Too Good to Be True? Fallacies in Evaluating Risk Factor Models

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2017-9
November 2017

Abstract: This paper is concerned with statistical inference and model evaluation in possibly misspecified and unidentified linear asset-pricing models estimated by maximum likelihood and one-step generalized method of moments. Strikingly, when spurious factors (that is, factors that are uncorrelated with 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. Furthermore, factors that are spurious are selected with high probability, while factors that are useful are driven out of the model. Although ignoring potential misspecification and lack of identification can be very problematic for models with macroeconomic factors, empirical specifications with traded factors (e.g., Fama and French, 1993, and Hou, Xue, and Zhang, 2015) do not suffer of the identification problems documented in this study.

JEL classification: G12, C12, C13

Key words: asset pricing, spurious risk factors, unidentified models, model misspecification, continuously updated GMM, maximum likelihood, goodness-of-fit, rank test

---

A part of this paper was circulated under the title "Spurious Inference in Unidentified Asset-Pricing Models." The authors thank Seung Ahn, Alex Horenstein, Lei Jiang, Robert Kimmel, Frank Kleibergen, Francisco Peñaranda, Enrique Sentana, Chu Zhang, seminar participants at Bocconi University, EDHEC, ESSEC, Federal Reserve Bank of New York, Boston University, Hong Kong University of Science and Technology, National University of Singapore, Queen Mary University of London, University of California-San Diego, University of Cantabria, University of Exeter, University of Geneva, University of Hong Kong, University of Lugano, University of Reading, University of Rome Tor Vergata, University of Southampton, Vanderbilt University, 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 China International Conference in Finance, the 2014 NBER-NSF Time-Series Conference, the 2014 Northern Finance Association Conference, the 2014 SoFiE Conference, the 2014 Tsinghua Finance Workshop, the 2015 Brunel Workshop in Empirical Finance, the 2015 Toulouse Financial Econometrics Conference, and the 2015 York Conference on Macroeconomic, Financial, and International Linkages for helpful discussions and comments. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to corresponding author: Nikolay Gospodinov, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309, 404-498-7892, nikolay.gospodinov@atl.frb.org; Raymond Kan, Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario M5S 3E6, Canada, kan@chass.utoronto.ca; or Cesare Robotti, Department of Finance, Terry College of Business, University of Georgia, B328 Amos Hall, 620 South Lumpkin Street, Athens, GA 30602, robotti@uga.edu.

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, 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, 2016, and Feng, Giglio, and Xiu, 2017) 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, for example, Shanken and Zhou (2007), Almeida and Garcia (2012, 2017), Peñaranda and Sentana (2015), Manresa, Peñaranda, and Sentana (2017), Barillas and Shanken (2017a, 2017b), and Ghosh, Julliard, and Taylor (2017) for some recent results on the ML and CU-GMM estimators for asset-pricing models.

that many of these macroeconomic factors may be irrelevant for pricing and explaining the cross-sectional variation in stock expected returns. Importantly, the inclusion of spurious factors (that is, factors that are uncorrelated with 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 (see, for instance, Gospodinov, Kan, and Robotti, 2014).

Under standard regularity conditions (that include global and local identification as well as correct model specification), the ML and CU-GMM estimators considered here, which are invariant to data scaling, reparameterizations and normalizations, etc. (Hall, 2005), are asymptotically well-behaved and efficient. However, we show in this paper that in the presence of spurious factors, the tests and goodness-of-fit measures 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 start with some numerical evidence on the widely studied static capital asset-pricing model (CAPM) with the market excess 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 test asset returns are the monthly gross returns on the popular value-weighted 25 Fama-French size and book-to-market ranked portfolios from January 1967 until December 2012.[2]

[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).

Table 1 about here

The first column of Table 1 reports some conventional statistics for evaluating the performance of the model over this sample period both in the SDF (using CU-GMM, Panel A) and beta-pricing (using ML, Panel B) frameworks. The statistics include 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 of statistical significance constructed using standard errors that assume correct model specification, and the pseudo-R^2s computed as the squared correlation between the fitted expected returns and average returns. In line with the results reported elsewhere in the literature, the market factor appears to be priced with a statistically significant risk premium. Also, consistent with the existing studies, the CAPM model is rejected by the data. This requires the use of misspecification-robust standard errors in constructing the t-statistics (see Gospodinov, Kan, and Robotti, 2017a). Finally, the pseudo-R^2 points to some, but not particularly strong, explanatory power.

We now add a factor, which we call the "*sp*" factor, to the CAPM model and, for the time being, we do not reveal its identity and construction method. It is important to stress that the test assets, the sample period, and the market factor remain unchanged: the only change is the addition of the "*sp*" factor to the model. The results from this specification of the model are presented in the third column (CAPM + "*sp*" factor) of Panels A and B in Table 1. Interestingly, the specification tests now suggest that the model is correctly specified. Even more strikingly, the pseudo-R^2 jumps from 22.77% to 99.28% in the SDF model estimated by CU-GMM and from 14.47% to 99.99% in the beta-pricing model estimated by ML. The *sp* factor is highly statistically significant while the market factor becomes insignificant. An applied researcher who is interested in selecting parsimonious statistical models may want to remove the market factor and re-estimate the model with the "*sp*" factor only.

The results from this third specification are reported in the last column of Panels A and B in Table 1. The results are striking. First, this one-factor model exhibits an almost perfect fit ($R^2 \approx 1$). Based on the specification tests, the model appears to be correctly specified. Finally, the "*sp*" factor is highly statistically significant and is deemed to be priced. Given this exceptional performance of the model, we now ask "What is this "*sp*" factor?" It turns out that this factor is generated as a standard normal random variable which is independent of returns! The results of this numerical exercise are completely spurious since the "*sp*" factor does not contribute to pricing by construction. In summary, an arbitrarily bad model with one or more spurious factors is concluded to be a correctly specified model with a spectacular fit and priced factors. Even worse, the priced factors that are highly correlated with the test asset returns are driven out (become statistically insignificant) when a spurious factor is included in the model.[3]

It turns out that this type of behavior is not specific to artificial models and also arises in well-known empirical asset-pricing models. To substantiate this claim, we consider three other popular asset-pricing models. The first 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 portfolio returns or return spreads and exhibit a relatively high correlation with the 25 Fama-French portfolio returns. The other two models are models with traded and non-traded 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 cg and cay ($cg \cdot cay$).

Table 2 about here

Table 2 reports results for these three models. For completeness, we also present results for the CAPM model. In addition to the statistics in Table 1, we include rank tests to determine whether the asset-pricing models are properly identified,[4] and two widely-used specification tests

[3] It should be noted that the results in Table 1 are based on one draw from the standard normal distribution. However, our conclusions are qualitatively similar when the analysis in the table is based on the average of 100,000 replications. Starting from the CAPM + "*sp*" factor specification, $\mathcal{J} = 22.36$ (p-value=0.4813) and $\mathcal{S} = 22.50$ (p-value=0.4806). The t-statistics for vw are 0.60 (p-value=0.3916) and -0.63 (p-value=0.4096) for CU-GMM and ML, respectively. As for the spurious factor "*sp*", the absolute values of the t-ratios are 4.75 (p-value=0.0001) and 4.76 (p-value=0.0001) for CU-GMM and ML, respectively. Finally, the average R^2s are 0.9769 for CU-GMM and 0.9946 for ML. Turning to the "*sp*" factor specification, $\mathcal{J} = 23.62$ (p-value=0.4715) and $\mathcal{S} = 23.69$ (p-value=0.4738). As for the spurious factor "*sp*", the absolute values of the t-ratios are 4.89 (p-value=0.0000) and 4.90 (p-value=0.0000) for CU-GMM and ML, respectively. Finally, the average R^2s are 0.9774 for CU-GMM and 0.9948 for ML.

[4] In the SDF framework, we use the rank test of Cragg and Donald (1997) to assess whether the second moment

based on non-invariant estimators (the HJ-distance test of Hansen and Jagannathan, 1997, and the generalized least squares (GLS) cross-sectional regression (CSR) 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 confirm the evidence from the models with artificial data above. The models that contain factors that are only weakly correlated with the test asset returns (C-LAB and CC-CAY), as reflected in the non-rejection of the null hypothesis of a reduced rank in Table 2, 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[5] and one could proceed with constructing significance tests based on standard errors derived under correct model specification. These t-tests indicate that the proposed non-traded factors (default premium in C-LAB and consumption growth and the *cay* 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 22.77% and 78.10% for CU-GMM, and 14.47% and 73.64% for ML, respectively.

For comparison, Figures 3 and 4 plot the average realized returns versus fitted returns based on the non-invariant (HJ-distance and GLS CSR, respectively) estimators for each model.

Figures 3 and 4 about here

In sharp contrast with 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)

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 these Cragg and Donald (1997) tests are provided in the empirical section of the paper.

[5] In a related paper, Gospodinov, Kan, and Robotti (2017b) show that the over-identifying restriction tests lack power under the alternative of misspecified models when spurious factors are present. More specifically, under fairly weak assumptions, they show that the specification tests have power equal to (or below) their size in reduced-rank asset-pricing models. This paper complements and extends the analysis in Gospodinov, Kan, and Robotti (2017b) by establishing the highly non-standard limiting behaviour of significance tests and cross-sectional R^2.

no longer exhibit a perfect fit. As a result, these non-invariant tests appear to be more robust to lack of identification and can detect model misspecification with a higher probability than their invariant counterparts.

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 2, 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 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 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 1 (for more details on the simulation design, see Section 3 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 is 98.3%, while the mean pseudo-R^2 is 23.5%. In sharp contrast, when a spurious 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 9.7% and the mean pseudo-R^2 is 99%. Strikingly, the rejection rate of the t-test for the spurious 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 uncorrelated with 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, 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 spurious factors tends to dominate and forces the goodness-of-fit statistic to approach one.

Some 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 (2016) 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 (2015) 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 studies the limiting behavior of the parameter estimates, t-statistics, and goodness-of-fit measures in the beta-pricing and SDF setups. Section 3 reports Monte Carlo simulation results. Section 4 presents our empirical findings. Section 5 summarizes our main conclusions and provides some practical recommendations. All proofs are relegated to the Online Appendix.

2 Stochastic Discount Factor and Beta-Pricing Model Representations

We first introduce the SDF and beta representations of an asset-pricing model. Let

$$y_t(\lambda) = x_t'\lambda \tag{1}$$

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.[6] 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,

$$E[e_t(\lambda^*)] = D\lambda^* - 1_N = 0_N, \tag{2}$$

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

$$E[Y_t] \equiv \begin{bmatrix} \mu_f \\ \mu_R \end{bmatrix} \tag{3}$$

and a positive-definite matrix

$$\mathrm{Var}[Y_t] \equiv V = \begin{bmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{bmatrix}, \tag{4}$$

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

$$\mu_R = 1_N\gamma_0^* + \beta\gamma_1^*, \tag{5}$$

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

$$\alpha = \mu_R - \beta\mu_f, \tag{6}$$

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

[6] 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 adapted to handle the case of excess returns, that is, the $q = 0_N$ case.

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

$$\gamma_0 = \frac{1}{\lambda_0 + \mu_f' \lambda_1}, \tag{7}$$

$$\gamma_1 = -\frac{V_f \lambda_1}{\lambda_0 + \mu_f' \lambda_1}. \tag{8}$$

The main statistics of interest in evaluating asset-pricing models are the t-tests for statistical significance of the λ_1 and γ_1 estimates,[7] 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).

2.1 Maximum Likelihood

We start with the more restrictive ML estimation of the beta-pricing model that imposes the joint normality assumption on Y_t. Combining equations (5) and (6), we arrive at the restriction

$$\alpha = 1_N \gamma_0^* + \beta(\gamma_1^* - \mu_f). \tag{9}$$

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

$$\hat{\gamma}^{ML} = \underset{\gamma}{\operatorname{argmin}} \frac{(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))'\hat{\Sigma}^{-1}(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))}{1 + \gamma_1'\hat{V}_f^{-1}\gamma_1}, \tag{10}$$

[7] 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.

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

$$\mathcal{S} = T \min_{\gamma} \frac{(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))'\hat{\Sigma}^{-1}(\hat{\alpha} - 1_N\gamma_0 - \hat{\beta}(\gamma_1 - \hat{\mu}_f))}{1 + \gamma_1'\hat{V}_f^{-1}\gamma_1}, \tag{11}$$

and is asymptotically distributed as $\mathcal{S} \stackrel{d}{\to} \chi^2_{N-K}$ under the null $H_0 : \alpha = 1_N\gamma_0 + \beta(\gamma_1 - \mu_f)$.[8]

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

$$\min_{v} \frac{v'\hat{G}'\hat{\Sigma}^{-1}\hat{G}v}{v'A(X'X/T)^{-1}A'v}, \tag{12}$$

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[9]

$$\hat{\Omega} = (\hat{G}'\hat{\Sigma}^{-1}\hat{G})^{-1}[A(X'X/T)^{-1}A']. \tag{13}$$

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

$$\hat{\gamma}_0^{ML} = -\frac{\hat{v}_1}{\hat{v}_2}, \tag{14}$$

$$\hat{\gamma}_{1,i}^{ML} = \hat{\mu}_{f,i} - \frac{\hat{v}_{i+2}}{\hat{v}_2}, \quad i = 1, \ldots, K-1. \tag{15}$$

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

$$\sqrt{T}\begin{bmatrix} \hat{\gamma}_0^{ML} - \gamma_0^* \\ \hat{\gamma}_1^{ML} - \gamma_1^* \end{bmatrix} \stackrel{d}{\to} \mathcal{N}\left(0_K, (1 + \gamma_1^{*\prime}V_f^{-1}\gamma_1^*)(B_1'\Sigma^{-1}B_1)^{-1} + \begin{bmatrix} 0 & 0_{K-1}' \\ 0_{K-1} & V_f \end{bmatrix}\right), \tag{16}$$

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

$$t(\hat{\gamma}_0^{ML}) = \frac{\sqrt{T}\hat{\gamma}_0^{ML}}{s(\hat{\gamma}_0^{ML})}, \tag{17}$$

$$t(\hat{\gamma}_{1,i}^{ML}) = \frac{\sqrt{T}\hat{\gamma}_{1,i}^{ML}}{s(\hat{\gamma}_{1,i}^{ML})}, \tag{18}$$

[8] 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 $\mathcal{LR}$), the *iid* normality assumption 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.

