Spurious Inference in Reduced-Rank Asset-Pricing Models

NIKOLAY GOSPODINOV, RAYMOND KAN, and CESARE ROBOTTI*

July 2016

*Gospodinov is from the Federal Reserve Bank of Atlanta. Kan is from the University of Toronto. Robotti is from Imperial College London. We thank Seung Ahn, Alex Horenstein, Lei Jiang, Robert Kimmel, Francisco Peñaranda, Enrique Sentana, Chu Zhang, seminar participants at CEIS Tor Vergata, EDHEC Business School, National University of Singapore, University of Cantabria, University of Geneva, University of Southampton, and Western University, as well as conference participants at the 2013 All-Georgia Finance Conference, the 2013 Metro-Atlanta Econometric Study Group Meeting, the 2013 Seventh International Conference on Computational and Financial Econometrics, the 2014 SoFiE Conference, the 2014 Tsinghua Finance Workshop, the 2014 NFA Conference, the 2014 NBER-NSF Time Series Conference, and the 2014 China International Conference in Finance for helpful discussions and comments. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada and the National Bank Financial of Canada. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Corresponding author: Nikolay Gospodinov, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street N.E., Atlanta, GA 30309, USA; E-mail: `Nikolay.Gospodinov@atl.frb.org`.

Spurious Inference in Reduced-Rank Asset-Pricing Models

Abstract

This paper studies some seemingly anomalous results that arise in possibly misspecified and unidentified linear asset-pricing models estimated by maximum likelihood and one-step generalized method of moments. Strikingly, when useless factors (that is, factors that are independent of the returns on the test assets) are present, the models exhibit perfect fit, as measured by the squared correlation between the model's fitted expected returns and the average realized returns, and the tests for correct model specification have asymptotic power that is equal to the nominal size. In other words, applied researchers will erroneously conclude that the model is correctly specified even when the degree of misspecification is arbitrarily large. We also derive the highly non-standard limiting behavior of these invariant estimators and their t-tests in the presence of identification failure. These results reveal the spurious nature of inference as factors that are useless are selected with high probability, while factors that are useful are driven out from the model. The practical relevance of our findings is demonstrated using simulations and an empirical application.

Keywords: Asset pricing; Irrelevant risk factors; Unidentified models; Model misspecification; Continuously-updated GMM; Maximum likelihood; Rank test; Test for over-identifying restrictions.

JEL classification numbers: C12; C13; G12.

1 Introduction and Motivation

The search for (theoretically justified or empirically motivated) risk factors that improve the pricing performance of various asset-pricing models has generated a large, and constantly growing, literature in financial economics. A typical empirical strategy involves the development of a structural asset-pricing model and the evaluation of the pricing ability of the proposed factors in the linearized version of the model using actual data. The resulting linear asset-pricing model can be estimated and tested using a beta representation or, alternatively, using a linear stochastic discount factor (SDF) representation. Given the appealing efficiency and invariance properties of the maximum likelihood (ML) and continuously-updated generalized method of moments (CU-GMM) estimators,[1] it seems natural to opt for these estimators when conducting statistical inference (estimation, testing, and model evaluation) in these linear asset-pricing models. It is often the case that a high correlation between the realized and fitted expected returns (in the beta representation framework) or statistically small model pricing errors (in the SDF framework) appear to be sufficient for the applied researcher to conclude that the model is well specified and proceed with testing for statistical significance of the risk premium parameters using the standard tools for inference. Many asset-pricing studies have followed this empirical strategy and collectively, the literature has identified a large set of macroeconomic and financial factors (see Harvey, Liu, and Zhu, 2013) that are believed to explain the cross-sectional variation of various portfolio returns, such as the returns on the 25 Fama-French size and book-to-market ranked portfolios.

Despite these advances in the asset-pricing literature, two observations that consistently emerge in empirical work might call for a more cautious approach to statistical validation and economic interpretation of asset-pricing models. First, all asset-pricing models should be viewed only as approximations to reality and, hence, potentially misspecified. There is plenty of empirical evidence, mainly based on non-invariant estimators, which suggests that the asset-pricing models used in practice are misspecified. This raises the concern of using standard errors, derived under the assumption of correct model specification, that tend to underestimate the degree of uncertainty that the researcher faces. Second, the macroeconomic factors in several asset-pricing specifications tend to be only weakly correlated with the portfolio returns. As a result, it is plausible to conjecture

[1] See Shanken and Zhou (2007) and Peñaranda and Sentana (2014) for some recent results on the ML and CU-GMM estimators, respectively, for asset-pricing models.

that many of these macroeconomic factors may be irrelevant for pricing and explaining the cross-sectional variation of stock returns. Importantly, the inclusion of useless factors (that is, factors that are independent of the returns on the test assets) leads to serious identification issues regarding the parameters associated with all risk factors and gives rise to a non-standard statistical inference. In a recent paper, Gospodinov, Kan, and Robotti (2014a) analyze the detrimental effects of misspecification and factor irrelevance (lack of identification) in estimation, testing, and evaluation of asset-pricing models using the Hansen and Jagannathan (HJ, 1997) distance. In this study, we show that the use of optimal and invariant estimators does not alleviate these inference problems and, somewhat surprisingly, makes them substantially worse.

The invariant (ML and CU-GMM) estimators considered here have a generic form

$$\hat{\theta} = \text{argmin}_{\theta}\ \bar{g}(\theta)'\hat{W}(\theta)^{-1}\bar{g}(\theta), \tag{1}$$

with model-implied choices of moment conditions $\bar{g}(\theta)$, a weighting matrix $\hat{W}(\theta)$, and a parameter vector of interest θ. This objective function makes the resulting estimator invariant to data scaling, reparameterizations and normalizations, curvature-altering and stationarity-inducing transformations, etc. (Hall, 2005). Under standard regularity conditions (that include global and local identification as well as correct model specification), these invariant estimators are asymptotically well-behaved and efficient. However, we show in this paper that in the presence of lack of identification and model misspecification, the tests based on these estimators could be highly misleading. In summary, we argue that the standard inference procedures based on the ML and CU-GMM estimators lead to spurious results that suggest that the model is correctly specified and the risk premium parameters are highly significant (that is, the risk factors are priced) when, in fact, the model is misspecified and the factors are irrelevant. The distorted nature of these results bears strong similarities to spurious regressions with nonstationary data (Granger and Newbold, 1974, among many others). Phillips (1989) makes an analogous observation regarding the estimators in partially identified (albeit correctly specified) linear structural models and time series spurious regressions. Phillips (1989, p. 201) points out that "both regressions share a fundamental indeterminacy" due to a contaminated signal arising from either lack of identification or strength of the noise component. We show that allowing for model misspecification further exacerbates the spuriousness of the results and renders them completely unreliable.

To illustrate the seriousness of the problem, we report some representative empirical results from four popular asset-pricing models. The test asset returns are the monthly gross returns on the widely used value-weighted 25 Fama-French size and book-to-market ranked portfolios from February 1959 until December 2012.[2] The first model is the static capital asset-pricing model (CAPM) with the market return (the return on the value-weighted NYSE-AMEX-NASDAQ stock market index in excess of the one-month T-bill rate, vw) as a risk factor. The second model is the three-factor model (FF3) of Fama and French (1993) with (i) the market excess return (vw), (ii) the return difference between portfolios of stocks with small and large market capitalizations (smb), and (iii) the return difference between portfolios of stocks with high and low book-to-market ratios (hml) as risk factors. It should be noted that all of these risk factors are traded and exhibit a relatively high correlation with the 25 Fama-French portfolio returns. The last two models are models with macroeconomic factors: the model (C-LAB) proposed by Jagannathan and Wang (1996) which, in addition to the market excess return, includes the growth rate in per capita labor income ($labor$) and the lagged default premium ($prem$, the yield spread between Baa and Aaa-rated corporate bonds) as risk factors; and the model (CC-CAY) proposed by Lettau and Ludvigson (2001) with risk factors that include the growth rate in real per capita nondurable consumption (cg), the lagged consumption-aggregate wealth ratio (cay), and an interaction term between these two factors ($cg \cdot cay$).

Table I about here

The SDF and beta representations of the four asset-pricing models are estimated by CU-GMM and ML, respectively. Table I reports results from the invariant tests of correct model specification (Hansen, Heaton, and Yaron's (1996) over-identifying restrictions test, $\mathcal{J}$, for CU-GMM and Shanken's (1985) Wald-type test, $\mathcal{S}$, for ML), the t-statistics for each factor computed using standard errors that assume correct model specification, as well as the pseudo-R^2s from regressing fitted expected returns on average returns. In addition, we include rank tests to determine whether the asset-pricing models are properly identified,[3] and two popular specification tests based on non-

[2]The results that we report in this section are largely unchanged when we augment the 25 Fama-French portfolio returns with additional test asset returns (for example, the 17 Fama-French industry portfolio returns) as recommended by Lewellen, Nagel, and Shanken (2010).

[3]In the SDF framework, we use the rank test of Cragg and Donald (1997) to assess whether the second moment matrix of the returns and the factors is of reduced rank. In the beta-pricing framework, we employ the rank test of Cragg and Donald (1997) to test whether the matrix of multivariate betas has a reduced rank. The details of the Cragg and Donald (1997) test in the beta-pricing framework can be found in Kan and Robotti (2012).

invariant estimators (the HJ-distance test and the generalized least squares (GLS) cross-sectional regression test of Shanken, 1985). Figures 1 and 2 visualize the cross-sectional goodness-of-fit of the models by plotting average realized returns versus fitted (by CU-GMM and ML, respectively) expected returns from each model.

Figures 1 and 2 about here

The results are striking. The models that contain factors that are only weakly correlated with the test asset returns (C-LAB and CC-CAY) exhibit an almost perfect fit. The specification tests based on the invariant estimators cannot reject the null of correct specification, which suggests that the models are well specified and one could proceed with constructing significance tests based on standard errors derived under correct model specification. These t-tests indicate that the proposed macroeconomic factors (labor growth and default premium in C-LAB and the interaction term in CC-CAY, for example) are highly statistically significant. Interestingly, benchmark models such as CAPM and FF3 do not perform nearly as well according to these statistical measures. The tests for correct model specification based on CU-GMM and ML suggest that both of these models are rejected by the data, and their associated pseudo-R^2s are 0.1999 and 0.7847 for CU-GMM, and 0.1346 and 0.7677 for ML, respectively.

In this paper, we show that, due to the combined effect of identification failure and model misspecification, the results for C-LAB and CC-CAY can be spurious. While some warning signs of these problems are already present in Table I, they are often ignored by applied researchers. For example, the rank tests provide strong evidence that C-LAB and CC-CAY are not identified, which violates the regularity conditions for consistency and asymptotic normality of the ML and CU-GMM estimators. Furthermore, the HJ-distance and GLS cross-sectional regression tests, that we show to possess much higher power than the $\mathcal{J}$ and $\mathcal{S}$ tests in the presence of identification failure, point to severe misspecification of all the considered asset-pricing models.

Another interesting observation that emerges from these results is that the factors with low correlations with the returns tend to drive out the factors that are highly correlated with the returns. For example, the highly significant market factor in CAPM turns insignificant with the inclusion of the labor growth and default premium in the C-LAB model. To further examine this point, we simulate data for the returns on the test assets and the market factor from a misspecified model that is calibrated to the CAPM as estimated in Table I (for more details on the simulation design, see Section 5 below). With a sample size of 600 time series observations, the rejection rate (at the 5% significance level) of the t-test (based on the CU-GMM estimator) of whether the market factor is priced or not is 93.6%, while the test for correct model specification rejects the null of correct specification 100% of the time. In sharp contrast, when a completely useless factor (generated as an independent standard normal random variable) is included in the model, the rejection rate of the t-test for the market factor drops to 9.5% and the specification test rejects the null of correct specification only 4% of the time. Strikingly, the rejection rate of the t-test for the useless factor is 100%. This example clearly illustrates the severity of the problem and the perils for inference based on invariant tests in unidentified models. In summary, an arbitrarily poor model with factors that are independent of the test asset returns would be deemed to be correctly specified with a spectacular fit and priced risk factors.

In addition to identifying a serious problem with invariant tests of asset-pricing models, our paper also provides a number of theoretical contributions. First, we demonstrate the numerical equivalence of the invariant (ML and CU-GMM) and rank restriction frameworks for estimation and model specification testing. This equivalence proves to be useful from both a computational and a statistical inference perspective. For instance, it allows us to show that, under model misspecification and rank deficiency, the specification tests have power that is equal to their size. Second, we characterize the limiting behavior of the invariant estimators and their t-statistics under model misspecification and identification failure. While we show that all estimators are inconsistent and asymptotically non-normal, the estimates associated with the factors that cause the rank deficiency diverge at rate root-T and the t-tests have a bimodal and heavy-tailed distribution. The explosive behavior of the estimates on the useless factors tends to dominate and forces the goodness-of-fit statistic to approach one.

Finally, it is useful to position our results in the existing literature. In the statistics and econometrics literature, the analysis of rank restrictions, model under-identification, and inference under model misspecification has generated substantial interest since the seminal work of Anderson (1951), Koopmans and Hood (1953), and Maasoumi and Phillips (1982), respectively. We contribute to this literature by developing the appropriate limiting theory for invariant estimators and reduced rank tests in unidentified and possibly misspecified models. On the other hand, some of the recent asset-pricing studies have also expressed concerns about the appropriateness of the pseudo-R^2 as a reliable goodness-of-fit measure. In models with excess returns and under some particular normalizations of the SDF, Burnside (2012) derives a similar behavior of the goodness-of-fit statistic for non-invariant GMM estimators. This result, however, is normalization and setup specific and alternative normalizations or models based on gross returns render the non-invariant estimators immune to the perfect fit problem. Furthermore, Kleibergen and Zhan (2013) show that a sizeable unexplained factor structure (generated by a low correlation between the observed proxy factors and the true unobserved factors) in a two-pass cross-sectional regression framework can also produce spuriously large values of the ordinary least squares (OLS) R^2 coefficient. Their results complement the findings of Lewellen, Nagel, and Shanken (2010) who criticize the use of the OLS R^2 coefficient by showing that it provides an overly positive assessment of the performance of the asset-pricing model. Despite the suggestive nature of these findings, model evaluation tests based on non-invariant estimators, which are the focus of the analysis in these studies, tend to be relatively robust to lack of identification as we show later in the paper. In contrast, for invariant estimators in unidentified asset-pricing models, the spurious perfect fit is pervasive regardless of the model structure (gross or excess returns), estimation framework (SDF or beta pricing), and chosen normalization.

The rest of the paper is organized as follows. Section 2 introduces the main notation and assumptions. Section 3 studies the limiting behavior of the parameter estimates, t-statistics, goodness-of-fit measures, and model specification tests in the beta-pricing and SDF setups. Section 4 presents results for non-invariant estimators that allow for some comparisons with the limiting behavior of the ML and CU-GMM estimators. Section 5 reports Monte Carlo simulation results. Section 6 summarizes our main conclusions and provides some practical recommendations. All proofs are relegated to the Appendix. Some supplementary results, that we refer to throughout the paper, are available in an Internet Appendix.

The paper adopts the following notation. We denote convergence in probability by $\xrightarrow{p}$ and convergence in distribution by $\xrightarrow{d}$. In addition, let $Z = (Z_1, \ldots, Z_n)'$ be a vector of n independent standard normal random variables, and let $\xi = (\xi_1, \ldots, \xi_n)'$ be a vector of n real numbers. Then, $F_n(\xi) = \sum_{i=1}^{n} \xi_i Z_i^2$ denotes a random variable which is distributed as a weighted sum of n independent chi-squared random variables with one degree of freedom.

2 Preliminaries

In this section, we first introduce the SDF and beta representations of an asset-pricing model. Next, we briefly describe the rank test of Cragg and Donald (1997) and lay out our main assumptions.

2.1 Stochastic Discount Factor and Beta-Pricing Model Representations

Let

$$y_t(\lambda) = x_t'\lambda \tag{2}$$

be a candidate SDF at time t, where $x_t = [1,\ f_t']'$, f_t is a $(K-1)$-vector of systematic risk factors, and $\lambda = [\lambda_0,\ \lambda_1']'$ is a K-vector of SDF parameters. Also, let R_t denote the gross returns on N $(N > K)$ test assets and $e_t(\lambda) = D_t\lambda - 1_N$, where $D_t = R_t x_t'$ and 1_N is an $N \times 1$ vector of ones.[4] When the asset-pricing model is correctly specified and well identified, there exists a unique $\lambda^* = [\lambda_0^*,\ \lambda_1^{*\prime}]'$ such that the pricing errors of the model are zero, that is,

$$E[e_t(\lambda^*)] = D\lambda^* - 1_N = 0_N, \tag{3}$$

where $D = E[R_t x_t']$.

Alternatively, we can express the linear asset-pricing model using the beta representation. Let $Y_t = [f_t',\ R_t']'$ with

$$E[Y_t] \equiv \begin{bmatrix} \mu_f \\ \mu_R \end{bmatrix} \tag{4}$$

and

$$\mathrm{Var}[Y_t] \equiv V = \begin{bmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{bmatrix}, \tag{5}$$

and $\gamma = [\gamma_0,\ \gamma_1']'$ be a K-vector of parameters. When the asset-pricing model is correctly specified and well identified, there exists a unique $\gamma^* = [\gamma_0^*,\ \gamma_1^{*\prime}]'$ such that

$$\mu_R = 1_N\gamma_0^* + \beta\gamma_1^*, \tag{6}$$

where $\beta = [\beta_1, \ldots, \beta_{K-1}] = V_{Rf}V_f^{-1}$ is an $N \times (K-1)$ matrix of the betas of the N assets. Also, define

$$\alpha = \mu_R - \beta\mu_f, \tag{7}$$

[4]When R_t is a vector of payoffs with initial cost $q \neq 0_N$, we just need to replace 1_N with q. In addition, the analysis in the paper can be easily adapted to handle the case of excess returns, that is, the $q = 0_N$ case.

and $\Sigma = V_R - V_{Rf}V_f^{-1}V_{fR}$.

There are two main reasons why the beta-pricing framework is very popular in the empirical asset-pricing literature. First, unlike the SDF coefficients λ, the parameters γ_0 and γ_1 have a direct interpretation of zero-beta rate and risk premium parameters, respectively. Second, the beta representation allows for conveniently measuring and plotting the goodness-of-fit as a model's expected returns versus average realized returns. To capitalize on these advantages, the SDF parameters can be transformed into the beta-pricing model parameters using the mapping

$$\gamma_0 = \frac{1}{\lambda_0 + \mu_f'\lambda_1}, \tag{8}$$

$$\gamma_1 = -\frac{V_f\lambda_1}{\lambda_0 + \mu_f'\lambda_1}. \tag{9}$$

The main statistics of interest in evaluating asset-pricing models are the t-tests for statistical significance of λ_1 and γ_1,[5] the goodness-of-fit statistic defined as the squared correlation between the realized and model-implied expected returns, and the statistics for correct model specification that test the validity of the asset-pricing model restrictions: $D\lambda = 1_N$ in the SDF representation, and $\mu_R = 1_N\gamma_0 + \beta\gamma_1$ in the beta-pricing setup. The limiting behavior of these statistics, which is the primary focus of our analysis below, is determined by the rank of the matrices $H \equiv [1_N,\ D]$ (in the SDF representation) and $G \equiv [1_N,\ B]$, where $B = [\alpha,\ \beta]$ (in the beta-pricing representation).

2.2 Rank Restriction Test and Assumptions

In the subsequent analysis, we rely repeatedly on the representation of the rank restriction test of Cragg and Donald (1997) as an invariant test. Let Π be a generic notation for an $N \times K$ matrix and $\pi = \text{vec}(\Pi)$. Under the null that Π is of (reduced) rank $K-1$, $H_0 : \text{rank}(\Pi) = K-1$, there exists a nonzero K vector c such that $\Pi c = 0_N$ with the normalization $c'c = 1$.[6] Suppose $\hat{\Pi}$ is an estimator of Π and assume that

$$\sqrt{T}\text{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} \mathcal{N}(0_{NK}, M), \tag{10}$$

[5] It should be stressed that in a multi-factor model, testing $H_0 : \gamma_{1,i} = 0$ is not the same as testing $H_0 : \lambda_{1,i} = 0$ for $i = 1, \ldots, K-1$. More importantly, acceptance or rejection of $\gamma_{1,i} = 0$ does not tell us whether the i-th factor makes an incremental contribution to the model's overall explanatory power, given the presence of the other factors. See Kan, Robotti, and Shanken (2013) for a discussion of this subtle point.

[6] See Cragg and Donald (1997), Robin and Smith (2000), and Kleibergen and Paap (2006) for a detailed analysis of rank restriction tests.

where M is a finite and positive-definite matrix. Then, the Cragg and Donald (1997) test of H_0 : rank$(\Pi) = K-1$ can be rewritten as an invariant test of the form (see, for example, Kleibergen and Mavroeidis, 2009, and Arellano, Hansen, and Sentana, 2012)

$$\mathcal{CD} = \min_{c:c'c=1} T\hat{\pi}'Q(c)'(Q(c)\hat{M}Q(c)')^{-1}Q(c)\hat{\pi}, \tag{11}$$

where $Q(c) = c' \otimes I_N$ and $\hat{M}$ is a consistent estimator of M.

For our main results, some of the following assumptions are needed.

Assumption 1. Assume that Y_t is a jointly stationary and ergodic process with finite fourth moments and a positive-definite covariance matrix V.

Assumption 2. Assume that $e_t(\lambda) - E[e_t(\lambda)]$ forms a martingale difference sequence and has a positive-definite covariance matrix $W_e(\lambda)$.

Assumption 3. Assume that Y_t is *iid* normally distributed.

Assumption 3 is restrictive but it is used only for the ML estimator in the beta-pricing model (Shanken, 1985) and not for the CU-GMM estimator in the SDF framework. Assumption 2, which is used for the CU-GMM estimator, is much weaker; it could be further relaxed to allow for serial correlation in $e_t(\lambda) - E[e_t(\lambda)]$, at the cost of a more cumbersome notation. In what follows, the model is said to be misspecified if there does not exist a γ such that $\mu_R = 1_N\gamma_0 + \beta\gamma_1$ holds (or a λ such that $D\lambda = 1_N$ holds).

3 Main Results for Invariant Estimators

In this section, we establish the numerical equivalence between the tests of correct model specification (in the beta-pricing and SDF setups) and the reduced rank test. In addition, we characterize the limiting behavior of the corresponding parameter estimates and t-tests under the assumptions of model misspecification and lack of identification (which arises when a useless factor is included in the model).

3.1 Maximum Likelihood

We start with the more restrictive ML estimation of the beta-pricing model that imposes the joint normality assumption on Y_t (Assumption 3). Combining equations (6) and (7), we arrive at the restriction

$$\alpha = 1_N\gamma_0^* + \beta(\gamma_1^* - \mu_f). \tag{12}$$

Then, the ML estimator of γ^* is defined as (see Shanken, 1992, and Shanken and Zhou, 2007)

$$\hat{\gamma}^{ML} = \underset{\gamma}{\text{argmin}} \frac{(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))'\hat{\Sigma}^{-1}(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))}{1 + \gamma_1'\hat{V}_f^{-1}\gamma_1}, \tag{13}$$

where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}_f$, $\hat{V}_f$, and $\hat{\Sigma}$ are the sample estimators of α, β, μ_f, V_f, and Σ, respectively. The test for correct model specification of Shanken (1985) is given by

$$\mathcal{S} = T \min_{\gamma} \frac{(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))'\hat{\Sigma}^{-1}(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))}{1 + \gamma_1'\hat{V}_f^{-1}\gamma_1}, \tag{14}$$

and is asymptotically distributed as $\mathcal{S} \stackrel{d}{\to} \chi^2_{N-K}$ under the null $H_0 : \alpha = 1_N\gamma_0 + \beta(\gamma_1 - \mu_f)$ and Assumptions 1 and 3.[7]

Note that the ML estimator can be recast as the invariant estimator in equation (1) with $\bar{g}(\gamma) = \sqrt{T}(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))$ and $\hat{W}(\gamma) = (1 + \gamma_1'\hat{V}_f^{-1}\gamma_1)\hat{\Sigma}$. Due to the special structure of this objective function, the ML estimator of γ^* can be obtained explicitly as the solution to an eigenvector problem. Let $v = [-\gamma_0,\ 1,\ -(\gamma_1 - \hat{\mu}_f)']'$ and $\hat{G} = [1_N,\ \hat{\alpha},\ \hat{\beta}]$, and noting that $\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f) = \hat{G}v$, we can write the objective function of the ML estimator as

$$\min_{v} \frac{v'\hat{G}'\hat{\Sigma}^{-1}\hat{G}v}{v'A(X'X/T)^{-1}A'v}, \tag{15}$$

where $A = [0_K,\ I_K]'$ and X is a $T \times K$ matrix with a typical row x_t'. Let $\hat{v}$ be the eigenvector associated with the largest eigenvalue of[8]

$$\hat{\Omega} = (\hat{G}'\hat{\Sigma}^{-1}\hat{G})^{-1}[A(X'X/T)^{-1}A']. \tag{16}$$

[7] Our limiting result for the $\mathcal{S}$ test is also applicable to the asymptotically equivalent likelihood ratio ($\mathcal{LR}$) test, which is given by $\mathcal{LR} = T\ln(1 + \mathcal{S}/T)$ (Shanken, 1985). Note also that in deriving the asymptotic distribution of $\mathcal{S}$ (and $\mathcal{LR}$), Assumption 3 can be relaxed to conditional normality of returns (conditional on f_t). In fact, the asymptotic result for $\mathcal{S}$ (and $\mathcal{LR}$) continues to hold under the more general case of conditional homoskedasticity.

[8] See also Zhou (1995) and Bekker, Dobbelstein, and Wansbeek (1996) for expressing the beta-pricing model as a reduced rank regression whose parameters are obtained as an eigenvalue problem.

Then, the ML estimator of γ^* can be constructed as

$$\hat{\gamma}_0^{ML} = -\frac{\hat{v}_1}{\hat{v}_2}, \tag{17}$$

$$\hat{\gamma}_{1,i}^{ML} = \hat{\mu}_{f,i} - \frac{\hat{v}_{i+2}}{\hat{v}_2}, \quad i = 1, \ldots, K-1. \tag{18}$$

When the model is correctly specified and B is of full column rank, we have that $Gv^* = 0_N$ for $v^* = [-\gamma_0^*,\ 1,\ -(\gamma_1^* - \hat{\mu}_f)']'$ and, under Assumptions 1 and 3,

$$\sqrt{T}\begin{bmatrix} \hat{\gamma}_0^{ML} - \gamma_0^* \\ \hat{\gamma}_1^{ML} - \gamma_1^* \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0_K, (1 + \gamma_1^{*\prime} V_f^{-1} \gamma_1^*)(B_1' \Sigma^{-1} B_1)^{-1} + \begin{bmatrix} 0 & 0_{K-1}' \\ 0_{K-1} & V_f \end{bmatrix}\right), \tag{19}$$

where $B_1 = [1_N,\ \beta]$. As a result, the t-statistics for statistical significance of γ_0 and $\gamma_{1,i}$ ($i = 1, \ldots, K-1$) are constructed as

$$t(\hat{\gamma}_0^{ML}) = \frac{\sqrt{T}\hat{\gamma}_0^{ML}}{s(\hat{\gamma}_0^{ML})}, \tag{20}$$

$$t(\hat{\gamma}_{1,i}^{ML}) = \frac{\sqrt{T}\hat{\gamma}_{1,i}^{ML}}{s(\hat{\gamma}_{1,i}^{ML})}, \tag{21}$$

where $s(\hat{\gamma}_0^{ML})$, $s(\hat{\gamma}_{1,1}^{ML}), \ldots, s(\hat{\gamma}_{1,K-1}^{ML})$ denote the square root of the diagonal elements of

$$(1 + \hat{\gamma}_1^{ML\prime} \hat{V}_f^{-1} \hat{\gamma}_1^{ML})(\hat{B}_1' \hat{\Sigma}^{-1} \hat{B}_1)^{-1} + \begin{bmatrix} 0 & 0_{K-1}' \\ 0_{K-1} & \hat{V}_f \end{bmatrix}, \tag{22}$$

and $\hat{B}_1 = [1_N,\ \hat{\beta}]$. Using the ML estimates $\hat{\gamma}_0^{ML}$ and $\hat{\gamma}_1^{ML}$, the ML estimate of β, $\hat{\beta}^{ML}$, and the fitted expected returns on the test assets, $\hat{\mu}_R^{ML}$, are obtained as

$$\hat{\beta}^{ML} = \hat{\beta} + \frac{[\hat{\alpha} - 1_N \hat{\gamma}_0^{ML} - \hat{\beta}(\hat{\gamma}_1^{ML} - \hat{\mu}_f)]\hat{\gamma}_1^{ML\prime} \hat{V}_f^{-1}}{1 + \hat{\gamma}_1^{ML\prime} \hat{V}_f^{-1} \hat{\gamma}_1^{ML}} \tag{23}$$

and

$$\hat{\mu}_R^{ML} = 1_N \hat{\gamma}_0^{ML} + \hat{\beta}^{ML} \hat{\gamma}_1^{ML}. \tag{24}$$

When the asset-pricing model holds, $\text{rank}(G) = K$, that is, G has a reduced rank. Note that $\text{rank}(G) = K$ if and only if $\text{rank}(P_1'B) = K - 1$, where P_1 is an $N \times (N-1)$ orthonormal matrix whose columns are orthogonal to 1_N. Instead of using the $\mathcal{S}$ test to test the asset-pricing model, one may want to use the Cragg and Donald (1997) rank test to test $\text{rank}(P_1'B) = K - 1$. Under Assumptions 1 and 3, the Cragg and Donald (1997) test statistic is given by

$$\begin{aligned} \mathcal{CD}_1 &= T \min_{c:c'c=1} (P_1'\hat{B}c)'[(c' \otimes P_1')((X'X/T)^{-1} \otimes \hat{\Sigma})(c \otimes P_1)]^{-1}(P_1'\hat{B}c) \\ &= T \min_{c:c'c=1} \frac{c'\hat{B}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{B}c}{c'[(X'X/T)^{-1}]c}, \end{aligned} \tag{25}$$

where $\hat{B} = [\hat{\alpha},\ \hat{\beta}]$. Under the null $H_0 : \text{rank}(P_1'B) = K-1$, we have

$$\mathcal{CD}_1 \stackrel{d}{\to} \chi^2_{N-K}. \tag{26}$$

Lemma 1 below establishes the numerical equivalence of the tests $\mathcal{S}$ and $\mathcal{CD}_1$ and the relationship between their corresponding estimators.

Lemma 1. *Consider the $\mathcal{S}$ and $\mathcal{CD}_1$ tests defined in (14) and (25), and let $\hat{\gamma}^{ML} = [\hat{\gamma}_0^{ML},\ \hat{\gamma}_1^{ML\prime}]'$ and $\hat{c} = [\hat{c}_1,\ \hat{c}_2']'$ denote the estimators that minimize (14) and (25), respectively. Then, under Assumptions 1 and 3, we have $\mathcal{S} = \mathcal{CD}_1$, $\hat{\gamma}_1^{ML} = -\frac{\hat{c}_2}{\hat{c}_1} + \hat{\mu}_f$, and $\hat{\gamma}_0^{ML} = 1_N'\hat{\Sigma}^{-1}(\hat{\mu}_R - \hat{\beta}\hat{\gamma}_1^{ML})/(1_N'\hat{\Sigma}^{-1}1_N)$.*

Proof. See Appendix.

Lemma 1 reveals that the $\mathcal{S}$ test is in fact a rank test of $H_0 : \text{rank}(P_1'B) = K-1$, that is, G has a reduced rank. While G has a reduced rank when the asset-pricing model is correctly specified, there are also misspecified models that can give rise to a reduced rank of G. For example, G can have a reduced rank when the model contains a factor that is independent of the returns (useless factor) or when the model contains two factors that are noisy (due to measurement error, for instance) versions of the same underlying factor. An example of this latter scenario is a consumption-based asset-pricing model whose empirical specification includes several noisy measures of consumption growth. It is of interest to investigate how the ML estimation behaves under these other scenarios. The following theorem characterizes the limiting behavior of the ML estimates $\hat{\gamma}^{ML}$, the t-statistics $t(\hat{\gamma}_0^{ML})$ and $t(\hat{\gamma}_{1,i}^{ML})$ $(i = 1, \ldots, K-1)$, the pseudo-R^2 statistic $R^2_{ML} = \text{Corr}(\hat{\mu}_R^{ML}, \hat{\mu}_R)^2$, and the specification test $\mathcal{S}$ in misspecified models that contain a useless factor. Without loss of generality, we assume that the useless factor is the last element of the vector f_t with $\beta_{K-1} = 0_N$ and is independent of the test asset returns and the other factors.[9] Let $z = [z_1, z_2, \ldots, z_K]' \sim \mathcal{N}(0_K, (G_1'\Sigma^{-1}G_1)^{-1}/\sigma^2_{f,K-1})$, where $G_1 = [1_N,\ \alpha,\ \beta_1, \ldots, \beta_{K-2}]$ and $\sigma^2_{f,K-1} = \text{Var}[f_{K-1,t}]$. In addition, let $\sigma_i^2 = \text{Var}[z_i]$, $\sigma_{ij} \equiv \text{Cov}[z_i, z_j]$, $\rho_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$, and define the random variables $\tilde{z}_2 \equiv z_2/\sigma_2 \sim \mathcal{N}(0,1)$, $x \sim \chi^2_{N-K}$, $q_i \sim \mathcal{N}(0,1)$, where x and q_i are independent of $\tilde{z}_2$, and $b_i = (x + \tilde{z}_2^2)/(x + \tilde{z}_2^2 + q_i^2)$ for $i = 1, \ldots, K-1$. Then, we have the following result.

Theorem 1. *Suppose that the model is misspecified and it contains a useless factor (that is,*

[9] Our analysis can be easily modified to deal with the case in which the betas of the factors are constant across assets instead of being equal to zero.

rank$(B) = K-1$). *Then, under Assumptions 1 and 3 and as* $T \to \infty$, *we have*

(a) *(i)* $\hat{\gamma}_0^{ML} \xrightarrow{d} -\frac{z_1}{z_2}$; *(ii)* $\hat{\gamma}_{1,i}^{ML} \xrightarrow{d} \mu_{f,i} - \frac{z_{i+2}}{z_2}$ *for* $i = 1, \ldots, K-2$; *and (iii)* $\frac{\hat{\gamma}_{1,K-1}^{ML}}{\sqrt{T}} \xrightarrow{d} \frac{1}{z_2}$;

(b) *(i)* $t(\hat{\gamma}_0^{ML}) \xrightarrow{d} -\left(\frac{\rho_{12}|\tilde{z}_2|}{\sqrt{1-\rho_{12}^2}} + q_1\right) b_1^{\frac{1}{2}}$; *(ii)* $t(\hat{\gamma}_{1,i}^{ML}) \xrightarrow{d} \left(\frac{\frac{\mu_{f,i}\sigma_2}{\sigma_{i+2}} - \rho_{i+2,2}}{\sqrt{1-\rho_{i+2,2}^2}}|\tilde{z}_2| - q_{i+1}\right) b_{i+1}^{\frac{1}{2}}$ *for* $i = 1, \ldots, K-2$; *and (iii)* $t^2(\hat{\gamma}_{1,K-1}^{ML}) \xrightarrow{d} \chi^2_{N-K+1}$;

(c) $R^2_{ML} \xrightarrow{p} 1$;

(d) $\lim_{T\to\infty} \Pr[\mathcal{S} > p_\eta] = \eta$, *where* η *denotes the significance level of the specification test and* p_η *is the* $100(1-\eta)$-th *percentile of* χ^2_{N-K}.

