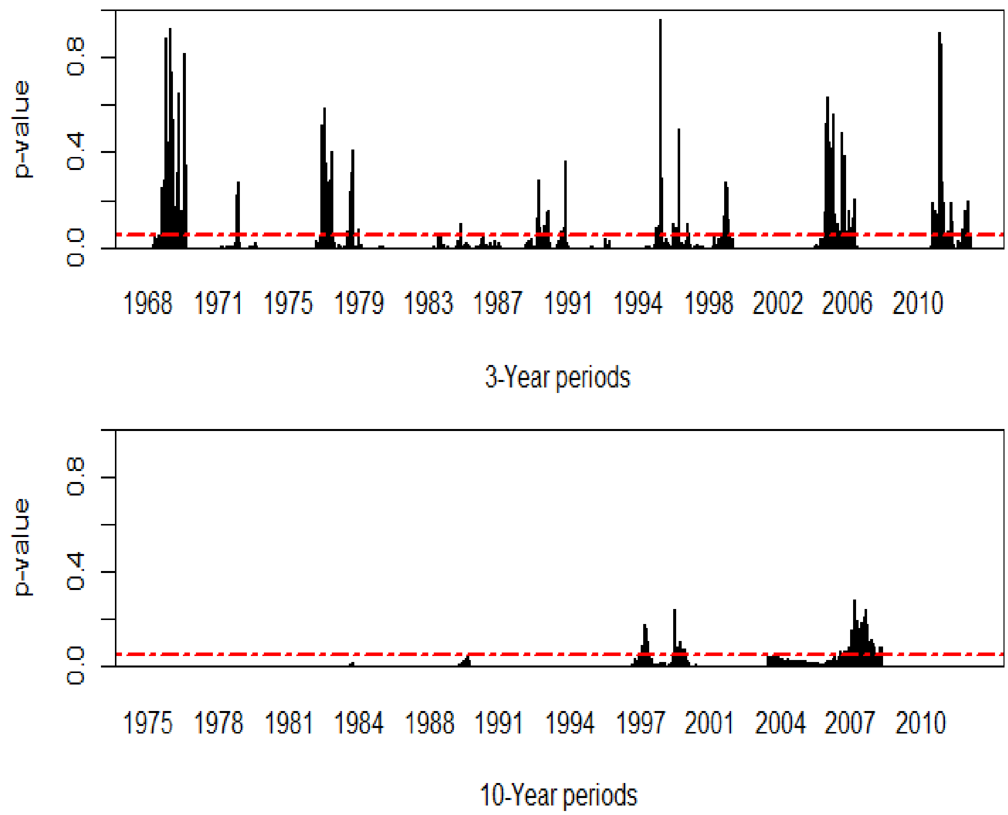


Figure 2
Specification testing for the liquidity-augmented Fama and French (2015) five-factor model

The figure presents the time series of p-values (black line) of $\mathcal{S}^*$ for the liquidity-augmented FF5. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the test. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's and Lubos Pástor's websites from January 1966 to December 2013.