[9] 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.

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

$$(1 + \hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML})(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1} + \begin{bmatrix} 0 & 0_{K-1}' \\ 0_{K-1} & \hat{V}_f \end{bmatrix}, \tag{19}$$

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

$$\hat{\beta}^{ML} = \hat{\beta} + \frac{[\hat{\alpha} - 1_N\hat{\gamma}_0^{ML} - \hat{\beta}(\hat{\gamma}_1^{ML} - \hat{\mu}_f)]\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}}{1 + \hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}} \tag{20}$$

and

$$\hat{\mu}_R^{ML} = 1_N\hat{\gamma}_0^{ML} + \hat{\beta}^{ML}\hat{\gamma}_1^{ML}. \tag{21}$$

Since the empirical evidence strongly suggests that linear asset-pricing models are misspecified (as emphasized in our empirical application and many papers in the literature), in the following analysis we present results only for the misspecified model case.[10] The following theorem (and Auxiliary Lemma 1 in the Online Appendix) characterizes the limiting behavior of the ML estimates $\hat{\gamma}^{ML}$, the t-statistics $t(\hat{\gamma}_0^{ML})$ and $t(\hat{\gamma}_{1,i}^{ML})$ $(i = 1, \ldots, K-1)$, and the pseudo-R^2 statistic $R^2_{ML} = \text{Corr}(\hat{\mu}_R^{ML}, \hat{\mu}_R)^2$ in misspecified models that contain a spurious factor.

Without loss of generality, we assume that the spurious 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.[11] Let $\bar{Z}_i$, $i = 0, \ldots, K-2$, denote a bounded random variable defined in the Online Appendix. Then, we have the following result.

Theorem 1. *Assume that Y_t is iid normally distributed. Suppose that the model is misspecified and it contains a spurious factor (that is,* $\text{rank}(B) = K-1$*). Then, as $T \to \infty$, we have*

(a) *(i) $t(\hat{\gamma}_0^{ML}) \xrightarrow{d} \bar{Z}_0$; (ii) $t(\hat{\gamma}_{1,i}^{ML}) \xrightarrow{d} \bar{Z}_i$ for $i = 1, \ldots, K-2$; and (iii) $t^2(\hat{\gamma}_{1,K-1}^{ML}) \xrightarrow{d} \chi^2_{N-K+1}$;*

(b) $R^2_{ML} \xrightarrow{p} 1$.

Proof. See the Online Appendix.

[10]The analytical and simulation results for the correctly specified model case are available from the authors upon request.

[11]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.

Theorem 1 establishes the limiting behavior of the t-tests, and pseudo-R^2 statistic, R^2_{ML}, in misspecified models with identification failure.[12] The Online Appendix also shows that when a spurious factor is present, the estimates on the useful factors are inconsistent and converge to ratios of normal random variables while the estimate for the spurious factor ($\hat{\gamma}^{ML}_{1,K-1}$) diverges at rate root-$T$ and the standardized estimator converges to the reciprocal of a normal random variable.[13] The t-tests for the useful factors converge to bounded random variables and, hence, are inconsistent. In fact, as our simulations illustrate, the tests $t(\hat{\gamma}^{ML}_{1,i})$ for $i = 1, \ldots, K-2$ tend to exhibit power that is close to their size. In contrast, the t-test for the spurious 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 (b) of Theorem 1 shows that the pseudo-R^2 of a misspecified model that contains a spurious factor approaches one. This leads to completely spurious inference as the spurious 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.

Figure 5 about here

To visualize the limiting behavior of the t-statistics in part(a), Figure 5 plots the limiting rejection rates of $t(\hat{\gamma}^{ML}_0)$, $t(\hat{\gamma}^{ML}_{1,i})$, and $t^2(\hat{\gamma}^{ML}_{1,K-1})$ as functions of $N-K$ when one uses the standard normal critical values. The sample quantities that enter the computation for the t-statistics for the constant term and the useful factor are calibrated to the CAPM model. Figure 5 confirms that both $t(\hat{\gamma}^{ML}_0)$ and $t(\hat{\gamma}^{ML}_{1,i})$ are inconsistent as their power does not go one. The over-rejection ($\hat{\gamma}^{ML}_{1,K-1}$) increases with $N-K$ and is effectively one when $N-K \geq 15$.

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 from the authors upon request), 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

[12]The t-statistics are constructed using a consistent estimator of the asymptotic variance for correctly specified models. For the form of the asymptotic variance of ML under misspecified models, see Gospodinov, Kan, and Robotti (2017a).

[13]Kan and Zhang (1999a) and Kleibergen (2009) also show that the estimate for the spurious 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 (2014) for models estimated via non-invariant (optimal and suboptimal) GMM.

noisy factors being the same as the asymptotic distribution for the spurious factor in part (a) of Theorem 1.

2.2 Continuously-Updated GMM

We now consider the more general GMM estimation of SDF and beta-pricing models. We assume that Y_t is a jointly stationary and ergodic process with a finite fourth moment, and $\mathrm{vec}(D_t - D)$ is a martingale difference sequence. The CU-GMM estimator of the SDF parameters λ^* is defined as (see Hansen, Heaton, and Yaron, 1996)

$$\hat{\lambda} = \underset{\lambda}{\mathrm{argmin}}\ \bar{e}(\lambda)'\hat{W}_e(\lambda)^{-1}\bar{e}(\lambda), \tag{22}$$

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

$$\hat{W}_e(\lambda) = \frac{1}{T}\sum_{t=1}^{T}(e_t(\lambda) - \bar{e}(\lambda))(e_t(\lambda) - \bar{e}(\lambda))'. \tag{23}$$

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

$$\mathcal{J} = T \min_{\lambda} \bar{e}(\lambda)'\hat{W}_e(\lambda)^{-1}\bar{e}(\lambda), \tag{24}$$

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

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

$$\sqrt{T}(\hat{\lambda} - \lambda^*) \xrightarrow{d} \mathcal{N}\left(0_K, (D'W_e(\lambda^*)^{-1}D)^{-1}\right), \tag{25}$$

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

$$t(\hat{\lambda}_0) = \frac{\sqrt{T}\hat{\lambda}_0}{s(\hat{\lambda}_0)}, \tag{26}$$

$$t(\hat{\lambda}_{1,i}) = \frac{\sqrt{T}\hat{\lambda}_{1,i}}{s(\hat{\lambda}_{1,i})}, \tag{27}$$

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 $(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}$ and $\hat{D}$ is a consistent estimate of D. In the Online Appendix, we show that the CU-GMM estimates of the SDF parameters λ^* can be used to obtain (in a computationally very efficient way) the CU-GMM estimates of γ^*, β, μ_f, and V_f. More specifically, let

$$w_t(\hat{\lambda}) = \frac{1 - (e_t(\hat{\lambda}) - \bar{e}(\hat{\lambda}))'\hat{W}_e(\hat{\lambda})^{-1}\bar{e}(\hat{\lambda})}{T}. \tag{28}$$

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

$$\hat{\mu}_f^{CU} = \sum_{t=1}^{T} w_t(\hat{\lambda}) f_t, \tag{29}$$

$$\hat{V}_f^{CU} = \sum_{t=1}^{T} w_t(\hat{\lambda}) f_t (f_t - \hat{\mu}_f^{CU})', \tag{30}$$

and

$$\hat{\beta}^{CU} = \sum_{t=1}^{T} w_t(\hat{\lambda}) R_t (f_t - \hat{\mu}_f^{CU})'(\hat{V}_f^{CU})^{-1}. \tag{31}$$

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

$$\hat{\gamma}_0^{CU} = \frac{1}{\hat{\lambda}_0 + \hat{\mu}_f^{CU\prime}\hat{\lambda}_1}, \tag{32}$$

$$\hat{\gamma}_1^{CU} = -\frac{\hat{V}_f^{CU}\hat{\lambda}_1}{\hat{\lambda}_0 + \hat{\mu}_f^{CU\prime}\hat{\lambda}_1}. \tag{33}$$

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, $R^2_{CU} = \mathrm{Corr}(\hat{\mu}_R^{CU}, \hat{\mu}_R)^2$, for CU-GMM.

Theorem 2 below establishes the limiting behavior of the t-tests of statistical significance (under $H_0 : \lambda_i = 0$) in misspecified models with identification failure. Note that the t-statistics are constructed using a consistent estimator of the asymptotic variance for correctly specified models, $(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}/T$.[14] Let $\tilde{Z}_i$, $i = 0, \ldots, K-2$, denote a bounded random variable defined in the Online Appendix.

Theorem 2. *Suppose that the model is misspecified and it contains a spurious factor (that is,* $\mathrm{rank}(D) = K - 1$*). In addition, assume that* Y_t *is a jointly stationary and ergodic process with a finite fourth moment,* $\{\mathrm{vec}(D_t - D) : t \geq 1\}$ *is a martingale difference sequence, and* $E[R_t R_t' | f_{t,K-1}] = E[R_t R_t'] = U$*. Then, as* $T \to \infty$*, we have*

[14]For the form of the asymptotic variance of CU-GMM under misspecified models, see Gospodinov, Kan, and Robotti (2017a).

(a) *(i)* $t(\hat{\lambda}_0) \xrightarrow{d} \tilde{Z}_0$ *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)* $t(\hat{\lambda}_{1,i}) \xrightarrow{d} \tilde{Z}_i$ *for* $i = 1, \ldots, K-2$*; and (iii)* $t^2(\hat{\lambda}_{1,K-1}) \xrightarrow{d} \chi^2_{N-K+1}$*;*

(b) $R^2_{CU} \xrightarrow{p} 1$.

Proof. See the Online Appendix.

Auxiliary Lemma 2 in the Online Appendix shows that when a spurious factor is present, the estimates on the useful factors ($\hat{\lambda}_{1,i}$ $(i = 1, \ldots, K-2)$) converge to ratios of normal random variables. The estimate for the spurious factor ($\hat{\lambda}_{1,K-1}$) diverges at rate root-T, and $\hat{\lambda}_{1,K-1}/\sqrt{T}$ converges to the reciprocal of a normal random variable.[15] Based on Theorem 2, the t-tests of statistical significance for the useful factors converge to bounded random variables. As our simulations illustrate, the tests $t(\hat{\lambda}_{1,i})$ for $i = 1, \ldots, K-2$ tend to exhibit power that is close to their size. In contrast, the t-test for the spurious factor will overreject substantially (with the probability of rejection rapidly approaching one as N increases) when $\mathcal{N}(0,1)$ critical values are used. Another interesting feature is that the asymptotic behavior of $\hat{\lambda}_0$ and $t(\hat{\lambda}_0)$ changes in a discontinuous fashion depending on whether the population mean of the spurious factor is zero or not. In the practically relevant case of a nonzero mean, $\hat{\lambda}_0$ and $t(\hat{\lambda}_0)$ inherit the limiting properties of $\hat{\lambda}_{1,K-1}$ and $t(\hat{\lambda}_{1,K-1})$ for the spurious factor. Similarly to the beta-pricing setup, part (b) of Theorem 2 shows that the pseudo-R^2 measure converges to one when one or more factors are spurious.

Figure 6 about here

The top graph in Figure 6 plots the limiting rejection rates of the t-statistics for a misspecified model with a spurious factor. When the model is correctly specified, the limiting distribution for the t-statistics for the useful factors is nonstandard but, unlike the misspecified model case, useful factors that are priced are maintained in the model with probability approaching one. Although less pronounced than in the misspecified model case, using $\mathcal{N}(0,1)$ critical values will lead to substantial overrejections of $H_0 : \lambda_{1,K-1} = 0$ for the spurious factor. This is revealed by the bottom graph in Figure 6.

[15] Kan and Zhang (1999a) and Kleibergen (2009) also show that the estimate for the spurious 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 (2014) for models estimated via non-invariant (optimal and suboptimal) GMM.

$\boxed{\text{Figure 7 about here}}$

The reason for the overrejection for the parameter on the spurious factor is clearly illustrated in Figure 7 which plots the limiting probability density functions of $t(\hat{\lambda}_{1,K-1})$ 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 $t(\hat{\lambda}_{1,K-1})$ under correctly specified models (which arises from the model's underidentification), using $\mathcal{N}(0,1)$ critical values will lead to an overrejection of the hypothesis that the spurious factor is not priced. This overrejection 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 arises in correctly specified models as well as in misspecified models, although the inference problems are more pronounced in the latter case.

3 Simulation Results

In this section, we undertake a Monte Carlo simulation experiment to study the empirical rejection rates of the 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 spurious factor, and (iii) a model with a constant term, a useful, and a spurious 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 1967:1–2012:12 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 spurious 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 $T = 200$, 600, and 1000, and all results are based on 100,000 Monte Carlo replications. We also report the limiting rejection probabilities (denoted by $T = \infty$) for the t-tests based on our asymptotic results in Section 2.

Table 3 about here

A 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 3. Again, as our theoretical analysis suggests, the empirical distribution of the pseudo-R^2 in models with a spurious 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 4 about here

Table 4 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 spurious 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 4 reveals that for models with a spurious factor, the t-tests will give rise to spurious results, suggesting that these completely irrelevant factors are priced. Moreover, the spurious 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 spurious factor and drop the useful factor from the model (see Panel C of Table 4).