Proof. See Appendix.

Theorem 1 establishes the limiting behavior of the parameter estimates, t-tests, and pseudo-R^2 statistic, R^2_{ML}, in misspecified models with identification failure. In addition, it characterizes the asymptotic power of the $\mathcal{S}$ test under rank deficiency. When a useless factor is present, the estimates on the useful factors are inconsistent and converge to ratios of normal random variables. The estimate for the useless factor ($\hat{\gamma}_{1,K-1}^{ML}$) diverges at rate root-$T$, and the standardized estimator converges to the reciprocal of a normal random variable.[10] The t-tests for the useful factors converge to bounded random variables and, hence, are inconsistent. In fact, as our simulations illustrate, the tests $t(\hat{\gamma}_{1,i}^{ML})$ for $i = 1, \ldots, K-2$ tend to exhibit power that is close to their size. In contrast, the t-test for the useless factor will over-reject substantially (with the probability of rejection rapidly approaching one as N increases) when $\mathcal{N}(0,1)$ critical values are used. Furthermore, part (c) of Theorem 1 shows that the pseudo-R^2 of a misspecified model that contains a useless factor approaches one. This leads to completely spurious inference as the useless factors do not contribute to the pricing performance of the model and yet the sample pseudo-R^2 would indicate that the model perfectly explains the cross-sectional variations in the expected returns on the test assets.

Finally, part (d) of Theorem 1 demonstrates that the specification test in the beta-pricing model has asymptotic power equal to its size when a useless factor is included in the model. This

[10] Kan and Zhang (1999a) and Kleibergen (2009) also show that the estimate for the useless factor diverges at rate root-T when employing non-invariant two-pass cross-sectional regression estimators. Similar results are documented by Kan and Zhang (1999b) and Gospodinov, Kan, and Robotti (2014a) for models estimated via non-invariant (optimal and suboptimal) GMM.

result follows directly from the numerical equivalence of $\mathcal{S}$ and $\mathcal{CD}_1$ and has two main implications. First, it suggests that the test for over-identifying restrictions will erroneously conclude, with a limiting probability of $1-\eta$, that a model with an arbitrarily large degree of misspecification is correctly specified. While this result applies to the case of a model with a useless factor, it also has implications for models with factors that are weakly correlated with the test asset returns. In finite samples, factors with low correlations with returns are indistinguishable from useless factors, so the $\mathcal{S}$ test has almost no power in rejecting such models.[11] Second, it indicates that the conventional inference on the parameter estimates $\hat{\gamma}^{ML}$ is likely to be distorted and highly misleading as the standard errors under correct specification will not account for the additional uncertainty arising from model misspecification. It should be stressed that, apart from the rates of convergence and the asymptotic distribution of $t^2(\hat{\gamma}^{ML}_{1,K-1})$, these ML results sharply differ from the results for non-invariant estimators. As further emphasized in Section 4 below, the considered non-invariant estimators appear to exhibit less sensitivity to lack of identification.

As previously discussed, the full rank condition on G may also be violated when the model includes two (or more) factors that are noisy versions of the same underlying factor. In this case (a proof for this result is available in Internet Appendix Section 2), the behavior of the parameter estimates, t-ratios, and pseudo-R^2 is the same as the one described in Theorem 1 with the limiting representations for the noisy factors being the same as the asymptotic distributions for the useless factor in parts (a) and (b) of Theorem 1. In Theorem 2 below, we present a general result for the asymptotic distribution of the specification test that covers both correctly specified and misspecified models that are possibly fully identified or under-identified of arbitrary order (that is, G has rank $K+1-r$ for an integer $r \geq 1$).

Theorem 2: *Suppose that the matrix G has a column rank $K+1-r$ ($r=1,2,\ldots$), that is, there exist r linear combinations of the columns of G that are equal to zero vectors. Then, under Assumptions 1 and 3 and as $T \to \infty$, we have*

$$\mathcal{S} \stackrel{d}{\to} w_r, \tag{27}$$

where w_r is the smallest eigenvalue of $W \sim \mathcal{W}_r(N-K-1+r, I_r)$, and $\mathcal{W}_r(N-K-1+r, I_r)$ denotes the Wishart distribution with $N-K-1+r$ degrees of freedom and a scaling matrix I_r.

[11] The result also suggests that if a model is rejected by the $\mathcal{S}$ test, one can easily "save" the model by adding measurement errors to the factors in the model.

Furthermore, $\Pr[w_r < a] \geq \Pr[x < a]$, *where* $x \sim \chi^2_{N-K}$.

Proof: See Appendix.

Theorem 2 shows that the limiting distribution of $\mathcal{S}$ depends only on r and not on whether the model is correctly specified or misspecified. Note that Theorem 2 nests several cases of interest depending on the value of r. When $r = 1$, that is, the model is correctly specified with a useful factor or the model is misspecified with a useless factor, we have $\mathcal{S} \stackrel{d}{\to} \chi^2_{N-K}$. This case subsumes the standard asymptotic approximation for identified models as well as the result in part (d) of Theorem 1. When the model is correctly specified with a useless factor or when the model is misspecified but it contains two useless factors, we have $r = 2$ and $\mathcal{S} \stackrel{d}{\to} w_2$, where w_2 is the smallest eigenvalue of $\mathcal{W}_2(N - K + 1, I_2)$. In this situation, as the second part of Theorem 2 suggests, the specification test, which is based on critical values from the χ^2_{N-K} distribution, will under-reject the null hypothesis. The following figure shows the limiting distribution of $\mathcal{S}$ for the case of $N - K = 7$.

Figure 3 about here

Finally, in Internet Appendix Section 1, we present limiting results for correctly specified models with identification failure. In this case, the parameter estimates for the useful factors ($\hat{\gamma}_0^{ML}$ and $\hat{\gamma}_{1,i}^{ML}$ for $i = 1, \ldots, K-2$) are consistent but have non-normal asymptotic distributions. The parameter estimate for the useless factor ($\hat{\gamma}_{1,K-1}^{ML}$) is inconsistent and is asymptotically distributed as a Cauchy random variable. The t-statistics are also asymptotically non-normal and using $\mathcal{N}(0,1)$ critical values will lead to under-rejections for the parameters on the useful factors and to over-rejections for the parameter on the useless factor.

Figure 4 about here

The reason for the over-rejection is clearly illustrated in Figure 4 which plots the limiting probability density functions of $t(\hat{\gamma}_{1,K-1}^{ML})$ under correctly specified and misspecified models ($N - K = 7$), along with the standard normal density. Given the bimodal shape and large variance of the probability density function of the limiting distribution of $t(\hat{\gamma}_{1,K-1}^{ML})$ under correctly specified models (which arises from the model's lack of identification), using $\mathcal{N}(0,1)$ critical values will lead to an over-rejection of the hypothesis of zero risk premium on the useless factor. This over-rejection is further exacerbated by model misspecification, as illustrated by the outward shift of the probability density function when the model is misspecified. Hence, with lack of identification, misleading inference also arises in correctly specified models although the inference problems are more pronounced in misspecified models.

3.2 Continuously-Updated GMM

We now consider the more general GMM estimation of SDF and beta-pricing models. The CU-GMM estimator of the SDF parameters λ^* is defined as (see Hansen, Heaton, and Yaron, 1996)

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}}\ \bar{e}(\lambda)'\hat{W}_e(\lambda)^{-1}\bar{e}(\lambda), \tag{28}$$

where $\bar{e}(\lambda) = T^{-1}\sum_{t=1}^{T} e_t(\lambda)$ and

$$\hat{W}_e(\lambda) = \frac{1}{T}\sum_{t=1}^{T}(e_t(\lambda) - \bar{e}(\lambda))(e_t(\lambda) - \bar{e}(\lambda))'. \tag{29}$$

The over-identifying restriction test of the asset-pricing model is given by

$$\mathcal{J} = T \min_{\lambda} \bar{e}(\lambda)'\hat{W}_e(\lambda)^{-1}\bar{e}(\lambda), \tag{30}$$

and $\mathcal{J} \stackrel{d}{\to} \chi^2_{N-K}$ when the asset-pricing model holds.

When the model is correctly specified and D is of full column rank, we have, under Assumptions 1 and 2, that (Hansen, 1982; Newey and Smith, 2004)

$$\sqrt{T}(\hat{\lambda} - \lambda^*) \stackrel{d}{\to} \mathcal{N}\left(0_K, (D'W_e(\lambda^*)^{-1}D)^{-1}\right), \tag{31}$$

where $W_e(\lambda^*) = E[e_t(\lambda^*)e_t(\lambda^*)']$. The t-statistics for statistical significance of λ_0 and $\lambda_{1,i}$ ($i = 1, \ldots, K-1$) are constructed as

$$t(\hat{\lambda}_0) = \frac{\sqrt{T}\hat{\lambda}_0}{s(\hat{\lambda}_0)}, \tag{32}$$

$$t(\hat{\lambda}_{1,i}) = \frac{\sqrt{T}\hat{\lambda}_{1,i}}{s(\hat{\lambda}_{1,i})}, \tag{33}$$

where the quantities $s(\hat{\lambda}_0)$, $s(\hat{\lambda}_{1,1}), \ldots, s(\hat{\lambda}_{1,K-1})$ denote the square root of the diagonal elements of $(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}$. Internet Appendix Section 3 describes how to use the CU-GMM estimates of the SDF parameters λ^* to obtain (in a computationally very efficient way) the CU-GMM estimates of γ^*, β, μ_f, and V_f. More specifically, let

$$w_t(\hat{\lambda}) = \frac{1 - (e_t(\hat{\lambda}) - \bar{e}(\hat{\lambda}))'\hat{W}_e(\hat{\lambda})^{-1}\bar{e}(\hat{\lambda})}{T}. \tag{34}$$

Then, the CU-GMM estimates of μ_f, V_f, and β are given by

$$\hat{\mu}_f^{CU} = \sum_{t=1}^{T} w_t(\hat{\lambda}) f_t, \tag{35}$$

$$\hat{V}_f^{CU} = \sum_{t=1}^{T} w_t(\hat{\lambda}) f_t (f_t - \hat{\mu}_f^{CU})', \tag{36}$$

and

$$\hat{\beta}^{CU} = \sum_{t=1}^{T} w_t(\hat{\lambda}) R_t (f_t - \hat{\mu}_f^{CU})'(\hat{V}_f^{CU})^{-1}. \tag{37}$$

These estimates are subsequently used to construct estimates of the risk premium parameters

$$\hat{\gamma}_0^{CU} = \frac{1}{\hat{\lambda}_0 + \hat{\mu}_f^{CU\prime}\hat{\lambda}_1}, \tag{38}$$

$$\hat{\gamma}_1^{CU} = -\frac{\hat{V}_f^{CU}\hat{\lambda}_1}{\hat{\lambda}_0 + \hat{\mu}_f^{CU\prime}\hat{\lambda}_1}. \tag{39}$$

The fitted (model-implied) expected returns, $\hat{\mu}_R^{CU} = 1_N \hat{\gamma}_0^{CU} + \hat{\beta}^{CU}\hat{\gamma}_1^{CU}$, are used to compute the pseudo-R^2 for CU-GMM.

When the asset-pricing model holds, $\text{rank}(H) = K$. Note that $\text{rank}(H) = K$ if and only if $\text{rank}(P_1'D) = K - 1$. Instead of using the $\mathcal{J}$ statistic to test the asset-pricing model, one may want to use the Cragg and Donald (1997) rank test to test $H_0 : \text{rank}(P_1'D) = K - 1$, which is given by

$$\mathcal{CD}_2 = T \min_{c:c'c=1} (P_1'\hat{D}c)'[(c' \otimes P_1')\hat{V}_d(c \otimes P_1)]^{-1}(P_1'\hat{D}c), \tag{40}$$

where $\hat{D}$ is the sample estimate of D and $\hat{V}_d = \frac{1}{T}\sum_{t=1}^{T} \text{vec}(D_t - \hat{D})\text{vec}(D_t - \hat{D})'$.

The following lemma establishes the numerical equivalence of the $\mathcal{J}$ and $\mathcal{CD}_2$ tests so that the CU-GMM specification test of the asset-pricing model is in essence a rank test.[12] The lemma also explains how to obtain the CU-GMM estimator of λ^* from the $\hat{c}$ that minimizes (40).

[12] We can also interpret Lemma 2 below in terms of Proposition 3 in Peñaranda and Sentana (2014), which states that the $\mathcal{J}$ statistic obtained from the N gross returns is numerically equivalent to the $\mathcal{J}$ statistic from $N - 1$ excess returns. The $\mathcal{CD}_2$ statistic is a quadratic form in $P_1'\hat{D}c$, where the $N \times (N - 1)$ orthonormal matrix P_1 is effectively transforming the N-vector of gross returns into an $(N - 1)$-vector of excess returns with zero costs. Therefore, we can interpret the $\mathcal{CD}_2$ statistic as the $\mathcal{J}$ statistic obtained from the $N - 1$ pricing errors for those excess returns.

LEMMA 2. *Consider the $\mathcal{J}$ and $\mathcal{CD}_2$ tests defined in (30) and (40), and let $\hat{\lambda}$ and $\hat{c}$ be the estimators that minimize (30) and (40), respectively. Then, under Assumptions 1 and 2, we have $\mathcal{J} = \mathcal{CD}_2$ and $\hat{\lambda} = [1_N'\hat{W}_e(\hat{c})^{-1}1_N/(1_N'\hat{W}_e(\hat{c})^{-1}\hat{D}\hat{c})]\hat{c}$.*

Proof. See Appendix.

As in the MLE case, H can have a reduced rank even when the model is misspecified. This can happen when the model contains a useless factor. It is of interest to understand how CU-GMM behaves under unidentified and misspecified models. The following theorem describes the limiting behavior of the CU-GMM estimates, their corresponding t-statistics, the pseudo-R^2, and the $\mathcal{J}$ test for misspecified models that contain a useless factor. Without loss of generality, we assume that the useless factor is ordered last with mean $\mu_{f,K-1}$ and variance $\sigma^2_{f,K-1}$. The CU-GMM estimator and the t-statistic corresponding to the useless factor are $\hat{\lambda}_{1,K-1}$ and $t(\hat{\lambda}_{1,K-1})$, respectively. Let $z = [z_1, z_2, \ldots, z_K]' \sim \mathcal{N}(0_K, \sigma^2_{f,K-1}(H_1'U^{-1}H_1)^{-1})$, where $U = E[R_tR_t']$, $H_1 = [1_N,\ D_1]$, $D_1 = [d_1,\ d_2, \ldots, d_{K-1}]$, and d_i ($i = 1, \ldots, K-1$) is the i-th column of D. Also, let $\sigma_i^2 = \mathrm{Var}[z_i]$, $\sigma_{ij} \equiv \mathrm{Cov}[z_i, z_j]$, $\rho_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$, and define the random variables $\tilde{z}_1 \equiv z_1/\sigma_1 \sim \mathcal{N}(0,1)$, $x \sim \chi^2_{N-K}$, $q_i \sim \mathcal{N}(0,1)$, where x and q_i are independent of $\tilde{z}_1$, and $b_i = (x + \tilde{z}_1^2)/(x + \tilde{z}_1^2 + q_i^2)$ for $i = 1, \ldots, K-1$.

THEOREM 3. *Suppose that the model is misspecified and it contains a useless factor (that is,* $\mathrm{rank}(D) = K-1$*). Then, under Assumptions 1 and 2 and as $T \to \infty$, we have*

(a) *(i)* $\hat{\lambda}_0 \xrightarrow{d} -\dfrac{z_2}{z_1}$ *if* $\mu_{f,K-1} = 0$ *or* $\dfrac{\hat{\lambda}_0}{\sqrt{T}} \xrightarrow{d} \dfrac{\mu_{f,K-1}}{z_1}$ *if* $\mu_{f,K-1} \neq 0$*; (ii)* $\hat{\lambda}_{1,i} \xrightarrow{d} -\dfrac{z_{i+2}}{z_1}$ *for* $i = 1, \ldots, K-2$*; and (iii)* $\dfrac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\dfrac{1}{z_1}$*;*

(b) *(i)* $t(\hat{\lambda}_0) \xrightarrow{d} -\left(\dfrac{\rho_{1,2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,2}^2}} + q_2\right) b_1^{\frac{1}{2}}$ *if* $\mu_{f,K-1} = 0$ *or* $t^2(\hat{\lambda}_0) \xrightarrow{d} \chi^2_{N-K+1}$ *if* $\mu_{f,K-1} \neq 0$*; (ii)* $t(\hat{\lambda}_{1,i}) \xrightarrow{d} -\left(\dfrac{\rho_{1,i+2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,i+2}^2}} + q_{i+2}\right) b_{i+1}^{\frac{1}{2}}$ *for* $i = 1, \ldots, K-2$*; and (iii)* $t^2(\hat{\lambda}_{1,K-1}) \xrightarrow{d} \chi^2_{N-K+1}$*;*

(c) $R^2_{CU} = \mathrm{Corr}(\hat{\mu}_R^{CU}, \hat{\mu}_R)^2 \xrightarrow{p} 1$*;*

(d) $\lim_{T\to\infty} \Pr[\mathcal{J} > p_\eta] = \eta$*, where η denotes the significance level of the test for over-identifying restrictions and p_η is the* $100(1-\eta)$-th *percentile of χ^2_{N-K}.*

Proof. See Appendix.

While the limiting distributions in Theorem 3 for the CU-GMM estimator are broadly consistent with those in Theorem 1 for MLE, some interesting differences emerge. In particular, the asymptotic behavior of $\hat{\lambda}_0$ and $t(\hat{\lambda}_0)$ depends on whether the population mean of the useless factor is zero or not. In the practically relevant case when this mean is nonzero, $\hat{\lambda}_0$ and $t(\hat{\lambda}_0)$ inherit the limiting properties of $\hat{\lambda}_{1,K-1}$ and $t(\hat{\lambda}_{1,K-1})$ for the useless factor. This provides another example of the highly irregular behavior of the invariant estimators in misspecified and unidentified models. Similarly to the beta-pricing setup, parts (c) and (d) of Theorem 3 show that the pseudo-R^2 measure converges to one and the asymptotic power of the CU-GMM specification test is equal to its size when one or more factors are independent of the test asset returns.

As for the beta-pricing model, it is desirable to characterize the asymptotic behavior of the specification test when the model is possibly unidentified of arbitrary order, that is, the matrix H has rank $K+1-r$ for an integer $r \geq 1$. Theorem 4 below establishes that the $\mathcal{J}$ test shares the same limiting distribution as the $\mathcal{S}$ test under some additional restrictions on the data.

Assumption 3'. Assume that

$$\sqrt{T}\text{vec}(\hat{B}-B) \stackrel{d}{\to} \mathcal{N}(0_{NK}, E[x_t x_t']^{-1} \otimes \Sigma). \tag{41}$$

Note that Assumption 3' imposes weaker restrictions on the data than Assumption 3. For example, the result in Assumption 3' holds under conditional homoskedasticity without any distributional requirements on the data.

Theorem 4: *Suppose that the matrix H has a column rank $K+1-r$ $(r=1,2,\ldots)$, that is, there exist r linear combinations of the columns of H that are equal to zero vectors. Then, under Assumptions 1, 2, and 3' and as $T\to\infty$, we have*

$$\mathcal{J} \stackrel{d}{\to} w_r, \tag{42}$$

where w_r is the smallest eigenvalue of $W \sim \mathcal{W}_r(N-K-1+r, I_r)$, and $\mathcal{W}_r(N-K-1+r, I_r)$ denotes the Wishart distribution with $N-K-1+r$ degrees of freedom and a scaling matrix I_r. Furthermore, $\Pr[w_r < a] \geq \Pr[x<a]$, where $x \sim \chi^2_{N-K}$.

Proof: See Appendix.

Theorem 4 nests the standard asymptotic approximation for identified models (Hansen, 1982) as well as the result in part (d) of Theorem 3. Note, however, that the χ^2_{N-K} limiting distribution in these two cases is a more general result and does not hinge on the validity of Assumption 3'. When the model is correctly specified with a useless factor or when the model is misspecified but it contains two useless factors, we have $\mathcal{J} \stackrel{d}{\to} w_2$ and the specification test, which is based on critical values from the χ^2_{N-K} distribution, tends to under-reject.

4 Non-Invariant Estimators

The results so far suggest that the invariant estimators (ML and CU-GMM) are extremely sensitive to model misspecification and lack of identification. It is instructive to study some popular non-invariant estimators (two-pass cross-sectional regression estimators in the beta-pricing setup and the HJ-distance estimators in the SDF setup) under misspecified and unidentified models, and compare their properties with the properties of the invariant estimators analyzed in Section 3. It is well known (see, for example, Shanken and Zhou, 2007) that the GLS cross-sectional regression estimator is the argument that minimizes the numerator of the ML objective function

$$\min_{\gamma} \frac{(\hat{\mu}_R - 1_N\gamma_0 - \hat{\beta}\gamma_1)'\hat{\Sigma}^{-1}(\hat{\mu}_R - 1_N\gamma_0 - \hat{\beta}\gamma_1)}{1 + \gamma_1'\hat{V}_f^{-1}\gamma_1}, \tag{43}$$

which suggests that the non-standard behavior of the ML estimator and test statistics documented above can be due to or exacerbated by the "denominator" problem.

Specifically, the GLS cross-sectional regression estimator of γ^* is given by

$$\hat{\gamma}^{GLS} = (\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\hat{B}_1'\hat{\Sigma}^{-1}\hat{\mu}_R. \tag{44}$$

The cross-sectional regression (CSR) test for correct model specification of Shanken (1985) is based on the statistic

$$\hat{\mathcal{Q}} = (\hat{\mu}_R - \hat{B}_1\hat{\gamma}^{GLS})'\hat{\Sigma}^{-1}(\hat{\mu}_R - \hat{B}_1\hat{\gamma}^{GLS}). \tag{45}$$

To characterize the asymptotic distribution of the CSR test, let

$$l_t(\gamma) = [R_t - \mu_R - \beta(f_t - \mu_f)][1 - \gamma_1'V_f^{-1}(f_t - \mu_f)], \tag{46}$$

$$S_l = E[l_t(\gamma^*)l_t(\gamma^*)'], \tag{47}$$

and $P_{\tilde{V}}$ be an $N \times (N-K)$ orthonormal matrix with columns orthogonal to $\tilde{V} = \Sigma^{-\frac{1}{2}}[1_N,\ V_{Rf}]$. Then, under Assumptions 1 and 2 and when the asset-pricing model is correctly specified and well identified, the limiting distribution of the CSR test is given by (Kan, Robotti, and Shanken, 2013)

$$T\hat{\mathcal{Q}} \stackrel{d}{\to} F_{N-K}(\xi), \tag{48}$$

where the ξ_i's are the eigenvalues of the matrix

$$P_{\tilde{V}}' \Sigma^{-\frac{1}{2}} S_l \Sigma^{-\frac{1}{2}} P_{\tilde{V}}. \tag{49}$$

Similarly, define the HJ-distance estimator of λ^* as

$$\hat{\lambda}^{HJ} = \underset{\lambda}{\operatorname{argmin}}\ \bar{e}(\lambda)'\hat{U}^{-1}\bar{e}(\lambda) = (\hat{D}'\hat{U}^{-1}\hat{D})^{-1}\hat{D}'\hat{U}^{-1}1_N, \tag{50}$$

where $\hat{U}$ is the sample estimate of $U = E[R_t R_t']$ defined above Theorem 3, and let

$$\hat{\delta}^2 = \bar{e}(\hat{\lambda}^{HJ})'\hat{U}^{-1}\bar{e}(\hat{\lambda}^{HJ}) \tag{51}$$

denote the sample squared HJ-distance. Then, under Assumptions 1 and 2 and when the asset-pricing model is correctly specified and well identified, the HJ-distance test, $T\hat{\delta}^2$, follows a weighted chi-squared limiting distribution (Jagannathan and Wang, 1996):

$$T\hat{\delta}^2 \stackrel{d}{\to} F_{N-K}(\xi), \tag{52}$$

where the ξ_i's are the nonzero eigenvalues of the matrix

$$S_d^{\frac{1}{2}} U^{-1} S_d^{\frac{1}{2}} - S_d^{\frac{1}{2}} U^{-1} D(D'U^{-1}D)^{-1} D'U^{-1} S_d^{\frac{1}{2}}, \tag{53}$$

and $S_d = E[e_t(\lambda^*)e_t(\lambda^*)']$. The next theorem characterizes the limiting behavior of the sample squared HJ-distance and CSR tests in the presence of a useless factor.

THEOREM 5. *Let* $\mathcal{B} \sim \text{Beta}\left(\frac{N-K}{2}, \frac{1}{2}\right)$ *with density* $f_{\mathcal{B}}(\cdot)$ *and* p_η *be the* $100(1-\eta)$-th *percentile of* χ^2_{N-K}.

(a) *Suppose that Assumptions 1 and 2 hold, the asset-pricing model is misspecified and it contains a useless factor. In addition, denote by* $\delta^2 = 1_N'U^{-1}1_N - 1_N'U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}1_N$ *the squared population HJ-distance. Then, we have*

$$\hat{\delta}^2 \stackrel{d}{\to} \delta^2\mathcal{B}, \tag{54}$$

and the limiting probability of rejecting the null hypothesis of correct model specification by the HJ-distance test of size η *is*

$$\int_0^1 P\left[\chi^2_{N-K+1} > \frac{p_\eta s}{1-s}\right] f_{\mathcal{B}}(s)\mathrm{d}s < 1. \tag{55}$$

(b) *Suppose that Assumption 1 holds and* $l_t(\gamma) - E[l_t(\gamma)]$ *forms a martingale difference sequence. In addition, suppose that the beta-pricing model is misspecified and it contains a useless factor. Finally, let* $G_2 = [1_N,\ \beta_1, \ldots, \beta_{K-2}]$ *and denote by* $\mathcal{Q} = \mu_R'\Sigma^{-1}\mu_R - \mu_R'\Sigma^{-1}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-1}\mu_R$ *the population CSR.*[13] *Then, we have*

$$\hat{\mathcal{Q}} \xrightarrow{d} \mathcal{Q}\mathcal{B}, \tag{56}$$

and the limiting probability of rejecting the null hypothesis of correct model specification by the CSR test of size η *is also given by (55).*

Proof. See Appendix.

While Theorem 5 shows that the HJ-distance and CSR tests are inconsistent under identification failure, the limiting probability of rejection of the null hypothesis is different from the ones for the invariant estimator-based tests. In particular, this limiting probability is a function of $N-K$ and is very close to one when $N-K$ is of the magnitude typically encountered in empirical work. Consider again the same test assets and asset-pricing models described in Section 1. Figures 5 and 6 visualize the cross-sectional goodness-of-fit of the models by plotting average realized returns versus fitted (by HJ-distance and GLS, respectively) expected returns from each model.

Figures 5 and 6 about here

In sharp contrast to the results for invariant estimators in Figures 1 and 2, the models that contain factors that are only weakly correlated with the test asset returns (C-LAB and CC-CAY) no longer exhibit a perfect fit, and the HJ-distance and CSR tests strongly reject the null of correct specification (see Panel A of Table I). As a result, these non-invariant tests appear to be more robust to lack of identification and can detect model misspecification with higher probability than their invariant counterparts.

[13] The explicit expression for Σ when the model contains a useless factor is provided in the proof of Theorem 5 (b).

5 Simulation Results

In this section, we undertake a Monte Carlo simulation experiment to study the empirical rejection rates of the specification and t-tests for the CU-GMM and ML estimators as well as the finite-sample distribution of the goodness-of-fit measure. We consider three linear models: (i) a model with a constant term and a useful factor, (ii) a model with a constant term and a useless factor, and (iii) a model with a constant term, a useful, and a useless factor. All three models are misspecified.

The returns on the test assets and the useful factor are drawn from a multivariate normal distribution. In all simulation designs, the covariance matrix of the simulated test asset returns is set equal to the sample covariance matrix from the 1959:2–2012:12 sample of monthly gross returns on the 25 Fama-French size and book-to-market ranked portfolios (from Kenneth French's website). The means of the simulated returns are set equal to the sample means of the actual returns, and they are not exactly linear in the chosen betas for the useful factors. As a result, the models are misspecified in all three cases. The mean and variance of the simulated useful factor are calibrated to the sample mean and variance of the value-weighted market excess return. The covariances between the useful factor and the returns are chosen based on the sample covariances estimated from the data. The useless factor is generated as a standard normal random variable which is independent of the returns and the useful factor. The time series sample size is $T = 200$, 400, 600, 800, 1000, and 3600, and all results are based on 100,000 Monte Carlo replications. We also report the limiting rejection probabilities (denoted by $T = \infty$) for the specification and t-tests based on our asymptotic results in Sections 3 and 4.

Table II about here

Table II presents the probabilities of rejection of the model specification tests based on invariant (CU-GMM and ML) and non-invariant (HJ-distance and GLS) estimators at the 10%, 5%, and 1% nominal levels. Table II.a reports the rejection probabilities of the $\mathcal{J}$ and HJ-distance tests for the SDF representation, while Table II.b includes the $\mathcal{S}$ and CSR tests for the beta-pricing model. Consistent with our theoretical results, the specification tests based on invariant estimators do not exhibit any power in the presence of a useless factor and the empirical rejection probabilities approach the nominal size under the alternative of a misspecified model. As a result, when a useless factor is included in the model, the researcher will conclude erroneously (with probability one minus the nominal size of the test) that the model is correctly specified even when the misspecification of the model is arbitrarily large. In contrast, the HJ-distance and the CSR tests, albeit still inconsistent, prove to be much more successful at detecting model misspecification.

Table III about here

An alternative and popular way to look at the performance of the model is to compute the squared correlation between the expected fitted returns of the model and the average realized returns. The distribution of this pseudo-R^2 is reported in Table III (III.a for CU-GMM and III.b for ML). Again, as our theoretical analysis suggests, the empirical distribution of the pseudo-R^2 in models with a useless factor collapses to 1 as the sample size gets large. For example, this measure will indicate a perfect fit for models that include a factor that is independent of the returns on the test assets. These spurious results should serve as a warning signal in applied work where many macroeconomic factors are only weakly correlated with the returns on the test assets.

Table IV about here

Finally, Table IV (IV.a for CU-GMM and IV.b for ML) presents the rejection probabilities of the t-tests of $H_0 : \lambda_{1,i} = 0$ and $H_0 : \gamma_{1,i} = 0$ (tests of statistical significance) for the useful and the useless factor in the SDF and beta representations of models (i), (ii), and (iii). The t-statistics are computed under the assumption that the model is correctly specified and are compared against the critical values from the standard normal distribution, as is commonly done in the literature. Table IV reveals that for models with a useless factor, the t-tests will give rise to spurious results, suggesting that these completely irrelevant factors are priced. Moreover, the useless factor (which, by construction, does not contribute to the pricing performance of the model) drives out the useful factor and leads to the grossly misleading conclusion to keep the useless factor and drop the useful factor from the model (see Panel C of Table IV.a and IV.b).

6 Concluding Remarks

In this paper, we study the limiting properties of some invariant tests of asset-pricing models, and show that the inference based on these tests can be spurious when the models are unidentified. In particular, we demonstrate that, in the presence of factors that are independent of the returns on the test assets (useless factors), the power of the specification tests based on invariant estimators is equal to their size, and the pseudo-R^2 that measures the distance between the fitted and realized average returns approaches one. As a consequence, an applied researcher would conclude with high probability that the model is correctly specified and proceed with constructing standard errors and test statistics that assume correct model specification. Since these statistics would not take into the account the extra uncertainty arising from potential model misspecification, the inference on the model parameters would be highly distorted and would manifest in highly significant estimates for factors that do not contribute to improved pricing.

The spurious results in these models arise from the combined effect of identification failure and model misspecification. It is important to stress that this is not an isolated problem limited to a particular sample (data frequency), test assets, and asset-pricing models. This suggests that the statistical evidence on the pricing ability of many macro factors and their usefulness in explaining the cross-section of asset returns should be interpreted with caution. Some warning signs about this problem (for example, the outcome of a rank test) are often ignored by applied researchers. While non-invariant estimators (HJ-distance non-optimal GMM and GLS two-pass cross-sectional regressions) also suffer from similar problems, the invariant (CU-GMM and ML) estimators turn out to be much more sensitive to model misspecification and lack of identification.

Given the severity of the inference problems associated with invariant estimators of possibly unidentified and misspecified asset-pricing models that we document in this paper, our recommendations for empirical practice can be summarized as follows. The lack of power of the specification tests in unidentified models suggests that the decision regarding the model specification should be augmented with additional diagnostics. For instance, the tests developed by Arellano, Hansen, and Sentana (2012) and Peñaranda and Sentana (2014) can detect if the lack of rejection of the model specification tests is genuine or is due to the presence of a useless factor. Importantly, any model should be subjected to a rank test which will provide evidence on whether the model parameters are identified or not. If the null hypothesis of a reduced rank is rejected, the researcher can proceed with the standard tools for inference in analyzing and evaluating the model. If the null of a reduced rank is not rejected, there are two possible ways to proceed. A first possibility is to work with non-invariant estimators (HJ-distance and cross-sectional regression estimators, for example) and pursue misspecification-robust inference that is asymptotically valid regardless of the degree of identification (see Gospodinov, Kan, and Robotti, 2014a). Second, if there is a strong preference for using invariant estimators, the researcher needs to estimate consistently the reduced rank of the model (as in Ahn, Horenstein, and Wang, 2010, for instance) and select a subset of the original factors such that its dimension is equal to the estimated rank, and the resulting model is identified. For example, one could perform a rank test for each possible combination of factors and choose the combination that delivers the largest rejection of the reduced rank hypothesis. This procedure would restore the standard inference although it may still need to be robustified against possible model misspecification as in Gospodinov, Kan, and Robotti (2013).

While the results in this paper are developed in the context of linear factor models, we conjecture that similar results characterize the limiting behavior of specification tests in a different or more general setup. For example, Cragg and Donald (1996) establish the inconsistency of the Anderson-Rubin test for over-identifying restrictions in unidentified linear instrumental variable models while Dovonon and Renault (2013) derive the asymptotic distribution of the test for over-identifying restrictions under lack of first-order identification. Furthermore, the CU-GMM is a member of the class of generalized empirical likelihood (GEL) estimators (Newey and Smith, 2004) that provides a unifying framework for assessing asset-pricing models (Almeida and Garcia, 2012). To the best of our knowledge, the statistical properties of the GEL estimators under model misspecification (Schennach, 2007) and potential identification failure (Otsu, 2006; Guggenberger and Smith, 2008) have been analyzed only in isolation. Establishing whether the results in this paper for the CU-GMM estimator carry over to other GEL estimators or constructing GEL-based tests of correct model specification that are robust to a complete or partial lack of identification is a promising direction for future research.