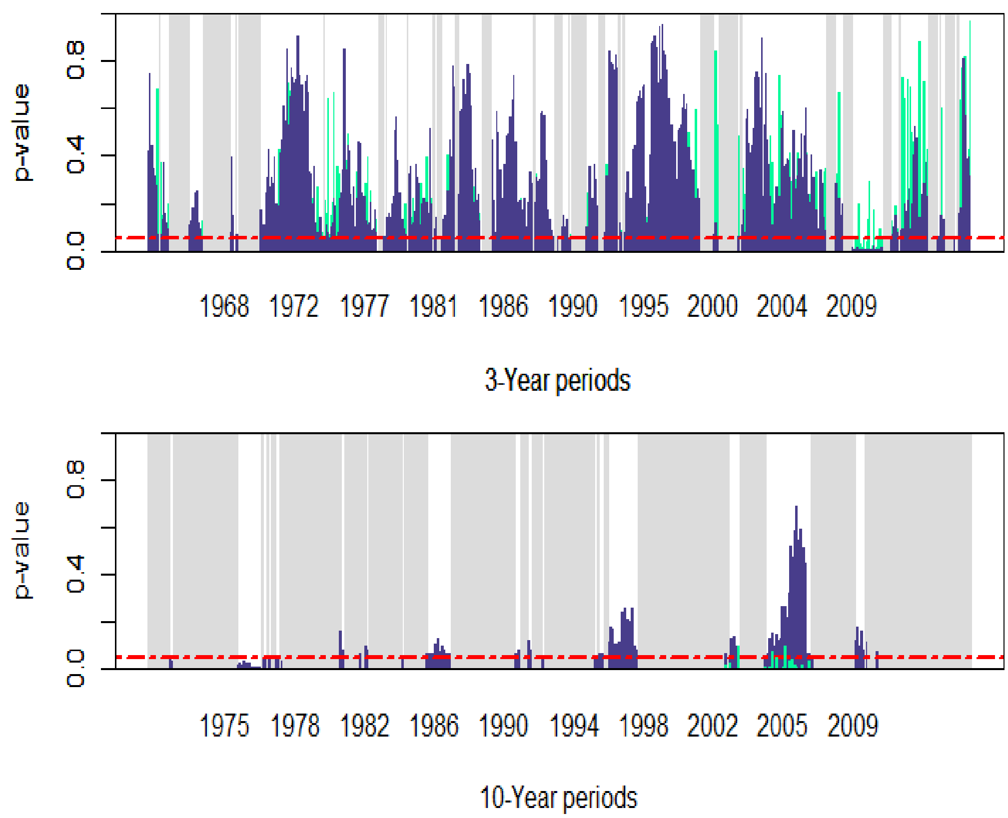


Figure 3
Specification testing for the Fama and French (2015) five-factor model using the Gibbons et al. (1989) and Gungor and Luger (2016) tests

The figure presents the time series of p-values of the GRS (blue line) and GL (green line) tests for FF5. Rolling time windows of three (top panel) and 10 years (bottom panel) are used. The dashed dotted red line denotes the 5% significance level of the tests. The grey bars are for the periods in which the GL test is inconclusive. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's website from January 1966 to December 2013.

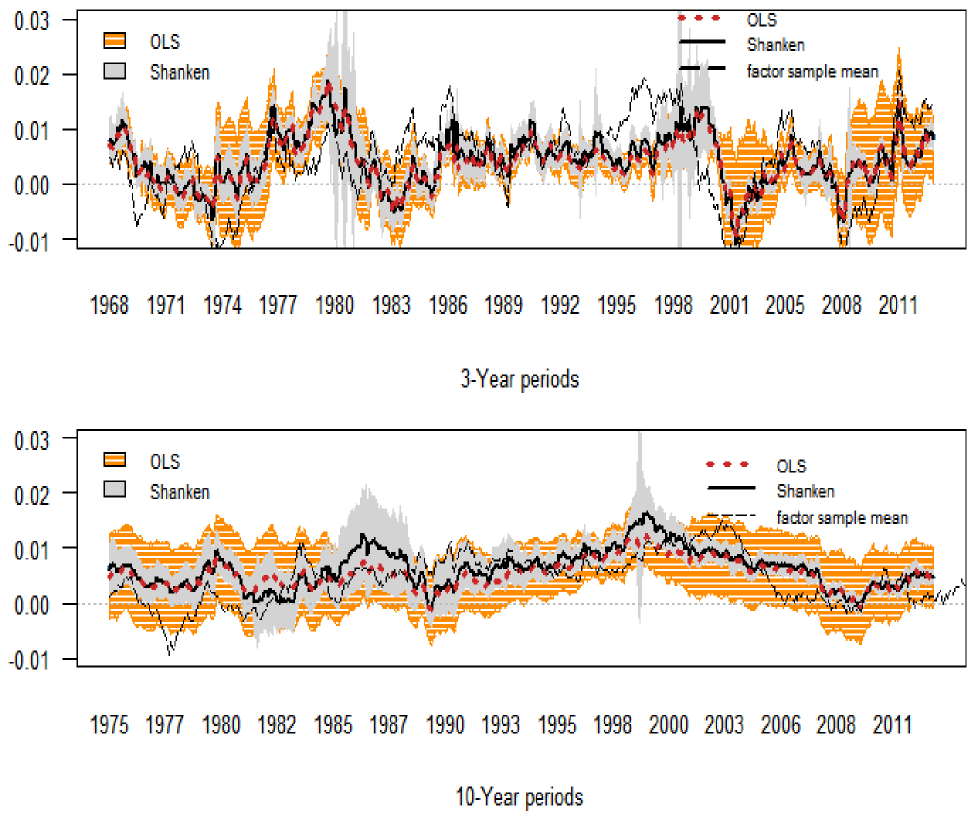


Figure 4
Estimates and confidence intervals for the market risk premium
The figure presents the estimates and the associated confidence intervals for the market risk premium from the Fama and French (2015) five-factor model. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-N standard errors of Theorem 5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-T standard errors. Finally, the dashed black line is for the rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's website from January 1966 to December 2013.