The lack of power of the specification tests, the spuriously high pseudo-R^2 values, and the perils of relying on the traditional t-tests of parameter significance in unidentified models suggest that the decision regarding the model specification should be augmented with additional diagnostics. One approach to restoring the validity of the standard inference is based on the following model reduction procedure.[16] We present the details of this procedure for matrix D in the SDF framework and the corresponding procedure for the beta-pricing framework is obtained by replacing D with the matrix B. First, the matrix D should be subjected to a rank test. If the null hypothesis of a reduced rank is rejected, the researcher can proceed with the standard specification test. If the null of a reduced rank is not rejected, the researcher needs to estimate consistently the reduced rank L of D. The estimation of the rank of D can be performed using the modified Bayesian information criterion (MBIC) of Ahn, Horenstein, and Wang (2017) by choosing the value of L (for $L = 1, \ldots, K-1$) that minimizes

$$MBIC(L) = \mathcal{CD}_D(L) - T^{0.2}(N-L)(K-L), \tag{34}$$

where $\mathcal{CD}_D(L)$ is the Cragg and Donald (1997) test of the null that the rank of D is equal to L.[17] It is worth pointing out that this step of the model reduction procedure can be implemented using any available rank test.[18] If the rank is estimated to be $1 \leq k \leq K-1$, construct $N \times k$ matrices $\tilde{D}$ by selecting all possible combinations of $k-1$ risk factors, $\tilde{f}$,[19] and perform a rank test on each $\tilde{D}$. Then, choose the $\tilde{f}$ that gives rise to the largest rejection of the reduced-rank hypothesis.

Table 5 reports the probabilities of retaining factors in the proposed model reduction procedure in the SDF framework. To minimize the probability of rejecting the null of a reduced rank when the true rank of D is deficient, we fix the significance level of the rank test on D to be 1% even though we allow for different significance levels for the subsequent specification test. In addition, we denote by P_A, P_B, and P_C the marginal probability of retaining the useful factors, the marginal probability of eliminating the spurious factors, and the joint probability of retaining the useful factors and eliminating the spurious factors, respectively. The reported probabilities are numerically identical for the correctly specified and misspecified versions of each model.

Table 5 about here

In order to make the simulation design more challenging for the model reduction procedure,

[16] A similar procedure is proposed in Burnside (2016) where the selected factors are linear combinations of the original factors. By contrast, our procedure selects individual factors and preserves their economic interpretation.

[17] In order to minimize the probability of rejecting the null of a reduced rank when the true rank of D is deficient and to guard the procedure against the selection of nearly spurious factors (see also Wright, 2003), we fix the level of the rank test on D to be the same and small (say 1%) for all levels of the subsequent specification test.

[18] In our simulations in Table 5, we employ the rank test of Kleibergen and Paap (2006) instead of the $\mathcal{CD}$ test since the former does not require numerical optimization.

[19] In our setup of gross returns on the test assets, the intercept is always included in the model.

we also consider, in addition to the three models described above, a model with a constant term, three useful, and two spurious factors. The results for all models suggest that our model selection procedure is very effective in retaining the useful factors and eliminating the spurious factors from the analysis. For sample sizes $T \geq 600$, the most challenging scenario of a model with three useful and two spurious factors retains (removes) the useful (spurious) factors with probability one.

4 Empirical Illustration

We evaluate the performance of several prominent asset-pricing models with traded and non-traded factors in light of our analytical and simulation results in Sections 2 and 3. First, we describe the data used in the empirical analysis and outline the different specifications of the asset-pricing models considered. Next, we present our results.

4.1 Data and Asset-Pricing Models

The return data are from Kenneth French's website and consist of the monthly value-weighted gross returns on the (i) 25 Fama-French size and book-to-market ranked portfolios, (ii) 25 Fama-French size and momentum ranked portfolios, and (iii) 32 Fama-French size, operating profitability, and investment ranked portfolios. To conserve space, we briefly summarize the results for other sets of test portfolio returns at the end of the section. The data are from January 1967 to December 2012 (552 monthly observations). The beginning date of our sample period is dictated by profitability and investment data availability.[20] We analyze six asset-pricing models starting with the conditional labor model (C-LAB) of Jagannathan and Wang (1996). This model incorporates measures of the return on human capital as well as the change in financial wealth and allows the conditional moments to vary with a state variable, *prem*, the lagged yield spread between Baa- and Aaa-rated corporate bonds from the Board of Governors of the Federal Reserve System. The SDF specification for this model is

$$y_t^{C-LAB} = \lambda_0 + \lambda_{vw}\, vw_t + \lambda_{labor}\, labor_t + \lambda_{prem}\, prem_t, \tag{35}$$

where vw is the excess return (in excess of the 1-month T-bill rate from Ibbotson Associates) on the value-weighted stock market index (NYSE-AMEX-NASDAQ) from Kenneth French's website,

[20] We thank Lu Zhang for sharing his data with us.

labor is the growth rate in per capita labor income, L, defined as the difference between total personal income and dividend payments, divided by the total population (from the Bureau of Economic Analysis). Following Jagannathan and Wang (1996), we use a 2-month moving average to construct the growth rate $labor_t = (L_{t-1}+L_{t-2})/(L_{t-2}+L_{t-3})-1$, for the purpose of minimizing the influence of measurement error. The corresponding cross-sectional specification is $\mu_R = 1_N\gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{labor}\gamma_{labor} + \beta_{prem}\gamma_{prem}$.

Our second model (CC-CAY) is a conditional version of the consumption CAPM due to Lettau and Ludvigson (2001) with

$$y_t^{CC-CAY} = \lambda_0 + \lambda_{cg}\, cg_t + \lambda_{cay}\, cay_{t-1} + \lambda_{cg\cdot cay}\, cg_t \cdot cay_{t-1}, \tag{36}$$

where cg is the growth rate in real per capita nondurable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis, and cay, the conditioning variable, is a consumption-aggregate wealth ratio.[21] This specification is obtained by scaling the constant term and the cg factor of a linearized consumption CAPM by a constant and cay. The beta-pricing specification is $\mu_R = 1_N\gamma_0 + \beta_{cg}\gamma_{cg} + \beta_{cay}\gamma_{cay} + \beta_{cg\cdot cay}\gamma_{cg\cdot cay}$.

The third model (ICAPM) is an empirical implementation of Merton's (1973) intertemporal extension of the CAPM based on Campbell (1996), who argues that innovations in state variables that forecast future investment opportunities should serve as the factors. The candidate SDF for the five-factor specification proposed by Petkova (2006) is

$$y_t^{ICAPM} = \lambda_0 + \lambda_{vw}\, vw_t + \lambda_{term}\, term_t + \lambda_{def}\, def_t + \lambda_{div}\, div_t + \lambda_{rf}\, rf_t, \tag{37}$$

where $term$ is the difference between the yields of 10- and 1-year government bonds (from the Board of Governors of the Federal Reserve System), def is the difference between the yields of long-term corporate Baa bonds and long-term government bonds (from Ibbotson Associates), div is the dividend yield on the Center for Research in Security Prices (CRSP) value-weighted stock market portfolio, and rf is the 1-month T-bill yield (from CRSP, Fama Risk Free Rates). The actual factors for $term$, def, div, and rf are their innovations from a VAR(1) system of seven state variables that also includes vw, smb, and hml (the market, size, and value factors of the three-factor model of Fama and French, 1993). As for the beta-pricing specification, we have $\mu_R = 1_N\gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{term}\gamma_{term} + \beta_{def}\gamma_{def} + \beta_{div}\gamma_{div} + \beta_{rf}\gamma_{rf}$.

[21] Following Vissing-Jørgensen and Attanasio (2003), we linearly interpolate the quarterly values of cay to permit analysis at the monthly frequency.

We complete our list of models with traded and non-traded factors by considering a specification (D-CCAPM), due to Yogo (2006), which highlights the cyclical role of durable consumption in asset pricing. The candidate SDF is given by

$$y_t^{D-CCAPM} = \lambda_0 + \lambda_{vw}\, vw_t + \lambda_{cg}\, cg_t + \lambda_{cgdur}\, cgdur_t, \tag{38}$$

where *cgdur* is the growth rate in real per capita durable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis. In beta form, we have $\mu_R = 1_N\gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{cg}\gamma_{cg} + \beta_{cgdur}\gamma_{cgdur}$.

Our fifth model (FF3), due to Fama and French (1993),

$$y_t^{FF3} = \lambda_0 + \lambda_{vw}\, vw_t + \lambda_{smb}\, smb_t + \lambda_{hml}\, hml_t, \tag{39}$$

includes two empirically motivated factors, *smb*, the return difference between portfolios of stocks with small and large market capitalizations, and *hml*, the return difference between portfolios of stocks with high and low book-to-market ratios (from Kenneth French's website). The familiar beta representation of the model is $\mu_R = 1_N\gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{smb}\gamma_{smb} + \beta_{hml}\gamma_{hml}$.

Finally, we consider a newly proposed empirical specification (HXZ), due to Hou, Xue, and Zhang (2015), which is built on the neoclassical *q*-theory of investment. The candidate SDF for this model is

$$y_t^{HXZ} = \lambda_0 + \lambda_{vw}\, vw_t + \lambda_{me}\, me_t + \lambda_{roe}\, roe_t + \lambda_{ia}\, ia_t, \tag{40}$$

where *me* is the difference between the return on a portfolio of small size stocks and the return on a portfolio of big size stocks, *roe* is the difference between the return on a portfolio of high profitability stocks and the return on a portfolio of low profitability stocks, and *ia* is the difference between the return on a portfolio of low investment stocks and the return on a portfolio of high investment stocks. This four-factor model has been shown to successfully explain many asset-pricing anomalies.[22] The corresponding beta representation of the model is $\mu_R = 1_N\gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{me}\gamma_{me} + \beta_{roe}\gamma_{roe} + \beta_{ia}\gamma_{ia}$.

[22] Empirical results for the five-factor model of Fama and French (2015), the three-factor model of Fama and French (1993) augmented with the momentum factor of Carhart (1997), and the three-factor model of Fama and French (1993) augmented with the non-traded liquidity factor of Pastor and Stambaugh (2003) are available from the authors upon request.

4.2 Results

Starting with C-LAB, we investigate whether this model is well identified. To this end, we employ the rank test of Cragg and Donald (1997) and consider two test statistics, $\mathcal{CD}_{SDF}$ in the SDF setup and $\mathcal{CD}_{Beta}$ in the beta setup. To understand how $\mathcal{CD}_{SDF}$ and $\mathcal{CD}_{Beta}$ are computed, let Π be an $N \times K$ matrix and denote its estimate by $\hat{\Pi}$. Assume that $\sqrt{T}\text{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} \mathcal{N}(0_{NK}, M)$, where M is a finite and positive-definite matrix. Under the null that Π is of (reduced) column 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 $c'c = 1$. The Cragg and Donald (1997) test of $H_0 : \text{rank}(\Pi) = K - 1$ takes the form

$$\mathcal{CD} = \min_{c:c'c=1} c'\hat{\Pi}'[c' \otimes I_N)\hat{M}(c \otimes I_N)]^{-1}\hat{\Pi}c, \tag{41}$$

where $\hat{\Pi}$ and $\hat{M}$ are consistent estimates of Π and M, respectively.

In the SDF framework, we have $\hat{\Pi} = \hat{D}$ and $\hat{M} = \frac{1}{T}\sum_{t=1}^{T} \text{vec}(D_t - \hat{D})\text{vec}(D_t - \hat{D})'$. We denote this rank test by $\mathcal{CD}_{SDF}$. In the beta-pricing model setup, we have $\hat{\Pi} = \hat{B} = [\hat{\alpha},\ \hat{\beta}]$ and $\hat{M} = \left[\left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \otimes I_N\right] \left[T^{-1}\sum_{t=1}^{T}(x_t x_t') \otimes (\hat{e}_t\hat{e}_t')\right] \left[\left(T^{-1}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \otimes I_N\right]$, where $\hat{e}_t = R_t - \hat{B}x_t$. The corresponding rank test in the beta-pricing framework is denoted by $\mathcal{CD}_{Beta}$.[23]

Table 6 about here

The outcomes of the two rank tests suggest that C-LAB is poorly identified across different sets of test assets. The p-values of these tests are large, ranging from 0.55 to 0.82 in the SDF setup and from 0.34 to 0.74 in the beta setup, and indicate that the null hypothesis of a deficient column rank for the D and B matrices cannot be rejected. Consistent with our analytical results, this identification failure results in the inability of the $\mathcal{J}$ and $\mathcal{S}$ specification tests to reject the models and in spuriously high pseudo-R^2s for CU-GMM and ML. Across panels, the pseudo-R^2 values for C-LAB (see the columns labeled "all" in the table) range from 0.98 to 0.99 for CU-GMM and are indistinguishable from 1 for ML. Based on $\mathcal{J}$, $\mathcal{S}$, and the CU-GMM and ML pseudo-R^2s, C-LAB appears to have a spectacular fit and a researcher would likely proceed with t-tests of parameter significance with standard errors computed under the assumption of correct model specification. This would lead us to conclude that the *labor* and *prem* factors are often priced in the cross-section

[23] In the model selection procedure described below, we employ this heteroskedasticity-robust version of the $\mathcal{CD}$ test for both ML and CSR GLS even though the ML estimation imposes homoskedasticity on the data.

of expected returns, as emphasized by the high traditional t-ratios on the *prem* and *labor* factors for CU-GMM and ML. Interestingly, the evidence of pricing for the market factor is rather weak, with traditional absolute t-ratio values ranging from 1.29 to 1.73 for CU-GMM and from 0.67 to 1.42 for ML. These empirical findings are again consistent with our methodological findings and reveal the spurious nature of inference as factors that are spurious are selected with high probability, while factors that are useful (such as the market factor) are driven out of the model.