Appendix: Proofs of Lemmas and Theorems

A.1 Proof of Lemma 1

Note that the $\hat{c}$ that minimizes (25) can be analytically obtained as the eigenvector associated with the smallest eigenvalue of

$$\tilde{\Omega} = (X'X/T)\hat{B}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{B}. \tag{A.1}$$

For the numerical equivalence of $\mathcal{S}$ and $\mathcal{CD}_1$ to hold, it is sufficient to show that $\hat{\Omega}$ and $\tilde{\Omega}^{-1}$ share the same nonzero eigenvalues. Using the formula for the inverse of a partitioned matrix, we obtain

$$\hat{\Omega} = (\hat{G}'\hat{\Sigma}^{-1}\hat{G})^{-1}\begin{bmatrix} 0 & 0_K' \\ 0_K & (X'X/T)^{-1} \end{bmatrix} = \begin{bmatrix} 0 & -(1_N'\hat{\Sigma}^{-1}1_N)^{-1}(1_N'\hat{\Sigma}^{-1}\hat{B})C \\ 0_K & C \end{bmatrix},$$

where

$$C = [\hat{B}'\hat{\Sigma}^{-1}\hat{B} - \hat{B}'\hat{\Sigma}^{-1}1_N(1_N'\hat{\Sigma}^{-1}1_N)^{-1}1_N'\hat{\Sigma}^{-1}\hat{B}]^{-1}(X'X/T)^{-1}. \tag{A.2}$$

Note that

$$I_N - \hat{\Sigma}^{-\frac{1}{2}}1_N(1_N'\hat{\Sigma}^{-1}1_N)^{-1}1_N'\hat{\Sigma}^{-\frac{1}{2}} = \hat{\Sigma}^{\frac{1}{2}}P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{\Sigma}^{\frac{1}{2}} \tag{A.3}$$

since $P_1'\hat{\Sigma}^{\frac{1}{2}}\hat{\Sigma}^{-\frac{1}{2}}1_N = 0_{N-1}$. Then, it follows that

$$\begin{aligned}\hat{B}'\hat{\Sigma}^{-1}\hat{B} - \hat{B}'\hat{\Sigma}^{-1}1_N(1_N'\hat{\Sigma}^{-1}1_N)^{-1}1_N'\hat{\Sigma}^{-1}\hat{B} &= \hat{B}'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}1_N(1_N'\hat{\Sigma}^{-1}1_N)^{-1}1_N'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{B} \\ &= \hat{B}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{B} \end{aligned} \tag{A.4}$$

and

$$\hat{\Omega} = \begin{bmatrix} 0 & -(1_N'\hat{\Sigma}^{-1}1_N)^{-1}(1_N'\hat{\Sigma}^{-1}\hat{B})\tilde{\Omega}^{-1} \\ 0_K & \tilde{\Omega}^{-1} \end{bmatrix}. \tag{A.5}$$

If ξ is a nonzero eigenvalue of $\tilde{\Omega}^{-1}$ and c is the corresponding eigenvector, then we have

$$\hat{\Omega}v = \xi v, \tag{A.6}$$

where $v = [-(1_N'\hat{\Sigma}^{-1}1_N)^{-1}(1_N'\hat{\Sigma}^{-1}\hat{B}c),\ c']'$. Therefore, $\hat{\Omega}$ and $\tilde{\Omega}^{-1}$ share the same nonzero eigenvalues. In particular, if $\hat{c} = [\hat{c}_1,\ \hat{c}_2']'$ is the eigenvector associated with the largest eigenvalue of $\tilde{\Omega}^{-1}$ or equivalently the eigenvector associated with the smallest eigenvalue of $\tilde{\Omega}$, then $\hat{v} = [-(1_N'\hat{\Sigma}^{-1}1_N)^{-1}(1_N'\hat{\Sigma}^{-1}\hat{B}\hat{c}),\ \hat{c}']'$ is proportional to the eigenvector associated with the largest eigenvalue of $\hat{\Omega}$. Finally, using (17) and (18), we obtain

$$\hat{\gamma}_1^{ML} = -\frac{\hat{c}_2}{\hat{c}_1} + \hat{\mu}_f, \tag{A.7}$$

$$\hat{\gamma}_0^{ML} = \frac{1_N'\hat{\Sigma}^{-1}\hat{B}\hat{c}}{(1_N'\hat{\Sigma}^{-1}1_N)\hat{c}_1} = \frac{1_N'\hat{\Sigma}^{-1}(\hat{\mu}_R - \hat{\beta}\hat{\gamma}_1^{ML})}{1_N'\hat{\Sigma}^{-1}1_N}. \tag{A.8}$$

This completes the proof.

A.2 Proof of Theorem 1

part (a): When the model is misspecified and contains a useless factor (ordered last), we have $Gv^* = 0_N$ for $v^* = [0_K', \ 1]'$. Let $\hat{v}$ be the eigenvector associated with the largest eigenvalue of

$$\hat{\Omega} = (\hat{G}'\hat{\Sigma}^{-1}\hat{G})^{-1}[A(X'X/T)^{-1}A']. \tag{A.9}$$

Define $\hat{\psi} = [\hat{\psi}_1, \ \hat{\psi}_2, \ldots, \hat{\psi}_K]'$ as

$$\hat{\psi}_i = -\frac{\hat{v}_i}{\hat{v}_{K+1}}, \quad i = 1, \ldots, K, \tag{A.10}$$

which is asymptotically equivalent to the estimator

$$\tilde{\psi} = (\hat{G}_1'\hat{\Sigma}^{-1}\hat{G}_1)^{-1}(\hat{G}_1'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}). \tag{A.11}$$

Since $\sqrt{T}\hat{\beta}_{K-1} \xrightarrow{d} \mathcal{N}(0_N, \Sigma/\sigma_{f,K-1}^2)$, we have

$$\sqrt{T}\tilde{\psi} \xrightarrow{d} \mathcal{N}(0_K, (G_1'\Sigma^{-1}G_1)^{-1}/\sigma_{f,K_1}^2), \tag{A.12}$$

and $\sqrt{T}\hat{\psi}$ also has the same asymptotic distribution. Therefore, we can write

$$\hat{\gamma}_0^{ML} = -\frac{\sqrt{T}\hat{\psi}_1}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} -\frac{z_1}{z_2}, \tag{A.13}$$

$$\hat{\gamma}_{1,i}^{ML} = \hat{\mu}_{f,i} - \frac{\sqrt{T}\hat{\psi}_{i+2}}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} \mu_{f,i} - \frac{z_{i+2}}{z_2}, \quad i = 1, \ldots, K-2, \tag{A.14}$$

$$\frac{\hat{\gamma}_{1,K-1}^{ML}}{\sqrt{T}} = \frac{\hat{\mu}_{f,K-1}}{\sqrt{T}} + \frac{1}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} \frac{1}{z_2}. \tag{A.15}$$

This completes the proof of part (a).

part (b): We start with the squared t-ratio of the useless factor, $t^2(\hat{\gamma}_{1,K-1}^{ML})$. Define $G_2 = [1_N,\ \beta_1, \ldots, \beta_{K-2}]$ and $\hat{G}_2 = [1_N,\ \hat{\beta}_1, \ldots, \hat{\beta}_{K-2}]$. Using the formula for the inverse of a partitioned matrix, we obtain

$$
\begin{aligned}
s^2(\hat{\gamma}_{1,K-1}^{ML}) &= (1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML})\left(\hat{\beta}_{K-1}'[\hat{\Sigma}^{-1}-\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_{K-1}\right)^{-1}+\hat{\sigma}_{f,K-1}^2 \\
&= \left(\frac{\hat{\gamma}_{1,K-1}^{ML}}{\hat{\sigma}_{f,K-1}}\right)^2\left(\hat{\beta}_{K-1}'[\hat{\Sigma}^{-1}-\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_{K-1}\right)^{-1}+O_p(T^{\frac{1}{2}}) \quad \text{(A.16)}
\end{aligned}
$$

by using the fact that $\hat{\gamma}_{1,i}^{ML} = O_p(1)$ for $i = 1, \ldots, K-2$ and $\hat{\gamma}_{1,K-1}^{ML} = O_p(T^{\frac{1}{2}})$. In addition, by defining u as follows:

$$
\sqrt{T}\hat{\sigma}_{f,K-1}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1} \xrightarrow{d} u \sim \mathcal{N}(0_N, I_N), \quad \text{(A.17)}
$$

we obtain

$$
\begin{aligned}
t^2(\hat{\gamma}_{1,K-1}^{ML}) &= \frac{T(\hat{\gamma}_{1,K-1}^{ML})^2\hat{\beta}_{K-1}'[\hat{\Sigma}^{-1}-\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2\hat{\Sigma}^{-1}]\hat{\beta}_{K-1}}{(\hat{\gamma}_{1,K-1}^{ML}/\hat{\sigma}_{f,K-1})^2}+O_p(T^{-\frac{1}{2}}) \\
&= u'[I_N-\hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]u+O_p(T^{-\frac{1}{2}}) \\
&\xrightarrow{d} u'[I_N-\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u \sim \chi^2_{N-K+1}. \quad \text{(A.18)}
\end{aligned}
$$

For the limiting distributions of $t(\hat{\gamma}_0^{ML})$ and $t(\hat{\gamma}_{1,i}^{ML})$, $i = 1, \ldots, K-2$, we use the formula for the inverse of a partitioned matrix to obtain the upper left $(K-1)\times(K-1)$ block of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$ as

$$
\begin{aligned}
&(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}+\frac{(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}}{\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}-\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}} \\
&= (G_2'\Sigma^{-1}G_2)^{-1}+\frac{(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}uu'\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}}{u'[I_N-\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u}+O_p(T^{-\frac{1}{2}}). \quad \text{(A.19)}
\end{aligned}
$$

Note that we can write

$$
I_N-\Sigma^{-\frac{1}{2}}G_1(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}} = I_N-\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}-hh', \quad \text{(A.20)}
$$

where

$$
h = \frac{[I_N-\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha}{\left(\alpha'\Sigma^{-\frac{1}{2}}[I_N-\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha\right)^{\frac{1}{2}}}. \quad \text{(A.21)}
$$

With this expression, we can write

$$
\begin{aligned}
u'[I_N-\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u &= u'[I_N-\Sigma^{-\frac{1}{2}}G_1(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}}]u+(h'u)^2 \\
&= x+\tilde{z}_2^2, \quad \text{(A.22)}
\end{aligned}
$$

where $x \sim \chi^2_{N-K}$ and it is independent of $\tilde{z}_2 \sim \mathcal{N}(0,1)$. To establish the last equality, we need to show that $h'u = \tilde{z}_2$. Denote by $\boldsymbol{\iota}_{m,i}$ an m-vector with its i-th element equals to one and zero elsewhere, and let $\sigma_{ij} \equiv \text{Cov}[z_i, z_j] = \boldsymbol{\iota}'_{K,i}(G_1'\Sigma^{-1}G_1)^{-1}\boldsymbol{\iota}_{K,j}/\sigma^2_{f,K-1}$. Using the formula for the inverse of a partitioned matrix, we obtain

$$
\begin{aligned}
z_2 &= \frac{1}{\sigma_{f,K-1}}\boldsymbol{\iota}'_{K,2}(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}}u \\
&= \frac{1}{\sigma_{f,K-1}}\frac{\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u}{\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha}. \qquad \text{(A.23)}
\end{aligned}
$$

It follows that

$$
\sigma_2^2 = \frac{1}{\sigma^2_{f,K-1}\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha} \qquad \text{(A.24)}
$$

and $h'u = z_2/\sigma_2 = \tilde{z}_2$.

Denote by w_i the i-th diagonal element of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$, $i = 1, \ldots, K-1$. Using (A.19), we have

$$
\begin{aligned}
w_i &\xrightarrow{d} \boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i} + \frac{\boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}uu'\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}}{x + \tilde{z}_2^2} \\
&= \boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}\left(1 + \frac{q_i^2}{x + \tilde{z}_2^2}\right), \qquad \text{(A.25)}
\end{aligned}
$$

where

$$
q_i = \frac{\boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}u}{[\boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}]^{\frac{1}{2}}} \sim \mathcal{N}(0,1). \qquad \text{(A.26)}
$$

Using the fact that $\text{Var}[u] = I_N$ and

$$
(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-1}G_2 = [\boldsymbol{\iota}_{K,1},\ \boldsymbol{\iota}_{K,3}, \ldots, \boldsymbol{\iota}_{K,K}], \qquad \text{(A.27)}
$$

it is straightforward to show that

$$
\begin{aligned}
\text{Cov}[z_1, q_1] &= \frac{\boldsymbol{\iota}'_{K,1}(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-1}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}}{\sigma_{f,K-1}[\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} \\
&= [\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma^2_{f,K-1}]^{\frac{1}{2}}, \qquad \text{(A.28)} \\
\text{Cov}[z_2, q_1] &= \frac{\boldsymbol{\iota}'_{K,2}(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-1}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}}{\sigma_{f,K-1}[\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} = 0. \qquad \text{(A.29)}
\end{aligned}
$$

From the formula for the inverse of a partitioned matrix, we have

$$
\frac{1}{\sigma^2_{f,K-1}}\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1} = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} = \sigma_1^2(1 - \rho_{12}^2). \qquad \text{(A.30)}
$$

It follows that

$$\text{Cov}\left[z_1 - \frac{\sigma_{12}}{\sigma_2^2}z_2, q_1\right] = [\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma^2_{f,K-1}]^{\frac{1}{2}} = \sigma_1\sqrt{1-\rho_{12}^2}. \tag{A.31}$$

Therefore, $z_1 - (\sigma_{12}/\sigma_2^2)z_2$ is perfectly correlated with q_1 and we can write

$$z_1 = \frac{\sigma_{12}}{\sigma_2^2}z_2 + \sqrt{1-\rho_{12}^2}\sigma_1 q_1 = \sigma_1\left(\rho_{12}\tilde{z}_2 + \sqrt{1-\rho_{12}^2}q_1\right). \tag{A.32}$$

Similarly,

$$z_{i+1} = \frac{\sigma_{i+1,2}}{\sigma_2^2}z_2 + \sqrt{1-\rho_{i+1,2}^2}\sigma_{i+1}q_i = \sigma_{i+1}\left(\rho_{i+1,2}\tilde{z}_2 + \sqrt{1-\rho_{i+1,2}^2}q_i\right), \quad i = 2,\ldots,K-1. \tag{A.33}$$

Let

$$b_i = \frac{x + \tilde{z}_2^2}{x + \tilde{z}_2^2 + q_i^2}, \quad i = 1,\ldots,K-1. \tag{A.34}$$

With the above results, we can now write the limiting distribution of the t-ratios as

$$\begin{aligned}
t(\hat{\gamma}_0^{ML}) &\xrightarrow{d} -\frac{z_1|z_2|b_1^{\frac{1}{2}}}{z_2[\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma^2_{f,K-1}]^{\frac{1}{2}}} \\
&= -\left(\frac{\rho_{12}|\tilde{z}_2|}{\sqrt{1-\rho_{12}^2}} + q_1\right)b_1^{\frac{1}{2}}, \qquad \text{(A.35)} \\
t(\hat{\gamma}_{1,i}^{ML}) &\xrightarrow{d} \frac{\left(\mu_{f,i} - \frac{z_{i+2}}{z_2}\right)|z_2|b_{i+1}^{\frac{1}{2}}}{[\boldsymbol{\iota}'_{K-1,i+1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i+1}/\sigma^2_{f,K-1}]^{\frac{1}{2}}} \\
&= \left(\frac{\frac{\mu_{f,i}\sigma_2}{\sigma_{i+2}} - \rho_{i+2,2}}{\sqrt{1-\rho_{i+2,2}^2}}|\tilde{z}_2| - q_{i+1}\right)b_{i+1}^{\frac{1}{2}}, \quad i = 1,\ldots,K-2. \qquad \text{(A.36)}
\end{aligned}$$

This completes the proof of part (b).

part (c): Let $\hat{e} = \hat{\mu}_R - 1_N\hat{\gamma}_0^{ML} - \hat{\beta}\hat{\gamma}_1^{ML}$ and note that the fitted (model-implied) expected returns can be rewritten as

$$\begin{aligned}
\hat{\mu}_R^{ML} &= 1_N\hat{\gamma}_0^{ML} + \hat{\beta}^{ML}\hat{\gamma}_1^{ML} \\
&= 1_N\hat{\gamma}_0^{ML} + \hat{\beta}\hat{\gamma}_1^{ML} + \hat{e}\frac{\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}{1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}} \\
&= \hat{\mu}_R - \hat{e} + \hat{e}\frac{\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}{1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}} \\
&= \hat{\mu}_R - \hat{e}\frac{1}{1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}. \qquad \text{(A.37)}
\end{aligned}$$

Using the result from part (a) that $\hat{\gamma}_{1,i}^{ML} = O_p(1)$ for $i = 1, \ldots, K-2$ and $\hat{\gamma}_{1,K-1}^{ML} = O_p(T^{\frac{1}{2}})$, we have $\hat{\mu}_R^{ML} - \hat{\mu}_R \xrightarrow{p} 0_N$ and

$$R^2_{ML} = \text{Corr}(\hat{\mu}_R^{ML}, \hat{\mu}_R)^2 \xrightarrow{p} 1 \tag{A.38}$$

as $T \to \infty$. This completes the proof of part (c).

part (d): Because B is of reduced rank due to the presence of a useless factor, $\mathcal{CD}_1 \xrightarrow{d} \chi^2_{N-K}$. From the numerical equivalence of $\mathcal{S}$ and $\mathcal{CD}_1$ in Lemma 1, it follows that $\mathcal{S} \xrightarrow{d} \chi^2_{N-K}$ even though the model is misspecified. This completes the proof of part (d).

A.3 Proof of Theorem 2

It follows from Lemma 1 that the $\mathcal{S}$ test is asymptotically distributed as T times the smallest eigenvalue of

$$\tilde{\Omega} = (X'X/T)\hat{B}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{B}. \tag{A.39}$$

Let L_f be a lower triangular matrix such that $L_f L_f' = V_f$ and define

$$L = \begin{bmatrix} 1 & 0'_{K-1} \\ \mu_f & L_f \end{bmatrix}. \tag{A.40}$$

Using that $(X'X)/T \xrightarrow{p} LL'$ and $\hat{\Sigma} \xrightarrow{p} \Sigma$, the $\mathcal{S}$ test has the same distribution as the smallest eigenvalue of

$$W_0 = TLL'\hat{B}'P_1(P_1'\Sigma P_1)^{-1}P_1'\hat{B}. \tag{A.41}$$

Let $\tilde{P}_1$ be an $N \times (N-1)$ orthonormal matrix such that $\tilde{P}_1'\Sigma^{-\frac{1}{2}}1_N = 0_{N-1}$. Then, we have

$$P_1(P_1'\Sigma P_1)^{-1}P_1' = \Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}} \tag{A.42}$$

and

$$W_0 = TLL'\hat{B}'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\hat{B}. \tag{A.43}$$

Define

$$Z = \sqrt{T}\tilde{P}_1'\Sigma^{-\frac{1}{2}}\hat{B}L \tag{A.44}$$

and

$$M = E[Z] = \sqrt{T}\tilde{P}_1'\Sigma^{-\frac{1}{2}}BL. \tag{A.45}$$

Then, under Assumptions 1 and 3,

$$\text{vec}(Z) \sim \mathcal{N}\left(\text{vec}(M), I_{(N-1)K}\right). \tag{A.46}$$

Since W_0 and $Z'Z$ share the same eigenvalues, the smallest eigenvalue of W_0 has the same limiting distribution as the smallest eigenvalue of $W_1 = Z'Z \sim \mathcal{W}_K(N-1, I_K, M'M)$, which has a noncentral Wishart distribution. Since B has rank $K - r$, there exists a $K \times r$ orthonormal matrix C_1 such that $MC_1 = 0_{(N-1)\times r}$. Let $C = [C_1,\ C_2]$ be a $K \times K$ orthonormal matrix, and define $\tilde{Z} = [\tilde{Z}_1,\ \tilde{Z}_2] = [ZC_1,\ ZC_2]$. Then, we have $E[\tilde{Z}_1] = 0_{(N-1)\times r}$ and $E[\tilde{Z}_2] \equiv \tilde{M}_2 = MC_2$. Using the fact that $W_1 = Z'Z$ and $W_2 = \tilde{Z}'\tilde{Z}$ share the same eigenvalues, it is sufficient to obtain the limiting distribution of the smallest eigenvalue of W_2 which is equal to the reciprocal of the largest eigenvalue of

$$W_2^{-1} = \begin{bmatrix} W_2^{11} & W_2^{12} \\ W_2^{21} & W_2^{22} \end{bmatrix}. \tag{A.47}$$

Using the formula for the inverse of a partitioned matrix, we have

$$W_2^{11} = \left(\tilde{Z}_1'[I_{N-1} - \tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1}\tilde{Z}_2']\tilde{Z}_1\right)^{-1} \sim \mathcal{W}_r(N-K-1+r, I_r)^{-1}, \tag{A.48}$$

$$W_2^{12} = -W_2^{11}\tilde{Z}_1'\tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1} = O_p(T^{-\frac{1}{2}}), \tag{A.49}$$

$$W_2^{22} = (\tilde{Z}_2'\tilde{Z}_2)^{-1} + (\tilde{Z}_2'\tilde{Z}_2)^{-1}(\tilde{Z}_2'\tilde{Z}_1)W_2^{11}(\tilde{Z}_1'\tilde{Z}_2)(\tilde{Z}_2'\tilde{Z}_2)^{-1} = O_p(T^{-1}). \tag{A.50}$$

Therefore, the limiting distribution of the largest eigenvalue of W_2^{-1} is the same as the limiting distribution of the largest eigenvalue of W_2^{11}. Equivalently, the smallest eigenvalue of $T\tilde{\Omega}$ has the same limiting distribution as the smallest eigenvalue of $W \sim \mathcal{W}_r(N-K-1+r, I_r)$.

We now show that $\Pr[w_r < a] \geq \Pr[x < a]$, where w_r is the smallest eigenvalue of W and $x \sim \chi^2_{N-K}$. By the Bartlett decomposition of a Wishart matrix, we can write

$$W = \begin{bmatrix} W_1 & W_1^{\frac{1}{2}}z \\ z'W_1^{\frac{1}{2}} & x + z'z \end{bmatrix}, \tag{A.51}$$

where $W_1 \sim \mathcal{W}_{r-1}(N-K-2+r, I_{r-1})$, $z \sim \mathcal{N}(0_{r-1}, I_{r-1})$, and they are independent of each other and x. Using the fact that the eigenvalues of W are the same as the reciprocal of the eigenvalues of W^{-1}, it follows that

$$w_r = \min_{\omega:\omega'\omega=1} \omega'W\omega = \left(\max_{\omega:\omega'\omega=1} \omega'W^{-1}\omega\right)^{-1} \leq ([0_{r-1}',\ 1]W^{-1}[0_{r-1}',\ 1]')^{-1} = x \sim \chi^2_{N-K}. \tag{A.52}$$

This completes the proof.

A.4 Proof of Lemma 2

Let $H_t = [1_N,\ D_t]$, $\hat{H} = \frac{1}{T}\sum_{t=1}^T H_t = [1_N,\ \hat{D}]$, $\hat{V}_h = \frac{1}{T}\sum_{t=1}^T \mathrm{vec}(H_t - \hat{H})\mathrm{vec}(H_t - \hat{H})'$, and $Q(\lambda) = [-1,\ \lambda'] \otimes I_N$. Then, we have

$$e_t(\lambda) = Q(\lambda)\mathrm{vec}(H_t), \tag{A.53}$$

$$\hat{W}_e(\lambda) = Q(\lambda)\hat{V}_h Q(\lambda)' = (\lambda' \otimes I_N)\hat{V}_d(\lambda \otimes I_N), \tag{A.54}$$

and

$$\mathcal{J}(\lambda) = T(\hat{D}\lambda - 1_N)'[(\lambda' \otimes I_N)\hat{V}_d(\lambda \otimes I_N)]^{-1}(\hat{D}\lambda - 1_N). \tag{A.55}$$

Let $P = [1_N/\sqrt{N},\ P_1]$ be an orthonormal matrix. Then, we can write

$$\begin{aligned}\mathcal{J}(\lambda) &= T(\hat{D}\lambda - 1_N)'[(\lambda' \otimes I_N)\hat{V}_d(\lambda \otimes I_N)]^{-1}(\hat{D}\lambda - 1_N)\\ &= T(\hat{D}\lambda - 1_N)'P(P'\hat{W}_e(\lambda)P)^{-1}P'(\hat{D}\lambda - 1_N)\\ &= T\begin{bmatrix}\frac{1_N'(\hat{D}\lambda - 1_N)}{\sqrt{N}}\\ P_1'\hat{D}\lambda\end{bmatrix}'\begin{bmatrix}\frac{1_N'\hat{W}_e(\lambda)1_N}{N} & \frac{1_N'\hat{W}_e(\lambda)P_1}{\sqrt{N}}\\ \frac{P_1'\hat{W}_e(\lambda)1_N}{\sqrt{N}} & P_1'\hat{W}_e(\lambda)P_1\end{bmatrix}^{-1}\begin{bmatrix}\frac{1_N'(\hat{D}\lambda - 1_N)}{\sqrt{N}}\\ P_1'\hat{D}\lambda\end{bmatrix}.\end{aligned} \tag{A.56}$$

Denote the matrix in the middle as

$$A \equiv \begin{bmatrix}\frac{1_N'\hat{W}_e(\lambda)1_N}{N} & \frac{1_N'\hat{W}_e(\lambda)P_1}{\sqrt{N}}\\ \frac{P_1'\hat{W}_e(\lambda)1_N}{\sqrt{N}} & P_1'\hat{W}_e(\lambda)P_1\end{bmatrix} = \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}. \tag{A.57}$$

Using the following formula for the inverse of a partitioned matrix

$$A^{-1} = \begin{bmatrix}0 & 0_{N-1}'\\ 0_{N-1} & A_{22}^{-1}\end{bmatrix} + \frac{1}{A_{11} - A_{12}A_{22}^{-1}A_{21}}\begin{bmatrix}-1\\ A_{22}^{-1}A_{21}\end{bmatrix}\begin{bmatrix}-1\\ A_{22}^{-1}A_{21}\end{bmatrix}', \tag{A.58}$$

we obtain

$$\mathcal{J}(\lambda) = \mathcal{CD}_2(\lambda) + \frac{T}{N(A_{11} - A_{12}A_{22}^{-1}A_{21})}[1_N'\hat{W}_e(\lambda)^{-1}P_1(P_1'\hat{W}_e(\lambda)^{-1}P_1)^{-1}P_1'\hat{D}\lambda - 1_N'(\hat{D}\lambda - 1_N)]^2, \tag{A.59}$$

where

$$\mathcal{CD}_2(\lambda) = T\lambda'\hat{D}'P_1[(\lambda' \otimes P_1')\hat{V}_d(\lambda \otimes P_1)]^{-1}P_1'\hat{D}\lambda. \tag{A.60}$$

The above identity suggests that $\mathcal{J}(\lambda) \leq \mathcal{CD}_2(\lambda)$, with the equality holding if and only if

$$1_N'\hat{W}_e(\lambda)^{-1}P_1(P_1'\hat{W}_e(\lambda)^{-1}P_1)^{-1}P_1'\hat{D}\lambda = 1_N'(\hat{D}\lambda - 1_N). \tag{A.61}$$

Note that $\mathcal{CD}_2(\lambda)$ is invariant to multiplying λ by a nonzero constant and, hence, we can only identify the λ that minimizes $\mathcal{CD}_2(\lambda)$ up to a scalar multiple. Consider the following normalization:

$$\hat{c} = \text{argmin}_{c:c'c=1}\mathcal{CD}_2(c). \tag{A.62}$$

Then, we have

$$\mathcal{CD}_2 \equiv \mathcal{CD}_2(\hat{c}) = \mathcal{CD}_2(a\hat{c}) \geq \mathcal{J}(a\hat{c}) \tag{A.63}$$

for any nonzero scalar a. To satisfy the equality in (A.61), we need to set a equal to

$$\hat{a} = \frac{N}{[1_N - P_1(P_1'\hat{W}_e(\hat{c})P_1)^{-1}P_1'\hat{W}_e(\hat{c})1_N]'\hat{D}\hat{c}} = \frac{1_N'\hat{W}_e(\hat{c})^{-1}1_N}{1_N'\hat{W}_e(\hat{c})^{-1}\hat{D}\hat{c}}, \tag{A.64}$$

where the last equality follows from

$$\begin{aligned} 1_N - P_1(P_1'\hat{W}_e(\hat{c})P_1)^{-1}P_1'\hat{W}_e(\hat{c})1_N &= \hat{W}_e(\hat{c})^{-\frac{1}{2}}[I_N - \hat{W}_e(\hat{c})^{\frac{1}{2}}P_1(P_1'\hat{W}_e(\hat{c})P_1)^{-1}P_1'\hat{W}_e(\hat{c})^{\frac{1}{2}}]\hat{W}_e(\hat{c})^{\frac{1}{2}}1_N \\ &= \hat{W}_e(\hat{c})^{-\frac{1}{2}}[\hat{W}_e(\hat{c})^{-\frac{1}{2}}1_N(1_N'\hat{W}_e(\hat{c})^{-1}1_N)^{-1}1_N'\hat{W}_e(\hat{c})^{-\frac{1}{2}}]\hat{W}_e(\hat{c})^{\frac{1}{2}}1_N \\ &= \frac{\hat{W}_e(\hat{c})^{-1}1_N(1_N'1_N)}{1_N'\hat{W}_e(\hat{c})^{-1}1_N} \\ &= \frac{N\hat{W}_e(\hat{c})^{-1}1_N}{1_N'\hat{W}_e(\hat{c})^{-1}1_N}. \end{aligned} \tag{A.65}$$

With this choice of a, we have $\mathcal{J} \equiv \mathcal{J}(\hat{a}\hat{c}) = \mathcal{CD}_2 \leq \mathcal{J}(\lambda)$ and hence $\hat{\lambda} = \hat{a}\hat{c}$. This completes the proof.

A.5 Proof of Theorem 3

part (a): We first perform an alternative parameterization of the problem. Let

$$g_t(c) = H_t \begin{bmatrix} c \\ 1 \end{bmatrix}. \tag{A.66}$$

When the useless factor is ordered last, we have that $E[R_t f_{K-1,t}] = \mu_R \mu_{f,K-1}$ and $H[c^{*\prime},\ 1]' = 0_N$, where

$$c^* = \begin{bmatrix} 0 \\ -\mu_{f,K-1} \\ 0_{K-2} \end{bmatrix}. \tag{A.67}$$

Consider the CU-GMM estimator of c^*:

$$\hat{c} = \text{argmin}_c \bar{g}(c)'\hat{W}_g(c)^{-1}\bar{g}(c), \tag{A.68}$$

where $\bar{g}(c) = \sum_{t=1}^{T} g_t(c)/T$ and

$$\hat{W}_g(c) = \frac{1}{T}\sum_{t=1}^{T}[g_t(c) - \bar{g}(c)][g_t(c) - \bar{g}(c)]'. \tag{A.69}$$

The asymptotic distribution of $\hat{c}$ is given by

$$\sqrt{T}(\hat{c} - c^*) \xrightarrow{d} \mathcal{N}\left(0_K, (H_1' S_g^{-1} H_1)^{-1}\right), \tag{A.70}$$

where

$$S_g = E[g_t(c^*) g_t(c^*)'] = E[R_t R_t'(f_{K-1,t} - \mu_{f,K-1})^2] = U\sigma^2_{f,K-1}, \tag{A.71}$$

and $U = E[R_t R_t']$. Note that $\hat{c}$ has the same asymptotic distribution as the estimator

$$\tilde{c} = (\hat{H}_1'\hat{U}^{-1}\hat{H}_1)^{-1}\hat{H}_1'\hat{U}^{-1}\hat{d}_K, \tag{A.72}$$

where $\hat{H}_1 = [1_N,\ \hat{D}_1]$ with $\hat{D}_1$ being the first $K-1$ columns of $\hat{D}$, and $\hat{d}_K$ being the last column of $\hat{D}$. Let

$$z \sim \mathcal{N}(0_K, \sigma^2_{f,K-1}(H_1' U^{-1} H_1)^{-1}). \tag{A.73}$$

Then, we have

$$\sqrt{T}\hat{c}_1 \xrightarrow{d} z_1, \tag{A.74}$$

$$\sqrt{T}(\hat{c}_2 + \mu_{f,K-1}) \xrightarrow{d} z_2, \tag{A.75}$$

$$\sqrt{T}\hat{c}_i \xrightarrow{d} z_i, \quad i = 3, \ldots, K. \tag{A.76}$$

Due to the invariance property of CU-GMM, we know that $[-1,\ \hat{\lambda}']$ is proportional to $[\hat{c}',\ 1]$. Then, it follows that

$$\hat{\lambda}_0 = -\frac{\hat{c}_2}{\hat{c}_1}, \tag{A.77}$$

$$\hat{\lambda}_{1,i} = -\frac{\hat{c}_{i+2}}{\hat{c}_1}, \quad i = 1, \ldots, K-2, \tag{A.78}$$

$$\hat{\lambda}_{1,K-1} = -\frac{1}{\hat{c}_1}. \tag{A.79}$$

Therefore, the limiting distributions of the $K-1$ elements of $\hat{\lambda}_1$ are given by

$$\hat{\lambda}_{1,i} \xrightarrow{d} -\frac{z_{i+2}}{z_1}, \quad i = 1, \ldots, K-2, \tag{A.80}$$

$$\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{z_1}. \tag{A.81}$$

The limiting distribution of $\hat{\lambda}_0$ depends on whether $\mu_{f,K-1} = 0$ or not. If $\mu_{f,K-1} = 0$, we have $\hat{\lambda}_0 \xrightarrow{d} -z_2/z_1$. If $\mu_{f,K-1} \neq 0$, we have $\hat{\lambda}_0/\sqrt{T} \xrightarrow{d} \mu_{f,K-1}/z_1$. This completes the proof of part (a).

part (b): It is easy to show that

$$\sigma_1^2 = \frac{\sigma_{f,K-1}^2}{1_N'[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]1_N} = \frac{\sigma_{f,K-1}^2}{\delta^2}, \tag{A.82}$$

where δ is the HJ-distance of the misspecified model. Then,

$$\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{\sigma_1 \tilde{z}_1} = -\frac{\delta}{\sigma_{f,K-1}\tilde{z}_1}, \tag{A.83}$$

where $\tilde{z}_1 = z_1/\sigma_1 \sim \mathcal{N}(0,1)$. In addition, using the same proof as in Theorem 1 part (b), we can write

$$z_{i+1} = \sigma_{i+1}\left(\rho_{1,i+1}\tilde{z}_1 + \sqrt{1-\rho_{1,i+1}^2}\, q_{i+1}\right), \tag{A.84}$$

where $q_{i+1} \sim \mathcal{N}(0,1)$ and it is independent of $\tilde{z}_1$.