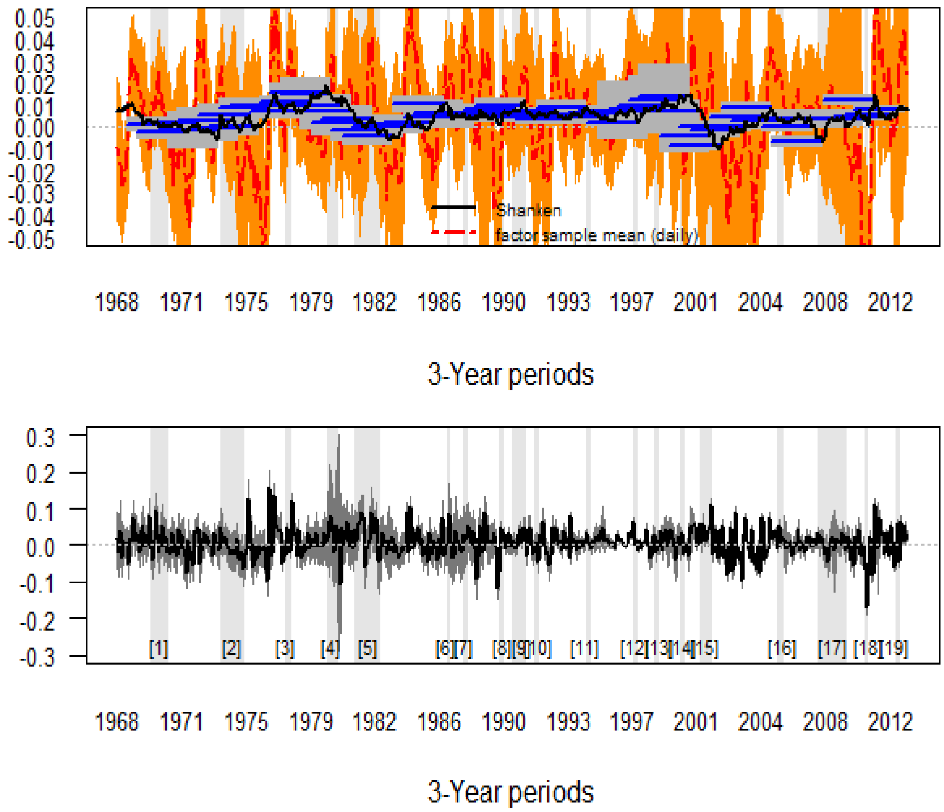


Figure 5
Estimates and confidence intervals for the time-varying market risk premium
The figure presents the estimates and the associated confidence intervals for the time-varying market risk premium from the Fama and French (2015) five-factor model based on our large-N methodology. The top panel reports the Shanken (1992) large-N estimates, expressed in terms of a single line (black line) and in terms of horizontal bars of length $T = 36$ observations (blue line), with the corresponding 95% confidence intervals based on the large-N standard errors of Theorem 5 (grey band). We also report the rolling sample mean (using fixed rolling windows of six months of daily data) of the market excess return (dashed dotted red line) and the corresponding 95% confidence interval (orange band). The bottom panel reports the modified Shanken (1992) estimator (black line) and the corresponding 95% confidence interval (grey band) based on the large-N standard errors of part (ii) of Theorem 3. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's website from January 1966 to December 2013. The daily data on the market excess return is from Kenneth French's website. The light grey bands correspond to the NBER recession dates and to various economic and financial crises. They are numbered as follows: [1] 1969:10-1970:11, [2] 1973:11-1975:3, [3] 1977:8-1977:11, [4] 1980:1-1980:7, [5] 1981:7-1982:11, [6] 1986:10-1986:12, [7] 1987:9-1987:11, [8] 1989:9-1989:12, [9] 1990:7-1991:3, [10] 1991:8-1992:12, [11] 1994:7-1994:10, [12] 1997:5-1997:9, [13] 1998:8-1998:10, [14] 2000:2-2000:4, [15] 2001:3-2001:11, [16] 2005:8-2005:11, [17] 2007:12-2009:6, [18] 2010:8-2010:10, [19] 2012:5-2012:7.