Applying the model reduction procedure, described in Section 3, to C-LAB reveals that only the market factor survives the selection procedure. Essentially, C-LAB reduces to CAPM and the $\mathcal{J}$ and $\mathcal{S}$ tests now have power to reject the model (see columns labeled "selected" in the table). In turn, the pseudo R^2s provide a completely different and more realistic assessment of the goodness-of-fit of the model, ranging from 0.23 to 0.54 for CU-GMM and from 0.11 to 0.14 for ML. The high misspecification-robust t-ratios on *vw* in Panel A suggest some strong pricing ability for the market factor when portfolios are formed on size and book-to-market.[24] In contrast, when considering portfolios formed on size and momentum, the evidence of pricing for *vw* is very limited, consistent with the uncontroversial finding that CAPM cannot explain the returns on portfolios formed on momentum. Panel C also shows that the pricing ability of *vw* is rather weak when employing misspecification-robust t-ratios and considering portfolios formed on size, operating profitability, and investment.

It should be noted that non-invariant estimators, such as the HJ-distance and CSR GLS estimators, provide a less optimistic picture of C-LAB compared to CU-GMM and ML. The p-values of HJD and GLS in Table 6 are always zero even before applying the model selection procedure just described. Therefore, even if HJD and GLS are inconsistent under identification failure (a proof of this claim is available from the authors upon request), they seem to be more robust to lack of identification and can detect model misspecification with higher probability than their invariant counterparts. In sharp contrast with the pseudo-R^2s for CU-GMM and ML, the pseudo-R^2s for the non-invariant estimators are relatively small and range from 0.03 to 0.66 for HJ-distance and from 0.07 to 0.69 for GLS CSR (see the columns labeled "all" in the table). Finally, after applying our model selection procedure, the pricing implications for *vw* (based on misspecification-robust

[24]See Gospodinov, Kan, and Robotti (2017a) for the derivation of misspecification-robust t-ratios for CU-GMM and ML.

t-ratios) are largely consistent across invariant and non-invariant estimators.[25]

The spurious nature of the results analyzed in this paper are probably best illustrated with CC-CAY in Table 7.

Table 7 about here

The rank tests in all three panels provide strong evidence that the models are unidentified. Ignoring the outcome of the rank tests would lead us to conclude that the models estimated by CU-GMM and ML are correctly specified and that scaled consumption growth, $cg \cdot cay$, is highly significant. However, none of the factors survive after applying the proposed model selection procedure since none of the factors (or a subset of factors) in this model satisfy the rank condition.

Tables 8 and 9 about here

The results for ICAPM and D-CCAPM in Tables 8 and 9 further reveal the fragility of statistical inference in models with factors that are only weakly correlated with the test asset returns. As for C-LAB in Table 6, only the market factor survives the selection procedure in D-CCAPM. While the factors *div* and *rf* are selected for some test assets in the final specification of ICAPM in Table 8, their t-statistics are all insignificant when constructed using misspecification-robust standard errors.

Tables 10 and 11 about here

Turning to models with traded factors only, the results for the rank tests in Tables 10 and 11 for FF3 and HXZ suggest that these models are well-identified at the 5% significance level, albeit misspecified (except for the CU-GMM estimator in Panel C of Table 11). We should note that even for models with traded factors, the inference based on non-invariant (HJ-distance and GLS CSR) estimators appears to be more stable and reliable than the inference based on invariant (CU-GMM and ML) estimators.

Overall, our empirical analysis reveals that, for models with non-traded factors, both the SDF and beta-pricing setups are affected by the weak identification problem and share similar pricing

[25] The misspecification-robust t-ratios for HJ-distance are provided in Kan and Robotti (2009), while the ones for GLS CSR can be found in Kan, Robotti, and Shanken (2013).

implications once our model selection procedure and misspecification-robust standard errors are used. It should also be noted that the ML results are based on the joint normality assumption on the factors and the returns. This assumption could be relaxed by adopting a quasi-maximum likelihood framework as, for example, in White (1994). Alternatively, a researcher could use CU-GMM to estimate the parameters of the asset-pricing model in beta-pricing form as described in the Online Appendix.

Finally, in unreported empirical investigations, we explored the performance of these six models using the (i) 25 Fama-French portfolios formed on size and short-term reversal, (ii) 25 Fama-French portfolios formed on size and long-term reversal, and (iii) 25 Fama-French portfolios formed on size and book-to-market plus 17 industry portfolios (all the test assets are from Kenneth French's website). The results based on these three additional sets of test asset returns are largely consistent with the results reported in the paper. However, at least in the SDF formulation of the model, the identification issues become more severe when using the 25 Fama-French portfolios formed on size and short-term reversal and the 25 Fama-French portfolios formed on size and long-term reversal, consistent with the relatively uncontroversial finding in the literature that all these models have problems in explaining short- and long-term reversal. In these latter cases, even models with traded factors only, such as FF3 and HXZ, start to exhibit some non-trivial patterns of weak identification.

Our main empirical findings can be summarized as follows. Models with non-traded factors are often poorly identified, and tend to produce highly misleading inference in terms of spuriously high statistical significance and lack of power in rejecting the null of correct model specification. In addition to the outcome of the rank tests, two observations cast doubts on the validity of the results for these models: (i) the difference between the t-statistics computed under the assumption of correct specification and the misspecification-robust t-statistics (with the misspecification-robust t-statistics being typically statistically insignificant), and (ii) the unrealistically high value of the pseudo-R^2. The models that perform the best are FF3 and especially HXZ where all the factors appear to contribute to pricing and are characterized by statistically significant risk premia. Out of the different sets of test portfolios, the portfolios formed on size and momentum, size and short- and long-term reversal appear to be the most challenging from a pricing perspective.

5 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. 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. 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, the researcher needs to estimate consistently the reduced rank of the model and select the combination of factors 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 (2017a). An alternative empirical strategy is to work with non-invariant estimators (HJ-distance and cross-sectional regression estimators) and pursue misspecification-robust inference that is asymptotically valid regardless of the degree of identification (see Gospodinov, Kan, and Robotti, 2014).[26]

[26] See also Bryzgalova (2016) and Feng, Giglio, and Xiu (2017) for newly proposed model-selection methods based on the lasso estimator in a two-pass setting.

References

[1] Ahn, Seung C., Alex R. Horenstein, and Na Wang, 2017, Beta matrix and common factors in stock returns, *Journal of Financial and Quantitative Analysis*, forthcoming.

[2] Almeida, Caio, and René Garcia, 2012, Assessing misspecified asset pricing models with empirical likelihood estimators, *Journal of Econometrics* 170, 519–537.

[3] Almeida, Caio, and René Garcia, 2017, Economic implications of nonlinear pricing kernels, *Management Science*, forthcoming.

[4] Barillas, Francisco, and Jay Shanken, 2017a, Which alpha?, *Review of Financial Studies* 30, 1316–1338.

[5] Barillas, Francisco, and Jay Shanken, 2017b, Comparing asset pricing models, *Journal of Finance*, forthcoming.

[6] Bekker, Paul, Pascal Dobbelstein, and Tom Wansbeek, 1996, The APT model as reduced-rank regression, *Journal of Business and Economic Statistics* 14, 199–202.

[7] Bryzgalova, Svetlana, 2016, Spurious factors in linear asset pricing models, Working Paper.

[8] Burnside, Craig, 2016, Identification and inference in linear stochastic discount factor models with excess returns, *Journal of Financial Econometrics* 14, 295–330.

[9] Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.

[10] Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

[11] Cragg, John G., and Stephen G. Donald, 1997, Inferring the rank of a matrix, *Journal of Econometrics* 76, 223–250.

[12] Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

[13] Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

[14] Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2017, Taming the factor zoo, Chicago Booth Research Paper No. 17-04.

[15] Ghosh, Anisha, Christian Julliard, and Alex P. Taylor, 2017, What is the consumption-CAPM missing? An information-theoretic framework for the analysis of asset pricing models, *Review of Financial Studies* 30, 442–504.

[16] Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2014, Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors, *Review of Financial Studies* 27, 2139–2170.

[17] Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2017a, Asymptotic variance approximations for invariant estimators in uncertain asset-pricing models, *Econometric Reviews*, forthcoming.

[18] Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2017b, Spurious Inference in reduced-rank asset-pricing models, *Econometrica* 85, 1613–1628.

[19] Granger, Clive W. J., and Paul Newbold, 1974, Spurious regressions in econometrics, *Journal of Econometrics* 2, 111–120.

[20] Hall, Alastair R., 2005, *Generalized Method of Moments*. Oxford University Press, Oxford.

[21] Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

[22] Hansen, Lars P., John Heaton, and Amir Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.

[23] Hansen, Lars P., and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

[24] Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, . . . and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

[25] Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

[26] Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

[27] Kan, Raymond, and Cesare Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies* 22, 3449–3490.

[28] Kan, Raymond, Cesare Robotti, and Jay Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance* 68, 2617–2649.

[29] Kan, Raymond and Chu Zhang, 1999a, Two-pass tests of asset pricing models with useless factors, *Journal of Finance* 54, 203–235.

[30] Kan, Raymond, and Chu Zhang, 1999b, GMM tests of stochastic discount factor models with useless factors, *Journal of Financial Economics* 54, 103–127.

[31] Kleibergen, Frank, 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149–173.

[32] Kleibergen, Frank, and Richard Paap, 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97–126.

[33] Kleibergen, Frank, and Zhaoguo Zhan, 2015, Unexplained factors and their effects on second pass R-squared's, *Journal of Econometrics* 189, 101–116.

[34] Lettau, Martin, and Sydney C. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.

[35] Lewellen, Jonathan W., Stefan Nagel, and Jay Shanken, 2010, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics* 96, 175–194.

[36] Manresa, Elena, Francisco Peñaranda, and Enrique Sentana, 2017, Empirical evaluation of overspecified asset pricing models, CEPR Discussion Paper No. DP12085.

[37] Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

[38] Newey, Whitney K., and Richard J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* 72, 219–255.

[39] Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

[40] Peñaranda, Francisco, and Enrique Sentana, 2015, A unifying approach to the empirical evaluation of asset pricing models, *Review of Economics and Statistics* 97, 412–435.

[41] Petkova, Ralitsa, 2006, Do the Fama-French factors proxy for innovations in predictive variables?, *Journal of Finance* 61, 581–612.

[42] Phillips, Peter C. B., 1989, Partially identified econometric models, *Econometric Theory* 5, 181–240.

[43] Shanken, Jay, 1985, Multivariate tests of the zero-beta CAPM, *Journal of Financial Economics* 14, 327–348.

[44] Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

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

[46] Vissing-Jørgensen, Annette, and Orazio P. Attanasio, 2003, Stock-market participation, intertemporal substitution, and risk-aversion, *American Economic Review Papers and Proceedings* 93, 383–391.

[47] White, Halbert L., 1994, *Estimation, Inference and Specification Analysis.* Cambridge University Press, New York.

[48] Wright, Jonathan H., 2003, Detecting lack of identification in GMM, *Econometric Theory* 19, 322–330.

[49] Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539–580.

[50] Zhou, Guofu, 1995, Small sample rank tests with applications to asset pricing, *Journal of Empirical Finance* 2, 71–93.

Table 1

Test Statistics for CAPM and CAPM Augmented with the "*sp*" Factor

The table reports test statistics for the CAPM, the CAPM augmented with the "*sp*" factor, and a model with the "*sp*" factor only. $\mathcal{J}$ 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: CU-GMM

	CAPM	CAPM + "*sp*" factor	"*sp*" factor
t_{vw}	5.28	0.85	
(p-value)	(0.0000)	(0.3954)	
t_{sp}		5.08	5.12
(p-value)		(0.0000)	(0.0000)
$\mathcal{J}$	62.20	25.53	25.96
(p-value)	(0.0000)	(0.2726)	(0.3029)
R^2	0.2277	0.9928	0.9938

Panel B: ML

	CAPM	CAPM + "*sp*" factor	"*sp*" factor
t_{vw}	−3.65	0.52	
(p-value)	(0.0003)	(0.6027)	
t_{sp}		−4.62	−4.64
(p-value)		(0.0000)	(0.0000)
$\mathcal{S}$	68.79	21.32	21.56
(p-value)	(0.0000)	(0.5011)	(0.5471)
R^2	0.1447	0.9999	1.0000

Table 2
Test Statistics for Various Asset-Pricing Models

The table reports test statistics for four asset-pricing models: CAPM, FF3, C-LAB, and CC-CAY. $\mathcal{CD}_{SDF}$ 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 GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares 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. The rows for the different factors report the t-tests of statistical significance 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 GLS Tests

	CAPM	FF3	C-LAB	CC-CAY
$\mathcal{CD}_{SDF}$	157.03	109.73	18.72	12.34
(p-value)	(0.0000)	(0.0000)	(0.6626)	(0.9499)
HJD	69.35	53.48	67.24	70.15
(p-value)	(0.0000)	(0.0005)	(0.0000)	(0.0009)
$\mathcal{CD}_{Beta}$	493.38	298.24	24.09	15.16
(p-value)	(0.0000)	(0.0000)	(0.3427)	(0.8552)
GLS	71.96	55.61	69.68	71.77
(p-value)	(0.0000)	(0.0004)	(0.0000)	(0.0009)

Panel B: CU-GMM

	CAPM	FF3	C-LAB	CC-CAY
$\mathcal{J}$	62.20	44.26	18.01	12.46
(p-value)	(0.0000)	(0.0022)	(0.6487)	(0.9264)
vw	5.28	4.78	1.73	
smb		−4.60		
hml		−2.93		
$labor$			0.17	
$prem$			−4.16	
cg				2.85
cay				1.79
$cg \cdot cay$				−3.17
R^2	0.2277	0.7783	0.9788	0.9457

Panel C: ML

	CAPM	FF3	C-LAB	CC-CAY
$\mathcal{S}$	68.79	51.05	20.87	13.85
(p-value)	(0.0000)	(0.0003)	(0.4672)	(0.8758)
vw	−3.65	−3.80	1.42	
smb		1.73		
hml		3.04		
$labor$			−3.14	
$prem$			−4.07	
cg				−2.23
cay				−0.77
$cg \cdot cay$				3.63
R^2	0.1447	0.7337	1.0000	0.9995

Table 3
Empirical Distribution of the R^2 Coefficient

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 and ML estimators, respectively. 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 1967:1–2012:12.