Using the fact that

$$\frac{e_t(\hat{\lambda})}{\sqrt{T}} = -\frac{R_t(f_{K-1,t} - \mu_{f,K-1})}{z_1} + O_p(T^{-\frac{1}{2}}), \tag{A.85}$$

we can show that

$$\frac{\hat{W}_e(\hat{\lambda})}{T} = \frac{\sigma_{f,K-1}^2}{z_1^2}U + o_p(1). \tag{A.86}$$

This allows us to show that the squared t-ratio of $\hat{\lambda}_{1,K-1}$ can be expressed as

$$t^2(\hat{\lambda}_{1,K-1}) = \frac{T\hat{\lambda}_{1,K-1}^2}{\boldsymbol{\iota}_{K,K}'(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{K,K}} = \frac{T\hat{d}_K'[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]\hat{d}_K}{\sigma_{f,K-1}^2} + o_p(1). \tag{A.87}$$

Let

$$\tilde{d}_K = \hat{d}_K - \hat{\mu}_R\mu_{f,K-1} = \frac{1}{T}\sum_{t=1}^{T} R_t(f_{K-1,t} - \mu_{f,K-1}). \tag{A.88}$$

Then, we have $\sqrt{T}\tilde{d}_K \xrightarrow{d} \mathcal{N}(0_N, \sigma_{f,K-1}^2 U)$. Since $\hat{\mu}_R\mu_{f,K-1} = \hat{D}_1[\mu_{f,K-1},\ 0_{K-2}']'$, it follows that

$$\begin{aligned} & T\hat{d}_K'[U^{-1} - U^{-1}\hat{D}_1(\hat{D}_1'U^{-1}\hat{D}_1)^{-1}\hat{D}_1'U^{-1}]\hat{d}_K \\ &= T\tilde{d}_K'[U^{-1} - U^{-1}\hat{D}_1(\hat{D}_1'U^{-1}\hat{D}_1)^{-1}\hat{D}_1'U^{-1}]\tilde{d}_K \\ &= T\tilde{d}_K'[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]\tilde{d}_K + o_p(1). \end{aligned} \tag{A.89}$$

Let P_U be an $N \times (N-K+1)$ orthonormal matrix with its columns orthogonal to $U^{-\frac{1}{2}}D_1$. Then, we have

$$\frac{1}{\sigma_{f,K-1}}\sqrt{T}P_U'U^{-\frac{1}{2}}\tilde{d}_K \xrightarrow{d} \mathcal{N}(0_{N-K+1}, I_{N-K+1}) \tag{A.90}$$

and

$$t^2(\hat{\lambda}_{1,K-1}) \xrightarrow{d} \chi^2_{N-K+1}. \tag{A.91}$$

For the derivation of the limiting distributions for $t(\hat{\lambda}_0)$ and $t(\hat{\lambda}_{1,i})$ $(i = 1, \ldots, K-2)$, we use the identity

$$I_N - U^{-\frac{1}{2}}H_1(H_1'U^{-1}H_1)^{-1}H_1'U^{-\frac{1}{2}} = I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}} - hh', \tag{A.92}$$

where

$$h = \frac{[I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}1_N}{\delta} = \frac{P_U P_U' U^{-\frac{1}{2}}1_N}{\delta} \tag{A.93}$$

and $h'h = 1$. Note that

$$\sqrt{T}h'U^{-\frac{1}{2}}\tilde{d}_K/\sigma_{f,K-1} = \sqrt{T}1_N'U^{-\frac{1}{2}}P_U P_U'U^{-\frac{1}{2}}\tilde{d}_K/(\sigma_{f,K-1}\delta) \xrightarrow{d} \tilde{z}_1 \sim \mathcal{N}(0,1), \tag{A.94}$$

$$T\tilde{d}_K'[I_N - U^{-\frac{1}{2}}H_1(H_1'U^{-1}H_1)^{-1}H_1'U^{-\frac{1}{2}}]\tilde{d}_K/\sigma^2_{f,K-1} \xrightarrow{d} x \sim \chi^2_{N-K}, \tag{A.95}$$

and they are independent of each other. Using the formula for the inverse of a partitioned matrix, we can show that

$$\sigma^2_{f,K-1}\boldsymbol{\iota}_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,i} = \sigma^2_{i+1} - \frac{\sigma_{1,i+1}}{\sigma_1^2} = \sigma^2_{i+1}(1-\rho^2_{1,i+1}). \tag{A.96}$$

In addition, we can easily show that for $i = 2, \ldots, K-1$

$$\frac{\sqrt{T}\boldsymbol{\iota}_{K-1,i}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K}{\sigma_{f,K-1}[\boldsymbol{\iota}_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,i}]^{\frac{1}{2}}} \xrightarrow{d} q_{i+1} \sim \mathcal{N}(0,1). \tag{A.97}$$

For $i = 1$, the result depends on whether $\mu_{f,K-1} = 0$ or not. If $\mu_{f,K-1} = 0$, we have $\sqrt{T}U^{-\frac{1}{2}}\hat{d}_K/\sigma_{f,K-1} \xrightarrow{d} \mathcal{N}(0_N, I_N)$ and hence

$$\frac{\sqrt{T}\boldsymbol{\iota}_{K-1,1}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K}{\sigma_{f,K-1}[\boldsymbol{\iota}_{K-1,1}'(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} \xrightarrow{d} q_2 \sim \mathcal{N}(0,1). \tag{A.98}$$

If $\mu_{f,K-1} \neq 0$, we have $\hat{d}_K \xrightarrow{p} \mu_R\mu_{f,K-1} = D_1[\mu_{f,K-1},\ 0_{K-2}']'$ and hence

$$\begin{aligned}\boldsymbol{\iota}_{K-1,1}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K/\sigma_{f,K-1} &\xrightarrow{p} \boldsymbol{\iota}_{K-1,1}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}D_1[\mu_{f,K-1},\ 0_{K-2}']'/\sigma_{f,K-1} \\ &= \mu_{f,K-1}/\sigma_{f,K-1}.\end{aligned} \tag{A.99}$$

Note that the q_i's are independent of $\tilde{z}_1$ and x.

Consider the upper left $(K-1)\times(K-1)$ submatrix of $(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}/T$, which has the same limit as

$$\frac{\sigma^2_{f,K-1}}{z_1^2}\left[(D_1'U^{-1}D_1)^{-1}+\frac{(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K\hat{d}_K'U^{-1}D_1(D_1'U^{-1}D_1)^{-1}}{\tilde{d}_K'U^{-\frac{1}{2}}[I_N-U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}\tilde{d}_K}\right]. \tag{A.100}$$

In particular, for $i=2,\ldots,K-1$, the i-th diagonal element of this matrix has a limiting distribution

$$\frac{\sigma^2_{f,K-1}\boldsymbol{\iota}_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,i}}{z_1^2}\left(1+\frac{q_{i+1}^2}{x+\tilde{z}_1^2}\right)=\frac{\sigma^2_{i+1}(1-\rho^2_{1,i+1})}{z_1^2}\left(1+\frac{q_{i+1}^2}{x+\tilde{z}_1^2}\right). \tag{A.101}$$

Let

$$b_{i+1}=\frac{x+\tilde{z}_1^2}{q_{i+1}^2+x+\tilde{z}_1^2}. \tag{A.102}$$

Then, we can write the limiting distribution of $t(\hat{\lambda}_{1,i})$ for $i=1,\ldots,K-2$ as

$$\begin{aligned} t(\hat{\lambda}_{1,i}) &\stackrel{d}{\to} -\frac{z_{i+2}/z_1}{\sigma_{i+2}\sqrt{1-\rho^2_{1,i+2}}|z_1|b_{i+1}^{-\frac{1}{2}}} \\ &= -\frac{|\tilde{z}_1|}{\tilde{z}_1}\frac{\left(\rho_{1,i+2}\tilde{z}_1+\sqrt{1-\rho^2_{1,i+2}}q_{i+2}\right)b_{i+1}^{\frac{1}{2}}}{\sqrt{1-\rho^2_{1,i+2}}} \\ &= -\left(\frac{\rho_{1,i+2}|\tilde{z}_1|}{\sqrt{1-\rho^2_{1,i+2}}}+q_{i+2}\right)b_{i+1}^{\frac{1}{2}}. \end{aligned} \tag{A.103}$$

The limiting distribution of $t(\hat{\lambda}_0)$ depends on whether $\mu_{f,K-1}=0$ or not. If $\mu_{f,K-1}=0$, we have a similar limiting expression

$$t(\hat{\lambda}_0)\stackrel{d}{\to}-\left(\frac{\rho_{1,2}|\tilde{z}_1|}{\sqrt{1-\rho^2_{1,2}}}+q_2\right)b_1^{\frac{1}{2}}. \tag{A.104}$$

If $\mu_{f,K-1}\neq 0$, we have

$$t^2(\hat{\lambda}_0)\stackrel{d}{\to}\frac{\frac{\mu^2_{f,K-1}}{z_1^2}}{\frac{\sigma^2_{f,K-1}}{z_1^2}\left[\frac{\mu^2_{f,K-1}}{\sigma^2_{f,K-1}(x+\tilde{z}_1^2)}\right]}=x+\tilde{z}_1^2\sim\chi^2_{N-K+1}. \tag{A.105}$$

This completes the proof of part (b).

part (c): The proof follows similar arguments as the proof of part (c) in Theorem 1 by replacing the expression for $\hat{\beta}^{ML}$ with the expression for $\hat{\beta}^{CU}$ and, to conserve space, is omitted.

part (d): Because D is of reduced rank due to the presence of a useless factor, $\mathcal{CD}_2 \stackrel{d}{\to} \chi^2_{N-K}$. From the numerical equivalence of the $\mathcal{J}$ and $\mathcal{CD}_2$ tests in Lemma 2, we have $\mathcal{J} \stackrel{d}{\to} \chi^2_{N-K}$ even when the model is misspecified. This completes the proof of part (d).

A.6 Proof of Theorem 4

First, note that

$$\hat{B} = \hat{D}\left(\frac{X'X}{T}\right)^{-1} \tag{A.106}$$

and the $\mathcal{CD}_1$ specification test in Section 3.1 can be rewritten as

$$\mathcal{CD}_1 = T\min_c \frac{c'\hat{B}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{B}c}{c'[(X'X/T)^{-1}]c} = T\min_c \frac{c'\hat{D}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{D}c}{c'(X'X/T)c}. \tag{A.107}$$

Also, recall that $\mathcal{CD}_1$ is numerically equivalent to $\mathcal{S}$ whose asymptotic distribution is presented in Theorem 2.

Using (A.106) and Assumption 3', we obtain

$$\sqrt{T}\text{vec}(P_1'\hat{D} - P_1'D) \stackrel{d}{\to} \mathcal{N}\left(0_{(N-1)K}, E[x_tx_t'] \otimes P_1'\Sigma P_1 + (I_K \otimes P_1'B)V_x(I_K \otimes B'P_1)\right), \tag{A.108}$$

where V_x is the asymptotic variance of $\sqrt{T}\text{vec}\left((X'X)/T - E[x_tx_t']\right)$. For any c such that $P_1'Dc = 0_{N-1}$, we have

$$\sqrt{T}P_1'\hat{D}c \stackrel{d}{\to} \mathcal{N}\left(0_{N-1}, c'E[x_tx_t']cP_1'\Sigma P_1\right) \tag{A.109}$$

since

$$\lim_{T\to\infty} P_1'B\left(\frac{X'X}{T}\right)c = P_1'BE[x_tx_t']c = P_1'Dc = 0_{N-1}. \tag{A.110}$$

Hence, under Assumption 3' and when H (or equivalently $P_1'D$) has a reduced rank, we have

$$\begin{aligned}\mathcal{CD}_2 &= T\min_c (P_1'\hat{D}c)'[(c' \otimes P_1')\hat{V}_d(c \otimes P_1)]^{-1}(P_1'\hat{D}c) \\ &= T\min_c \frac{c'\hat{D}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{D}c}{c'(X'X/T)c} + o_p(1) = \mathcal{CD}_1 + o_p(1).\end{aligned} \tag{A.111}$$

Therefore, the $\mathcal{J}$ test, which is numerically equivalent to the $\mathcal{CD}_2$ test (see Lemma 2), shares the same asymptotic distribution of the $\mathcal{S}$ test in Theorem 2. This completes the proof.

A.7 Proof of Theorem 5

part (a): The proof closely follows the proof of part (b) of Theorem 3 in Gospodinov, Kan, and Robotti (2014b) and is provided here for completeness.[14] Let $\tilde{D} = U^{-\frac{1}{2}}D_1$, $\tilde{1}_N = U^{-\frac{1}{2}}1_N$, P_U be an $N \times (N-K+1)$ orthonormal matrix whose columns are orthogonal to $\tilde{D}$ so that $P_U P_U' = I_N - \tilde{D}(\tilde{D}'\tilde{D})^{-1}\tilde{D}'$, $u \sim \mathcal{N}(0_N, I_N)$, and $w = P_U' u \sim \mathcal{N}(0_{N-K+1}, I_{N-K+1})$. When the model contains a useless factor, we have (see Gospodinov, Kan, and Robotti, 2014b)

$$-\hat{U}^{-\frac{1}{2}}\bar{e}(\hat{\lambda}^{HJ}) \stackrel{d}{\to} P_U[I_{N-K+1} - w(w'w)^{-1}w']P_U'\tilde{1}_N, \tag{A.112}$$

and the limiting distribution of $\hat{\delta}^2$ is given by

$$\begin{aligned}\hat{\delta}^2 &\stackrel{d}{\to} \tilde{1}_N' P_U[I_{N-K+1} - w(w'w)^{-1}w']P_U'\tilde{1}_N \\ &= (\tilde{1}_N' P_U P_U' \tilde{1}_N)\frac{w'[I_{N-K+1} - P_U'\tilde{1}_N(\tilde{1}_N' P_U P_U'\tilde{1}_N)^{-1}\tilde{1}_N' P_U]w}{w'w} = \delta^2\mathcal{B},\end{aligned} \tag{A.113}$$

where

$$\mathcal{B} = \frac{w'[I_{N-K+1} - P_U'\tilde{1}_N(\tilde{1}_N' P_U P_U'\tilde{1}_N)^{-1}\tilde{1}_N' P_U]w}{w'w} \sim \text{Beta}\left(\frac{N-K}{2}, \frac{1}{2}\right) \tag{A.114}$$

and it is independent of w. Furthermore, the estimated weights $\hat{\xi}_i$ for the weighted chi-squared distribution are asymptotically distributed as (see Gospodinov, Kan, and Robotti, 2014b)

$$\frac{\hat{\xi}_i}{T} \stackrel{d}{\to} \frac{\delta^2(1-\mathcal{B})}{w'w}. \tag{A.115}$$

When comparing $T\hat{\delta}^2$ with $F_{N-K}(\hat{\xi})$, we are effectively comparing $\mathcal{B}$ with $(1-\mathcal{B})/(w'w)\chi^2_{N-K}$, and we will reject $H_0 : \delta^2 = 0$ when

$$w'w > \frac{p_\eta \mathcal{B}}{1-\mathcal{B}}. \tag{A.116}$$

Note that $w'w \sim \chi^2_{N-K+1}$ and it is independent of $\mathcal{B}$. Therefore, the limiting probability of rejection for a test with size η is

$$\int_0^1 P\left[\chi^2_{N-K+1} > \frac{p_\eta s}{1-s}\right] f_{\mathcal{B}}(s)\mathrm{d}s. \tag{A.117}$$

This completes the proof of part (a).

[14] The results in Gospodinov, Kan, and Robotti (2014b) are derived under the assumption that the mean and the variance of the useless factor are 0 and 1, respectively. However, since the sample pricing errors and the sample HJ-distance are invariant to affine transformations of the factors, their results also apply when the useless factor has a generic mean and a generic variance.

part (b): Denote by $f_t = [f_{1,t}, f_{2,t}, \ldots, f_{K-1,t}]' \equiv [\tilde{f}_t',\ f_{K-1,t}]'$ the $(K-1)$-vector of useful and useless factors, where the last factor is assumed to be useless. We denote by $\mu_{\tilde{f}}$ and $V_{\tilde{f}}$ the mean and covariance matrix of the useful factors $\tilde{f}_t$. Since $\hat{\mathcal{Q}}$ is invariant to linear transformations of the factors, without loss of generality we assume that the useless factor $f_{K-1,t}$ has zero mean and unit variance. The GLS CSR estimator of γ^* is defined as

$$\hat{\gamma}^{GLS} = (\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\hat{B}_1'\hat{\Sigma}^{-1}\hat{\mu}_R, \tag{A.118}$$

where $\hat{B}_1 = [\hat{G}_2,\ \hat{\beta}_{K-1}]$ with $\hat{G}_2 = [1_N,\ \hat{\beta}_1, \ldots, \hat{\beta}_{K-2}]$. Note that the estimator in (A.118) can be obtained equivalently by running an OLS regression of $\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R$ on $\hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2$ and $\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}$. To obtain $\hat{\gamma}^{GLS}_{1,K-1}$, we can project $\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R$ and $\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}$ on $\hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2$, and then regress the residuals from the first projection onto the residuals from the second projection. It follows that

$$\hat{\gamma}^{GLS}_{1,K-1} = \frac{\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R}{\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}}. \tag{A.119}$$

Similarly, the first $(K-1)$ sub-vector of $\hat{\gamma}^{GLS}$, denoted by $\hat{\gamma}^{GLS}_u \equiv [\hat{\gamma}_0,\ \hat{\gamma}_{1,1}, \ldots, \hat{\gamma}_{1,K-2}]'$, is obtained by projecting $\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R$ and $\hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2$ on $\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}$, and then regressing the residuals from the first projection onto the residuals from the second projection, which yields

$$\begin{aligned}\hat{\gamma}^{GLS}_u &= (\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}(\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{\beta}_{K-1})^{-1}\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2)^{-1} \\ &\quad \times \hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}(\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{\beta}_{K-1})^{-1}\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R. \end{aligned} \tag{A.120}$$

We adopt the following simplifying notation. Let $\tilde{G}_2 = \Sigma^{-\frac{1}{2}}G_2$, $\tilde{\mu}_R = \Sigma^{-\frac{1}{2}}\mu_R$, $M = I_N - \tilde{G}_2(\tilde{G}_2'\tilde{G}_2)^{-1}\tilde{G}_2'$, $u \sim \mathcal{N}(0_N, I_N)$, and denote by P_Σ an $N \times (N-K+1)$ orthonormal matrix whose columns are orthogonal to $\tilde{G}_2$ so that $P_\Sigma P_\Sigma' = I_N - \tilde{G}_2(\tilde{G}_2'\tilde{G}_2)^{-1}\tilde{G}_2'$. In addition, let $w = P_\Sigma' u \sim N(0_{N-K+1}, I_{N-K+1})$. Then, we have[15]

$$\sqrt{T}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1} \xrightarrow{d} u, \tag{A.121}$$

$$\sqrt{T}P_\Sigma'\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1} \xrightarrow{d} w, \tag{A.122}$$

$$\hat{M} = I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}} \xrightarrow{p} M = P_\Sigma P_\Sigma', \tag{A.123}$$

$$\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R \xrightarrow{p} \tilde{\mu}_R, \tag{A.124}$$

[15]Note that (A.121) holds under conditional heteroskedasticity because the returns are assumed to be independent of $f_{K-1,t}$.

and

$$T^{-\frac{1}{2}}\hat{\gamma}_{1,K-1}^{GLS} = \frac{(\sqrt{T}\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}})\hat{M}\hat{\Sigma}^{-\frac{1}{2}}\hat{\mu}_R}{(\sqrt{T}\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}})\hat{M}(\sqrt{T}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1})} \stackrel{d}{\to} \frac{u'M\tilde{\mu}_R}{u'Mu} = \frac{\sqrt{\mathcal{Q}}s}{w'w}, \tag{A.125}$$

where $\mathcal{Q} = \tilde{\mu}_R' P_\Sigma P_\Sigma' \tilde{\mu}_R$ and $s = \tilde{\mu}_R' P_\Sigma P_\Sigma' u/\sqrt{\mathcal{Q}} \sim \mathcal{N}(0,1)$. Similarly, since $\hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2 \stackrel{p}{\to} \tilde{G}_2$, we have

$$\hat{\gamma}_u^{GLS} \stackrel{d}{\to} (\tilde{G}_2'[I_N - u(u'u)^{-1}u']\tilde{G}_2)^{-1}\tilde{G}_2'[I_N - u(u'u)^{-1}u']\tilde{\mu}_R = O_p(1). \tag{A.126}$$

Using the identity

$$I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{B}_1(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\hat{B}_1'\hat{\Sigma}^{-\frac{1}{2}} = \hat{M} - \hat{M}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1}(\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}}\hat{M}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1})^{-1}\hat{\beta}_{K-1}'\hat{\Sigma}^{-\frac{1}{2}}\hat{M}, \tag{A.127}$$

(A.122)–(A.124), and (A.127), we obtain

$$\hat{\Sigma}^{-\frac{1}{2}}\hat{e} = \hat{\Sigma}^{-\frac{1}{2}}[\hat{\mu}_R - \hat{B}_1(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\hat{B}_1'\hat{\Sigma}^{-1}\hat{\mu}_R] \stackrel{d}{\to} P_\Sigma[I_{N-K+1} - w(w'w)^{-1}w']P_\Sigma'\tilde{\mu}_R. \tag{A.128}$$

After simplification, we have

$$\hat{\mathcal{Q}} = \hat{e}'\hat{\Sigma}^{-1}\hat{e} \stackrel{d}{\to} \mathcal{Q}\mathcal{B}, \tag{A.129}$$

where

$$\mathcal{B} = \frac{w'[I_{N-K+1} - P_\Sigma'\tilde{\mu}_R(\tilde{\mu}_R'P_\Sigma P_\Sigma'\tilde{\mu}_R)^{-1}\tilde{\mu}_R'P_\Sigma]w}{w'w} \sim \text{Beta}\left(\frac{N-K}{2}, \frac{1}{2}\right) \tag{A.130}$$

and it is independent of w. To characterize the asymptotic distribution of the CSR test, we let $P_{\tilde{V}}$ be an $N \times (N-K)$ orthonormal matrix with columns orthogonal to $\tilde{V} = \Sigma^{-\frac{1}{2}}[1_N,\ V_{Rf}]$ and denote by $\hat{P}_{\tilde{V}}$ its sample counterpart. In addition, denote by $\hat{\mu}_f$ and $\hat{V}_f$ the sample mean and covariance matrix of f_t, and let $\hat{\epsilon}_t = (R_t - \hat{\mu}_R) - \hat{\beta}(f_t - \hat{\mu}_f)$. When testing $H_0 : \mathcal{Q} = 0$, we compare $T\hat{\mathcal{Q}}$ with $F_{N-K}(\hat{\xi})$, where the estimated weights, $\hat{\xi}_i$, are the eigenvalues of the matrix

$$\hat{P}_{\tilde{V}}'\hat{\Sigma}^{-\frac{1}{2}}\hat{S}_l\hat{\Sigma}^{-\frac{1}{2}}\hat{P}_{\tilde{V}}, \tag{A.131}$$

with

$$\hat{S}_l = \frac{1}{T}\sum_{t=1}^{T}\hat{l}_t\hat{l}_t' \tag{A.132}$$

and

$$\hat{l}_t = \hat{\epsilon}_t[1 - \hat{\gamma}^{GLS\prime}\hat{V}_f^{-1}(f_t - \hat{\mu}_f)]. \tag{A.133}$$

We first need to determine the limiting behavior of $\hat{S}_l/T$ when the model contains a useless factor. Using that

$$\hat{V}_f^{-1} \stackrel{p}{\to} \begin{bmatrix} V_{\tilde{f}}^{-1} & 0_{K-2} \\ 0_{K-2}' & 1 \end{bmatrix}, \tag{A.134}$$

we have

$$
\begin{aligned}
\frac{\hat{l}_t}{\sqrt{T}} &= \frac{1}{\sqrt{T}}\hat{\epsilon}_t[1-\hat{\gamma}^{GLS\prime}\hat{V}_f^{-1}(f_t-\hat{\mu}_f)] \\
&= \epsilon_t\left[\frac{1}{\sqrt{T}}-\frac{\hat{\gamma}_u^{GLS\prime}}{\sqrt{T}}V_{\tilde{f}}^{-1}(\tilde{f}_t-\hat{\mu}_{\tilde{f}})-\frac{\hat{\gamma}_{1,K-1}^{GLS}}{\sqrt{T}}f_{K-1,t}\right]+o_p(1) \\
&\stackrel{d}{\to} -\epsilon_t\frac{f_{K-1,t}\sqrt{\mathcal{Q}}s}{w'w},
\end{aligned}
\tag{A.135}
$$

where $\epsilon_t = R_t-\mu_R-\beta(f_t-\mu_f)$. It follows that

$$
\frac{\hat{S}_l}{T} \stackrel{d}{\to} \frac{\mathcal{Q}s^2}{(w'w)^2}\Sigma. \tag{A.136}
$$

This implies that

$$
\frac{\hat{P}_{\tilde{V}}'\hat{\Sigma}^{-\frac{1}{2}}\hat{S}_l\hat{\Sigma}^{-\frac{1}{2}}\hat{P}_{\tilde{V}}}{T} \stackrel{d}{\to} \frac{\mathcal{Q}s^2}{(w'w)^2}I_{N-K} \tag{A.137}
$$

and

$$
\frac{\hat{\xi}_i}{T} \stackrel{d}{\to} \frac{\mathcal{Q}s^2}{(w'w)^2} = \frac{\mathcal{Q}(1-\mathcal{B})}{w'w}. \tag{A.138}
$$

Following similar arguments as in the proof of part (a), we obtain the limiting probability of rejection for the CSR test. This completes the proof of part (b).

References

[1] Ahn, S. C., A. R. Horenstein, and N. Wang (2010): "Determining the Rank of the Beta Matrix in a Factor Model with Factor-Candidate Regressors," Working Paper.

[2] Almeida, C., and R. Garcia (2012): "Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators," *Journal of Econometrics*, 170, 519–537.

[3] Anderson, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, 327–351.

[4] Arellano, M., L. P. Hansen, and E. Sentana (2012): "Underidentification?," *Journal of Econometrics*, 170, 256–280.

[5] Bekker, P., P. Dobbelstein, and T. Wansbeek (1996): "The APT Model as Reduced-Rank Regression," *Journal of Business and Economic Statistics*, 14, 199–202.

[6] Burnside, C. A. (2012): "Identification and Inference in Linear Stochastic Discount Factor Models," Working Paper.

[7] Cragg, J. G., and S. G. Donald (1996): "Testing Overidentifying Restrictions in Unidentified Models," Working Paper.

[8] Cragg, J. G., and S. G. Donald (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250.

[9] Dovonon, P., and E. Renault (2013): "Testing for Common Conditionally Heteroskedastic Factors," *Econometrica*, 81, 2561–2586.

[10] 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.

[11] Gallant, R. A. (1987): *Nonlinear Statistical Models*. Wiley, New York.

[12] Gospodinov, N., R. Kan, and C. Robotti (2013): "Misspecification-Robust Standard Errors for the Continuously-Updated GMM Estimator," Working Paper.

[13] Gospodinov, N., R. Kan, and C. Robotti (2014a): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *Review of Financial Studies*, 27, 2139–2170.

[14] Gospodinov, N., R. Kan, and C. Robotti (2014b): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," Internet Appendix.

[15] Granger, C. W. J., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

[16] Guggenberger, P., and R. J. Smith (2008): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," *Journal of Econometrics*, 142, 134–161.

[17] Hall, A. R. (2005): *Generalized Method of Moments*. Oxford University Press, Oxford.

[18] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

[19] Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

[20] Hansen, L. P., and R. Jagannathan (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.

[21] Harvey, C. R., Y. Liu, and H. Zhu (2013): "...And the Cross-Section of Expected Returns," Working Paper.

[22] Jagannathan, R., and Z. Wang (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

[23] Kan, R., and C. Robotti (2012): "Evaluation of Asset Pricing Models Using Two-Pass Cross-Sectional Regressions," in *Handbook of Computational Finance*, Chapter 9, ed. by J. C. Duan, J. E. Gentle, and W. K. Härdle, Berlin Heidelberg: Springer-Verlag, 223–251.

[24] 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.

[25] Kan, R., and C. Zhang (1999a): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203–235.

[26] Kan, R., and C. Zhang (1999b): "GMM Tests of Stochastic Discount Factor Models with Useless Factors," *Journal of Financial Economics*, 54, 103–127.

[27] Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149–173.

[28] Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics*, 27, 293–311.

[29] Kleibergen, F., and R. Paap (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

[30] Kleibergen, F., and Z. Zhan (2013): "Unexplained Factors and Their Effects on Second Pass R-squared's and t-tests," Working Paper.

[31] Koopmans, T. C., and W. C. Hood (1953): "The Estimation of Simultaneous Linear Economic Relationships," in *Studies in Econometric Method* (Cowles Commission Monograph No. 14), ed. by W. C. Hood and T. C. Koopmans, New York: Wiley, 112–199.

[32] Lettau, M., and S. C. Ludvigson (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1287.

[33] Lewellen, J. W., S. Nagel, and J. Shanken (2010): "A Skeptical Appraisal of Asset-Pricing Tests," *Journal of Financial Economics*, 96, 175–194.

[34] Maasoumi, E., and P. C. B. Phillips (1982): "On the Behavior of Inconsistent Instrumental Variable Estimators," *Journal of Econometrics*, 19, 183–201.

[35] Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

[36] Otsu, T. (2006): "Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models under Weak Identification," *Econometric Theory*, 22, 513–527.

[37] Peñaranda, F., and E. Sentana (2014): "A Unifying Approach to the Empirical Evaluation of Asset Pricing Models," *Review of Economics and Statistics*, forthcoming.

[38] Phillips, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

[39] Robin, J.-M., and R. J. Smith (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175.

[40] Schennach, S. M. (2007): "Point Estimation with Exponentially Tilted Empirical Likelihood," *Annals of Statistics*, 35, 634–672.

[41] Shanken, J. (1985): "Multivariate Tests of the Zero-beta CAPM," *Journal of Financial Economics*, 14, 327–348.

[42] Shanken, J. (1992): "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1–33.

[43] 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.

[44] Zhou, G. (1995): "Small Sample Rank Tests with Applications to Asset Pricing," *Journal of Empirical Finance*, 2, 71–93.

Table I
Test Statistics for Various Asset-Pricing Models

The table reports test statistics for the four asset-pricing models (CAPM, FF3, C-LAB, and CC-CAY) described in Section 1. $\mathcal{CD}_{SDF}$ and $\mathcal{CD}_{Beta}$ denote the Cragg and Donald (1997) test for the null of a reduced rank in the SDF and beta-pricing setups, respectively. HJD and CSR denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the GLS cross-sectional regression test of Shanken (1985). $\mathcal{J}$ denotes Hansen, Heaton, and Yaron's (1996) test for over-identifying restrictions based on the CU-GMM estimator. $\mathcal{S}$ denotes Shanken's (1985) Wald-type test of correct model specification based on the ML estimator. t_x denotes the t-test of statistical significance for the parameter associated with factor x, with standard errors computed under the assumption of correct model specification. Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: Rank, HJD, and CSR Tests

	CAPM	FF3	C-LAB	CC-CAY
$\mathcal{CD}_{SDF}$	133.43	86.18	20.82	10.44
(p-value)	(0.0000)	(0.0000)	(0.5320)	(0.9818)
HJD	67.67	51.15	66.36	69.15
(p-value)	(0.0000)	(0.0024)	(0.0000)	(0.0005)
$\mathcal{CD}_{Beta}$	421.73	183.61	24.05	13.27
(p-value)	(0.0000)	(0.0000)	(0.3448)	(0.9255)
CSR	69.96	53.00	68.60	70.56
(p-value)	(0.0000)	(0.0020)	(0.0000)	(0.0005)

Panel B: CU-GMM

	CAPM	FF3	C-LAB	CC-CAY
$\mathcal{J}$	64.58	45.10	20.58	10.57
(p-value)	(0.0000)	(0.0017)	(0.4848)	(0.9705)
t_{vw}	4.29	3.92	−0.93	
t_{smb}		−4.23		
t_{hml}		−2.01		
t_{labor}			4.32	
t_{prem}			2.82	
t_{cg}				1.46
t_{cay}				0.85
$t_{cg \cdot cay}$				−3.22
R^2	0.1999	0.7847	0.9595	0.9952

Panel C: ML

	CAPM	FF3	C-LAB	CC-CAY
$\mathcal{S}$	67.66	49.06	23.10	11.58
(p-value)	(0.0000)	(0.0005)	(0.3388)	(0.9503)
t_{vw}	−3.24	−3.43	−1.34	
t_{smb}		2.08		
t_{hml}		2.33		
t_{labor}			2.81	
t_{prem}			4.21	
t_{cg}				−0.90
t_{cay}				0.76
$t_{cg \cdot cay}$				3.45
R^2	0.1346	0.7677	0.9997	0.9994

Table II.a
Rejection Rates of Specification Tests (SDF Setup)

The table presents the rejection rates of specification tests under misspecified models. $\mathcal{J}$ denotes Hansen, Heaton, and Yaron's (1996) test for over-identifying restrictions based on the CU-GMM estimator. HJD denotes the Hansen and Jagannathan's (1997) distance test of correct model specification. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns for the period 1959:2–2012:12. The $\mathcal{J}$ test statistic is compared with the critical values from a χ^2_{N-K} distribution. The HJ-distance test is compared with the critical values from a weighted chi-squared distribution as in Jagannathan and Wang (1996). The rejection rates for the limiting case ($T = \infty$) in Panels B and C are based on the results in part (d) of Theorem 3 and part (a) of Theorem 5.

	$\mathcal{J}$ Test			HJD Test		
T	10%	5%	1%	10%	5%	1%
Panel A: Model with a Useful Factor Only						
200	0.900	0.831	0.635	0.915	0.856	0.690
400	0.996	0.991	0.963	0.997	0.992	0.967
600	1.000	1.000	0.999	1.000	1.000	0.999
800	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000
3600	1.000	1.000	1.000	1.000	1.000	1.000
∞	1.000	1.000	1.000	1.000	1.000	1.000
Panel B: Model with a Useless Factor Only						
200	0.130	0.063	0.011	0.877	0.810	0.637
400	0.118	0.058	0.011	0.976	0.961	0.910
600	0.113	0.057	0.011	0.990	0.984	0.965
800	0.111	0.055	0.011	0.994	0.990	0.979
1000	0.110	0.055	0.011	0.995	0.993	0.985
3600	0.103	0.052	0.010	0.998	0.997	0.995
∞	0.100	0.050	0.010	0.999	0.998	0.997
Panel C: Model with a Useful and a Useless Factor						
200	0.101	0.046	0.006	0.829	0.747	0.548
400	0.111	0.054	0.009	0.963	0.940	0.864
600	0.110	0.054	0.010	0.985	0.976	0.950
800	0.108	0.054	0.010	0.991	0.986	0.971
1000	0.105	0.053	0.011	0.993	0.990	0.980
3600	0.102	0.051	0.010	0.997	0.997	0.993
∞	0.100	0.050	0.010	0.999	0.998	0.996

Table II.b
Rejection Rates of Specification Tests (Beta-Pricing Setup)

The table presents the rejection rates of specification tests under misspecified models. $\mathcal{S}$ denotes Shanken's (1985) Wald-type test of correct model specification based on the ML estimator. CSR denotes Shanken's (1985) cross-sectional regression test of correct model specification based on the non-invariant GLS estimator. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns for the period 1959:2–2012:12. The $\mathcal{S}$ test statistic is compared with the critical values from a χ^2_{N-K} distribution. The CSR test is compared with the critical values from a weighted chi-squared distribution as shown in Section 4. The rejection rates for the limiting case ($T = \infty$) in Panels B and C are based on the results in part (d) of Theorem 1 and part (b) of Theorem 5.