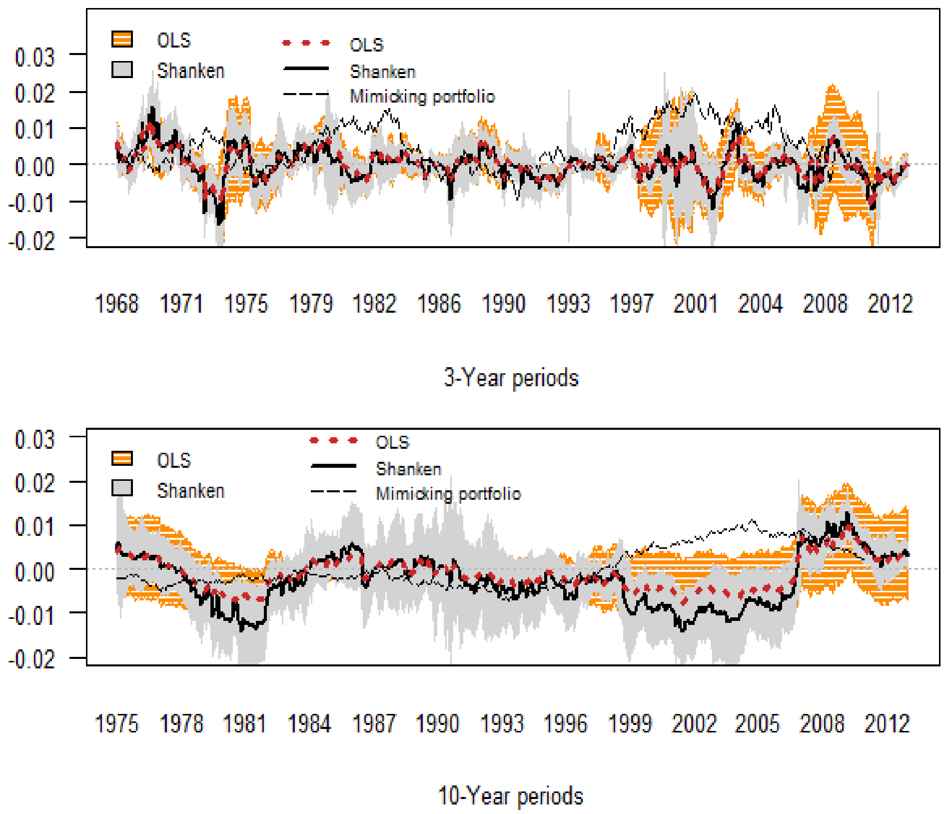


Figure 6
Estimates and confidence intervals for the liquidity risk premium

The figure presents the estimates and the associated confidence intervals for the liquidity risk premium from the liquidity-augmented Fama and French (2015) five-factor model. The bold black line is for the Shanken (1992) estimator. The corresponding grey band represents the 95% confidence intervals based on the large-N standard errors of Theorem 5. We also report the OLS CSR estimator (dotted red line) and the corresponding 95% confidence interval (striped orange band) based on the traditional large-T standard errors. Finally, the dashed black line is for the mimicking portfolio rolling factor sample mean. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's and Lubos Pástor's websites from January 1966 to December 2013.