T	mean	std	1%	5%	10%	25%	50%	75%	90%	95%	99%
					Panel A: Model with a Useful Factor Only						
					CU-GMM						
200	0.304	0.249	0.000	0.003	0.013	0.076	0.257	0.496	0.678	0.759	0.868
600	0.235	0.185	0.000	0.004	0.015	0.074	0.202	0.364	0.506	0.582	0.702
1000	0.217	0.156	0.000	0.009	0.027	0.089	0.195	0.322	0.438	0.506	0.617
					ML						
200	0.234	0.222	0.000	0.002	0.006	0.041	0.165	0.383	0.582	0.677	0.808
600	0.186	0.169	0.000	0.002	0.007	0.042	0.141	0.293	0.437	0.521	0.652
1000	0.171	0.141	0.000	0.003	0.012	0.053	0.142	0.260	0.376	0.445	0.566
					Panel B: Model with a Spurious Factor Only						
					CU-GMM						
200	0.906	0.125	0.303	0.679	0.789	0.893	0.948	0.973	0.984	0.988	0.993
600	0.990	0.012	0.938	0.970	0.979	0.988	0.994	0.996	0.998	0.998	0.999
1000	0.996	0.004	0.981	0.990	0.993	0.996	0.998	0.999	0.999	0.999	1.000
					ML						
200	0.958	0.109	0.371	0.807	0.903	0.969	0.992	0.998	1.000	1.000	1.000
600	0.997	0.006	0.972	0.989	0.993	0.997	0.999	1.000	1.000	1.000	1.000
1000	0.999	0.002	0.992	0.996	0.998	0.999	1.000	1.000	1.000	1.000	1.000
					Panel C: Model with a Useful and a Spurious Factor						
					CU-GMM						
200	0.909	0.121	0.341	0.684	0.795	0.897	0.950	0.974	0.985	0.989	0.994
600	0.990	0.012	0.944	0.971	0.980	0.989	0.994	0.997	0.998	0.998	0.999
1000	0.997	0.004	0.982	0.990	0.993	0.996	0.998	0.999	0.999	0.999	1.000
					ML						
200	0.956	0.110	0.370	0.792	0.896	0.968	0.992	0.998	1.000	1.000	1.000
600	0.997	0.006	0.973	0.989	0.993	0.997	0.999	1.000	1.000	1.000	1.000
1000	0.999	0.002	0.992	0.997	0.998	0.999	1.000	1.000	1.000	1.000	1.000

Table 4
Rejection Rates of t-tests

The table presents the rejection rates of t-tests of statistical significance under misspecified models for the CU-GMM estimator in the SDF setup and the ML estimator in the beta setup, respectively. 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 1967:1–2012:12. The t-statistics with standard errors computed under the assumption of correct model specification 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 (a) of Theorems 1 and 2.

	CU-GMM						ML					
	useful			spurious			useful			spurious		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Model with a Useful Factor Only												
200	0.908	0.883	0.828	–	–	–	0.698	0.616	0.442	–	–	–
600	0.988	0.983	0.968	–	–	–	0.959	0.934	0.848	–	–	–
1000	0.999	0.998	0.996	–	–	–	0.996	0.992	0.971	–	–	–
∞	1.000	1.000	1.000	–	–	–	1.000	1.000	1.000	–	–	–
Panel B: Model with a Spurious Factor Only												
200	–	–	–	0.997	0.996	0.994	–	–	–	0.996	0.996	0.993
600	–	–	–	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
∞	–	–	–	1.000	1.000	1.000	–	–	–	1.000	1.000	1.000
Panel C: Model with a Useful and a Spurious Factor												
200	0.335	0.246	0.121	0.994	0.992	0.988	0.271	0.185	0.075	0.992	0.991	0.986
600	0.170	0.097	0.024	1.000	1.000	1.000	0.171	0.099	0.025	1.000	1.000	1.000
1000	0.142	0.075	0.015	1.000	1.000	1.000	0.152	0.083	0.019	1.000	1.000	1.000
∞	0.105	0.049	0.007	1.000	1.000	1.000	0.124	0.062	0.011	1.000	1.000	1.000

Table 5
Probabilities of Retaining Factors in Model Reduction Procedure

The table presents the probabilities of retaining factors in our proposed model reduction procedure in the SDF framework. The results are reported 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 1967:1–2012:12. The level of the rank test on D is 1%. P_A, P_B, and P_C are the marginal probability of retaining the useful factors, the marginal probability of eliminating the spurious factors, and the joint probability of retaining the useful factors and eliminating the spurious factors, respectively.

T	P_A	P_B	P_C
Panel A: 1 Useful Factor Only			
200	1.000	–	1.000
600	1.000	–	1.000
1000	1.000	–	1.000
Panel B: 1 Spurious Factor Only			
200	–	0.962	0.962
600	–	0.985	0.985
1000	–	0.987	0.987
Panel C: 1 Useful and 1 Spurious Factor			
200	0.992	0.958	0.958
600	1.000	0.986	0.986
1000	1.000	0.988	0.988
Panel D: 3 Useful and 2 Spurious Factors			
200	0.993	0.989	0.983
600	1.000	1.000	1.000
1000	1.000	1.000	1.000

Table 6
Test Statistics for C-LAB

The table reports test statistics for C-LAB. $\mathcal{CD}_{SDF}$ 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. $\mathcal{J}$ 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. HJD and GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares cross-sectional regression test of Shanken (1985), respectively. The rows for the different factors report the t-tests of statistical significance with standard errors computed under the assumption of correct model specification and the misspecification-robust t-tests (in square brackets). Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: 25 Portfolios Formed on Size and Book-to-Market

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	18.72 (0.6626)		$\mathcal{CD}_{Beta}$ (p-value)	24.09 (0.3427)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	18.01 (0.6487)	62.20 (0.0000)	$\mathcal{S}$ (p-value)	20.87 (0.4672)	68.79 (0.0000)
R^2	0.9788	0.2277	R^2	1.0000	0.1447
vw	1.73 [0.68]	5.28 [3.13]	vw	1.42 [0.01]	−3.65 [−2.92]
$labor$	0.17 [0.10]	–	$labor$	−3.14 [−0.01]	–
$prem$	−4.16 [−0.51]	–	$prem$	−4.07 [−0.01]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	67.24 (0.0000)	69.35 (0.0000)	GLS (p-value)	69.68 (0.0000)	71.96 (0.0000)
R^2	0.1280	0.0815	R^2	0.1111	0.0993
vw	3.05 [2.21]	3.35 [3.10]	vw	−2.66 [−2.09]	−3.14 [−2.97]
$labor$	0.93 [0.41]	–	$labor$	−1.05 [−0.47]	–
$prem$	−0.48 [−0.19]	–	$prem$	0.56 [0.23]	–

Table 6 (cont'd)

Panel B: 25 Portfolios Formed on Size and Momentum

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	15.93 (0.8194)		$\mathcal{CD}_{Beta}$ (p-value)	17.45 (0.7383)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	15.70 (0.7865)	99.86 (0.0000)	$\mathcal{S}$ (p-value)	17.71 (0.6674)	105.80 (0.0000)
R^2	0.9896	0.2421	R^2	1.0000	0.1128
vw	−1.45 [−0.29]	3.44 [1.23]	*vw*	−0.67 [−0.00]	−0.95 [−0.68]
labor	2.47 [0.28]	–	*labor*	2.82 [0.00]	–
prem	3.97 [0.00]	–	*prem*	−4.07 [−0.01]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	95.07 (0.0000)	103.54 (0.0000)	GLS (p-value)	97.23 (0.0000)	106.09 (0.0000)
R^2	0.6642	0.0702	R^2	0.6890	0.0963
vw	0.14 [0.11]	0.91 [0.82]	*vw*	−0.49 [−0.42]	−0.75 [−0.68]
labor	−1.56 [−0.76]	–	*labor*	1.89 [0.98]	–
prem	0.64 [0.27]	–	*prem*	−0.89 [−0.37]	–

Panel C: 32 Portfolios Formed on Size, Operating Profitability, and Investment

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	27.43 (0.5487)		$\mathcal{CD}_{Beta}$ (p-value)	30.35 (0.3966)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	27.42 (0.4957)	146.89 (0.0000)	$\mathcal{S}$ (p-value)	25.56 (0.5972)	159.18 (0.0000)
R^2	0.9869	0.5370	R^2	1.0000	0.1055
vw	1.29 [0.01]	10.20 [1.69]	*vw*	0.73 [0.06]	−2.61 [−1.68]
labor	5.23 [0.01]	–	*labor*	−4.86 [−0.06]	–
prem	1.48 [0.01]	–	*prem*	−1.74 [−0.06]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	157.26 (0.0000)	157.32 (0.0000)	GLS (p-value)	161.93 (0.0000)	161.99 (0.0000)
R^2	0.0317	0.0527	R^2	0.0679	0.0717
vw	2.17 [1.85]	2.16 [1.92]	*vw*	−1.88 [−1.69]	−1.90 [−1.71]
labor	0.19 [0.07]	–	*labor*	−0.18 [−0.07]	–
prem	−0.08 [−0.03]	–	*prem*	0.07 [0.03]	–

Table 7
Test Statistics for CC-CAY

The table reports test statistics for CC-CAY. $\mathcal{CD}_{SDF}$ 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. $\mathcal{J}$ 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. HJD and GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares cross-sectional regression test of Shanken (1985), respectively. The rows for the different factors report the t-tests of statistical significance with standard errors computed under the assumption of correct model specification and the misspecification-robust t-tests (in square brackets). Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: 25 Portfolios Formed on Size and Book-to-Market

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	12.34 (0.9499)		$\mathcal{CD}_{Beta}$ (p-value)	15.16 (0.8552)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	12.46 (0.9264)	81.90 (0.0000)	$\mathcal{S}$ (p-value)	13.85 (0.8758)	81.90 (0.0000)
R^2	0.9457	–	R^2	0.9995	–
cg	2.55 [0.12]	–	cg	-2.23 [-0.12]	–
cay	1.79 [0.11]	–	cay	-0.77 [-0.04]	–
$cg \cdot cay$	-3.17 [-0.13]	–	$cg \cdot cay$	3.63 [0.19]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	70.15 (0.0009)	79.72 (0.0000)	GLS (p-value)	71.77 (0.0009)	81.90 (0.0000)
R^2	0.1100	–	R^2	0.0475	–
cg	-0.80 [-0.49]	–	cg	0.70 [0.40]	–
cay	-1.06 [-0.55]	–	cay	1.34 [0.70]	–
$cg \cdot cay$	-1.77 [-0.91]	–	$cg \cdot cay$	1.84 [0.94]	–

Table 7 (cont'd)

Panel B: 25 Portfolios Formed on Size and Momentum

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	18.57 (0.6717)		$\mathcal{CD}_{Beta}$ (p-value)	28.58 (0.1574)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	17.90 (0.6553)	106.65 (0.0000)	$\mathcal{S}$ (p-value)	20.79 (0.4721)	106.65 (0.0000)
R^2	0.9967	–	R^2	0.9977	–
cg	−2.49 [−0.44]	–	cg	−0.63 [−0.05]	–
cay	0.40 [0.14]	–	cay	−0.20 [−0.01]	–
$cg \cdot cay$	−3.93 [−0.49]	–	$cg \cdot cay$	4.69 [0.16]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	72.81 (0.0095)	104.35 (0.0000)	GLS (p-value)	73.91 (0.0098)	106.65 (0.0000)
R^2	0.1032	–	R^2	0.0368	–
cg	−1.61 [−1.24]	–	cg	1.71 [1.27]	–
cay	−3.14 [−2.14]	–	cay	3.59 [2.51]	–
$cg \cdot cay$	−1.45 [−0.91]	–	$cg \cdot cay$	1.64 [1.01]	–

Panel C: 32 Portfolios Formed on Size, Operating Profitability, and Investment

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	22.83 (0.7843)		$\mathcal{CD}_{Beta}$ (p-value)	27.04 (0.5695)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	24.79 (0.6392)	165.59 (0.0000)	$\mathcal{S}$ (p-value)	23.68 (0.6985)	165.59 (0.0000)
R^2	0.9951	–	R^2	0.9999	–
cg	4.12 [0.05]	–	cg	−4.21 [−0.15]	–
cay	−1.23 [−0.06]	–	cay	1.21 [0.05]	–
$cg \cdot cay$	−4.10 [−0.05]	–	$cg \cdot cay$	4.04 [0.10]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	160.03 (0.0000)	161.67 (0.0000)	GLS (p-value)	163.80 (0.0000)	165.59 (0.0000)
R^2	0.0273	–	R^2	0.0009	–
cg	0.43 [0.18]	–	cg	−0.37 [−0.16]	–
cay	−0.92 [−0.34]	–	cay	1.07 [0.41]	–
$cg \cdot cay$	−0.68 [−0.29]	–	$cg \cdot cay$	0.72 [0.31]	–

Table 8
Test Statistics for ICAPM

The table reports test statistics for ICAPM. $\mathcal{CD}_{SDF}$ 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. $\mathcal{J}$ 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. HJD and GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares cross-sectional regression test of Shanken (1985), respectively. The rows for the different factors report the t-tests of statistical significance with standard errors computed under the assumption of correct model specification and the misspecification-robust t-tests (in square brackets). Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: 25 Portfolios Formed on Size and Book-to-Market