	$\mathcal{S}$ Test			CSR Test		
T	10%	5%	1%	10%	5%	1%
	Panel A: Model with a Useful Factor Only					
200	0.914	0.857	0.697	0.919	0.862	0.702
400	0.997	0.992	0.968	0.997	0.993	0.969
600	1.000	1.000	0.999	1.000	1.000	0.999
800	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000
3600	1.000	1.000	1.000	1.000	1.000	1.000
∞	1.000	1.000	1.000	1.000	1.000	1.000
	Panel B: Model with a Useless Factor Only					
200	0.168	0.096	0.025	0.880	0.815	0.648
400	0.135	0.073	0.018	0.976	0.961	0.911
600	0.124	0.066	0.015	0.990	0.984	0.966
800	0.117	0.062	0.014	0.994	0.990	0.980
1000	0.114	0.060	0.013	0.996	0.993	0.986
3600	0.105	0.052	0.011	0.998	0.997	0.995
∞	0.100	0.050	0.010	0.999	0.998	0.997
	Panel C: Model with a Useful and a Useless Factor					
200	0.148	0.082	0.020	0.833	0.752	0.558
400	0.128	0.068	0.016	0.964	0.941	0.868
600	0.119	0.063	0.014	0.985	0.976	0.949
800	0.113	0.058	0.013	0.991	0.986	0.972
1000	0.112	0.057	0.012	0.993	0.989	0.979
3600	0.104	0.052	0.010	0.998	0.997	0.994
∞	0.100	0.050	0.010	0.999	0.998	0.996

Table III.a
Empirical Distribution of the R^2 Coefficient (CU-GMM)

The table presents the empirical distribution of the pseudo-R^2 computed as the squared correlation between the realized and fitted expected returns based on the CU-GMM estimator. The results are based on 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns for the period 1959:2–2012:12.

T	mean	std	1%	5%	10%	25%	50%	75%	90%	95%	99%
					Panel A: Model with a Useful Factor Only						
200	0.298	0.247	0.000	0.003	0.012	0.073	0.251	0.483	0.669	0.755	0.871
400	0.237	0.204	0.000	0.002	0.010	0.056	0.188	0.378	0.543	0.628	0.757
600	0.214	0.181	0.000	0.003	0.011	0.058	0.176	0.335	0.481	0.563	0.692
800	0.203	0.164	0.000	0.004	0.013	0.063	0.171	0.311	0.443	0.518	0.638
1000	0.196	0.152	0.000	0.004	0.017	0.069	0.169	0.295	0.414	0.483	0.600
3600	0.172	0.089	0.012	0.041	0.062	0.105	0.164	0.230	0.293	0.332	0.404
					Panel B: Model with a Useless Factor Only						
200	0.900	0.125	0.342	0.658	0.770	0.883	0.944	0.971	0.983	0.988	0.993
400	0.973	0.040	0.809	0.912	0.942	0.970	0.985	0.992	0.995	0.996	0.998
600	0.989	0.015	0.929	0.966	0.976	0.987	0.993	0.996	0.998	0.998	0.999
800	0.994	0.008	0.963	0.982	0.987	0.993	0.996	0.998	0.999	0.999	0.999
1000	0.996	0.005	0.978	0.989	0.992	0.995	0.997	0.999	0.999	0.999	1.000
3600	1.000	0.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000
					Panel C: Model with a Useful and a Useless Factor						
200	0.903	0.125	0.325	0.667	0.779	0.889	0.946	0.973	0.984	0.988	0.993
400	0.974	0.039	0.810	0.916	0.945	0.972	0.986	0.992	0.995	0.996	0.998
600	0.989	0.015	0.933	0.968	0.978	0.988	0.993	0.996	0.998	0.998	0.999
800	0.994	0.007	0.967	0.983	0.988	0.993	0.996	0.998	0.999	0.999	0.999
1000	0.996	0.004	0.980	0.989	0.993	0.996	0.998	0.999	0.999	0.999	1.000
3600	1.000	0.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000

Table III.b
Empirical Distribution of the R^2 Coefficient (ML)

The table presents the empirical distribution of the pseudo-R^2 computed as the squared correlation between the realized and fitted expected returns based on the ML estimator. The results are based on 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns for the period 1959:2–2012:12.

T	mean	std	1%	5%	10%	25%	50%	75%	90%	95%	99%
					Panel A: Model with a Useful Factor Only						
200	0.231	0.220	0.000	0.002	0.006	0.040	0.161	0.376	0.577	0.674	0.806
400	0.194	0.188	0.000	0.001	0.005	0.034	0.135	0.309	0.482	0.578	0.719
600	0.178	0.167	0.000	0.002	0.006	0.036	0.130	0.280	0.429	0.514	0.651
800	0.169	0.152	0.000	0.002	0.007	0.040	0.130	0.263	0.393	0.469	0.601
1000	0.163	0.140	0.000	0.002	0.009	0.045	0.131	0.249	0.367	0.438	0.561
3600	0.143	0.082	0.006	0.026	0.043	0.081	0.133	0.195	0.256	0.294	0.367
					Panel B: Model with a Useless Factor Only						
200	0.940	0.141	0.150	0.703	0.852	0.955	0.988	0.998	1.000	1.000	1.000
400	0.989	0.035	0.877	0.958	0.976	0.991	0.998	1.000	1.000	1.000	1.000
600	0.996	0.009	0.961	0.985	0.991	0.997	0.999	1.000	1.000	1.000	1.000
800	0.998	0.004	0.981	0.992	0.995	0.998	0.999	1.000	1.000	1.000	1.000
1000	0.999	0.003	0.988	0.995	0.997	0.999	1.000	1.000	1.000	1.000	1.000
3600	1.000	0.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
					Panel C: Model with a Useful and a Useless Factor						
200	0.942	0.135	0.216	0.701	0.852	0.956	0.989	0.998	1.000	1.000	1.000
400	0.989	0.036	0.873	0.959	0.977	0.992	0.998	1.000	1.000	1.000	1.000
600	0.996	0.010	0.963	0.986	0.991	0.997	0.999	1.000	1.000	1.000	1.000
800	0.998	0.004	0.983	0.993	0.996	0.998	1.000	1.000	1.000	1.000	1.000
1000	0.999	0.002	0.990	0.996	0.997	0.999	1.000	1.000	1.000	1.000	1.000
3600	1.000	0.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table IV.a
Rejection Rates of t-tests (CU-GMM)

The table presents the rejection rates of t-tests of statistical significance under misspecified models in the SDF setup. The null hypothesis is that the parameter of interest is equal to zero. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns for the period 1959:2–2012:12. The t-statistics are compared with the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) in Panels B and C are based on the asymptotic distributions in part (b) of Theorem 3.

	useful			useless		
T	10%	5%	1%	10%	5%	1%
	Panel A: Model with a Useful Factor Only					
200	0.849	0.814	0.732	–	–	–
400	0.906	0.879	0.812	–	–	–
600	0.953	0.936	0.889	–	–	–
800	0.977	0.968	0.939	–	–	–
1000	0.989	0.985	0.969	–	–	–
3600	1.000	1.000	1.000	–	–	–
∞	1.000	1.000	1.000	–	–	–
	Panel B: Model with a Useless Factor Only					
200	–	–	–	0.997	0.996	0.994
400	–	–	–	1.000	1.000	1.000
600	–	–	–	1.000	1.000	1.000
800	–	–	–	1.000	1.000	1.000
1000	–	–	–	1.000	1.000	1.000
3600	–	–	–	1.000	1.000	1.000
∞	–	–	–	1.000	1.000	1.000
	Panel C: Model with a Useful and a Useless Factor					
200	0.330	0.239	0.114	0.993	0.992	0.986
400	0.210	0.130	0.041	1.000	1.000	0.999
600	0.167	0.095	0.023	1.000	1.000	1.000
800	0.148	0.080	0.017	1.000	1.000	1.000
1000	0.136	0.070	0.014	1.000	1.000	1.000
3600	0.108	0.051	0.007	1.000	1.000	1.000
∞	0.100	0.045	0.006	1.000	1.000	1.000

Table IV.b
Rejection Rates of t-tests (ML)

The table presents the rejection rates of t-tests of statistical significance under misspecified models in the beta-pricing setup. The null hypothesis is that the parameter of interest is equal to zero. The results are reported for different levels of significance (10%, 5%, and 1%) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the returns are generated from a multivariate normal distribution with means and covariance matrix calibrated to the 25 size and book-to-market Fama-French portfolio returns for the period 1959:2–2012:12. The t-statistics are compared with the critical values from a standard normal distribution. The rejection rates for the limiting case ($T = \infty$) in Panels B and C are based on the asymptotic distributions in part (b) of Theorem 1.

	useful			useless		
T	10%	5%	1%	10%	5%	1%
Panel A: Model with a Useful Factor Only						
200	0.597	0.508	0.336	–	–	–
400	0.778	0.703	0.527	–	–	–
600	0.887	0.834	0.690	–	–	–
800	0.944	0.913	0.810	–	–	–
1000	0.974	0.955	0.887	–	–	–
3600	1.000	1.000	1.000	–	–	–
∞	1.000	1.000	1.000	–	–	–
Panel B: Model with a Useless Factor Only						
200	–	–	–	0.992	0.990	0.985
400	–	–	–	1.000	1.000	1.000
600	–	–	–	1.000	1.000	1.000
800	–	–	–	1.000	1.000	1.000
1000	–	–	–	1.000	1.000	1.000
3600	–	–	–	1.000	1.000	1.000
∞	–	–	–	1.000	1.000	1.000
Panel C: Model with a Useful and a Useless Factor						
200	0.261	0.176	0.070	0.987	0.984	0.976
400	0.189	0.113	0.033	1.000	0.999	0.999
600	0.163	0.090	0.022	1.000	1.000	1.000
800	0.149	0.080	0.017	1.000	1.000	1.000
1000	0.142	0.076	0.016	1.000	1.000	1.000
3600	0.120	0.058	0.010	1.000	1.000	1.000
∞	0.113	0.054	0.008	1.000	1.000	1.000

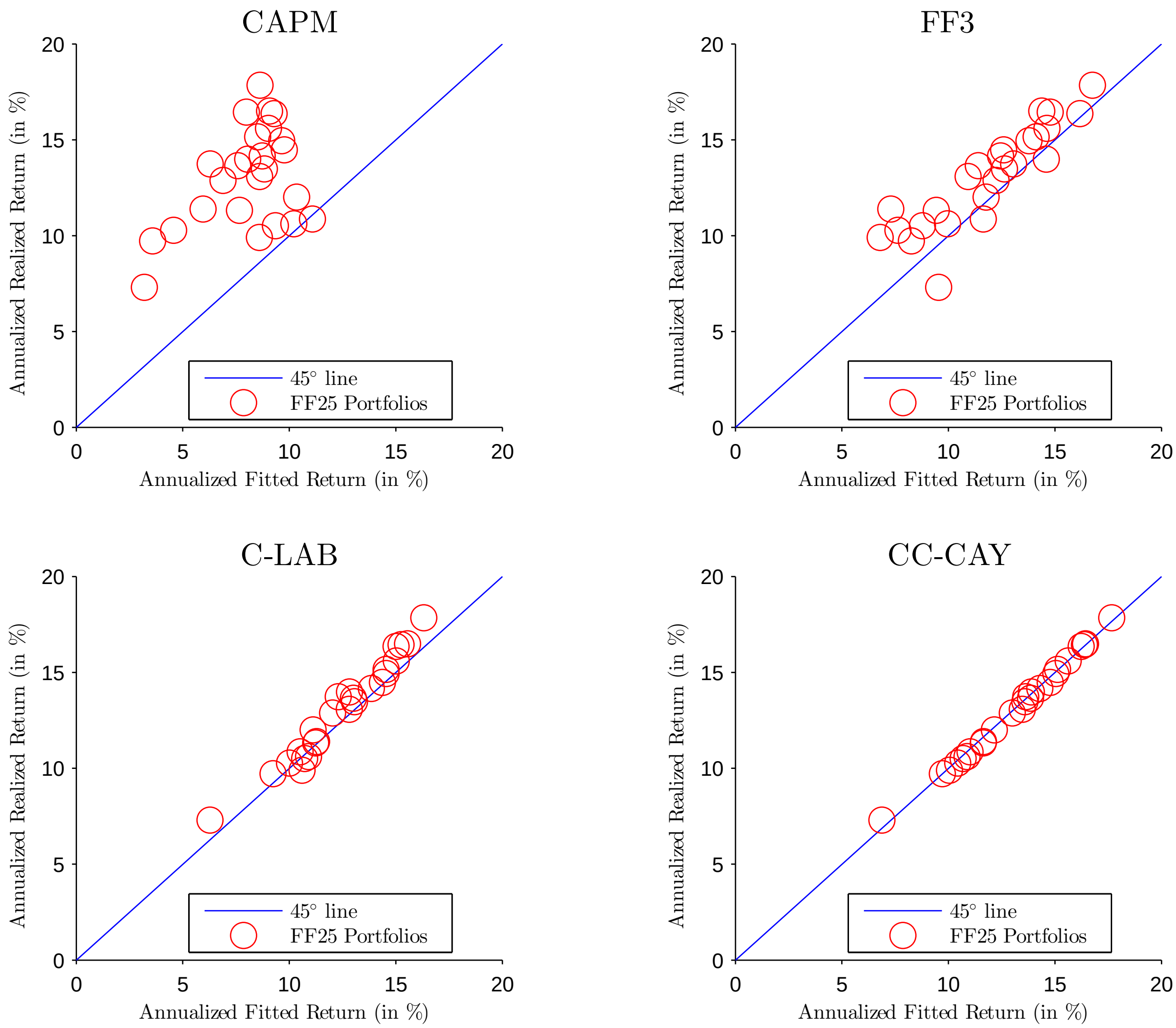

Figure 1. Realized vs. Fitted (by CU-GMM) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by CU-GMM) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

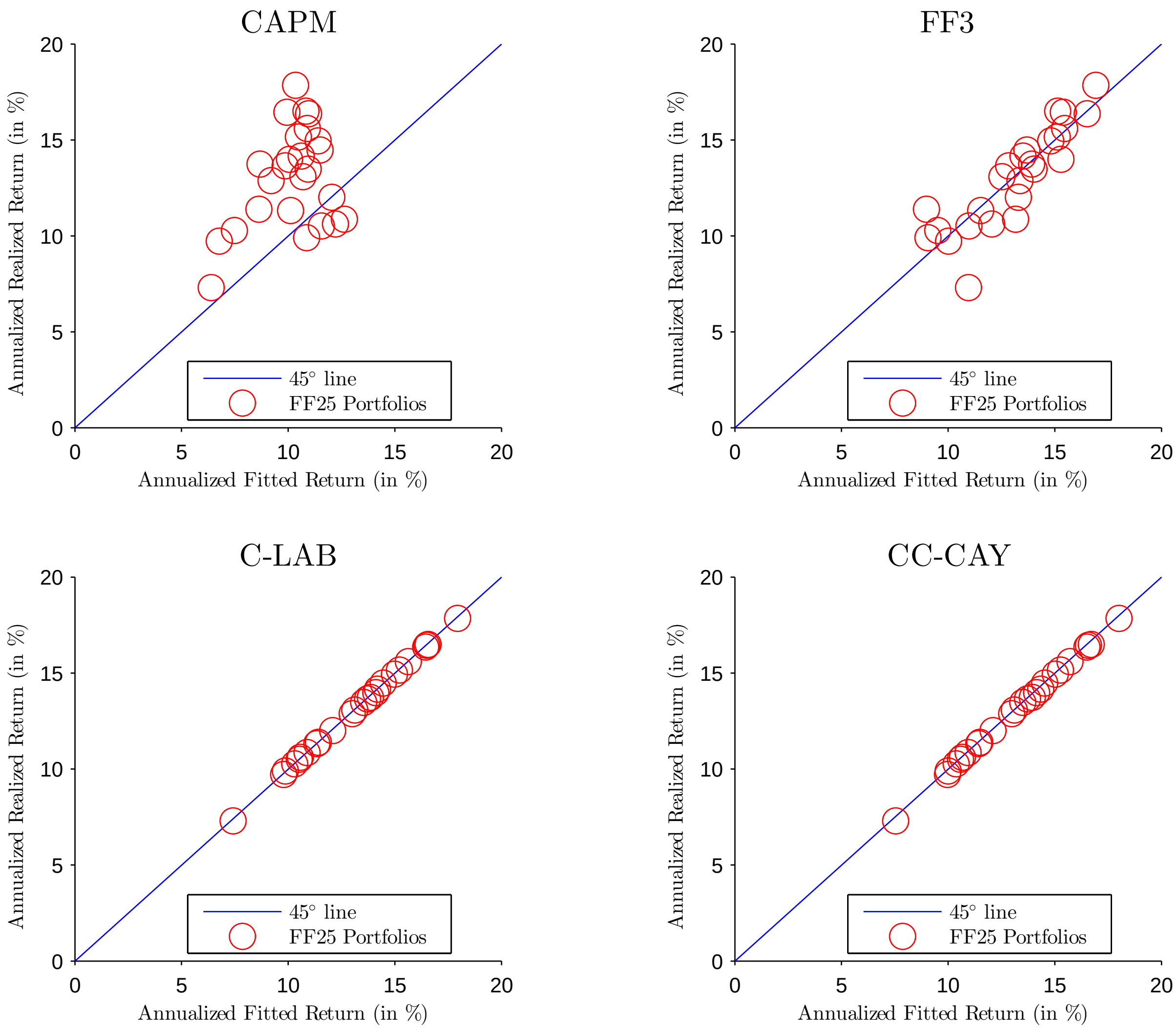

Figure 2. Realized vs. Fitted (by ML) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by ML) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

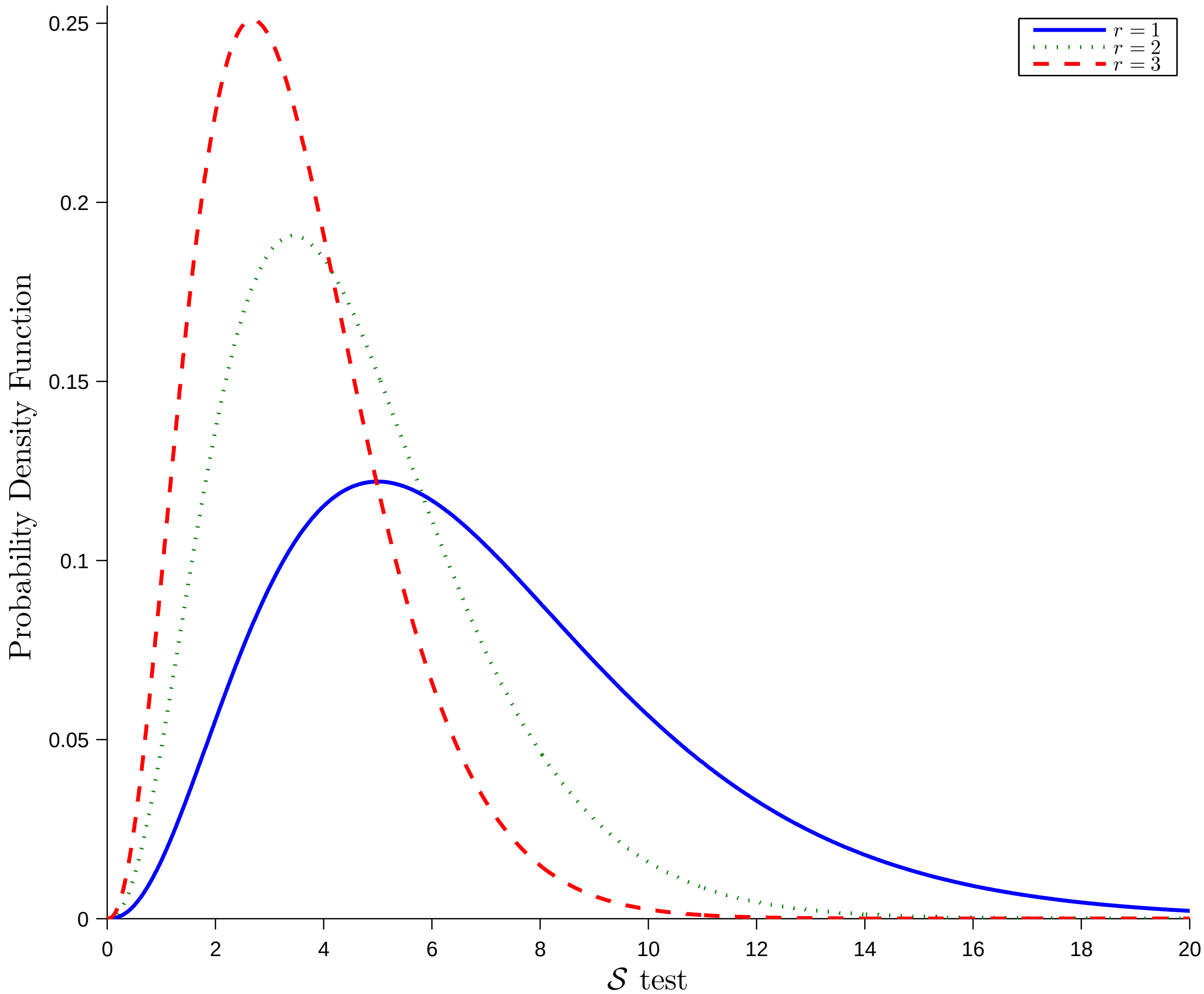

Figure 3. Limiting Distribution of the Specification Test $\mathcal{S}$. The figure plots the asymptotic distributions of $\mathcal{S}$ presented in Theorem 2 for $r = 1$, 2, and 3 (for $N - K = 7$).

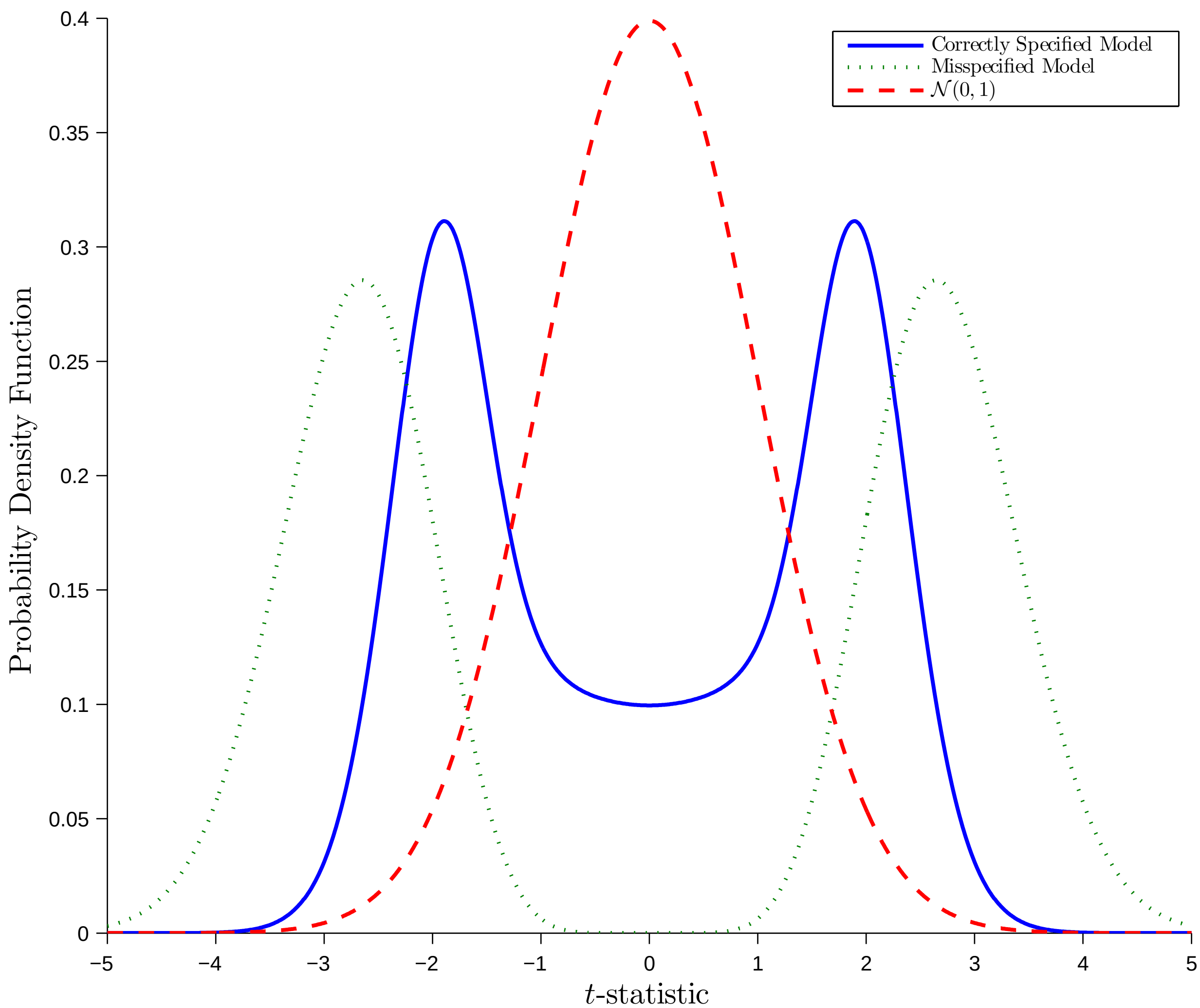

Figure 4. Limiting Distributions of $t(\hat{\gamma}^{ML}_{1,K-1})$ under Correctly Specified and Misspecified Models. The figure plots the limiting densities of $t(\hat{\gamma}^{ML}_{1,K-1})$ for correctly specified and misspecified models that contain a useless factor (for $N-K=7$), along with the standard normal density.

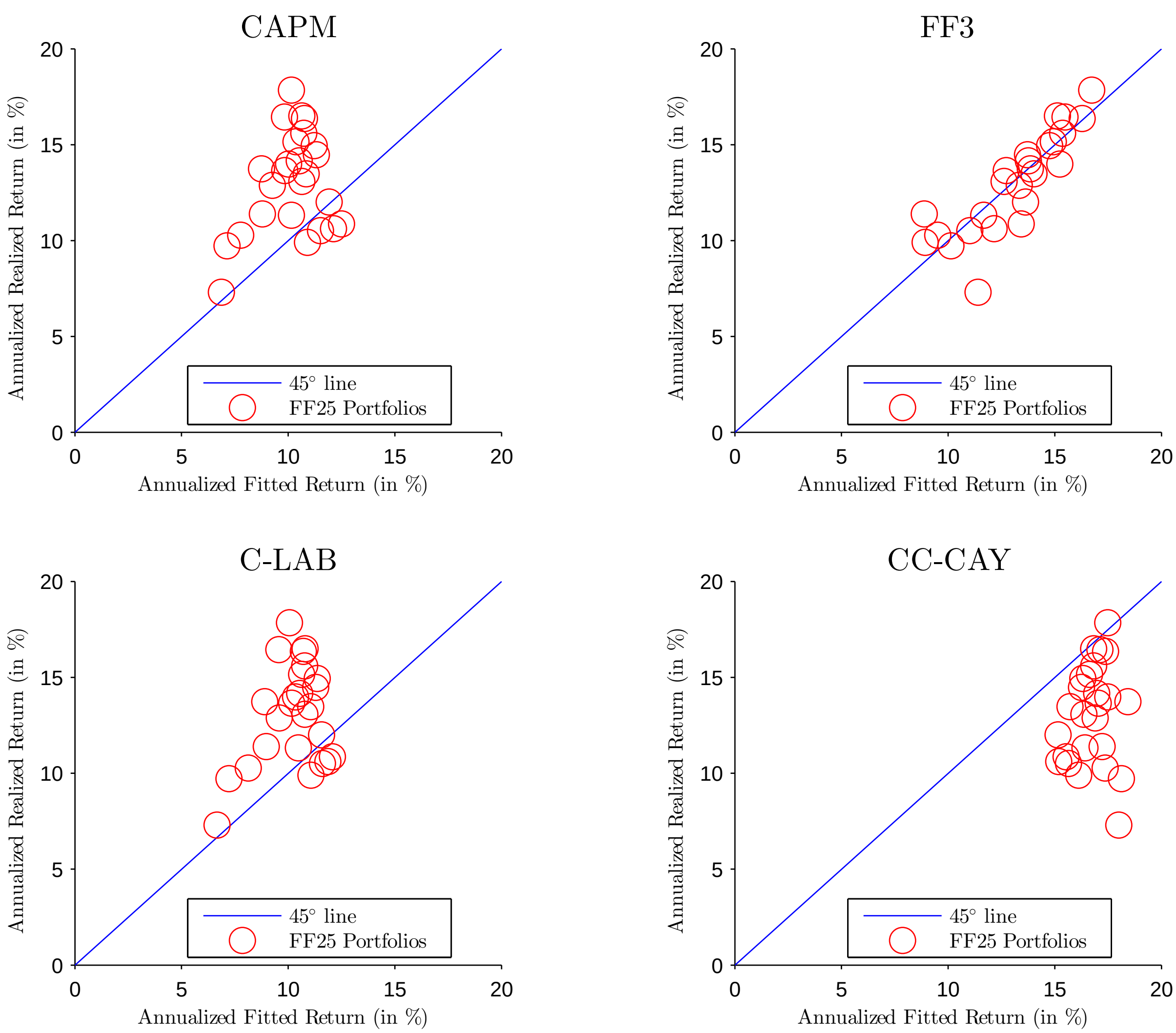

Figure 5. Realized vs. Fitted (by HJ-Distance) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by HJ-distance) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

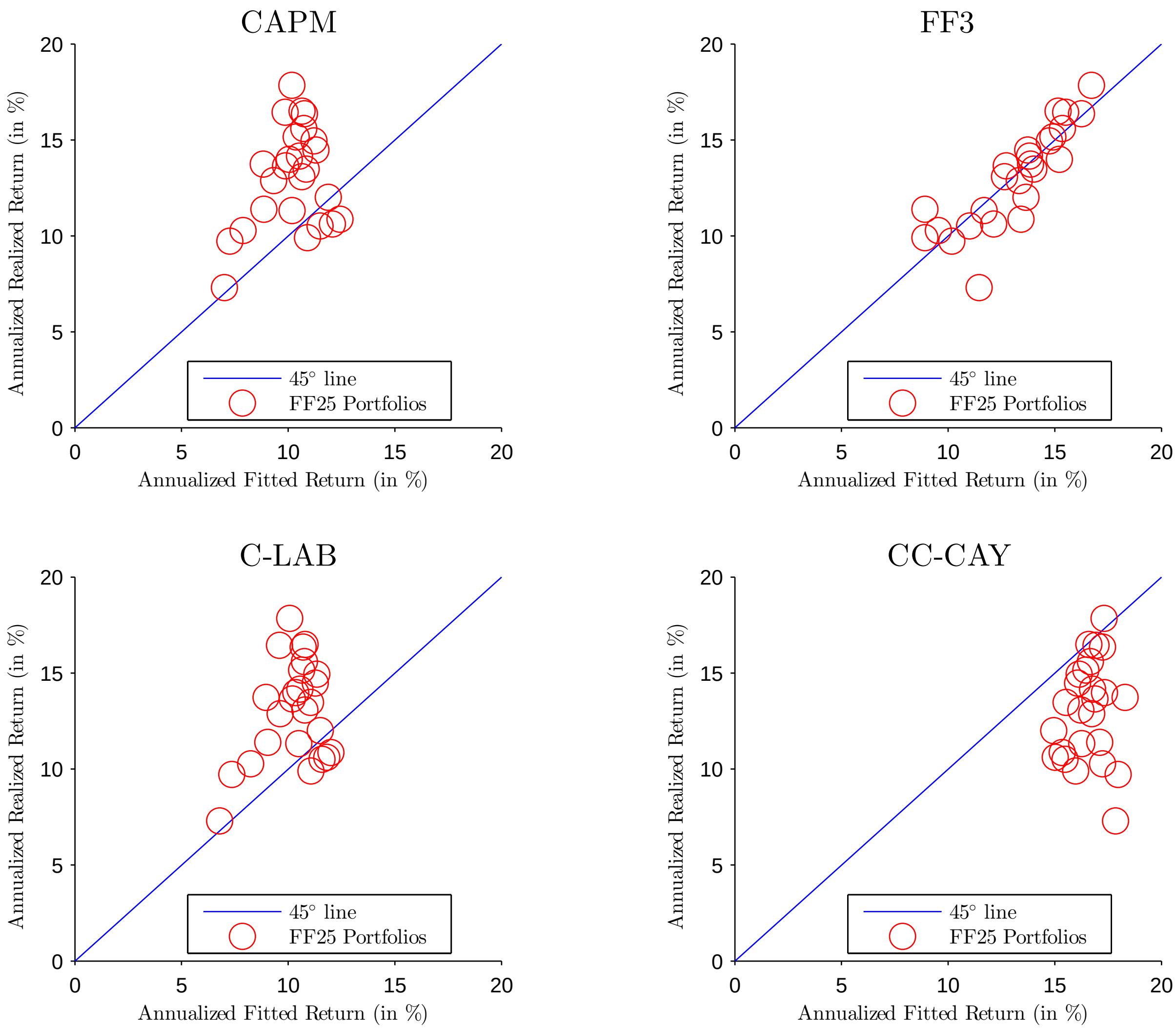

Figure 6. Realized vs. Fitted (by GLS) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by GLS) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

Internet Appendix for

“Spurious Inference in Reduced-Rank Asset-Pricing Models”

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

July 2016

This Internet Appendix contains some supplementary results that are referred to in the main paper. Under the assumption of correct model specification, Section 1 derives the limiting distributions of the ML and CU-GMM estimators, and their t-statistics, in the beta-pricing and SDF frameworks, respectively. Section 2 develops the asymptotic distributions for misspecified models with two noisy factors, whose linear combination is a useless factor. Section 3 discusses a computationally efficient approach to the CU-GMM estimation of the beta-pricing model. We refer the reader to the main paper for the notation used here.

1 Correctly Specified Model

In this section, we derive the limiting distributions of the ML and CU-GMM estimators (and their t-statistics) when the model is correctly specified and contains a useless factor. Note that in this case, there are two linear combinations that give rise to a reduced rank for matrices G and H, respectively.