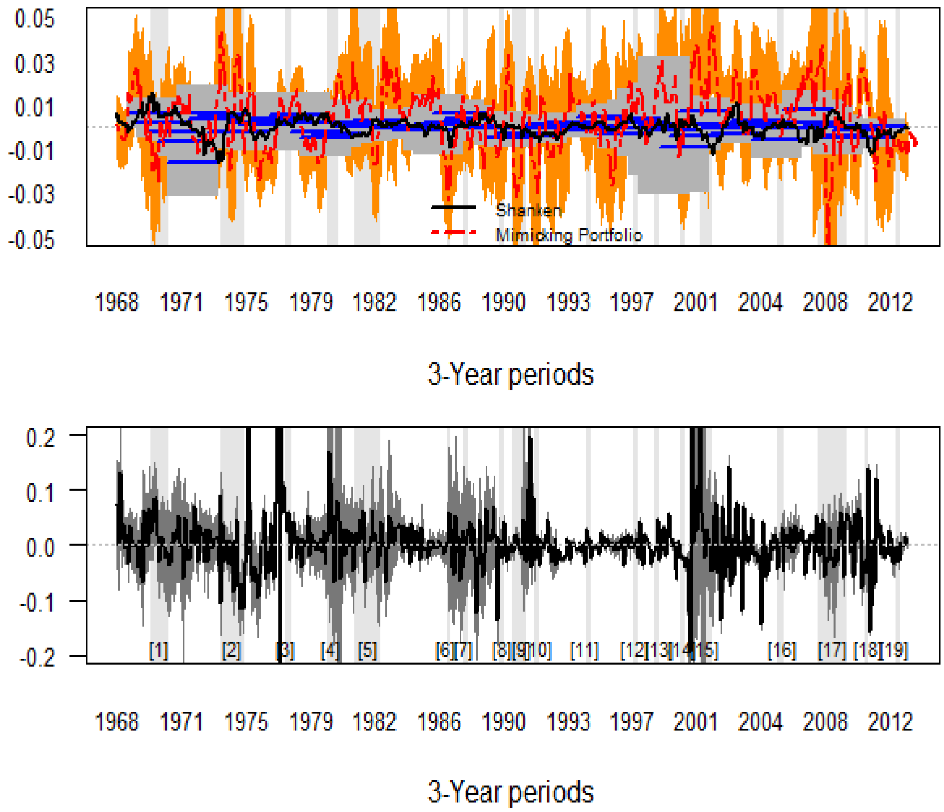
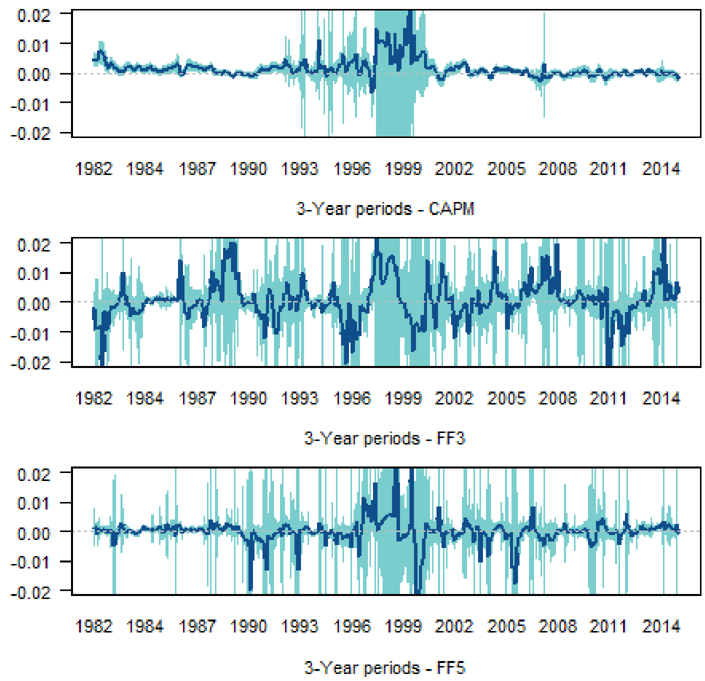
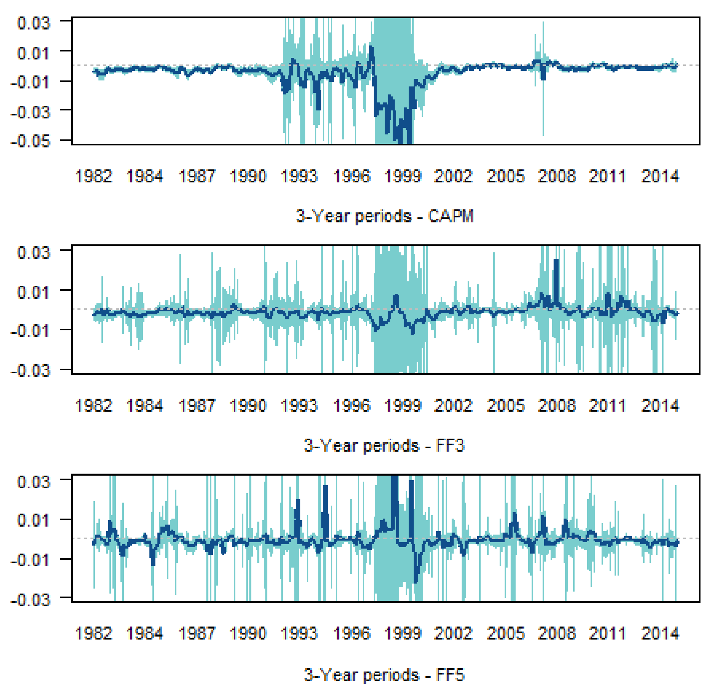