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	20.25 (0.4425)		$\mathcal{CD}_{Beta}$ (p-value)	29.61 (0.0764)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	21.84 (0.2922)	62.20 (0.0000)	$\mathcal{S}$ (p-value)	21.46 (0.3118)	40.74 (0.0088)
R^2	0.9702	0.2277	R^2	0.9942	0.9736
vw	0.39 [0.16]	5.28 [3.13]	vw	1.79 [0.61]	1.37 [0.41]
$term$	−3.39 [−1.26]	–	$term$	4.78 [1.05]	–
def	−1.20 [−0.48]	–	def	1.16 [0.41]	–
div	0.17 [0.07]	–	div	−2.14 [−0.60]	–
rf	2.57 [0.49]	–	rf	−3.19 [−0.90]	−6.31 [−0.88]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	61.70 (0.0015)	69.35 (0.0000)	GLS (p-value)	63.72 (0.0016)	69.37 (0.0001)
R^2	0.3497	0.0815	R^2	0.3692	0.1360
vw	0.11 [0.07]	3.35 [3.10]	vw	−1.77 [−1.38]	−2.49 [−2.10]
$term$	−1.85 [−1.11]	–	$term$	2.13 [1.16]	–
def	0.37 [0.25]	–	def	−0.23 [−0.14]	–
div	−0.44 [−0.27]	–	div	1.00 [0.64]	–
rf	0.25 [0.18]	–	rf	−1.66 [−0.91]	−1.48 [−0.90]

Table 8 (cont'd)

Panel B: 25 Portfolios Formed on Size and Momentum

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	15.30 (0.7588)		$\mathcal{CD}_{Beta}$ (p-value)	19.01 (0.5212)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	13.50 (0.8122)	101.51 (0.0000)	$\mathcal{S}$ (p-value)	24.55 (0.1757)	41.39 (0.0050)
R^2	0.9940	0.2128	R^2	0.9976	0.9950
vw	3.41 [1.14]	–	vw	2.02 [0.62]	0.58 [0.27]
$term$	0.15 [0.10]	–	$term$	3.41 [0.57]	–
def	3.78 [1.10]	–	def	−3.72 [−0.60]	–
div	3.41 [1.16]	−2.98 [−0.85]	div	−3.44 [−0.66]	−0.84 [−0.24]
rf	1.29 [0.73]	–	rf	−1.63 [−0.49]	−6.44 [−0.63]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	98.03 (0.0000)	104.20 (0.0000)	GLS (p-value)	99.90 (0.0000)	100.48 (0.0000)
R^2	0.0001	0.0495	R^2	0.1048	0.0653
vw	1.93 [1.01]	–	vw	0.17 [0.13]	0.02 [0.02]
$term$	−0.18 [−0.10]	–	$term$	1.05 [0.56]	–
def	0.59 [0.31]	–	def	−0.25 [−0.13]	–
div	1.73 [0.89]	−0.40 [−0.34]	div	−1.33 [−0.74]	−1.19 [−0.83]
rf	1.10 [0.68]	–	rf	−1.35 [−0.84]	−1.33 [−0.82]

Panel C: 32 Portfolios Formed on Size, Operating Profitability, and Investment

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	21.34 (0.7699)		$\mathcal{CD}_{Beta}$ (p-value)	23.97 (0.6321)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	21.32 (0.7252)	146.89 (0.0000)	$\mathcal{S}$ (p-value)	18.39 (0.8611)	159.18 (0.0000)
R^2	0.9954	0.5370	R^2	0.9996	0.1055
vw	−1.26 [−0.02]	10.20 [1.69]	vw	−2.42 [−0.45]	−2.61 [−1.68]
$term$	3.21 [0.03]	–	$term$	−3.64 [−0.47]	–
def	4.28 [0.02]	–	def	−3.58 [−0.46]	–
div	−1.43 [−0.02]	–	div	2.70 [0.43]	–
rf	2.49 [0.02]	–	rf	1.55 [0.35]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	147.87 (0.0000)	157.32 (0.0000)	GLS (p-value)	152.35 (0.0000)	161.99 (0.0000)
R^2	0.1584	0.0527	R^2	0.2528	0.0717
vw	2.10 [0.98]	2.16 [1.92]	vw	−1.62 [−0.98]	−1.90 [−1.71]
$term$	1.89 [0.78]	–	$term$	−1.56 [−0.61]	–
def	1.13 [0.47]	–	def	−0.79 [−0.32]	–
div	1.70 [0.79]	–	div	0.23 [0.10]	–
rf	0.92 [0.43]	–	rf	0.30 [0.14]	–

Table 9
Test Statistics for D-CCAPM

The table reports test statistics for D-CCAPM. $\mathcal{CD}_{SDF}$ 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. $\mathcal{J}$ 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. HJD and GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares cross-sectional regression test of Shanken (1985), respectively. The rows for the different factors report the t-tests of statistical significance with standard errors computed under the assumption of correct model specification and the misspecification-robust t-tests (in square brackets). Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: 25 Portfolios Formed on Size and Book-to-Market

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	25.35 (0.2808)		$\mathcal{CD}_{Beta}$ (p-value)	35.33 (0.0358)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	20.63 (0.4820)	62.20 (0.0000)	$\mathcal{S}$ (p-value)	28.37 (0.1299)	68.79 (0.0000)
R^2	0.9368	0.2277	R^2	0.9762	0.1447
vw	3.06 [2.10]	5.28 [3.13]	vw	−2.17 [−1.77]	−3.65 [−2.92]
cg	−4.73 [−1.55]	–	cg	5.31 [1.77]	–
$cgdur$	0.94 [0.52]	–	$cgdur$	2.35 [0.54]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	59.65 (0.0026)	69.35 (0.0000)	GLS (p-value)	61.96 (0.0023)	71.96 (0.0000)
R^2	0.4297	0.0815	R^2	0.3642	0.0993
vw	3.69 [3.18]	3.35 [3.10]	vw	−3.09 [−2.93]	−3.14 [−2.97]
cg	−2.42 [−1.66]	–	cg	2.61 [1.69]	–
$cgdur$	−0.49 [−0.36]	–	$cgdur$	1.09 [0.79]	–

Table 9 (cont'd)

Panel B: 25 Portfolios Formed on Size and Momentum

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	20.37 (0.5599)		$\mathcal{CD}_{Beta}$ (p-value)	25.86 (0.2579)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	20.11 (0.5145)	99.86 (0.0000)	$\mathcal{S}$ (p-value)	30.15 (0.0889)	105.80 (0.0000)
R^2	0.9870	0.2421	R^2	0.9926	0.1128
vw	1.69 [0.26]	3.44 [1.23]	*vw*	−0.81 [−0.62]	−0.95 [−0.68]
cg	−4.27 [−0.25]	–	*cg*	−0.07 [−0.00]	–
cgdur	3.27 [0.22]	–	*cgdur*	5.52 [0.22]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	92.57 (0.0000)	103.54 (0.0000)	GLS (p-value)	95.14 (0.0000)	106.09 (0.0000)
R^2	0.2193	0.0702	R^2	0.0143	0.0963
vw	1.76 [1.33]	0.91 [0.82]	*vw*	−1.10 [−0.96]	−0.75 [−0.68]
cg	−1.72 [−1.00]	–	*cg*	2.21 [1.24]	–
cgdur	−1.48 [−0.81]	–	*cgdur*	2.04 [1.07]	–

Panel C: 32 Portfolios Formed on Size, Operating Profitability, and Investment

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	21.37 (0.8452)		$\mathcal{CD}_{Beta}$ (p-value)	23.84 (0.7366)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	21.29 (0.8131)	146.89 (0.0000)	$\mathcal{S}$ (p-value)	29.82 (0.3717)	159.18 (0.0000)
R^2	0.9963	0.5370	R^2	0.9990	0.1055
vw	2.91 [0.09]	10.20 [1.69]	*vw*	0.08 [0.00]	−2.61 [−1.68]
cg	−4.60 [−0.09]	–	*cg*	−2.89 [−0.02]	–
cgdur	1.03 [0.07]	–	*cgdur*	4.02 [0.08]	–
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	149.60 (0.0000)	157.32 (0.0000)	GLS (p-value)	154.38 (0.0000)	161.99 (0.0000)
R^2	0.6335	0.0527	R^2	0.5117	0.0717
vw	2.23 [1.54]	2.16 [1.92]	*vw*	−2.14 [−1.83]	−1.90 [−1.71]
cg	−0.69 [−0.32]	–	*cg*	1.07 [0.50]	–
cgdur	−1.95 [−1.05]	–	*cgdur*	2.32 [1.13]	–

Table 10
Test Statistics for FF3

The table reports test statistics for FF3. $\mathcal{CD}_{SDF}$ 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. $\mathcal{J}$ 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. HJD and GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares cross-sectional regression test of Shanken (1985), respectively. The rows for the different factors report the t-tests of statistical significance with standard errors computed under the assumption of correct model specification and the misspecification-robust t-tests (in square brackets). Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: 25 Portfolios Formed on Size and Book-to-Market

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	109.73 (0.0000)		$\mathcal{CD}_{Beta}$ (p-value)	298.24 (0.0000)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	44.26 (0.0022)	44.26 (0.0022)	$\mathcal{S}$ (p-value)	51.05 (0.0003)	51.05 (0.0003)
R^2	0.7783	0.7783	R^2	0.7337	0.7337
vw	4.78 [3.33]	4.78 [3.33]	vw	−3.80 [−3.03]	−3.80 [−3.03]
smb	−4.60 [−3.48]	−4.60 [−3.48]	smb	1.73 [1.72]	1.73 [1.72]
hml	−2.93 [−2.11]	−2.93 [−2.11]	hml	3.04 [3.03]	3.04 [3.03]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	53.48 (0.0005)	53.48 (0.0005)	GLS (p-value)	55.61 (0.0004)	55.61 (0.0004)
R^2	0.6890	0.6890	R^2	0.6901	0.6901
vw	3.24 [2.95]	3.24 [2.95]	vw	−3.29 [−3.02]	−3.29 [−3.02]
smb	−3.32 [−3.26]	−3.32 [−3.26]	smb	1.73 [1.73]	1.73 [1.73]
hml	−1.96 [−1.96]	−1.96 [−1.96]	hml	3.04 [3.04]	3.04 [3.04]

Table 10 (cont'd)

Panel B: 25 Portfolios Formed on Size and Momentum

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	35.09 (0.0379)		$\mathcal{CD}_{Beta}$ (p-value)	64.60 (0.0000)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	33.71 (0.0389)	68.40 (0.0000)	$\mathcal{S}$ (p-value)	77.55 (0.0000)	77.55 (0.0000)
R^2	0.9869	0.7269	R^2	0.8805	0.8805
vw	4.71 [0.70]	6.94 [3.17]	*vw*	−5.32 [−1.76]	−5.32 [−1.76]
smb	−4.67 [−0.73]	−6.86 [−3.48]	*smb*	4.06 [2.84]	4.06 [2.84]
hml	5.58 [0.68]	–	*hml*	−4.63 [−1.48]	−4.63 [−1.48]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	90.29 (0.0000)	92.77 (0.0000)	GLS (p-value)	93.49 (0.0000)	93.49 (0.0000)
R^2	0.4837	0.2594	R^2	0.4934	0.4934
vw	2.36 [1.80]	2.20 [1.83]	*vw*	−1.88 [−1.48]	−1.88 [−1.48]
smb	−3.06 [−2.72]	−3.18 [−2.90]	*smb*	2.99 [2.76]	2.99 [2.76]
hml	1.43 [1.05]	–	*hml*	−1.30 [−0.95]	−1.30 [−0.95]

Panel C: 32 Portfolios Formed on Size, Operating Profitability, and Investment

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	80.26 (0.0000)		$\mathcal{CD}_{Beta}$ (p-value)	160.90 (0.0000)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	80.20 (0.0000)	80.20 (0.0000)	$\mathcal{S}$ (p-value)	133.50 (0.0000)	133.50 (0.0000)
R^2	0.9492	0.9492	R^2	0.5981	0.5981
vw	−8.64 [−0.02]	−8.64 [−0.02]	*vw*	−0.46 [−0.20]	−0.46 [−0.20]
smb	4.71 [0.02]	4.71 [0.02]	*smb*	0.94 [0.88]	0.94 [0.88]
hml	−7.86 [−0.02]	−7.86 [−0.02]	*hml*	4.66 [2.85]	4.66 [2.85]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	138.51 (0.0000)	138.51 (0.0000)	GLS (p-value)	141.97 (0.0000)	141.97 (0.0000)
R^2	0.5766	0.5766	R^2	0.5394	0.5394
vw	0.60 [0.48]	0.60 [0.48]	*vw*	−0.92 [−0.77]	−0.92 [−0.77]
smb	−2.24 [−2.12]	−2.24 [−2.12]	*smb*	1.12 [1.09]	1.12 [1.09]
hml	−3.45 [−2.87]	−3.45 [−2.87]	*hml*	3.96 [3.46]	3.96 [3.46]

Table 11
Test Statistics for HXZ

The table reports test statistics for HXZ. $\mathcal{CD}_{SDF}$ 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. $\mathcal{J}$ 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. HJD and GLS denote the tests of correct model specification based on the distance measure of Hansen and Jagannathan (1997) and on the generalized least squares cross-sectional regression test of Shanken (1985), respectively. The rows for the different factors report the t-tests of statistical significance with standard errors computed under the assumption of correct model specification and the misspecification-robust t-tests (in square brackets). Finally, R^2 denotes the squared correlation coefficient between the fitted expected returns and the average realized returns.