A. Beta-Pricing Model

To derive the limiting behavior of the ML parameter estimates and associated t-statistics in the beta-pricing framework, we first need to characterize the limiting distribution of the eigenvector associated with the smallest eigenvalue of

$$T\tilde{\Omega} = (X'X)\hat{B}'P_1(P_1'\hat{\Sigma}P_1)^{-1}P_1'\hat{B}. \tag{IA.1}$$

Let L_f be a lower triangular matrix such that $L_f L_f' = V_f$ and define

$$L = \begin{bmatrix} 1 & 0_{K-1}' \\ \mu_f & L_f \end{bmatrix}. \tag{IA.2}$$

Using that $(X'X)/T \xrightarrow{p} LL'$ and $\hat{\Sigma} \xrightarrow{p} \Sigma$, the $\mathcal{S}$ test has the same distribution as the smallest eigenvalue of

$$W_0 = TLL'\hat{B}'P_1(P_1'\Sigma P_1)^{-1}P_1'\hat{B}. \tag{IA.3}$$

Let $\tilde{P}_1$ be an $N \times (N-1)$ orthonormal matrix such that $\tilde{P}_1'\Sigma^{-\frac{1}{2}}1_N = 0_{N-1}$. Then, we have

$$P_1(P_1'\Sigma P_1)^{-1}P_1' = \Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}} \tag{IA.4}$$

and

$$W_0 = TLL'\hat{B}'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\hat{B}. \tag{IA.5}$$

Define

$$Z = \sqrt{T}\tilde{P}_1'\Sigma^{-\frac{1}{2}}\hat{B}L \tag{IA.6}$$

and

$$M = E[Z] = \sqrt{T}\tilde{P}_1'\Sigma^{-\frac{1}{2}}BL. \tag{IA.7}$$

Then, under Assumptions 1 and 3 in the paper,

$$\text{vec}(Z) \sim \mathcal{N}\left(\text{vec}(M), I_{(N-1)K}\right). \tag{IA.8}$$

Since W_0 and $Z'Z$ share the same eigenvalues, the smallest eigenvalue of W_0 has the same limiting distribution as the smallest eigenvalue of $W_1 = Z'Z \sim \mathcal{W}_K(N-1, I_K, M'M)$, which has a noncentral Wishart distribution. Since B has rank $K-r$, there exists a $K \times r$ orthonormal matrix C_1 such that $MC_1 = 0_{(N-1)\times r}$. Let $C = [C_1,\ C_2]$ be a $K \times K$ orthonormal matrix, and define $\tilde{Z} = [\tilde{Z}_1,\ \tilde{Z}_2] = [ZC_1,\ ZC_2]$. We then have $E[\tilde{Z}_1] = 0_{(N-1)\times r}$ and $E[\tilde{Z}_2] \equiv \tilde{M}_2 = MC_2$. Using the fact that $W_1 = Z'Z$ and $W_2 = \tilde{Z}'\tilde{Z}$ share the same eigenvalues, we can derive the limiting distribution of the smallest eigenvalue of W_2 which is equal to the reciprocal of the largest eigenvalue of

$$W_2^{-1} = \begin{bmatrix} W_2^{11} & W_2^{12} \\ W_2^{21} & W_2^{22} \end{bmatrix}. \tag{IA.9}$$

Using the formula for the inverse of a partitioned matrix, we have

$$\begin{aligned}
W_2^{11} &= \left(\tilde{Z}_1'[I_{N-1} - \tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1}\tilde{Z}_2]\tilde{Z}_1\right)^{-1} \\
&\stackrel{d}{\to} \tilde{Z}_1'[I_{N-1} - \tilde{M}_2(\tilde{M}_2'\tilde{M}_2)^{-1}\tilde{M}_2']\tilde{Z}_1 \sim \mathcal{W}_r(N-K+r-2, I_r)^{-1}, && \text{(IA.10)} \\
\sqrt{T}W_2^{12} &= -\sqrt{T}W_2^{11}\tilde{Z}_1'\tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1} \\
&\stackrel{d}{\to} -W_2^{11}\tilde{Z}_1'\tilde{M}_2(\tilde{M}_2'\tilde{M}_2)^{-1}, && \text{(IA.11)} \\
TW_2^{22} &= T(\tilde{Z}_2'\tilde{Z}_2)^{-1} + T(\tilde{Z}_2'\tilde{Z}_2)^{-1}(\tilde{Z}_2'\tilde{Z}_1)W_2^{11}(\tilde{Z}_1'\tilde{Z}_2)(\tilde{Z}_2'\tilde{Z}_2)^{-1} \\
&\stackrel{d}{\to} (\tilde{M}_2'\tilde{M}_2)^{-1} + (\tilde{M}_2'\tilde{M}_2)^{-1}(\tilde{M}_2'\tilde{Z}_1)W_2^{11}(\tilde{Z}_1'\tilde{M}_2)(\tilde{M}_2'\tilde{M}_2)^{-1}. && \text{(IA.12)}
\end{aligned}$$

Therefore, the limiting distribution of the largest eigenvalue of W_2^{-1} is the same as the limiting distribution of the largest eigenvalue of W_2^{11}. Equivalently, the smallest eigenvalue of $T\tilde{\Omega}$ has the same limiting distribution as w_r, the smallest eigenvalue of $W \sim \mathcal{W}_r(N-K-1+r, I_r)$.

Let $q = [q_1', \; q_2']'$ be the eigenvector associated with the largest eigenvalue of W_2^{-1}, where q_1 denotes the first r elements of q and q_2 denotes the last $K-r$ elements of q. Asymptotically, q_1 has the same distribution as the eigenvector associated with the largest eigenvalue of W_2^{11}, which has a uniform distribution over the r-dimensional sphere. In addition, q_2 is $O_p(T^{-\frac{1}{2}})$. To obtain the limiting distribution of $\sqrt{T}q_2$, we use the fact that

$$\begin{bmatrix} W_2^{11} & W_2^{12} \\ W_2^{21} & W_2^{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \frac{1}{w_r} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, \tag{IA.13}$$

which implies that

$$W_2^{21} q_1 + W_2^{22} q_2 = \frac{1}{w_r} q_2. \tag{IA.14}$$

Since $W_2^{22} = O_p(T^{-1})$, we have

$$\sqrt{T} q_2 = \sqrt{T} w_r W_2^{21} q_1 + O_p(T^{-\frac{1}{2}}) \xrightarrow{d} -w_r (\tilde{M}_2' \tilde{M}_2)^{-1} \tilde{M}_2' \tilde{Z}_1 W_2^{11} q_1. \tag{IA.15}$$

Then, using the fact that $W_2^{11} q_1 = \frac{1}{w_r} q_1 + O_p(T^{-1})$, we obtain

$$\sqrt{T} q_2 = -(\tilde{M}_2' \tilde{M}_2)^{-1} \tilde{M}_2' \tilde{Z}_1 q_1 + O_p(T^{-1}). \tag{IA.16}$$

It follows that $Cq = C_1 q_1 + C_2 q_2 \xrightarrow{d} C_1 q_1$ is the limiting distribution of the eigenvector associated with the smallest eigenvalue of W_2. It is then easy to show that the limiting distribution of the eigenvector associated with the smallest eigenvalue of $T\tilde{\Omega}$ is proportional to h, where

$$h = LCq = LC_1 q_1 + LC_2 q_2 \xrightarrow{d} LC_1 q_1. \tag{IA.17}$$

Having derived the limiting distribution of the eigenvector associated with the smallest eigenvalue of W_2, we are now ready to discuss the special case when the model is correctly specified and contains a useless factor ($r = 2$). Without loss of generality, we assume that the last factor $f_{K-1,t} = g_t$ is a useless factor with mean μ_g and variance σ_g^2. In addition, let $\tilde{f}_t$ be the first $K-2$ factors and denote their mean and covariance matrix by $\mu_{\tilde{f}}$ (with i-th element $\mu_{f,i}$ for $i = 1, \ldots, K-2$) and $V_{\tilde{f}}$ (with (i,i)-element $\sigma_{f,i}^2$ for $i = 1, \ldots, K-2$), respectively. Partition β as $\beta = [\beta_1, \; \beta_2]$, where β_1 denotes the betas of the N assets with respect to $\tilde{f}_t$ and $\beta_2 = 0_N$ denotes the betas of the N assets with respect to g_t. Since the model is correctly specified, we have $\alpha = 1_N \gamma_0^* + \beta_1 \phi_1^*$, where $\tilde{\gamma}_1^*$ (with i-th element $\gamma_{1,i}^*$ for $i = 1, \ldots, K-2$) is the vector of pseudo-true risk premia on $\tilde{f}_t$

and $\phi_1^* = (\tilde{\gamma}_1^* - \mu_{\tilde{f}})$ (with i-th element $\phi_{1,i}^*$ for $i = 1, \ldots, K-2$). It follows that

$$\tilde{P}_1' \Sigma^{-\frac{1}{2}} B \begin{bmatrix} 1 & 0 \\ -\phi_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix} = 0_{(N-1)\times 2}. \tag{IA.18}$$

Since

$$\tilde{P}_1' \Sigma^{-\frac{1}{2}} B L C_1 = 0_{(N-1)\times 2}, \tag{IA.19}$$

we have

$$\begin{aligned} C_1 &= L^{-1} \begin{bmatrix} 1 & 0 \\ -\phi_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ -\phi_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix}' (L')^{-1} L^{-1} \begin{bmatrix} 1 & 0 \\ -\phi_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix} \right)^{-\frac{1}{2}} \\ &= L^{-1} \begin{bmatrix} 1 & 0 \\ -\phi_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a + \frac{\mu_g^2}{\sigma_g^2} & -\frac{\mu_g}{\sigma_g^2} \\ -\frac{\mu_g}{\sigma_g^2} & \frac{1}{\sigma_g^2} \end{bmatrix}^{-\frac{1}{2}}, \end{aligned} \tag{IA.20}$$

where $a = 1 + \tilde{\gamma}_1^{*\prime} V_{\tilde{f}}^{-1} \tilde{\gamma}_1^*$. Using the fact that

$$\begin{bmatrix} a + \frac{\mu_g^2}{\sigma_g^2} & -\frac{\mu_g}{\sigma_g^2} \\ -\frac{\mu_g}{\sigma_g^2} & \frac{1}{\sigma_g^2} \end{bmatrix}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{a}} & 0 \\ \frac{\mu_g}{\sqrt{a}} & \sigma_g \end{bmatrix} \tag{IA.21}$$

and

$$L^{-1} = \begin{bmatrix} 1 & 0_{K-1}' \\ -L_f^{-1}\mu_f & L_f^{-1} \end{bmatrix}, \tag{IA.22}$$

an explicit expression for C_1 is given by

$$C_1 = \begin{bmatrix} \frac{1}{\sqrt{a}} & 0 \\ -\frac{1}{\sqrt{a}} L_{\tilde{f}}^{-1} \tilde{\gamma}_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix}, \tag{IA.23}$$

where $L_{\tilde{f}} L_{\tilde{f}}' = V_{\tilde{f}}$. For C_2, we have

$$C_2 = \begin{bmatrix} \tilde{P}_2 \\ 0_{K-2}' \end{bmatrix}, \tag{IA.24}$$

where $\tilde{P}_2$ is a $(K-1)\times(K-2)$ orthonormal matrix with its columns orthogonal to $[1,\ -(L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)']$. To obtain an explicit expression for $\tilde{P}_2$, note that

$$[1,\ -(L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)'] \begin{bmatrix} (L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)' \\ I_{K-2} \end{bmatrix} = 0_{K-2}', \tag{IA.25}$$

and hence we can write

$$\tilde{P}_2 = \begin{bmatrix} (L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)' \\ I_{K-2} \end{bmatrix} \left(\begin{bmatrix} (L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)' \\ I_{K-2} \end{bmatrix}' \begin{bmatrix} (L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)' \\ I_{K-2} \end{bmatrix} \right)^{-\frac{1}{2}} = \begin{bmatrix} (L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)' \\ I_{K-2} \end{bmatrix} A^{-\frac{1}{2}}, \tag{IA.26}$$

where

$$A = I_{K-2} + L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*\tilde{\gamma}_1^{*\prime}L_{\tilde{f}}^{-1\prime}. \quad \text{(IA.27)}$$

In addition, h is given by

$$h = LCq \stackrel{d}{\to} LC_1q_1 = \begin{bmatrix} 1 & 0 \\ -\phi_1^* & 0_{K-2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{a}} & 0 \\ \frac{\mu_g}{\sqrt{a}} & \sigma_g \end{bmatrix} \begin{bmatrix} q_{1,1} \\ q_{1,2} \end{bmatrix} = \begin{bmatrix} \frac{q_{1,1}}{\sqrt{a}} \\ -\frac{q_{1,1}}{\sqrt{a}}\phi_1^* \\ \frac{\mu_g}{\sqrt{a}}q_{1,1} + \sigma_g q_{1,2} \end{bmatrix}, \quad \text{(IA.28)}$$

where $q_{1,1}$ and $q_{1,2}$ are the first and second elements of q_1, respectively. The ML risk premia estimates are then given by

$$\hat{\gamma}_{1,i}^{ML} = -\frac{h_{i+1}}{h_1} + \hat{\mu}_{f,i} \stackrel{p}{\to} \gamma_{1,i}^*, \quad i = 1, \ldots, K-2, \quad \text{(IA.29)}$$

$$\hat{\gamma}_{1,K-1}^{ML} = -\frac{h_K}{h_1} + \hat{\mu}_g \stackrel{d}{\to} -\frac{\sigma_g\sqrt{a}q_{1,2}}{q_{1,1}}. \quad \text{(IA.30)}$$

Finally, since $\alpha - \beta_1\phi_1^* = 1_N\gamma_0^*$, we have

$$\hat{\gamma}_0^{ML} = \frac{1_N'\hat{\Sigma}^{-1}\hat{B}h}{h_1 1_N'\hat{\Sigma}^{-1}1_N} \stackrel{p}{\to} \gamma_0^*. \quad \text{(IA.31)}$$

The above results show that $\hat{\gamma}_0^{ML}$ and $\hat{\gamma}_{1,i}^{ML}$ $(i = 1, \ldots, K-2)$ for the useful factors are consistent estimators of γ_0^* and $\gamma_{1,i}^*$, respectively, while $\hat{\gamma}_{1,K-1}^{ML}$ for the useless factor has a limiting Cauchy distribution.

The explicit expressions for Z and $\tilde{Z}_1$ are then given by

$$\begin{aligned} Z &= \sqrt{T}\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}[\hat{\alpha},\ \hat{\beta}_1,\ \hat{\beta}_2] \begin{bmatrix} 1 & 0_{K-2}' & 0 \\ \mu_{\tilde{f}} & L_{\tilde{f}} & 0_{K-2} \\ \mu_g & 0_{K-2}' & \sigma_g \end{bmatrix} \\ &= [\sqrt{T}\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}(\hat{\alpha} + \hat{\beta}_1\mu_{\tilde{f}} + \hat{\beta}_2\mu_g),\ \sqrt{T}\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\hat{\beta}_1 L_{\tilde{f}},\ \sqrt{T}\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\hat{\beta}_2\sigma_g], \quad \text{(IA.32)} \\ \tilde{Z}_1 &= ZC_1 \\ &= \left[\frac{1}{\sqrt{a}}\sqrt{T}\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}(\hat{\alpha} - \hat{\beta}_1\phi_1^* + \hat{\beta}_2\mu_g),\ \sqrt{T}\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\hat{\beta}_2\sigma_g\right]. \quad \text{(IA.33)} \end{aligned}$$

Using the fact that $\beta_2 = 0_N$ and $\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\alpha = \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1\phi_1^*$, the mean of $\tilde{Z}_2/\sqrt{T}$ is given by

$$\begin{aligned}
\tilde{M}_2 &= \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}[\alpha,\ \beta_1,\ \beta_2]\begin{bmatrix} 1 & 0'_{K-2} & 0 \\ \mu_{\tilde{f}} & L_{\tilde{f}} & 0_{K-2} \\ \mu_g & 0'_{K-2} & \sigma_g \end{bmatrix}\begin{bmatrix} \tilde{P}_2 \\ 0'_{K-2} \end{bmatrix} \\
&= \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}[\beta_1\phi_1^*,\ \beta_1]\begin{bmatrix} 1 & 0'_{K-2} \\ \mu_{\tilde{f}} & L_{\tilde{f}} \end{bmatrix}\tilde{P}_2 \\
&= \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1[\tilde{\gamma}_1^*,\ L_{\tilde{f}}]\tilde{P}_2 \\
&= \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1 L_{\tilde{f}}[L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*,\ I_{K-2}]\tilde{P}_2 \\
&= \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1 L_{\tilde{f}}A^{\frac{1}{2}}. \qquad \text{(IA.34)}
\end{aligned}$$

It follows that

$$\begin{aligned}
\tilde{Z}_2(\tilde{Z}_2'\tilde{Z}_2)^{-1}\tilde{Z}_2' &\xrightarrow{p} \tilde{M}_2(\tilde{M}_2'\tilde{M}_2)^{-1}\tilde{M}_2' \\
&= \tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1. \qquad \text{(IA.35)}
\end{aligned}$$

We now derive the limiting distribution of $\sqrt{T}(\hat{\phi}_{1,i} - \phi_{1,i}^*)$ for $i = 1, \ldots, K-2$. Note that when the model does not contain a useless factor, we have

$$\sqrt{T}(\hat{\phi}_{1,i} - \phi_{1,i}^*) \xrightarrow{d} \mathcal{N}\left(0, (1 + \tilde{\gamma}_1^{*\prime}V_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1{}'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i}\right). \qquad \text{(IA.36)}$$

In contrast, the following analysis will show that introducing a useless factor in the model makes the limiting distribution of $\sqrt{T}(\hat{\phi}_{1,i} - \phi_{1,i}^*)$ non-normal and substantially more volatile.

To study the limiting distribution of $\sqrt{T}(\hat{\phi}_{1,i} - \phi_{1,i}^*)$, we need the terms in h that are $O_p(T^{-\frac{1}{2}})$. Let

$$\sqrt{T}LC_2q_2 \xrightarrow{d} u \equiv -\begin{bmatrix} 1 & 0'_{K-1} \\ \mu_f & L_f \end{bmatrix}\begin{bmatrix} (L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*)' \\ I_{K-2} \\ 0'_{K-2} \end{bmatrix}A^{-1}L_{\tilde{f}}^{-1}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_1q_1. \qquad \text{(IA.37)}$$

Using the fact that

$$A^{-1} = I_{K-2} - \frac{L_{\tilde{f}}^{-1}\tilde{\gamma}_1^*\tilde{\gamma}_1^{*\prime}L_{\tilde{f}}^{-1\prime}}{a} \qquad \text{(IA.38)}$$

and after some algebra, we can show that

$$u = -\begin{bmatrix} \frac{\tilde{\gamma}_1^{*\prime}V_{\tilde{f}}^{-1}}{a} \\ I_{K-2} - \frac{\phi_1^*\tilde{\gamma}_1^{*\prime}V_{\tilde{f}}^{-1}}{a} \\ \frac{\mu_g\tilde{\gamma}_1^{*\prime}V_{\tilde{f}}^{-1}}{a} \end{bmatrix}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_1q_1. \qquad \text{(IA.39)}$$

It follows that for $i = 1, \ldots, K-2$,

$$
\begin{aligned}
\sqrt{T}(\hat{\phi}_{1,i} - \phi^*_{1,i}) &= \sqrt{T}\left(-\frac{h_{i+1}}{h_1} - \phi^*_{1,i}\right) \\
&= \frac{\sqrt{T}\frac{q_{1,1}}{\sqrt{a}}\phi^*_{1,i} - u_{i+1}}{\frac{q_{1,1}}{\sqrt{a}} + \frac{u_1}{\sqrt{T}}} - \sqrt{T}\phi^*_{1,i} \\
&\stackrel{d}{\to} -\frac{\sqrt{a}}{q_{1,1}}(u_{i+1} + \phi^*_{1,i}u_1) \\
&= \frac{\sqrt{a}}{q_{1,1}}\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_1 q_1. \qquad \text{(IA.40)}
\end{aligned}
$$

Since $\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_1$ is independent of W_2^{11}, it is also independent of q_1. It follows that for $i = 1, \ldots, K-2$,

$$
\sqrt{T}(\hat{\phi}_{1,i} - \phi^*_{1,i}) \stackrel{d}{\to} \left(\frac{a}{q_{1,1}^2}\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i}\right)^{\frac{1}{2}} z_1, \qquad \text{(IA.41)}
$$

where $z_1 \sim \mathcal{N}(0,1)$ and $q_{1,1}^2 \sim \text{Beta}(1/2, 1/2)$. In addition, using the asymptotic independence between $\hat{\phi}_1$ and $\hat{\mu}_{\tilde{f}}$, we have

$$
\begin{aligned}
\sqrt{T}(\hat{\gamma}_{1,i}^{ML} - \gamma^*_{1,i}) &= \sqrt{T}(\hat{\phi}_{1,i} - \phi^*_{1,i}) + \sqrt{T}(\hat{\mu}_{f,i} - \mu_{f,i}) \\
&\stackrel{d}{\to} \left(\frac{a}{q_{1,1}^2}\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i} + \sigma^2_{f,i}\right)^{\frac{1}{2}} z, \qquad \text{(IA.42)}
\end{aligned}
$$

where $z \sim \mathcal{N}(0,1)$ and it is independent of $q_{1,1}^2$.

To derive the limiting distribution of $\sqrt{T}(\hat{\gamma}_0^{ML} - \gamma_0^*)$, we state and prove the following identity. Let $G_2 = [1_N,\ \beta_1]$. Then, we have

$$
\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-1} \equiv \frac{1_N'\Sigma^{-1} - 1_N'\Sigma^{-1}\beta_1(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}}{1_N'\Sigma^{-1}1_N}. \qquad \text{(IA.43)}
$$

The proof of (IA.43) is based on the formula for the inverse of a partitioned matrix:

$$
\begin{aligned}
\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1} &= \left[\frac{1}{1_N'\Sigma^{-1}1_N} + \frac{1_N'\Sigma^{-1}\beta_1(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-1}1_N}{(1_N'\Sigma^{-1}1_N)^2},\right. \\
&\qquad \left. -\frac{1_N'\Sigma^{-1}\beta_1(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}}{1_N'\Sigma^{-1}1_N}\right]. \qquad \text{(IA.44)}
\end{aligned}
$$

Post-multiplying the above expression by

$$
G_2'\Sigma^{-1} = \begin{bmatrix} 1_N'\Sigma^{-1} \\ \beta_1'\Sigma^{-1} \end{bmatrix} \qquad \text{(IA.45)}
$$

delivers the desired result.

Let $\hat{\phi}_2 = \hat{\gamma}_{1,K-1}^{ML} - \hat{\mu}_g$. Using the fact that $\hat{\Sigma} \xrightarrow{p} \Sigma$,

$$\hat{\gamma}_0^{ML} = \frac{1_N'\hat{\Sigma}^{-1}(\hat{\alpha} - \hat{\beta}_1\hat{\phi}_1 - \hat{\beta}_2\hat{\phi}_2)}{1_N'\hat{\Sigma}^{-1}1_N}, \tag{IA.46}$$

$$\gamma_0^* = \frac{1_N'\Sigma^{-1}(\alpha - \beta_1\phi_1^*)}{1_N'\Sigma^{-1}1_N}, \tag{IA.47}$$

we can write

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_0^{ML} - \gamma_0^*) &= \frac{1_N'\Sigma^{-1}}{1_N'\Sigma^{-1}1_N}\left[\sqrt{T}(\hat{\alpha} - \hat{\beta}_1\hat{\phi}_1 - \alpha + \beta_1\phi_1^*) - \sqrt{T}\hat{\beta}_2\hat{\phi}_2\right] + O_p(T^{-\frac{1}{2}}) \\
&= \frac{1_N'\Sigma^{-1}}{1_N'\Sigma^{-1}1_N}\left[\sqrt{T}(\hat{\alpha} - \hat{\beta}_1\hat{\phi}_1 + \hat{\beta}_2\hat{\mu}_g - \alpha + \beta_1\phi_1^*) - \sqrt{T}\hat{\beta}_2\hat{\gamma}_{1,K-1}^{ML}\right] + O_p(T^{-\frac{1}{2}}) \\
&= \frac{1_N'\Sigma^{-1}}{1_N'\Sigma^{-1}1_N}\left[\sqrt{T}(\hat{\alpha} - \hat{\beta}_1\phi_1^* + \hat{\beta}_2\mu_g - \alpha + \beta_1\phi_1^*) - \sqrt{T}\hat{\beta}_1(\hat{\phi}_1 - \phi_1^*)\right. \\
&\quad \left. + \sqrt{T}\hat{\beta}_2(\hat{\mu}_g - \mu_g) - \sqrt{T}\hat{\beta}_2\hat{\gamma}_{1,K-1}^{ML}\right] + O_p(T^{-\frac{1}{2}}).
\end{aligned} \tag{IA.48}$$

Since $1_N'\Sigma^{-1}\hat{\beta}_1 \xrightarrow{p} 1_N'\Sigma^{-1}\beta_1$ and $\sqrt{T}\hat{\beta}_2(\hat{\mu}_g - \mu_g) \xrightarrow{p} 0$, we can write

$$\sqrt{T}(\hat{\gamma}_0^{ML} - \gamma_0^*) \xrightarrow{d} \frac{1_N'\Sigma^{-1}}{1_N'\Sigma^{-1}1_N}\left[\sqrt{T}(\hat{\alpha} - \hat{\beta}_1\phi_1^* + \hat{\beta}_2\mu_g - \alpha + \beta_1\phi_1^*) - \sqrt{T}\beta_1(\hat{\phi}_1 - \phi_1^*) - \sqrt{T}\hat{\beta}_2\hat{\gamma}_{1,K-1}^{ML}\right]. \tag{IA.49}$$

From (IA.40), we obtain

$$\begin{aligned}
\sqrt{T}(\hat{\phi}_1 - \phi_1^*) &\xrightarrow{d} \sqrt{a}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1a} \\
&\quad + \frac{\sqrt{a}q_{1,2}}{q_{1,1}}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1b},
\end{aligned} \tag{IA.50}$$

where $\tilde{Z}_1 = [\tilde{Z}_{1a},\ \tilde{Z}_{1b}]$, with $\tilde{Z}_{1a} = \sqrt{T/a}\tilde{P}_1'\Sigma^{-\frac{1}{2}}(\hat{\alpha} - \hat{\beta}_1\phi_1^* + \hat{\beta}_2\mu_g)$ and $\tilde{Z}_{1b} = \sqrt{T}\sigma_g\tilde{P}_1'\Sigma^{-\frac{1}{2}}\hat{\beta}_2$.

Plugging (IA.50) into (IA.49) and using (IA.30), (IA.43), together with the fact $\tilde{P}_1'\Sigma^{-\frac{1}{2}}(\alpha - \beta_1\phi_1^*) = 0_{N-1}$, we can write

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_0^{ML} - \gamma_0^*) &\xrightarrow{d} \sqrt{T}\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-1}(\hat{\alpha} - \hat{\beta}_1\phi_1^* + \hat{\beta}_2\mu_g - \alpha + \beta_1\phi_1^*) \\
&\quad + \frac{\sqrt{a}q_{1,2}}{q_{1,1}}\sqrt{T}\sigma_g\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-1}\hat{\beta}_2.
\end{aligned} \tag{IA.51}$$

Let

$$\frac{\sqrt{T}\sigma_g\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-1}\hat{\beta}_2}{[\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} \xrightarrow{d} u_0, \tag{IA.52}$$

$$\frac{\sqrt{T}\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-1}(\hat{\alpha} - \hat{\beta}_1\phi_1^* + \hat{\beta}_2\mu_g - \alpha + \beta_1\phi_1^*)}{\sqrt{a}[\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} \xrightarrow{d} v_0. \tag{IA.53}$$

We can easily verify that $u_0 \sim \mathcal{N}(0,1)$, $v_0 \sim \mathcal{N}(0,1)$, and they are independent of each other and of $q_1 = [q_{1,1},\ q_{1,2}]'$, w_1, and w_2, where w_1 and w_2 (with $w_1 \geq w_2$) are the two eigenvalues of $\mathcal{W}_2(N-K+1, I_2)$. Using u_0 and v_0, we can now write

$$\sqrt{T}(\hat{\gamma}_0^{ML} - \gamma_0^*) \xrightarrow{d} \sqrt{a}[\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}} \left(v_0 + \frac{q_{1,2}}{q_{1,1}}u_0\right). \tag{IA.54}$$

Note that the first term, $\sqrt{a}[\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}v_0$, represents the limiting distribution of $\sqrt{T}(\hat{\gamma}_0^{ML} - \gamma_0^*)$ when the model does not contain a useless factor, while the second term reflects the extra noise that is brought in by the introduction of the useless factor.

We now consider the t-ratio of $\hat{\gamma}_{1,K-1}^{ML}$. We first note that

$$1 + \hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML} \xrightarrow{d} a + \frac{(\hat{\gamma}_{1,K-1}^{ML})^2}{\sigma_g^2} = a + \frac{aq_{1,2}^2}{q_{1,1}^2} = \frac{a}{q_{1,1}^2}. \tag{IA.55}$$

In addition, let $\hat{B}_1 = [1_N,\ \hat{\beta}_1,\ \hat{\beta}_2]$ and $\hat{G}_2 = [1_N,\ \hat{\beta}_1]$. We then have

$$\begin{aligned} & \frac{\boldsymbol{\iota}'_{K,K}(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\boldsymbol{\iota}_{K,K}}{T} \\ = & \left(T\hat{\beta}_2'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_2\right)^{-1} \\ \xrightarrow{d} & \ \sigma_g^2\left(\boldsymbol{\iota}'_{r,2}\tilde{Z}_1'[I_{N-1} - \tilde{M}_2(\tilde{M}_2'\tilde{M}_2)^{-1}\tilde{M}_2']\tilde{Z}_1\boldsymbol{\iota}_{r,2}\right)^{-1} \\ = & \ \frac{\sigma_g^2}{q_{1,1}^2 w_1 + q_{1,2}^2 w_2}. \end{aligned} \tag{IA.56}$$

It follows that

$$\begin{aligned} \frac{s^2(\hat{\gamma}_{1,K-1}^{ML})}{T} &= (1 + \hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML})\frac{\boldsymbol{\iota}'_{K,K}(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\boldsymbol{\iota}_{K,K}}{T} + \frac{\hat{\sigma}_g^2}{T} \\ &\xrightarrow{d} \frac{a}{q_{1,1}^2}\frac{\sigma_g^2}{(q_{1,1}^2 w_1 + q_{1,2}^2 w_2)}, \end{aligned} \tag{IA.57}$$

and the limiting distribution of $t(\hat{\gamma}_{1,K-1}^{ML})$ is given by

$$t(\hat{\gamma}_{1,K-1}^{ML}) \xrightarrow{d} -q_{1,2}(q_{1,1}^2 w_1 + q_{1,2}^2 w_2)^{\frac{1}{2}}. \tag{IA.58}$$

For the limiting distribution of the $t(\hat{\gamma}_{1,i}^{ML})$, for $i = 1, \ldots, K-2$, we first define the following random variables:

$$\frac{\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1b}}{[\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i}]^{\frac{1}{2}}} \xrightarrow{d} u_i, \tag{IA.59}$$

$$\frac{\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1a}}{[\boldsymbol{\iota}'_{K-2,i}(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i}]^{\frac{1}{2}}} \xrightarrow{d} v_i. \tag{IA.60}$$

It is easy to show that $u_i \sim \mathcal{N}(0,1)$, $v_i \sim \mathcal{N}(0,1)$, and they are independent of each other and of W_2^{11} (and hence they are independent of w_1, w_2, and q_1). Let $\sqrt{T}(\hat{\mu}_{f,i} - \mu_{f,i}) \stackrel{d}{\to} \sigma_{f,i} n_i$, where $n_i \sim \mathcal{N}(0,1)$. Then, we can write the numerator of $t(\hat{\gamma}_{1,i}^{ML})$ as

$$\begin{aligned}
\sqrt{T}(\hat{\gamma}_{1,i}^{ML} - \gamma_{1,i}^*) &= \sqrt{T}(\hat{\phi}_{1,i} - \phi_{1,i}^*) + \sqrt{T}(\hat{\mu}_{f,i} - \mu_{f,i}) \\
&\stackrel{d}{\to} \sqrt{a}\boldsymbol{\iota}_{K-2,i}'(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1a} \\
&\quad + \frac{\sqrt{a}q_{1,2}}{q_{1,1}}\boldsymbol{\iota}_{K-2,i}'(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1b} + \sigma_{f,i}n_i \\
&= \sqrt{a}[\boldsymbol{\iota}_{K-2,i}'(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i}]^{\frac{1}{2}}\left(v_i + \frac{q_{1,2}}{q_{1,1}}u_i\right) + \sigma_{f,i}n_i.
\end{aligned} \tag{IA.61}$$

By defining $c_i = \boldsymbol{\iota}_{K-2,i}'(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1)^{-1}\boldsymbol{\iota}_{K-2,i}$, we can also write the above expression as

$$\sqrt{T}(\hat{\gamma}_{1,i}^{ML} - \gamma_{1,i}^*) = \frac{q_{1,2}}{q_{1,1}}\sqrt{ac_i}u_i + \left(ac_i + \sigma_{f,i}^2\right)^{\frac{1}{2}} z_i, \tag{IA.62}$$

where $z_i \sim \mathcal{N}(0,1)$ and it is independent of u_i, q_1, w_1, and w_2.

For the denominator of $t(\hat{\gamma}_{1,i}^{ML})$, we first derive the limiting distribution of the $(i+1,i+1)$-element of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$ for $i = 1,\ldots,K-2$. Using the formula for the inverse of a partitioned matrix, we obtain

$$\begin{aligned}
&\boldsymbol{\iota}_{K,i+1}'(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)\boldsymbol{\iota}_{K,i+1} \\
=\ & \boldsymbol{\iota}_{K-2,i}'\left(\hat{\beta}_1'\hat{\Sigma}^{-\frac{1}{2}}\tilde{P}_1[I_{N-1} - \tilde{P}_1'\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_2(\hat{\beta}_2'\hat{\Sigma}^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_2)^{-1}\hat{\beta}_2'\hat{\Sigma}^{-\frac{1}{2}}\tilde{P}_1]\tilde{P}_1'\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_1\right)^{-1}\boldsymbol{\iota}_{K-2,i} \\
\stackrel{d}{\to}\ & \boldsymbol{\iota}_{K-2,i}'\left(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1[I_{N-1} - \tilde{Z}_{1b}(\tilde{Z}_{1b}'\tilde{Z}_{1b})^{-1}\tilde{Z}_{1b}']\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1\right)^{-1}\boldsymbol{\iota}_{K-2,i} \\
=\ & \boldsymbol{\iota}_{K-2,i}'\left(\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1 - \frac{\beta_1'\Sigma^{-\frac{1}{2}}\tilde{P}_1\tilde{Z}_{1b}\tilde{Z}_{1b}'\tilde{P}_1'\Sigma^{-\frac{1}{2}}\beta_1}{\tilde{Z}_{1b}'\tilde{Z}_{1b}}\right)^{-1}\boldsymbol{\iota}_{K-2,i} \\
=\ & c_i + \frac{c_i u_i^2}{q_{1,1}^2 w_1 + q_{1,2}^2 w_2},
\end{aligned} \tag{IA.63}$$

where the last equality follows from the identity

$$\left(A - \frac{1}{b}cc'\right)^{-1} \equiv A^{-1} + \frac{A^{-1}cc'A^{-1}}{b - c'A^{-1}c}. \tag{IA.64}$$

Combining these results and using (IA.55), we obtain

$$t(\hat{\gamma}_{1,i}^{ML}) \stackrel{d}{\to} \frac{\left(1 + \frac{\sigma_{f,i}^2}{ac_i}\right)^{\frac{1}{2}} z_i + \frac{q_{1,2}}{q_{1,1}}u_i}{\left[\frac{\sigma_{f,i}^2}{ac_i} + \frac{1}{q_{1,1}^2}\left(1 + \frac{u_i^2}{q_{1,1}^2w_1 + q_{1,2}^2w_2}\right)\right]^{\frac{1}{2}}}, \tag{IA.65}$$

for $i = 1, \dots K-2$. Since the term in the denominator is greater than

$$\frac{\sigma_{f,i}^2}{ac_i} + \frac{1}{q_{1,1}^2} = \frac{\sigma_{f,i}^2}{ac_i} + \frac{q_{1,1}^2 + q_{1,2}^2}{q_{1,1}^2} = 1 + \frac{\sigma_{f,i}^2}{ac_i} + \frac{q_{1,2}^2}{q_{1,1}^2}, \tag{IA.66}$$

which is the variance of the numerator, $t(\hat{\gamma}_{1,i}^{ML})$ is dominated by $\mathcal{N}(0,1)$.