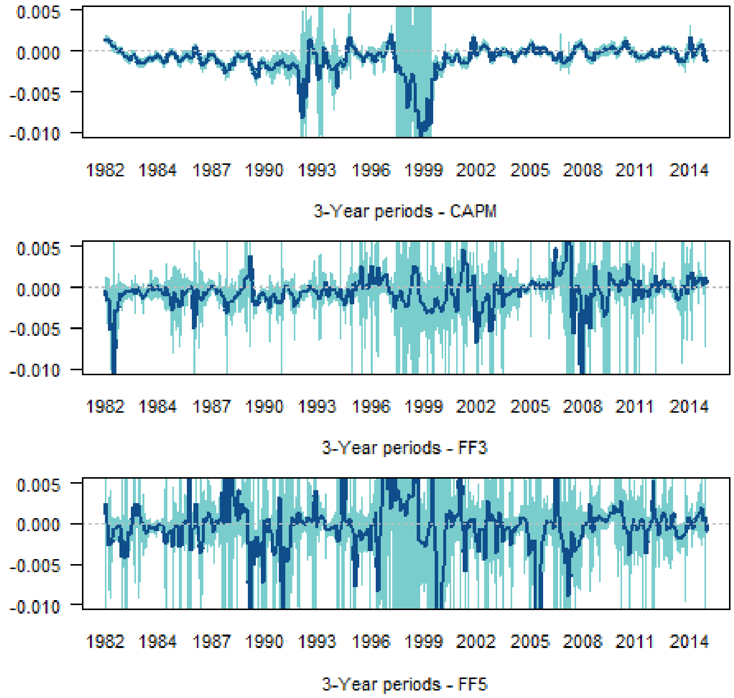
Figure 7
Estimates and confidence intervals for the time-varying liquidity risk premium
The figure presents the estimates and the associated confidence intervals for the time-varying liquidity risk premium from the liquidity-augmented Fama and French (2015) five-factor model based on our large-N methodology. The top panel reports the Shanken (1992) large-N estimates, expressed in terms of a single line (black line) and in terms of horizontal bars of length $T = 36$ observations (blue line), with the corresponding 95% confidence intervals based on the large-N standard errors of Theorem 5 (grey band). We also report the rolling sample mean (using fixed rolling windows of six months) of the corresponding mimicking portfolio excess return (dashed dotted red line) and the corresponding 95% confidence interval (orange band). The bottom panel reports the modified Shanken (1992) estimator (black line) and the corresponding 95% confidence interval (grey band) based on the large-N standard errors of part (ii) of Theorem 3. We use monthly return data on individual stocks from CRSP and factor data from Kenneth French's and Lubos Pástor's websites from January 1966 to December 2013. The light grey bands correspond to the NBER recession dates and to various economic and financial crises. They are numbered as follows: [1] 1969:10-1970:11, [2] 1973:11-1975:3, [3] 1977:8-1977:11, [4] 1980:1-1980:7, [5] 1981:7-1982:11, [6] 1986:10-1986:12, [7] 1987:9-1987:11, [8] 1989:9-1989:12, [9] 1990:7-1991:3, [10] 1991:8-1992:12, [11] 1994:7-1994:10, [12] 1997:5-1997:9, [13] 1998:8-1998:10, [14] 2000:2-2000:4, [15] 2001:3-2001:11, [16] 2005:8-2005:11, [17] 2007:12-2009:6, [18] 2010:8-2010:10, [19] 2012:5-2012:7.