Panel A: 25 Portfolios Formed on Size and Book-to-Market

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	42.60 (0.0035)		$\mathcal{CD}_{Beta}$ (p-value)	93.98 (0.0000)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	42.11 (0.0027)	42.11 (0.0027)	$\mathcal{S}$ (p-value)	50.72 (0.0002)	50.72 (0.0002)
R^2	0.8162	0.8162	R^2	0.7607	0.7607
vw	2.89 [2.12]	2.89 [2.12]	vw	−3.29 [−2.25]	−3.29 [−2.25]
me	−4.86 [−3.44]	−4.86 [−3.44]	me	2.53 [2.37]	2.53 [2.37]
roe	−2.30 [−1.19]	−2.30 [−1.19]	roe	1.65 [1.03]	1.65 [1.03]
ia	−1.73 [−1.00]	−1.73 [−1.00]	ia	2.72 [2.05]	2.72 [2.05]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	54.01 (0.0003)	54.01 (0.0003)	GLS (p-value)	56.09 (0.0003	56.09 (0.0003)
R^2	0.7204	0.7204	R^2	0.6938	0.6938
vw	2.23 [1.96]	2.23 [1.96]	vw	−2.98 [−2.67]	−2.98 [−2.67]
me	−3.62 [−3.42]	−3.62 [−3.42]	me	2.38 [2.34]	2.38 [2.34]
roe	−1.23 [−1.06]	−1.23 [−1.06]	roe	1.24 [1.06]	1.24 [1.06]
ia	−1.18 [−1.07]	−1.18 [−1.07]	ia	2.54 [2.35]	2.54 [2.35]

Table 11 (cont'd)

Panel B: 25 Portfolios Formed on Size and Momentum

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	37.19 (0.0160)		$\mathcal{CD}_{Beta}$ (p-value)	51.59 (0.0002)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	30.62 (0.0604)	67.90 (0.0000)	$\mathcal{S}$ (p-value)	51.56 (0.0001)	51.56 (0.0001)
R^2	0.9694	0.7263	R^2	0.9347	0.9347
vw	−3.98 [−1.51]	7.04 [3.45]	vw	3.21 [0.75]	3.21 [0.75]
me	−3.41 [−1.28]	−6.93 [−3.68]	me	3.80 [3.60]	3.80 [3.60]
roe	−4.67 [−2.06]	–	roe	3.71 [1.81]	3.71 [1.81]
ia	−4.47 [−1.67]	–	ia	3.35 [0.66]	3.35 [0.66]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	64.98 (0.0008)	93.45 (0.0000)	GLS (p-value)	65.79 (0.0009)	65.79 (0.0009)
R^2	0.9086	0.2628	R^2	0.8784	0.8784
vw	−0.82 [−0.60]	2.02 [1.67]	vw	0.74 [0.58]	0.74 [0.58]
me	−4.24 [−4.20]	−3.13 [−2.89]	me	3.46 [3.41]	3.46 [3.41]
roe	−3.30 [−3.17]	–	roe	3.23 [3.17]	3.23 [3.17]
ia	−1.08 [−0.79]	–	ia	0.96 [0.71]	0.96 [0.71]

Panel C: 32 Portfolios Formed on Size, Operating Profitability, and Investment

CU-GMM and HJ-Distance			ML and GLS CSR		
$\mathcal{CD}_{SDF}$ (p-value)	50.50 (0.0057)		$\mathcal{CD}_{Beta}$ (p-value)	122.10 (0.0000)	
CU-GMM			ML		
Factors	all	selected	Factors	all	selected
$\mathcal{J}$ (p-value)	28.54 (0.3836)	28.54 (0.3836)	$\mathcal{S}$ (p-value)	69.36 (0.0000)	69.36 (0.0000)
R^2	0.9472	0.9472	R^2	0.8810	0.8810
vw	−4.90 [−2.58]	−4.90 [−2.58]	vw	2.94 [1.80]	2.94 [1.80]
me	−1.67 [−0.94]	−1.67 [−0.94]	me	3.59 [3.40]	3.59 [3.40]
roe	−6.46 [−3.48]	−6.46 [−3.48]	roe	6.74 [4.45]	6.74 [4.45]
ia	−6.27 [−3.66]	−6.27 [−3.66]	ia	6.42 [5.65]	6.42 [5.65]
HJ-Distance			GLS CSR		
Factors	all	selected	Factors	all	selected
HJD (p-value)	101.22 (0.0001)	101.22 (0.0001)	GLS (p-value)	102.44 (0.0001)	102.44 (0.0001)
R^2	0.8325	0.8325	R^2	0.7499	0.7499
vw	−1.99 [−1.85]	−1.99 [−1.85]	vw	0.90 [0.82]	0.90 [0.82]
me	−4.03 [−3.87]	−4.03 [−3.87]	me	2.97 [2.86]	2.97 [2.86]
roe	−4.24 [−4.26]	−4.24 [−4.26]	roe	4.63 [4.56]	4.63 [4.56]
ia	−4.41 [−4.36]	−4.41 [−4.36]	ia	5.73 [5.57]	5.73 [5.57]

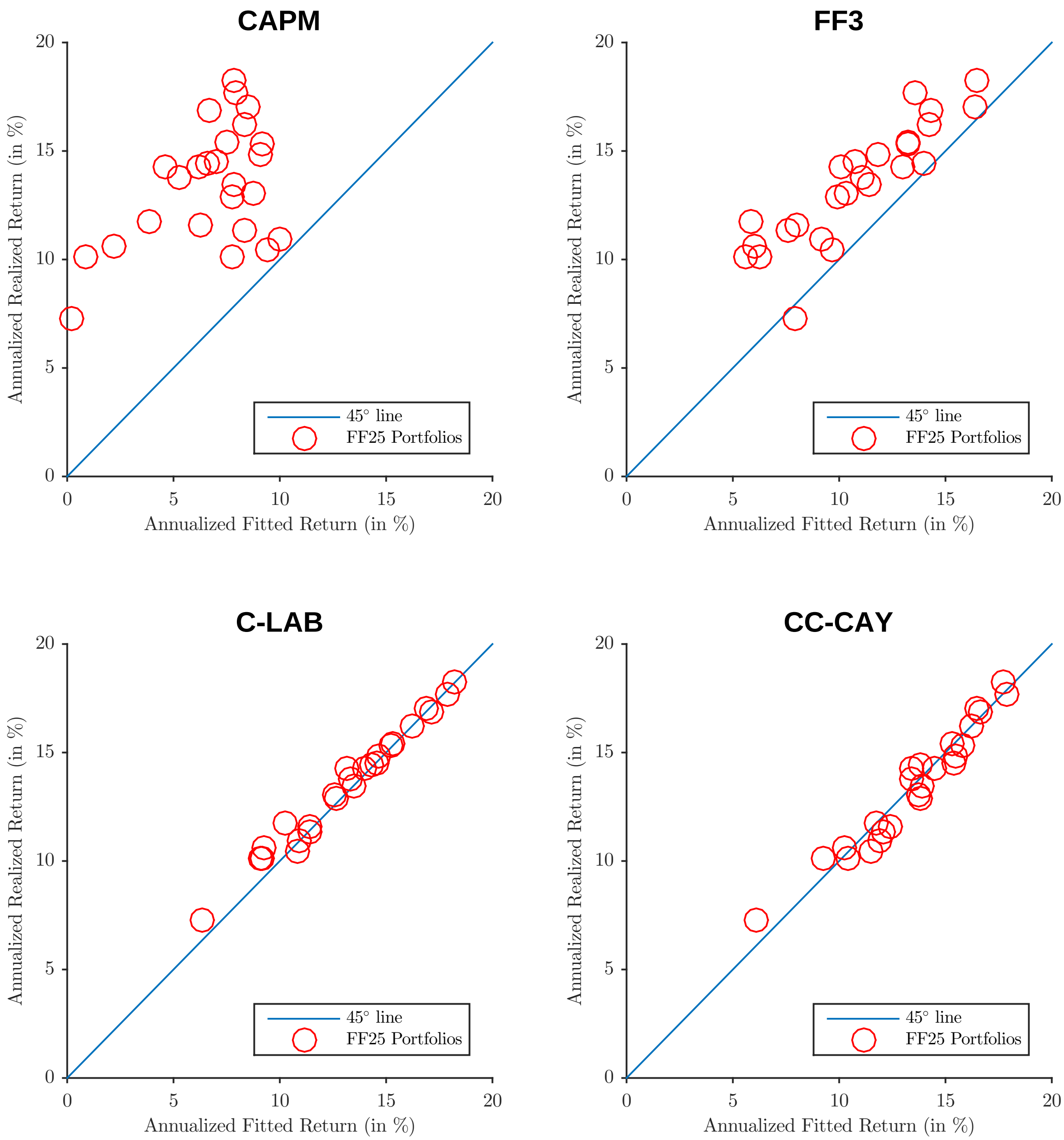


Figure 1. Realized vs. Fitted (by CU-GMM) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by CU-GMM) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

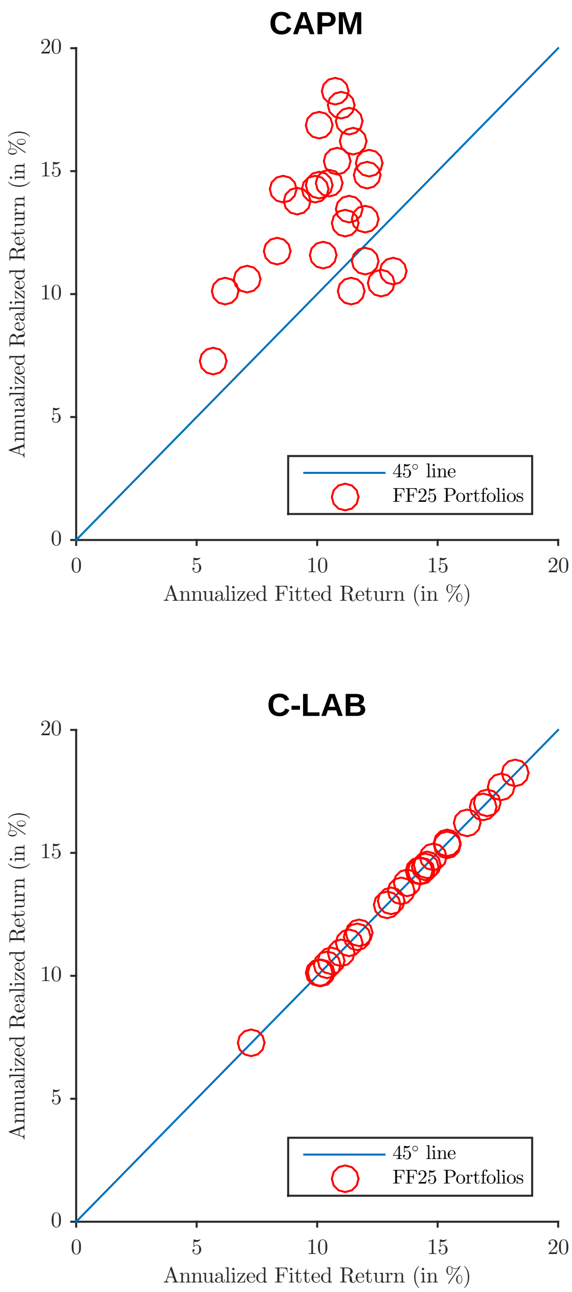


Figure 2. Realized vs. Fitted (by ML) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by ML) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

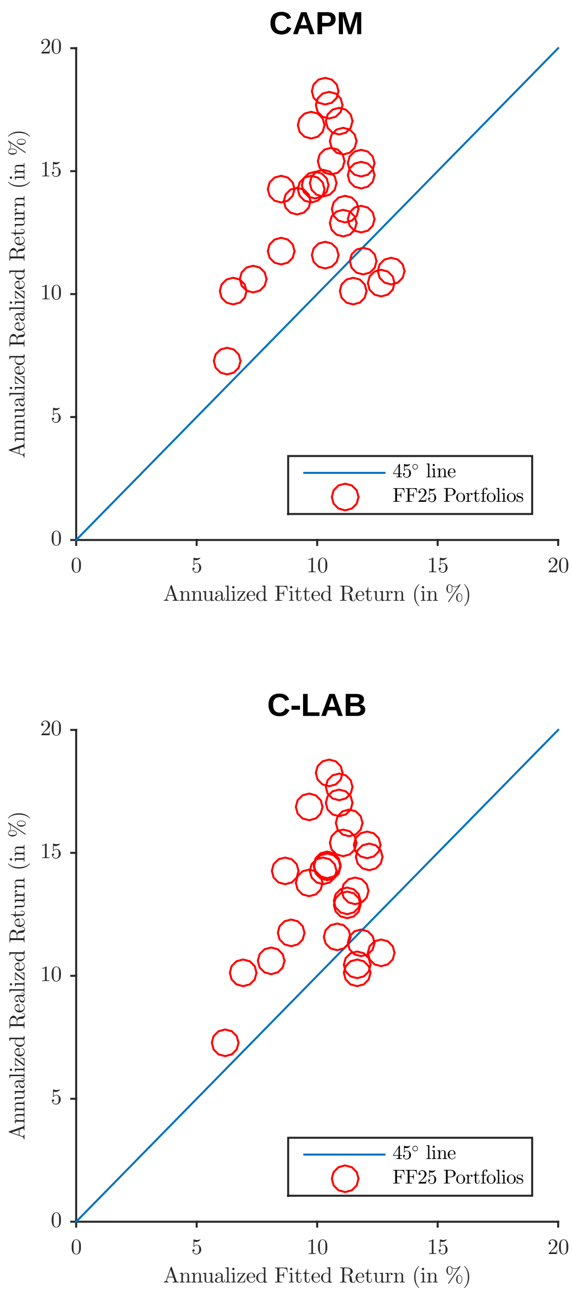


Figure 3. Realized vs. Fitted (by HJ-Distance) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by HJ-distance) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