Finally, we study the limiting distribution of $t(\hat{\gamma}_0^{ML})$. Using the formula for the inverse of a partitioned matrix, we can write

$$\begin{aligned} \boldsymbol{\iota}_{K,1}'(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}\boldsymbol{\iota}_{K,1} &= \boldsymbol{\iota}_{K-1,1}'(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\boldsymbol{\iota}_{K-1,1} \\ &+ \frac{[\boldsymbol{\iota}_{K-1,1}'(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_2]^2}{\hat{\beta}_2'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_2} \\ &\stackrel{d}{\to} \boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1} \\ &+ \boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}\frac{u_0^2}{q_{1,1}^2 w_1 + q_{1,2}^2 w_2}, \end{aligned} \tag{IA.67}$$

where u_0 is defined in (IA.52) and the last equality is obtained by using (IA.56). Then, using (IA.55), we obtain

$$t(\hat{\gamma}_0^{ML}) \stackrel{d}{\to} \frac{q_{1,1}v_0 + q_{1,2}u_0}{\left(1 + \frac{u_0^2}{q_{1,1}^2 w_1 + q_{1,2}^2 w_2}\right)^{\frac{1}{2}}}, \tag{IA.68}$$

which is dominated by $\mathcal{N}(0,1)$.

B. SDF Representation

For simplicity and ease of exposition, we consider the case in which the model is correctly specified and contains only a useless factor, that is, $\mu_R = \frac{1}{\lambda_0^*}1_N$, where λ_0^* is the pseudo-true value of λ_0. Suppose that g_t is a useless factor with mean μ_g and variance σ_g^2, and define $\tilde{D} = P_1'\hat{D}$ and $\tilde{d} = \text{vec}(\tilde{D})$. Under the null hypothesis of correct model specification, it is easy to show that

$$\sqrt{T}\tilde{d} \stackrel{d}{\to} \mathcal{N}(0_{2(N-1)}, V_{\tilde{d}}), \tag{IA.69}$$

where

$$V_{\tilde{d}} = \begin{bmatrix} 1 & \mu_g \\ \mu_g & \mu_g^2 + \sigma_g^2 \end{bmatrix} \otimes \tilde{U}, \tag{IA.70}$$

and $\tilde{U} = E[\tilde{R}_t\tilde{R}_t'] \equiv E[P_1'R_tR_t'P_1] = P_1'UP_1$.

It follows that

$$\begin{aligned}
\mathcal{J} &\stackrel{d}{\to} T \min_{c:c'c=1} c'\tilde{D}' \left(c' \begin{bmatrix} 1 & \mu_g \\ \mu_g & \mu_g^2+\sigma_g^2 \end{bmatrix} c\tilde{U} \right)^{-1} \tilde{D}c \\
&= \min_{c:c'c=1} \frac{Tc'\tilde{D}'\tilde{U}^{-1}\tilde{D}c}{c' \begin{bmatrix} 1 & \mu_g \\ \mu_g & \mu_g^2+\sigma_g^2 \end{bmatrix} c} \\
&= \min_{b:b'b=1} Tb' \begin{bmatrix} 1 & 0 \\ -\frac{\mu_g}{\sigma_g} & \frac{1}{\sigma_g} \end{bmatrix} \tilde{D}'\tilde{U}^{-1}\tilde{D} \begin{bmatrix} 1 & -\frac{\mu_g}{\sigma_g} \\ 0 & \frac{1}{\sigma_g} \end{bmatrix} b \\
&\stackrel{d}{\to} \min_{b:b'b=1} b'\tilde{Z}'\tilde{Z}b, \qquad \text{(IA.71)}
\end{aligned}$$

where

$$\sqrt{T}\tilde{U}^{-\frac{1}{2}}\tilde{D} \begin{bmatrix} 1 & -\frac{\mu_g}{\sigma_g} \\ 0 & \frac{1}{\sigma_g} \end{bmatrix} \stackrel{d}{\to} \tilde{Z} = [\tilde{Z}_1,\ \tilde{Z}_2]. \qquad \text{(IA.72)}$$

Note that $\tilde{Z}_1 \sim N(0_{N-1}, I_{N-1})$, $\tilde{Z}_2 \sim N(0_{N-1}, I_{N-1})$, and they are independent of each other. It follows that $\mathcal{J} \stackrel{d}{\to} \xi_2$, where ξ_2 is the smallest eigenvalue of $\tilde{Z}'\tilde{Z} \sim \mathcal{W}_2(N-1, I_2)$. In addition, $\hat{b} = [\hat{b}_1,\ \hat{b}_2]'$ that minimizes $b'\tilde{Z}'\tilde{Z}b$ has a uniform distribution over the unit circle. Finally, the $\hat{c}$ that minimizes $\mathcal{J}$ is proportional to

$$\hat{c} \propto \begin{bmatrix} 1 & -\frac{\mu_g}{\sigma_g} \\ 0 & \frac{1}{\sigma_g} \end{bmatrix} \hat{b} = \begin{bmatrix} \hat{b}_1 - \frac{\mu_g}{\sigma_g}\hat{b}_2 \\ \frac{\hat{b}_2}{\sigma_g} \end{bmatrix}. \qquad \text{(IA.73)}$$

Let

$$\hat{W}_e(\hat{c}) = (\hat{c}' \otimes I_N)\hat{V}_{\hat{d}}(\hat{c} \otimes I_N) \stackrel{d}{\to} (\hat{c}' \otimes I_N)V_{\hat{d}}(\hat{c} \otimes I_N), \qquad \text{(IA.74)}$$

where[1]

$$V_{\hat{d}} = \begin{bmatrix} V_R & \mu_g U \\ \mu_g U & (\mu_g^2+\sigma_g^2)U \end{bmatrix}. \qquad \text{(IA.75)}$$

It follows that

$$\hat{W}_e(\hat{c}) \stackrel{d}{\to} hU - \hat{c}_1^2\mu_R\mu_R', \qquad \text{(IA.76)}$$

where

$$h = \hat{c}' \begin{bmatrix} 1 & \mu_g \\ \mu_g & \mu_g^2+\sigma_g^2 \end{bmatrix} \hat{c}, \qquad \text{(IA.77)}$$

and the inverse of $\hat{W}_e(\hat{c})$ has the following limiting distribution:

$$\hat{W}_e(\hat{c})^{-1} \stackrel{d}{\to} \frac{1}{h}\left[U^{-1} + \frac{\hat{c}_1^2 U^{-1}\mu_R\mu_R' U^{-1}}{h - \hat{c}_1^2\mu_R' U^{-1}\mu_R}\right]. \qquad \text{(IA.78)}$$

[1]Note that unlike $V_{\tilde{d}}$, $V_{\hat{d}}$ does not have a Kronecker structure.

Using the fact that $\lambda_0^*\mu_R = 1_N$, we obtain

$$1_N'\hat{W}_e(\hat{c})^{-1} \xrightarrow{d} \frac{1}{h}\left[1_N'U^{-1} + \frac{\hat{c}_1^2(1_N'U^{-1}1_N)1_N'U^{-1}}{\lambda_0^{*2}h - \hat{c}_1^2 1_N'U^{-1}1_N}\right] = \left(\frac{\lambda_0^{*2}}{\lambda_0^{*2}h - \hat{c}_1^2 1_N'U^{-1}1_N}\right)1_N'U^{-1}. \quad \text{(IA.79)}$$

The CU-GMM estimates of λ are related to $\hat{c}$ by the following equation:

$$\hat{\lambda} = \frac{1_N'\hat{W}_e(\hat{c})^{-1}1_N}{1_N'\hat{W}_e(\hat{c})^{-1}\hat{D}\hat{c}}\hat{c} \xrightarrow{d} \frac{1_N'U^{-1}1_N}{1_N'U^{-1}\hat{\mu}_R\hat{c}_1 + 1_N'U^{-1}\hat{d}_2\hat{c}_2}\begin{bmatrix}\hat{c}_1\\ \hat{c}_2\end{bmatrix}, \quad \text{(IA.80)}$$

where $\hat{d}_2 = \sum_{t=1}^T R_t g_t/T$. Note that $1_N'U^{-1}\hat{d}_2$ and $1_N'U^{-1}\hat{\mu}_R$ are asymptotically independent of $\tilde{Z}$, and hence they are also independent of $\hat{c}$. There are two cases to consider: $\mu_g = 0$ and $\mu_g \neq 0$. When $\mu_g = 0$, the term $1_N'U^{-1}\hat{\mu}_R \xrightarrow{p} 1_N'U^{-1}1_N/\lambda_0^*$ dominates, and we have

$$\hat{\lambda} \xrightarrow{d} \begin{bmatrix}\lambda_0^*\\ \frac{\lambda_0^*\hat{c}_2}{\hat{c}_1}\end{bmatrix} = \begin{bmatrix}\lambda_0^*\\ \frac{\lambda_0^*\hat{b}_2}{\sigma_g\hat{b}_1}\end{bmatrix}. \quad \text{(IA.81)}$$

Therefore, $\hat{\lambda}_0$ is a consistent estimator of λ_0^*. In addition, using the fact that $\sqrt{T}\hat{d}_2 \xrightarrow{d} N(0_N, \sigma_g^2 U)$, we have

$$\begin{aligned}\sqrt{T}(\hat{\lambda}_0 - \lambda_0^*) &\xrightarrow{d} -\frac{\lambda_0^{*2}\sqrt{T}1_N'U^{-1}\hat{d}_2\hat{c}_2}{\hat{c}_1(1_N'U^{-1}1_N)}\\ &\xrightarrow{d} -\lambda_0^{*2}(1_N'U^{-1}1_N)^{-\frac{1}{2}}zr,\end{aligned} \quad \text{(IA.82)}$$

where $\sqrt{T}(1_N'U^{-1}\hat{d}_2)/[(1_N'U^{-1}1_N)^{\frac{1}{2}}\sigma_g] \xrightarrow{d} z \sim N(0,1)$, $r = \hat{b}_2/\hat{b}_1 \sim t_1$, and z and r are independent of each other. Note that the moments of the limiting distribution of $\sqrt{T}(\hat{\lambda}_0 - \lambda_0^*)$ do not exist.

When $\mu_g \neq 0$, we have $\hat{d}_2 \xrightarrow{p} \mu_R\mu_g = 1_N\mu_g/\lambda_0^*$ and hence

$$1_N'U^{-1}\hat{\mu}_R\hat{c}_1 + 1_N'U^{-1}\hat{d}_2\hat{c}_2 \xrightarrow{d} 1_N'U^{-1}1_N\frac{\hat{c}_1 + \mu_g\hat{c}_2}{\lambda_0^*}. \quad \text{(IA.83)}$$

It follows that

$$\hat{\lambda} \xrightarrow{d} \frac{\lambda_0^*}{\hat{c}_1 + \mu_g\hat{c}_2}\begin{bmatrix}\hat{c}_1\\ \hat{c}_2\end{bmatrix} = \lambda_0^*\begin{bmatrix}1 - \frac{\mu_g\hat{b}_2}{\sigma_g\hat{b}_1}\\ \frac{\hat{b}_2}{\sigma_g\hat{b}_1}\end{bmatrix}. \quad \text{(IA.84)}$$

Note that when the useless factor has a non-zero mean, $\hat{\lambda}_0$ will not be a consistent estimator of λ_0^*.

We now turn our attention to the t-ratios of $\hat{\lambda}_0$ and $\hat{\lambda}_1$. Let $e_t(\hat{\lambda}) = R_t(\hat{\lambda}_0 + \hat{\lambda}_1 g_t) - 1_N$. Then, we have

$$\bar{e}(\hat{\lambda}) = \frac{1}{T}\sum_{t=1}^{T} e_t(\hat{\lambda}) = \hat{\mu}_R\hat{\lambda}_0 + \hat{d}_2\hat{\lambda}_1 - 1_N \xrightarrow{d} \mu_R\lambda_0^* - 1_N = 0_N, \tag{IA.85}$$

$$\begin{aligned}
\hat{W}_e(\hat{\lambda}) &= \frac{1}{T}\sum_{t=1}^{T}\left(e_t(\hat{\lambda}) - \bar{e}(\hat{\lambda})\right)\left(e_t(\hat{\lambda}) - \bar{e}(\hat{\lambda})\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T} R_tR_t'(\hat{\lambda}_0 + \hat{\lambda}_1 g_t)^2 - (1_N + \hat{e})(1_N + \hat{e})' \\
&\xrightarrow{d} \frac{1}{T}\sum_{t=1}^{T}\lambda_0^{*2}\left(1 + \frac{(g_t - \mu_g)\hat{b}_2}{\sigma_g\hat{b}_1}\right)^2 R_tR_t' - (1_N + \hat{e})(1_N + \hat{e})' \\
&\xrightarrow{d} \lambda_0^{*2}\left(1 + \frac{\hat{b}_2^2}{\hat{b}_1^2}\right)U - 1_N 1_N'.
\end{aligned} \tag{IA.86}$$

It follows that[2]

$$(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1} \xrightarrow{d} a\left[(\hat{D}'U^{-1}\hat{D})^{-1} - \frac{(\hat{D}'U^{-1}\hat{D})^{-1}(\hat{D}'U^{-1}1_N)(1_N'U^{-1}\hat{D})(\hat{D}'U^{-1}\hat{D})^{-1}}{a - 1_N'U^{-1}1_N + 1_N'U^{-1}\hat{D}(\hat{D}'U^{-1}\hat{D})^{-1}\hat{D}'U^{-1}1_N}\right], \tag{IA.89}$$

where $a = \lambda_0^{*2}[1 + (\hat{b}_2/\hat{b}_1)^2]$. Note that when $\mu_g \neq 0$, we have

$$1_N'U^{-1}\hat{D}(\hat{D}'U^{-1}\hat{D})^{-1}\hat{D}'U^{-1}1_N \xrightarrow{p} 1_N'U^{-1}1_N, \tag{IA.90}$$

$$\frac{1}{\sqrt{T}}(\hat{D}'U^{-1}\hat{D})^{-1}\hat{D}'U^{-1}1_N \xrightarrow{p} 0_2. \tag{IA.91}$$

As a result,

$$\frac{(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}}{T} \xrightarrow{d} \lambda_0^{*2}\left(1 + \frac{\hat{b}_2^2}{\hat{b}_1^2}\right)\frac{(\hat{D}'U^{-1}\hat{D})^{-1}}{T}. \tag{IA.92}$$

In addition, we have

$$\begin{aligned}
\frac{\boldsymbol{\iota}_{2,1}'(\hat{D}'U^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{2,1}}{T} &= \frac{1}{T\hat{\mu}_R'U^{-\frac{1}{2}}[I_N - U^{-\frac{1}{2}}\hat{d}_2(\hat{d}_2'U^{-1}\hat{d}_2)^{-1}\hat{d}_2'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}\hat{\mu}_R} \\
&\xrightarrow{d} \frac{\mu_g^2}{\sigma_g^2\tilde{Z}_2'\tilde{Z}_2}.
\end{aligned} \tag{IA.93}$$

[2] When $W = aU - 1_N 1_N'$, where a is a scalar, we have

$$W^{-1} = \frac{1}{a}\left(U^{-1} + \frac{U^{-1}1_N 1_N'U^{-1}}{a - 1_N'U^{-1}1_N}\right), \tag{IA.87}$$

$$(D'W^{-1}D)^{-1} = a\left[(D'U^{-1}D)^{-1} - \frac{(D'U^{-1}D)^{-1}(D'U^{-1}1_N)(1_N'U^{-1}D)(D'U^{-1}D)^{-1}}{a - 1_N'U^{-1}1_N + 1_N'U^{-1}D(D'U^{-1}D)^{-1}(D'U^{-1}1_N)}\right]. \tag{IA.88}$$

The proof of this limiting result is nontrivial. Let $\tilde{1}_N = U^{-\frac{1}{2}}1_N/(1_N'U^{-1}1_N)^{\frac{1}{2}}$ and $P_{\tilde{1}_N}$ be an $N \times (N-1)$ orthonormal matrix with its columns orthogonal to $\tilde{1}_N$. It follows that

$$P_{\tilde{1}_N}P_{\tilde{1}_N}' = I_N - U^{-\frac{1}{2}}1_N(1_N'U^{-1}1_N)^{-1}1_N'U^{-\frac{1}{2}} = U^{\frac{1}{2}}P_1(P_1'UP_1)^{-1}P_1'U^{\frac{1}{2}}. \quad \text{(IA.94)}$$

Now, we can write

$$\hat{D}'U^{-1}\hat{D} = \hat{D}'U^{-\frac{1}{2}}[\tilde{1}_N,\ P_{\tilde{1}_N}][\tilde{1}_N,\ P_{\tilde{1}_N}]'U^{-\frac{1}{2}}\hat{D}. \quad \text{(IA.95)}$$

Let

$$\sqrt{T}\tilde{1}_N'U^{-\frac{1}{2}}\hat{\mu}_R \xrightarrow{p} \sqrt{T}(1_N'U^{-1}1_N)^{\frac{1}{2}}/\lambda_0^* \equiv \sqrt{T}v, \quad \text{(IA.96)}$$

$$\sqrt{T}\tilde{1}_N'U^{-\frac{1}{2}}\hat{d}_2 \xrightarrow{p} \sqrt{T}(1_N'U^{-1}1_N)^{\frac{1}{2}}\mu_g/\lambda_0^* \equiv \sqrt{T}\mu_g v, \quad \text{(IA.97)}$$

where $v = (1_N'U^{-1}1_N)^{\frac{1}{2}}/\lambda_0^*$. In addition, it is easy to show that

$$\sqrt{T}P_{\tilde{1}_N}'U^{-\frac{1}{2}}\hat{\mu}_R \xrightarrow{d} \tilde{Z}_1, \quad \text{(IA.98)}$$

$$\sqrt{T}P_{\tilde{1}_N}'U^{-\frac{1}{2}}\hat{d}_2 \xrightarrow{d} \mu_g\tilde{Z}_1 + \sigma_g\tilde{Z}_2. \quad \text{(IA.99)}$$

It follows that

$$T(\hat{D}'U^{-1}\hat{D}) \xrightarrow{d} \begin{bmatrix} \tilde{Z}_1'\tilde{Z}_1 + Tv^2 & \tilde{Z}_1'(\mu_g\tilde{Z}_1 + \sigma_g\tilde{Z}_2) + \mu_g Tv^2 \\ \tilde{Z}_1'(\mu_g\tilde{Z}_1 + \sigma_g\tilde{Z}_2) + \mu_g Tv^2 & (\mu_g\tilde{Z}_1 + \sigma_g\tilde{Z}_2)'(\mu_g\tilde{Z}_1 + \sigma_g\tilde{Z}_2) + \mu_g^2 Tv^2 \end{bmatrix}, \quad \text{(IA.100)}$$

and the $(1,1)$-element of $(\hat{D}'U^{-1}\hat{D})^{-1}/T$ has the following limiting distribution:

$$\frac{\mu_g^2(Tv^2 + \tilde{Z}_1'\tilde{Z}_1) + 2\mu_g\sigma_g\tilde{Z}_1'\tilde{Z}_2 + \sigma_g^2\tilde{Z}_2'\tilde{Z}_2}{\sigma_g^2[(Tv^2 + \tilde{Z}_1'\tilde{Z}_1)\tilde{Z}_2'\tilde{Z}_2 - (\tilde{Z}_1'\tilde{Z}_2)^2]}. \quad \text{(IA.101)}$$

As $\tilde{Z}_1$ and $\tilde{Z}_2$ are $O_p(1)$, all the terms that are not associated with v can be ignored, and we have

$$\frac{\boldsymbol{\iota}_{2,1}'(\hat{D}'U^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{2,1}}{T} \xrightarrow{d} \frac{\mu_g^2}{\sigma_g^2\tilde{Z}_2'\tilde{Z}_2}. \quad \text{(IA.102)}$$

Similarly, we have

$$\begin{aligned} \frac{\boldsymbol{\iota}_{2,2}'(\hat{D}'U^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{2,2}}{T} &= \frac{1}{T\hat{d}_2'U^{-\frac{1}{2}}[I_N - U^{-\frac{1}{2}}\hat{\mu}_R(\hat{\mu}_R'U^{-1}\hat{\mu}_R)^{-1}\hat{\mu}_R'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}\hat{d}_2} \\ &\xrightarrow{d} \frac{1}{\sigma_g^2\tilde{Z}_2'\tilde{Z}_2}. \end{aligned} \quad \text{(IA.103)}$$

For the t-ratio of $\hat{\lambda}_0$, we have

$$t(\hat{\lambda}_0) \xrightarrow{d} \left(1 - \frac{\mu_g\hat{b}_2}{\sigma_g\hat{b}_1}\right)\frac{\sigma_g|\hat{b}_1|}{|\mu_g|}(\tilde{Z}_2'\tilde{Z}_2)^{\frac{1}{2}} = \left(1 - \frac{\mu_g\hat{b}_2}{\sigma_g\hat{b}_1}\right)\frac{\sigma_g|\hat{b}_1|}{|\mu_g|}(\xi_1\hat{b}_1^2 + \xi_2\hat{b}_2^2)^{\frac{1}{2}}, \quad \text{(IA.104)}$$

where $\xi_1 > \xi_2 > 0$ are the two eigenvalues of $\tilde{Z}'\tilde{Z}$ and $[\hat{b}_1,\ \hat{b}_2]'$ is the eigenvector of $\tilde{Z}'\tilde{Z}$ associated with ξ_2. Note that the limiting distribution of $t(\hat{\lambda}_0)$ is not symmetric around zero.

For the t-ratio of $\hat{\lambda}_1$, we have

$$t(\hat{\lambda}_1) \xrightarrow{d} \hat{b}_2(\tilde{Z}_2'\tilde{Z}_2)^{\frac{1}{2}} = \hat{b}_2(\xi_1\hat{b}_1^2 + \xi_2\hat{b}_2^2)^{\frac{1}{2}}. \tag{IA.105}$$

When $\mu_g = 0$, the limiting distribution of $t(\hat{\lambda}_1)$ stays the same. For $\hat{\lambda}_0$, we derive the limiting distribution of its centered t-ratio, $t_c(\hat{\lambda}_0)$. When $\mu_g = 0$, we have $\sqrt{T}\tilde{1}_N' U^{-\frac{1}{2}}\hat{d}_2 \xrightarrow{d} \sigma_g z$, and we can show that

$$T(\hat{D}'U^{-1}\hat{D}) \xrightarrow{d} \begin{bmatrix} \tilde{Z}_1'\tilde{Z}_1 + Tv^2 & \sigma_g\tilde{Z}_1'\tilde{Z}_2 + \sqrt{T}\sigma_g vz \\ \sigma_g\tilde{Z}_1'\tilde{Z}_2 + \sqrt{T}\sigma_g vz & \sigma_g^2\tilde{Z}_2'\tilde{Z}_2 + \sigma_g^2 z^2 \end{bmatrix}. \tag{IA.106}$$

Using the same type of derivations as before, we have

$$\boldsymbol{\iota}_{2,1}'(\hat{D}'U^{-1}\hat{D})^{-1}\hat{D}'U^{-1}1_N \xrightarrow{p} \lambda_0^*, \tag{IA.107}$$

$$1_N'U^{-1}\hat{D}(\hat{D}'U^{-1}\hat{D})^{-1}\hat{D}'U^{-1}1_N \xrightarrow{p} 1_N'U^{-1}1_N, \tag{IA.108}$$

$$\boldsymbol{\iota}_{2,1}'(\hat{D}'U^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{2,1} \xrightarrow{d} \frac{\lambda_0^{*2}(z^2 + \tilde{Z}_2'\tilde{Z}_2)}{(1_N'U^{-1}1_N)(\tilde{Z}_2'\tilde{Z}_2)}, \tag{IA.109}$$

$$\boldsymbol{\iota}_{2,1}'(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{2,1} \xrightarrow{d} \lambda_0^{*4}\left(1 + \frac{\hat{b}_2^2}{\hat{b}_1^2}\right)\frac{z^2 + \tilde{Z}_2'\tilde{Z}_2}{(1_N'U^{-1}1_N)\tilde{Z}_2'\tilde{Z}_2} - \lambda_0^{*2}. \tag{IA.110}$$

As a result, the centered t-ratio of $\hat{\lambda}_0$ is given by

$$t_c(\hat{\lambda}_0) \xrightarrow{d} \frac{z\hat{b}_2}{\left[\frac{z^2 + \tilde{Z}_2'\tilde{Z}_2}{\tilde{Z}_2'\tilde{Z}_2} - \frac{\hat{b}_1^2(1_N'U^{-1}1_N)}{\lambda_0^{*2}}\right]^{\frac{1}{2}}}, \tag{IA.111}$$

with $\tilde{Z}_2'\tilde{Z}_2 = \hat{b}_1^2\xi_1 + \hat{b}_2^2\xi_2$. It is easy to see that $t_c(\hat{\lambda}_0)$ is symmetrically distributed around zero.

2 Misspecified Model: Two Noisy Factors

The full rank condition may also be violated when the model includes two factors that are noisy (due to measurement error, for instance) versions of the same underlying factor. Denote the first $K-3$ factors by the vector $f_{1,t}$ and suppose that the last two factors of $f_t = [f_{1,t}', f_{K-2,t}, f_{K-1,t}]'$ are noisy counterparts of each other, that is, $f_{K-2,t} = f_t^0 + \eta_{1,t}$ and $f_{K-1,t} = f_t^0 + \eta_{2,t}$, where f_t^0 is some underlying (but unobserved) factor, and $\eta_{1,t}$ and $\eta_{2,t}$ are two independent measurement errors with mean zero and $\text{Var}[\eta_{1,t}] = \sigma_{\eta_1}^2$ and $\text{Var}[\eta_{2,t}] = \sigma_{\eta_2}^2$.

A. Beta-Pricing Model

From the definition of noisy factors, it follows that

$$\mathrm{Cov}[R_t, f_{K-2,t}] = \mathrm{Cov}[R_t, f_{K-1,t}]. \tag{IA.112}$$

Partition $\beta = [\beta_1,\ \beta_2,\ \beta_3]$. Using the formula for the inverse of a partitioned matrix, we can easily show that

$$\sigma^2_{\eta_1}\beta_2 = \sigma^2_{\eta_2}\beta_3. \tag{IA.113}$$

Without loss of generality, we assume $\sigma^2_{\eta_2} \neq 0$. This will give us

$$\beta_3 = \frac{\sigma^2_{\eta_1}}{\sigma^2_{\eta_2}}\beta_2. \tag{IA.114}$$

In particular, if $\sigma^2_{\eta_1} = 0$, we have $\beta_3 = 0_N$, and the results are identical to the ones for the useless factor case.

If the model is misspecified, we have $Gv^* = 0_N$ for $v^* = [0'_{K-1}, -\sigma^2_{\eta_1}/\sigma^2_{\eta_2},\ 1]'$. Let $\hat{v}$ be the eigenvector associated with the largest eigenvalue of

$$\hat{\Omega} = (\hat{G}'\hat{\Sigma}^{-1}\hat{G})^{-1}[A(X'X/T)^{-1}A']. \tag{IA.115}$$

Define $\hat{\psi} = [\hat{\psi}_1, \ldots, \hat{\psi}_K]'$ as

$$\hat{\psi}_i = -\frac{\hat{v}_i}{\hat{v}_{K+1}}, \quad i = 1, \ldots, K, \tag{IA.116}$$

which is asymptotically equivalent to the estimator

$$\tilde{\psi} = (\hat{G}'_1\hat{\Sigma}^{-1}\hat{G}_1)^{-1}(\hat{G}'_1\hat{\Sigma}^{-1}\hat{\beta}_3), \tag{IA.117}$$

where $\hat{G}_1 = [1_N,\ \hat{\alpha},\ \hat{\beta}_1,\ \hat{\beta}_2]$. Let $\psi^* = [0'_{K-1},\ \sigma^2_{\eta_1}/\sigma^2_{\eta_2}]'$ and $G_1 = [1_N,\ \alpha,\ \beta_1,\ \beta_2]$. Since

$$\sqrt{T}(\hat{\beta}_3 - (\sigma^2_{\eta_1}/\sigma^2_{\eta_2})\hat{\beta}_2) \xrightarrow{d} \mathcal{N}\left(0_N, \frac{\sigma^2_{\eta_1} + \sigma^2_{\eta_2}}{\sigma^4_{\eta_2}}\Sigma\right), \tag{IA.118}$$

it follows that

$$\sqrt{T}(\tilde{\psi} - \psi^*) \xrightarrow{d} \mathcal{N}\left(0_K, \frac{\sigma^2_{\eta_1} + \sigma^2_{\eta_2}}{\sigma^4_{\eta_2}}(G'_1\Sigma^{-1}G_1)^{-1}\right), \tag{IA.119}$$

and $\hat{\psi}$ also has the same asymptotic distribution. Specifically, using the fact that

$$\sqrt{T}\frac{\hat{\sigma}^2_{\eta_2}}{\sqrt{\hat{\sigma}^2_{\eta_1} + \hat{\sigma}^2_{\eta_2}}}\hat{\Sigma}^{-\frac{1}{2}}(\hat{\beta}_3 - (\sigma^2_{\eta_1}/\sigma^2_{\eta_2})\hat{\beta}_2) \xrightarrow{d} u \sim \mathcal{N}(0_N, I_N), \tag{IA.120}$$

where $\hat{\sigma}^2_{\eta_1}$ and $\hat{\sigma}^2_{\eta_2}$ are the estimated variances of $\eta_{1,t}$ and $\eta_{2,t}$, respectively, it follows that

$$\sqrt{T}(\hat{\psi} - \psi^*) \xrightarrow{d} z \sim \mathcal{N}\left(0_K, \frac{\sigma^2_{\eta_1} + \sigma^2_{\eta_2}}{\sigma^4_{\eta_2}}(G_1'\Sigma^{-1}G_1)^{-1}\right). \tag{IA.121}$$

Therefore, assuming $\sigma^2_{\eta_1} > 0$, we can write

$$\hat{\gamma}_0^{ML} = -\frac{\sqrt{T}\hat{\psi}_1}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} -\frac{z_1}{z_2}, \tag{IA.122}$$

$$\hat{\gamma}_{1,i}^{ML} = \hat{\mu}_{f,i} - \frac{\sqrt{T}\hat{\psi}_{i+2}}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} \mu_{f,i} - \frac{z_{i+2}}{z_2}, \quad i = 1, \ldots, K-3, \tag{IA.123}$$

$$\frac{\hat{\gamma}_{1,K-2}^{ML}}{\sqrt{T}} = \frac{\hat{\mu}_{f,K-2}}{\sqrt{T}} - \frac{\hat{\psi}_K}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} -\frac{\sigma^2_{\eta_1}}{\sigma^2_{\eta_2} z_2}, \tag{IA.124}$$

$$\frac{\hat{\gamma}_{1,K-1}^{ML}}{\sqrt{T}} = \frac{\hat{\mu}_{f,K-1}}{\sqrt{T}} + \frac{1}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} \frac{1}{z_2}. \tag{IA.125}$$

For the asymptotic distributions of the t-ratios, we first show that $t^2(\hat{\gamma}_{1,K-2}^{ML}) \xrightarrow{d} \chi^2_{N-K+1}$ and $t^2(\hat{\gamma}_{1,K-1}^{ML}) \xrightarrow{d} \chi^2_{N-K+1}$. Let $\hat{G}_2 = [1_N,\ \hat{\beta}_1,\ \hat{\beta}_2]$. Using the formula for the inverse of a partitioned matrix, we obtain

$$s^2(\hat{\gamma}_{1,K-1}^{ML}) = (1 + \hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML})\left(\hat{\beta}_3'[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_3\right)^{-1} + \hat{\sigma}^2_{f^0}, \tag{IA.126}$$

where $\hat{\sigma}^2_{f^0}$ is the sample estimate of $\sigma^2_{f^0}$, the population variance f_t^0. Since $\hat{\gamma}_{1,i}^{ML} = O_p(1)$ for $i = 1, \ldots, K-3$, $\hat{\gamma}_{1,K-2}^{ML} = O_p(T^{\frac{1}{2}})$, and $\hat{\gamma}_{1,K-1}^{ML} = O_p(T^{\frac{1}{2}})$, it follows that

$$\frac{\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}{T} \xrightarrow{d} \frac{\sigma^2_{\eta_1} + \sigma^2_{\eta_2}}{\sigma^4_{\eta_2} z_2^2} \tag{IA.127}$$

and

$$s^2(\hat{\gamma}_{1,K-1}^{ML}) = \frac{\sigma^2_{\eta_1} + \sigma^2_{\eta_2}}{\sigma^4_{\eta_2} z_2^2}\left(\hat{\beta}_3'[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_3\right)^{-1} + O_p(T^{\frac{1}{2}}). \tag{IA.128}$$

Writing

$$\sqrt{T}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_3 = \sqrt{T}\left(\frac{\sigma^2_{\eta_1}}{\sigma^2_{\eta_2}}\right)\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_2 + \frac{\sqrt{\sigma^2_{\eta_1} + \sigma^2_{\eta_2}}}{\sigma^2_{\eta_2}}u + O_p(T^{-\frac{1}{2}}) \tag{IA.129}$$

and using that

$$\hat{\beta}_2'\hat{\Sigma}^{-\frac{1}{2}}[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_2 \tag{IA.130}$$

is identically zero because $\hat{\beta}_2$ is in the span of the column space of $\hat{G}_2$, we have

$$\begin{aligned} t^2(\hat{\gamma}_{1,K-1}^{ML}) &= \frac{T(\hat{\gamma}_{1,K-1}^{ML})^2\hat{\beta}_3'[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_3}{(\sigma_{\eta_1}^2 + \sigma_{\eta_2}^2)/(\sigma_{\eta_2}^4 z_2^2)} + O_p(T^{-\frac{1}{2}}) \\ &= u'[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]u + O_p(T^{-\frac{1}{2}}) \\ &\stackrel{d}{\to} u'[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u \sim \chi^2_{N-K+1}. \end{aligned} \tag{IA.131}$$

The proof for the limiting distribution of $t^2(\hat{\gamma}_{1,K-2}^{ML})$ is similar.