0.02
0.01
0.00
-0.01
-0.02
1982 1984 1987 1990 1993 1996 1999 2002 2005 2008 2011 2014
3-Year periods - CAPM
3-Year periods - FF3
3-Year periods - FF5

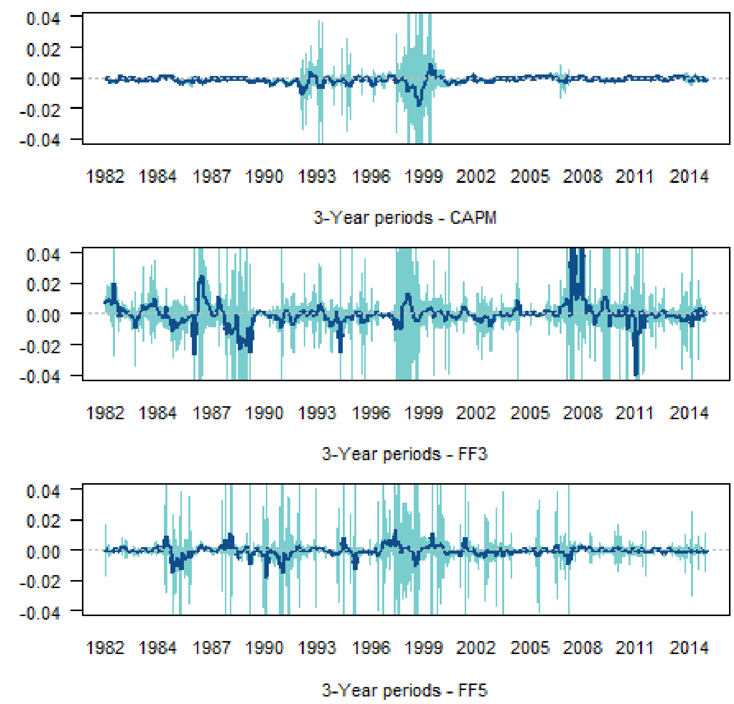
B/M


0.03
0.01
-0.01
-0.03
-0.05
1982 1984 1987 1990 1993 1996 1999 2002 2005 2008 2011 2014
3-Year periods - CAPM
3-Year periods - FF3
3-Year periods - FF5

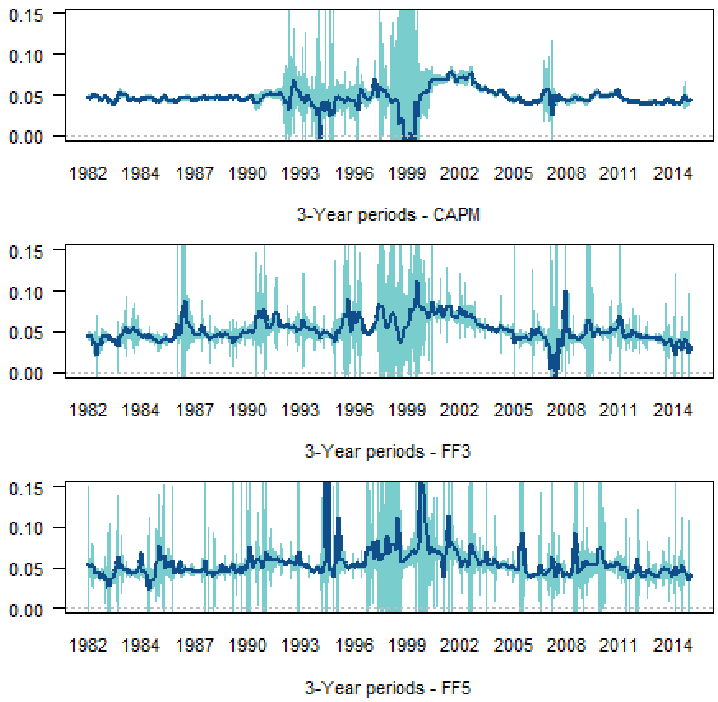
ASSGR



OPERPROF



MCAPIT



MOM6

Figure 8
Estimates and confidence intervals for the characteristic premia

The figure presents estimates (blue line) of the characteristic premia on the book-to-market ratio, *B/M*, asset growth, *ASSGR*, operating profitability, *OPERPROF*, market capitalization, *MCAPIT*, and six-month momentum, *MOM6*, and the associated confidence intervals based on Theorem 7 (light blue band), for the CAPM, the Fama and French (FF3, 1993) three-factor model, and the Fama and French (FF5, 2015) five-factor model. The data is from DeMiguel et al. (2018) and Kenneth French's website (from January 1980 to December 2015).