We next proceed with the derivation of the limiting distributions of $t(\hat{\gamma}_0^{ML})$ and $t(\hat{\gamma}_{1,i}^{ML})$, $i = 1,\ldots,K-3$. Let $G_2 = [1_N,\ \beta_1,\ \beta_2]$ and $\hat{G}_2 = [1_N,\ \hat{\beta}_1,\ \hat{\beta}_2]$. Using the formula for the inverse of a partitioned matrix, we obtain the upper left $(K-1)\times(K-1)$ block of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$ as

$$\begin{aligned} & (\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1} + \frac{(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_3\hat{\beta}_3'\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}}{\hat{\beta}_3'\hat{\Sigma}^{-1}\hat{\beta}_3 - \hat{\beta}_3'\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_3} \\ = \ & (G_2'\Sigma^{-1}G_2)^{-1} + \frac{(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}uu'\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}}{u'[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u} + O_p(T^{-\frac{1}{2}}). \end{aligned} \tag{IA.132}$$

In particular, the upper left $(K-2)\times(K-2)$ block of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$ is $O_p(1)$. Note that we can write

$$I_N - \Sigma^{-\frac{1}{2}}G_1(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}} = I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}} - hh', \tag{IA.133}$$

where

$$h = \frac{[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha}{\left(\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha\right)^{\frac{1}{2}}}. \tag{IA.134}$$

Let $\sigma_{ij} \equiv \mathrm{Cov}[z_i, z_j] = \boldsymbol{\iota}_{K,i}'(G_1'\Sigma^{-1}G_1)^{-1}\boldsymbol{\iota}_{K,j}/\sigma_{f^0}^2$. The same proof as in the useless factor case allows us to show that (1) $h'u = z_2/\sigma_2 \equiv \tilde{z}_2 \sim \mathcal{N}(0,1)$, and (2) $u'[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u = x + \tilde{z}_2^2$, where $x \sim \chi^2_{N-K}$ and is independent of $\tilde{z}_2$.

Denote by w_i the i-th diagonal element of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$, $i = 1,\ldots,K-1$. Using (IA.132), we have

$$\begin{aligned} w_i &\stackrel{d}{\to} \boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i} + \frac{\boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}uu'\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}}{x + \tilde{z}_2^2} \\ &= \boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}\left(1 + \frac{q_i^2}{x + \tilde{z}_2^2}\right), \end{aligned} \tag{IA.135}$$

where

$$q_i = \frac{\boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}u}{[\boldsymbol{\iota}'_{K-1,i}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}]^{\frac{1}{2}}} \sim \mathcal{N}(0,1). \tag{IA.136}$$

Again, the same proof as in the useless factor case in the paper allows us to establish that

$$z_1 = \frac{\sigma_{12}}{\sigma_2^2}z_2 + \sqrt{1-\rho_{12}^2}\sigma_1 q_1 = \sigma_1\left(\rho_{12}\tilde{z}_2 + \sqrt{1-\rho_{12}^2}q_1\right), \tag{IA.137}$$

$$z_{i+1} = \frac{\sigma_{i+1,2}}{\sigma_2^2}z_2 + \sqrt{1-\rho_{i+1,2}^2}\sigma_{i+1}q_i = \sigma_{i+1}\left(\rho_{i+1,2}\tilde{z}_2 + \sqrt{1-\rho_{i+1,2}^2}q_i\right), \quad i = 2,\ldots,K-2. \tag{IA.138}$$

Let

$$b_i = \frac{x+\tilde{z}_2^2}{x+\tilde{z}_2^2+q_i^2}, \quad i = 1,\ldots,K-2. \tag{IA.139}$$

With the above results, we can write the limiting distribution of the t-ratios as

$$\begin{aligned} t(\hat{\gamma}_0^{ML}) &\stackrel{d}{\to} -\frac{z_1|z_2|b_1^{\frac{1}{2}}}{z_2[\boldsymbol{\iota}'_{K-1,1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma_{f0}^2]^{\frac{1}{2}}} \\ &= -\left(\frac{\rho_{12}|\tilde{z}_2|}{\sqrt{1-\rho_{12}^2}} + q_1\right)b_1^{\frac{1}{2}}, \end{aligned} \tag{IA.140}$$

$$\begin{aligned} t(\hat{\gamma}_{1,i}^{ML}) &\stackrel{d}{\to} \frac{\left(\mu_{f,i} - \frac{z_{i+2}}{z_2}\right)|z_2|b_{i+1}^{\frac{1}{2}}}{[\boldsymbol{\iota}'_{K-1,i+1}(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i+1}/\sigma_{f0}^2]^{\frac{1}{2}}} \\ &= \left(\frac{\frac{\mu_{f,i}\sigma_2}{\sigma_{i+2}} - \rho_{i+2,2}}{\sqrt{1-\rho_{i+2,2}^2}}|\tilde{z}_2| - q_{i+1}\right)b_{i+1}^{\frac{1}{2}}, \quad i = 1,\ldots,K-3. \end{aligned} \tag{IA.141}$$

B. SDF Representation

First note that the last two columns of D are $d_{K-1} = E[R_t f_{K-2,t}] = E[R_t f_{K-1,t}] = d_K$. Therefore, we have

$$H[0'_{K-1},\ -1,\ 1]' = 0_N. \tag{IA.142}$$

Since H has rank K, we know $\mathcal{J} \stackrel{d}{\to} \chi^2_{N-K}$ because of the numerical equivalence between the $\mathcal{J}$test and the $\mathcal{CD}_2$ test.

To obtain the limiting distribution of $\hat{\lambda}$, we first perform an alternative parameterization of the problem. Let

$$g_t(c) = H_t\begin{bmatrix} c \\ 1 \end{bmatrix}, \tag{IA.143}$$

where $H_t = [1_N,\ D_t]$. When there are two noisy factors, we have $H[c^{*\prime},\ 1]' = 0_N$, where

$$c^* = \begin{bmatrix} 0_{K-1} \\ -1 \end{bmatrix}. \tag{IA.144}$$

Consider the CU-GMM estimator of c^*:

$$\hat{c} = \text{argmin}_c \bar{g}(c)' \hat{W}_g(c)^{-1} \bar{g}(c), \tag{IA.145}$$

where $\bar{g}(c) = \sum_{t=1}^T g_t(c)/T$ and

$$\hat{W}_g(c) = \frac{1}{T} \sum_{t=1}^T [g_t(c) - \bar{g}(c)][g_t(c) - \bar{g}(c)]'. \tag{IA.146}$$

The asymptotic distribution of $\hat{c}$ is given by

$$\sqrt{T}(\hat{c} - c^*) \stackrel{d}{\to} \mathcal{N}\left(0_K, (H_1' S_g^{-1} H_1)^{-1}\right), \tag{IA.147}$$

where D_1 is the first $K-1$ columns of D, $H_1 = [1_N,\ D_1]$, and

$$S_g = E[g_t(c^*) g_t(c^*)'] = E[R_t R_t' (\eta_{2,t} - \eta_{1,t})^2] = U \sigma_g^2, \tag{IA.148}$$

with $U = E[R_t R_t']$ and $\sigma_g^2 = \text{Var}[\eta_{2,t} - \eta_{1,t}] = \sigma_{\eta_1}^2 + \sigma_{\eta_2}^2$.

Let $\hat{H}_1 = [1_N,\ \hat{D}_1]$ and $\hat{d}_K$ be the last column of $\hat{D}$. Note that $\hat{c}$ has the same asymptotic distribution as

$$\tilde{c} = (\hat{H}_1' \hat{U}^{-1} \hat{H}_1)^{-1} \hat{H}_1' \hat{U}^{-1} \hat{d}_K. \tag{IA.149}$$

Let

$$z \sim \mathcal{N}(0_K, \sigma_g^2 (H_1' U^{-1} H_1)^{-1}). \tag{IA.150}$$

Then, we have

$$\sqrt{T}\hat{c}_i \stackrel{d}{\to} z_i, \quad i = 1, \ldots, K-1, \tag{IA.151}$$

$$\sqrt{T}(\hat{c}_K + 1) \stackrel{d}{\to} z_K. \tag{IA.152}$$

Due to the invariance property of CU-GMM, we know $[-1,\ \hat{\lambda}']$ is proportional to $[\hat{c}',\ 1]$. It then follows that

$$\hat{\lambda}_0 = -\frac{\hat{c}_2}{\hat{c}_1}, \tag{IA.153}$$

$$\hat{\lambda}_{1,i} = -\frac{\hat{c}_{i+2}}{\hat{c}_1}, \quad i = 1, \ldots, K-2, \tag{IA.154}$$

$$\hat{\lambda}_{1,K-1} = -\frac{1}{\hat{c}_1}. \tag{IA.155}$$

Therefore, the limiting distribution of $\hat{\lambda}$ is given by (for $i = 1, \ldots, K-3$)

$$\hat{\lambda}_0 \xrightarrow{d} -\frac{z_2}{z_1}, \tag{IA.156}$$

$$\hat{\lambda}_{1,i} \xrightarrow{d} -\frac{z_{i+2}}{z_1}, \tag{IA.157}$$

$$\frac{\hat{\lambda}_{1,K-2}}{\sqrt{T}} \xrightarrow{d} \frac{1}{z_1}, \tag{IA.158}$$

$$\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{z_1}. \tag{IA.159}$$

We now turn our attention to the t-ratios. Let $\sigma_{i,j} \equiv \mathrm{Cov}[z_i, z_j]$ and $\rho_{i,j} = \sigma_{i,j}/(\sigma_i \sigma_j)$. It is easy to show that

$$\sigma_1^2 = \frac{\sigma_g^2}{1_N'[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]1_N} = \frac{\sigma_g^2}{\delta^2}, \tag{IA.160}$$

where δ is the HJ-distance of the misspecified model. Then,

$$\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{\sigma_1 \tilde{z}_1} = -\frac{\delta}{\sigma_g \tilde{z}_1}, \tag{IA.161}$$

where $\tilde{z}_1 = z_1/\sigma_1 \sim \mathcal{N}(0, 1)$. Similarly,

$$\frac{\hat{\lambda}_{1,K-2}}{\sqrt{T}} \xrightarrow{d} \frac{\delta}{\sigma_g \tilde{z}_1}. \tag{IA.162}$$

In addition, using the same proof as in the paper, we can write

$$z_{i+1} = \sigma_{i+1}\left(\rho_{1,i+1}\tilde{z}_1 + \sqrt{1 - \rho_{1,i+1}^2}\, q_{i+1}\right), \tag{IA.163}$$

where $q_{i+1} \sim \mathcal{N}(0, 1)$ and it is independent of $\tilde{z}_1$. Using the fact that

$$\frac{e_t(\hat{\lambda})}{\sqrt{T}} = \frac{R_t(\eta_{1,t} - \eta_{2,t})}{z_1} + O_p(T^{-\frac{1}{2}}), \tag{IA.164}$$

we can show that

$$\frac{\hat{W}_e(\hat{\lambda})}{T} \xrightarrow{d} \frac{\sigma_g^2}{z_1^2} U. \tag{IA.165}$$

Therefore, the squared t-ratio of $\hat{\lambda}_{1,K-1}$ has the following limiting distribution:

$$t^2(\hat{\lambda}_{1,K-1}) = \frac{T\hat{\lambda}_{1,K-1}^2}{\boldsymbol{\iota}_{K,K}'(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}\boldsymbol{\iota}_{K,K}} \xrightarrow{d} \frac{T\hat{d}_K'[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]\hat{d}_K}{\sigma_g^2}. \tag{IA.166}$$

Now, note that

$$\begin{aligned}\hat{d}_K &= \frac{1}{T}\sum_{t=1}^{T} R_t(f_t^0 + \eta_{2,t}) \\ &= \frac{1}{T}\sum_{t=1}^{T} R_t(f_t^0 + \eta_{1,t}) + \frac{1}{T}\sum_{t=1}^{T} R_t(\eta_{2,t} - \eta_{1,t}) \\ &= \hat{d}_{K-1} + \frac{1}{T}\sum_{t=1}^{T} R_t(\eta_{2,t} - \eta_{1,t}). \end{aligned} \tag{IA.167}$$

Let P_U be an $N \times (N-K+1)$ orthonormal matrix with its columns orthogonal to $U^{-\frac{1}{2}}D_1$ and $\hat{P}_U$ be an $N \times (N-K+1)$ orthonormal matrix with its columns orthogonal to $\hat{U}^{-\frac{1}{2}}\hat{D}_1$. Then, it follows that

$$\sqrt{T}\hat{P}_U'\hat{U}^{-\frac{1}{2}}\hat{d}_K = \hat{P}_U'\hat{U}^{-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} R_t(\eta_{2,t} - \eta_{1,t}) \stackrel{d}{\to} \mathcal{N}(0_{N-K+1}, \sigma_g^2 I_{N-K+1}). \tag{IA.168}$$

Since $\sqrt{T}P_U'U^{-\frac{1}{2}}\hat{d}_K$ has the same limiting distribution as $\sqrt{T}\hat{P}_U'\hat{U}^{-\frac{1}{2}}\hat{d}_K$, we have

$$\frac{1}{\sigma_g}\sqrt{T}P_U'U^{-\frac{1}{2}}\hat{d}_K \stackrel{d}{\to} \mathcal{N}(0_{N-K+1}, I_{N-K+1}) \tag{IA.169}$$

and

$$t^2(\hat{\lambda}_{1,K-1}) \stackrel{d}{\to} \chi^2_{N-K+1}. \tag{IA.170}$$

Similarly, we have $t^2(\hat{\lambda}_{1,K-2}) \stackrel{d}{\to} \chi^2_{N-K+1}$.

To derive the limiting distributions of $t(\hat{\lambda}_0)$ and $t(\hat{\lambda}_{1,i})$ for $i = 1, \ldots, K-3$, we use the identity

$$I_N - U^{-\frac{1}{2}}H_1(H_1'U^{-1}H_1)^{-1}H_1'U^{-\frac{1}{2}} = I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}} - hh', \tag{IA.171}$$

where

$$h = \frac{[I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}1_N}{\delta} = \frac{P_U P_U' U^{-\frac{1}{2}}1_N}{\delta} \tag{IA.172}$$

and $h'h = 1$. Note that

$$\begin{aligned}\sqrt{T}h'U^{-\frac{1}{2}}\hat{d}_K/\sigma_g = \sqrt{T}1_N'U^{-\frac{1}{2}}P_U P_U'U^{-\frac{1}{2}}\hat{d}_K/(\sigma_g\delta) &\stackrel{d}{\to} \tilde{z}_1 \sim \mathcal{N}(0,1), &\text{(IA.173)}\\ T\hat{d}_K'[I_N - U^{-\frac{1}{2}}H_1(H_1'U^{-1}H_1)^{-1}H_1'U^{-\frac{1}{2}}]\hat{d}_K/\sigma_g^2 &\stackrel{d}{\to} x \sim \chi^2_{N-K}, &\text{(IA.174)}\end{aligned}$$

and they are independent of each other. Using the formula for the inverse of a partitioned matrix, we can show that

$$\sigma_g^2\boldsymbol{\iota}_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,i} = \sigma_{i+1}^2 - \frac{\sigma_{1,i+1}}{\sigma_1^2} = \sigma_{i+1}^2(1 - \rho_{1,i+1}^2). \tag{IA.175}$$

In addition, we can easily show that for $i = 1, \ldots, K-2$

$$\frac{\sqrt{T}\boldsymbol{\iota}'_{K-1,i}(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K}{\sigma_g[\boldsymbol{\iota}'_{K-1,i}(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,i}]^{\frac{1}{2}}} \stackrel{d}{\to} q_{i+1} \sim \mathcal{N}(0,1). \tag{IA.176}$$

Note that the q_i's are independent of $\tilde{z}_1$ and x.

Consider the upper left $(K-1) \times (K-1)$ submatrix of $(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}/T$, which has the same limit as

$$\frac{\sigma_g^2}{z_1^2}\left[(D_1'U^{-1}D_1)^{-1} + \frac{(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K\hat{d}_K'U^{-1}D_1(D_1'U^{-1}D_1)^{-1}}{\hat{d}_K'U^{-\frac{1}{2}}[I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}\hat{d}_K}\right]. \tag{IA.177}$$

In particular, for $i = 1, \ldots, K-2$, the i-th diagonal element of this matrix has a limiting distribution

$$\frac{\sigma_g^2\boldsymbol{\iota}'_{K-1,i}(D_1'U^{-1}D_1)^{-1}\boldsymbol{\iota}_{K-1,i}}{z_1^2}\left(1 + \frac{q_{i+1}^2}{x + \tilde{z}_1^2}\right) = \frac{\sigma_{i+1}^2(1-\rho_{1,i+1}^2)}{z_1^2}\left(1 + \frac{q_{i+1}^2}{x + \tilde{z}_1^2}\right). \tag{IA.178}$$

Let

$$b_{i+1} = \frac{x + \tilde{z}_1^2}{q_{i+1}^2 + x + \tilde{z}_1^2}. \tag{IA.179}$$

Then, we can write the limiting distribution of $t(\hat{\lambda}_{1,i})$ for $i = 1, \ldots, K-3$, as

$$\begin{aligned} t(\hat{\lambda}_{1,i}) \quad &\stackrel{d}{\to} \quad -\frac{z_{i+2}/z_1}{\sigma_{i+2}\sqrt{1-\rho_{1,i+2}^2}|z_1|b_{i+1}^{-\frac{1}{2}}} \\ &= \quad -\frac{|\tilde{z}_1|}{\tilde{z}_1}\frac{\left(\rho_{1,i+2}\tilde{z}_1 + \sqrt{1-\rho_{1,i+2}^2}q_{i+2}\right)b_{i+1}^{\frac{1}{2}}}{\sqrt{1-\rho_{1,i+2}^2}} \\ &= \quad -\left(\frac{\rho_{1,i+2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,i+2}^2}} + q_{i+2}\right)b_{i+1}^{\frac{1}{2}}. \end{aligned} \tag{IA.180}$$

Similarly,

$$t(\hat{\lambda}_0) \stackrel{d}{\to} -\left(\frac{\rho_{1,2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,2}^2}} + q_2\right)b_1^{\frac{1}{2}}. \tag{IA.181}$$

3 CU-GMM Estimation of the Beta-Pricing Model

Let $\phi = [\gamma_0,\ \gamma_1',\ \beta_1', \ldots, \beta_{K-1}',\ \mu_f',\ \text{vech}(V_f)']'$ denote the vector of parameters of interest. Define the moment conditions

$$g_t(\phi) = \begin{pmatrix} R_t - (1_N\gamma_0 + \beta\gamma_1) - \beta(f_t - \mu_f) \\ [R_t - (1_N\gamma_0 + \beta\gamma_1) - \beta(f_t - \mu_f)] \otimes f_t \\ f_t - \mu_f \\ \text{vech}\left((f_t - \mu_f)(f_t - \mu_f)' - V_f\right) \end{pmatrix} \tag{IA.182}$$

and note that $E[g_t(\phi)] = 0_{(N+1)K+K(K-1)/2-1}$. Let also $\bar{g}(\phi) = T^{-1}\sum_{t=1}^{T} g_t(\phi)$ and

$$\hat{W}_g(\phi) = \frac{1}{T}\sum_{t=1}^{T}(g_t(\phi) - \bar{g}(\phi))(g_t(\phi) - \bar{g}(\phi))'. \tag{IA.183}$$

Then, the CU-GMM estimator of ϕ is defined as

$$\hat{\phi} = \text{argmin}_{\phi}\ \bar{g}(\phi)'\hat{W}_g(\phi)^{-1}\bar{g}(\phi). \tag{IA.184}$$

The problem with implementing this CU-GMM estimator is that the parameter vector ϕ is highly dimensional especially when the number of test assets N is large. Peñaranda and Sentana (2014) show that CU-GMM delivers numerically identical estimates in the beta-pricing and SDF setups.[3] By augmenting $\bar{e}(\lambda)$ in the SDF representation with additional (just-identified) moment conditions for μ_f, V_f, and β, the CU-GMM estimate of the augmented parameter vector $\theta = [\lambda_0,\ \lambda_1',\ \beta_1',\ldots,\beta_{K-1}',\ \mu_f',\ \text{vech}(V_f)']'$ becomes numerically identical to the CU-GMM estimate of ϕ in the beta-pricing model. However, the estimation of θ can be performed in a sequential manner which offers substantial computational advantages. The following theorem presents a general result for this sequential estimation.

THEOREM A.1. *Let $\theta = (\theta_1',\ \theta_2')'$, where θ_1 is $K_1 \times 1$ and θ_2 is $K_2 \times 1$, and*

$$E[g_t(\theta)] = \begin{bmatrix} E[g_{1t}(\theta_1)] \\ E[g_{2t}(\theta)] \end{bmatrix} = \begin{bmatrix} 0_{N_1} \\ 0_{N_2} \end{bmatrix}, \tag{IA.185}$$

where $g_{1t}(\theta_1)$ is $N_1 \times 1$ and $g_{2t}(\theta)$ is $N_2 \times 1$, with $N_1 > K_1$ and $N_2 = K_2$. Define the estimators

$$\tilde{\theta}_1 = \text{argmin}_{\theta_1}\ \bar{g}_1(\theta_1)'\hat{W}_{11}(\theta_1)^{-1}\bar{g}_1(\theta_1), \tag{IA.186}$$

$$\hat{\theta} \equiv \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \text{argmin}_{\theta}\ \bar{g}(\theta)'\hat{W}(\theta)^{-1}\bar{g}(\theta), \tag{IA.187}$$

where $\bar{g}_1(\theta_1) = \frac{1}{T}\sum_{t=1}^{T} g_{1t}(\theta_1)$, $\hat{W}_{11}(\theta_1) = \frac{1}{T}\sum_{t=1}^{T}(g_{1t}(\theta_1) - \bar{g}_1(\theta_1))(g_{1t}(\theta_1) - \bar{g}_1(\theta_1))'$, $\bar{g}(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$, and $\hat{W}(\theta) = \frac{1}{T}\sum_{t=1}^{T}(g_t(\theta) - \bar{g}(\theta))(g_t(\theta) - \bar{g}(\theta))'$. The corresponding tests for over-identifying restrictions are given by

$$\mathcal{J}(\tilde{\theta}_1) = T\bar{g}_1(\tilde{\theta}_1)'\hat{W}_{11}(\tilde{\theta}_1)^{-1}\bar{g}_1(\tilde{\theta}_1), \tag{IA.188}$$

$$\mathcal{J}(\hat{\theta}) = T\bar{g}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\bar{g}(\hat{\theta}). \tag{IA.189}$$

[3] Shanken and Zhou (2007) show that under some particular Kronecker structure for the weighting matrix $\hat{W}_g$, the GMM estimator of the beta-pricing model is numerically identical to the MLE.

Then, $\tilde{\theta}_1 = \hat{\theta}_1$*, and* $\mathcal{J}(\tilde{\theta}_1) = \mathcal{J}(\hat{\theta})$*.*

Proof. Let

$$\tilde{D}_{11}(\theta_1) = \frac{1}{T}\sum_{t=1}^{T} \tilde{w}_t(\theta_1)\frac{\partial g_{1t}(\theta_1)}{\partial \theta_1'}, \tag{IA.190}$$

where

$$\tilde{w}_t(\theta_1) = 1 - \bar{g}_1(\theta_1)'\hat{W}_{11}(\theta_1)^{-1}[g_{1t}(\theta_1) - \bar{g}_1(\theta_1)]. \tag{IA.191}$$

The first-order conditions for the smaller system are given by

$$\tilde{D}_{11}(\tilde{\theta}_1)'\hat{W}_{11}(\tilde{\theta}_1)^{-1}\bar{g}_1(\tilde{\theta}_1) = 0_{N_1}. \tag{IA.192}$$

Similarly, we define

$$\hat{D}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \hat{w}_t(\theta)\frac{\partial g_t(\theta)}{\partial \theta'} \equiv \begin{bmatrix} \hat{D}_{11}(\theta) & 0_{N_1\times N_2} \\ \hat{D}_{21}(\theta) & \hat{D}_{22}(\theta) \end{bmatrix}, \tag{IA.193}$$

where

$$\hat{w}_t(\theta) = 1 - \bar{g}(\theta)'\hat{W}(\theta)^{-1}[g_t(\theta) - \bar{g}(\theta)]. \tag{IA.194}$$

The first-order conditions for the larger system are given by

$$\hat{D}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\bar{g}(\hat{\theta}) = 0_{N_1+N_2}. \tag{IA.195}$$

Let

$$\hat{W}(\theta)^{-1} = \begin{bmatrix} \hat{W}^{11}(\theta) & \hat{W}^{12}(\theta) \\ \hat{W}^{21}(\theta) & \hat{W}^{22}(\theta) \end{bmatrix}. \tag{IA.196}$$

Suppressing the dependence on the parameters in $\hat{D}(\hat{\theta})$ and $\hat{W}(\hat{\theta})$, the first-order conditions for the larger system can be written as

$$\begin{aligned} 0_{N_1+N_2} &= \hat{D}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\bar{g}(\hat{\theta}) \\ &= \begin{bmatrix} (\hat{D}_{11}'\hat{W}^{11} + \hat{D}_{21}'\hat{W}^{21})\bar{g}_1(\hat{\theta}_1) + (\hat{D}_{11}'\hat{W}^{12} + \hat{D}_{21}'\hat{W}^{22})\bar{g}_2(\hat{\theta}) \\ \hat{D}_{22}'\hat{W}^{21}\bar{g}_1(\hat{\theta}_1) + \hat{D}_{22}'\hat{W}^{22}\bar{g}_2(\hat{\theta}) \end{bmatrix}. \end{aligned} \tag{IA.197}$$

When $N_2 = K_2$, $\hat{D}_{22}$ and $\hat{W}^{22}$ are invertible with probability one. Using the second subset of the first-order conditions, we obtain

$$\bar{g}_2(\hat{\theta}) = -(\hat{W}^{22})^{-1}\hat{W}^{21}\bar{g}_1(\hat{\theta}_1). \tag{IA.198}$$

Plugging this equation into the first subset of first-order conditions, we obtain

$$\begin{aligned} 0_{N_1} &= (\hat{D}_{11}'\hat{W}^{11} + \hat{D}_{21}'\hat{W}^{21})\bar{g}_1(\hat{\theta}_1) - (\hat{D}_{11}'\hat{W}^{12} + \hat{D}_{21}'\hat{W}^{22})(\hat{W}^{22})^{-1}\hat{W}^{21}\bar{g}_1(\hat{\theta}_1) \\ &= \hat{D}_{11}(\hat{\theta}_1)'\hat{W}_{11}(\hat{\theta}_1)^{-1}\bar{g}_1(\hat{\theta}_1), \end{aligned} \quad \text{(IA.199)}$$

where the last identity is obtained by using the partitioned matrix inverse formula, which implies that

$$\hat{W}_{11}(\theta_1)^{-1} = \hat{W}^{11}(\theta) - \hat{W}^{12}(\theta)\hat{W}^{22}(\theta)^{-1}\hat{W}^{21}(\theta). \quad \text{(IA.200)}$$

In addition, defining $\bar{g}_2(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\theta)$ and using (IA.198), we have

$$\begin{aligned} \hat{w}_t(\hat{\theta}) &= 1 - \bar{g}'(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\begin{bmatrix} g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta}) - \bar{g}_2(\hat{\theta}) \end{bmatrix} \\ &= 1 - [\bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{21}(\hat{\theta}),\ \ \bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{22}(\hat{\theta})]\begin{bmatrix} g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta}) - \bar{g}_2(\hat{\theta}) \end{bmatrix} \\ &= 1 - \bar{g}_1(\hat{\theta}_1)'\hat{W}_{11}(\hat{\theta}_1)^{-1}[g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1)] \\ &= \tilde{w}_t(\hat{\theta}_1), \end{aligned} \quad \text{(IA.201)}$$

which only depends on $\hat{\theta}_1$. Therefore, we have $\hat{D}_{11}(\hat{\theta}_1) = \tilde{D}_{11}(\hat{\theta}_1)$ and (IA.199) is identical to the first-order conditions for the smaller system. It follows that $\hat{\theta}_1 = \tilde{\theta}_1$. Finally, using (IA.198) and (IA.200), it is possible to show that

$$\begin{aligned} \mathcal{J}(\hat{\theta}) &= T[\bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta})\bar{g}_1(\hat{\theta}_1) + 2\bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta})\bar{g}_2(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{22}(\hat{\theta})\bar{g}_2(\hat{\theta})] \\ &= T\bar{g}_1(\hat{\theta}_1)'\hat{W}_{11}(\hat{\theta}_1)^{-1}\bar{g}_1(\hat{\theta}_1) \\ &= T\bar{g}_1(\tilde{\theta}_1)'\hat{W}_{11}(\tilde{\theta}_1)^{-1}\bar{g}_1(\tilde{\theta}_1) \\ &= \mathcal{J}(\tilde{\theta}_1). \end{aligned} \quad \text{(IA.202)}$$

This completes the proof of Theorem A.1.

Theorem A.1 establishes that for CU-GMM, adding a new set of just-identified moment conditions to the original system does not alter the estimates of the original parameters as well as the test for over-identifying restrictions. This result has implications for speeding up the optimization problem in the CU-GMM estimation. The key is to discard the subset of moment conditions that are exactly identified and only perform the over-identifying restriction test on the remaining smaller set of moment conditions. This will lead to fewer moment conditions and parameters in the system, which is highly desirable when performing numerical optimization. The following lemma demonstrates how to solve for $\hat{\theta}_2$ after $\tilde{\theta}_1$ is obtained from the smaller system.

LEMMA A.1. *Let*

$$r_t(\hat{\theta}) = \bar{g}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}[g_t(\hat{\theta}) - \bar{g}(\hat{\theta})] \tag{IA.203}$$

and

$$r_{1t}(\tilde{\theta}_1) = \bar{g}_1(\tilde{\theta}_1)'\hat{W}_{11}(\tilde{\theta}_1)^{-1}[g_{1t}(\tilde{\theta}_1) - \bar{g}_1(\tilde{\theta}_1)]. \tag{IA.204}$$

The estimate $\hat{\theta}_2$ is given by the solution to

$$\frac{1}{T}\sum_{t=1}^{T} g_{2t}(\tilde{\theta}_1, \hat{\theta}_2)[1 - r_{1t}(\tilde{\theta}_1)] = 0_{K_2} \tag{IA.205}$$

and $r_t(\hat{\theta}) = r_{1t}(\tilde{\theta}_1)$. Furthermore, if g_{2t}, conditional on θ_1, is linear in θ_2, that is,

$$g_{2t}(\theta_1, \theta_2) = h_{1t}(\theta_1) - h_{2t}(\theta_1)\theta_2, \tag{IA.206}$$

where h_{1t} and h_{2t} are functions of the data and θ_1, then

$$\hat{\theta}_2 = \left(\sum_{t=1}^{T} h_{2t}(\tilde{\theta}_1)[1 - r_{1t}(\tilde{\theta}_1)]\right)^{-1} \sum_{t=1}^{T} h_{1t}(\tilde{\theta}_1)[1 - r_{1t}(\tilde{\theta}_1)]. \tag{IA.207}$$

Proof. Using the formula for the inverse of a partitioned matrix, we have $-(\hat{W}^{22})^{-1}\hat{W}^{21} = \hat{W}_{21}\hat{W}_{11}^{-1}$. Plugging this in (IA.198) and noting that $\hat{\theta}_1 = \tilde{\theta}_1$, we obtain

$$\bar{g}_2(\tilde{\theta}_1, \hat{\theta}_2) = \hat{W}_{21}(\tilde{\theta}_1, \hat{\theta}_2)\hat{W}_{11}(\tilde{\theta}_1)^{-1}\bar{g}_1(\tilde{\theta}_1). \tag{IA.208}$$

This is a system of K_2 equations with K_2 unknowns. Using the expression for $r_{1t}(\tilde{\theta}_1)$, we can write (IA.208) as

$$\begin{aligned} \bar{g}_2(\tilde{\theta}_1, \hat{\theta}_2) &= \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\tilde{\theta}_1, \hat{\theta}_2) r_{1t}(\tilde{\theta}_1) \\ \Rightarrow \quad 0_{K_2} &= \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\tilde{\theta}_1, \hat{\theta}_2)[1 - r_{1t}(\tilde{\theta}_1)]. \end{aligned} \tag{IA.209}$$

For the larger system, we have

$$\begin{aligned}
r_t(\hat{\theta}) &= \begin{bmatrix} \bar{g}_1(\hat{\theta}_1) \\ \bar{g}_2(\hat{\theta}) \end{bmatrix}' \begin{bmatrix} \hat{W}^{11}(\hat{\theta}) & \hat{W}^{12}(\hat{\theta}) \\ \hat{W}^{21}(\hat{\theta}) & \hat{W}^{22}(\hat{\theta}) \end{bmatrix} \begin{bmatrix} g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta}) - \bar{g}_2(\hat{\theta}) \end{bmatrix} \\
&= \left[\bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{21}(\hat{\theta}),\ \bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{22}(\hat{\theta})\right] \begin{bmatrix} g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta}) - \bar{g}_2(\hat{\theta}) \end{bmatrix} \\
&= \bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta})[g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1)] - \bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta})(\hat{W}^{22}(\hat{\theta}))^{-1}\hat{W}^{21}(\hat{\theta})[g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1)] \\
&= \bar{g}_1(\hat{\theta}_1)'\hat{W}_{11}^{-1}(\hat{\theta}_1)[g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1)] \\
&= r_{1t}(\tilde{\theta}_1),
\end{aligned} \tag{IA.210}$$

where the third equality follows from (IA.198), the fourth equality follows from the formula for the inverse of a partitioned matrix, and the last equality follows because $\hat{\theta}_1 = \tilde{\theta}_1$. The expression for $\hat{\theta}_2$ can be obtained by plugging $g_{2t}(\theta_1, \theta_2) = h_{1t}(\theta_1) - h_{2t}(\theta_1)\theta_2$ into (IA.209) and solving for $\hat{\theta}_2$. This completes the proof of Lemma A.1.

Lemma A.1 shows that when g_{2t} is linear in θ_2, $\hat{\theta}_2$ has a closed-form solution. When $h_{2t}(\theta_1) = I_{K_2}$, which is the case of the asset-pricing models considered in this paper, we have

$$\hat{\theta}_2 = \frac{\sum_{t=1}^{T} h_{1t}(\tilde{\theta}_1)[1 - r_{1t}(\tilde{\theta}_1)]}{\sum_{t=1}^{T}[1 - r_{1t}(\tilde{\theta}_1)]}. \tag{IA.211}$$

Adding an extra set of just-identified moment conditions proves to be straightforward since $r_t(\hat{\theta}) = r_{1t}(\tilde{\theta}_1)$ and r_t does not need to be recomputed for the larger system.

References

[1] Peñaranda, F., and E. Sentana (2014): "A Unifying Approach to the Empirical Evaluation of Asset Pricing Models," *Review of Economics and Statistics*, forthcoming.

[2] 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.