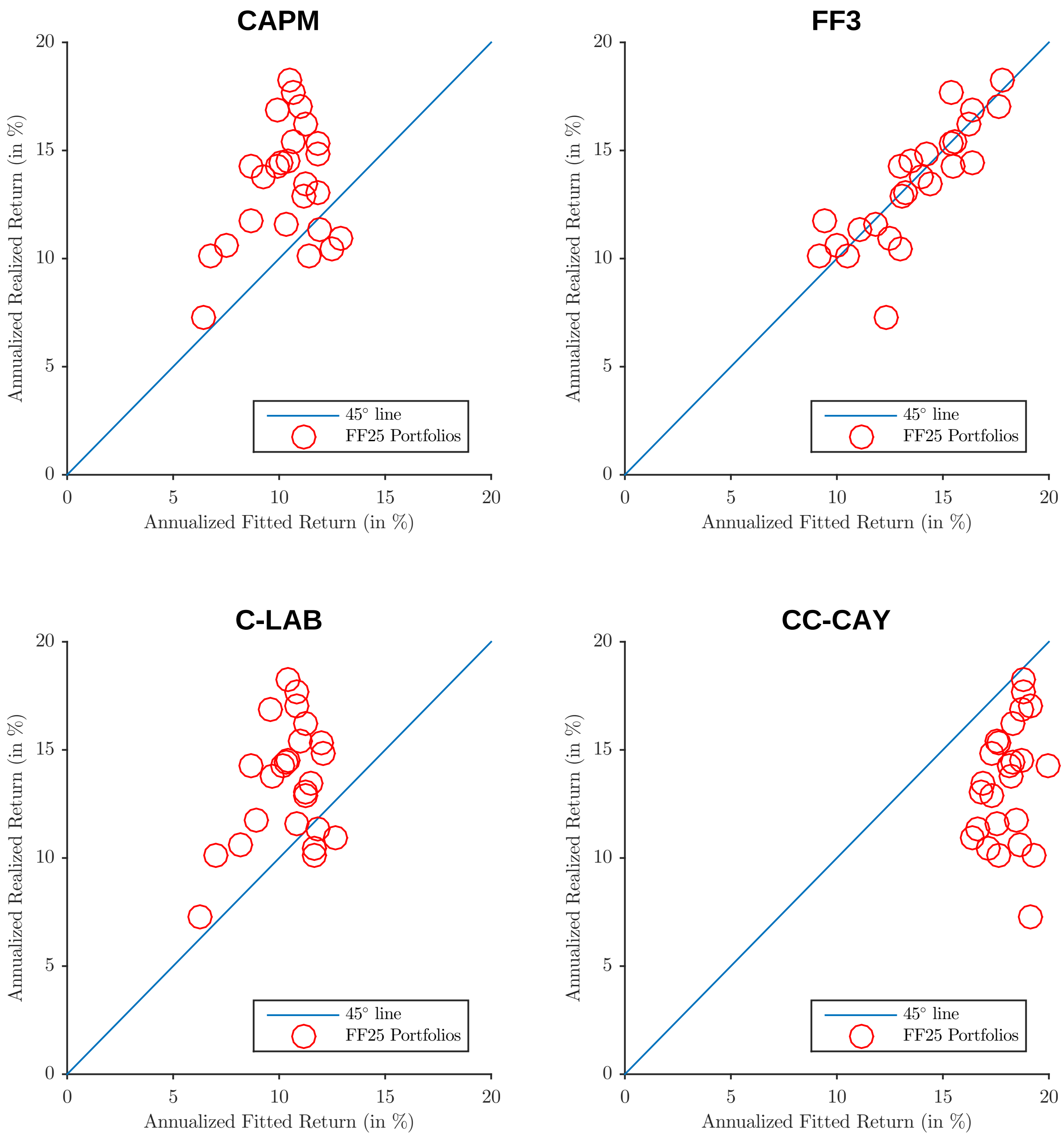


Figure 4. Realized vs. Fitted (by GLS) Returns: 25 Fama-French Portfolios. The figure shows the average realized returns versus fitted expected returns (by GLS) for each of the 25 Fama-French portfolios for CAPM, FF3, C-LAB, and CC-CAY.

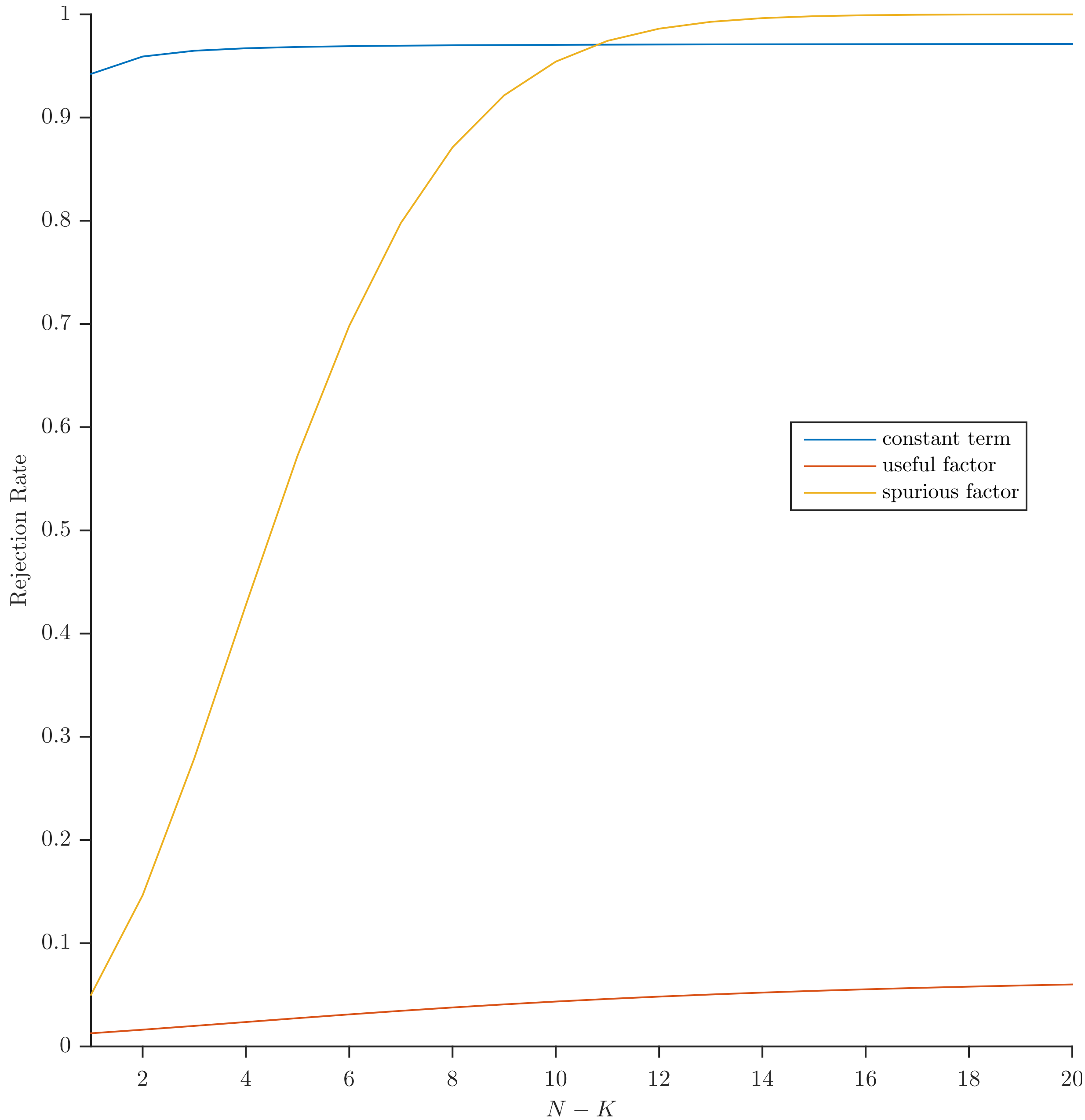
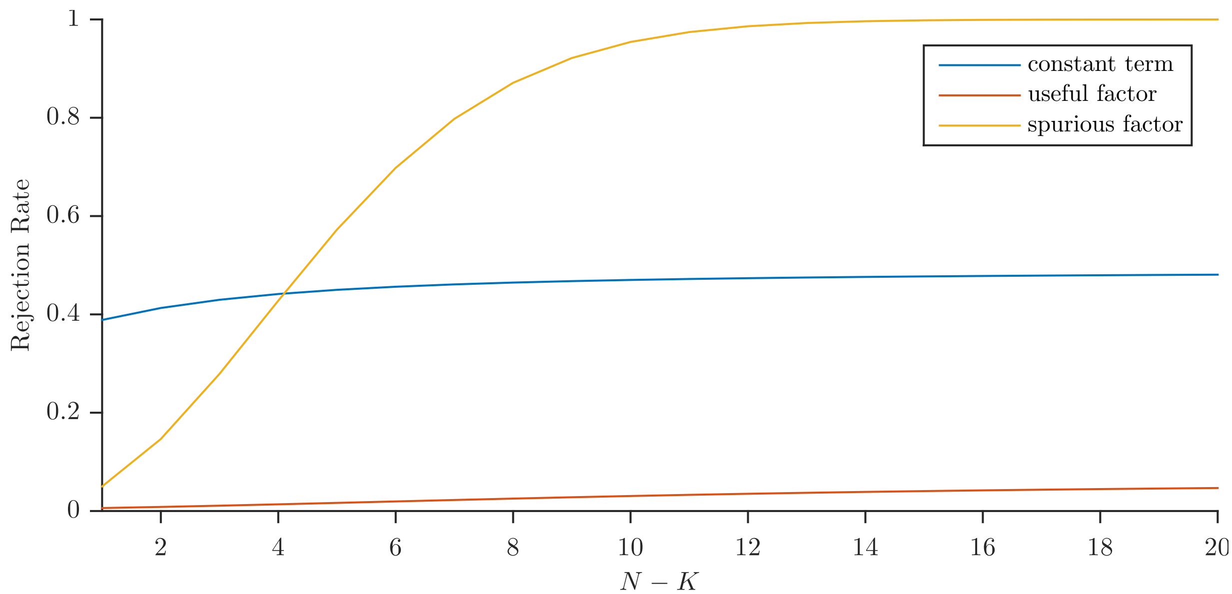


Figure 5. Limiting Rejection Rates of t-tests of Statistical Significance under Misspecified Models for MLE. The figure plots the limiting rejection rates of $t(\hat{\gamma}_0^{ML})$, $t(\hat{\gamma}_{1,i}^{ML})$, and $t(\hat{\gamma}_{1,K-1}^{ML})$ as functions of $N-K$ when one uses the standard normal critical values.



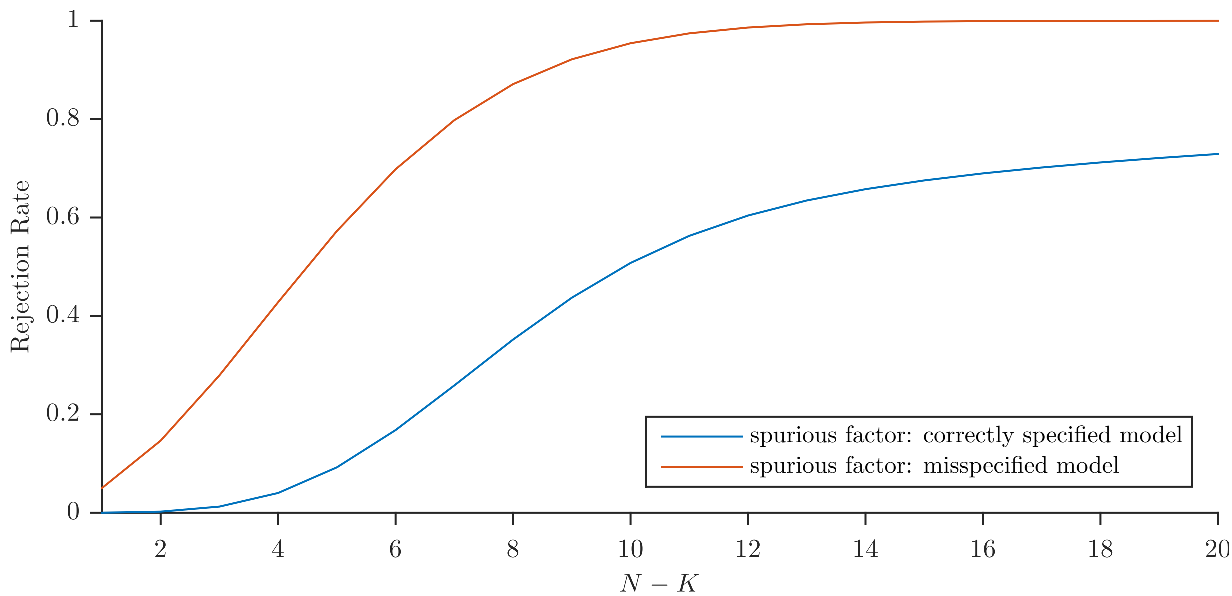


Figure 6. Limiting Rejection Rates of *t*-tests of Statistical Significance for CU-GMM. The top graph plots the limiting rejection rates under misspecified models of $t(\hat{\lambda}_0)$ (with $\mu_{f,K-1} = 0$), $t(\hat{\lambda}_{1,i})$, and $t(\hat{\lambda}_{1,K-1})$ as functions of $N - K$ when one uses the standard normal critical values. The bottom graph plots the limiting rejection rates under correctly specified and misspecified models of $t(\hat{\lambda}_{1,K-1})$ as functions of $N - K$ when one uses the standard normal critical values.

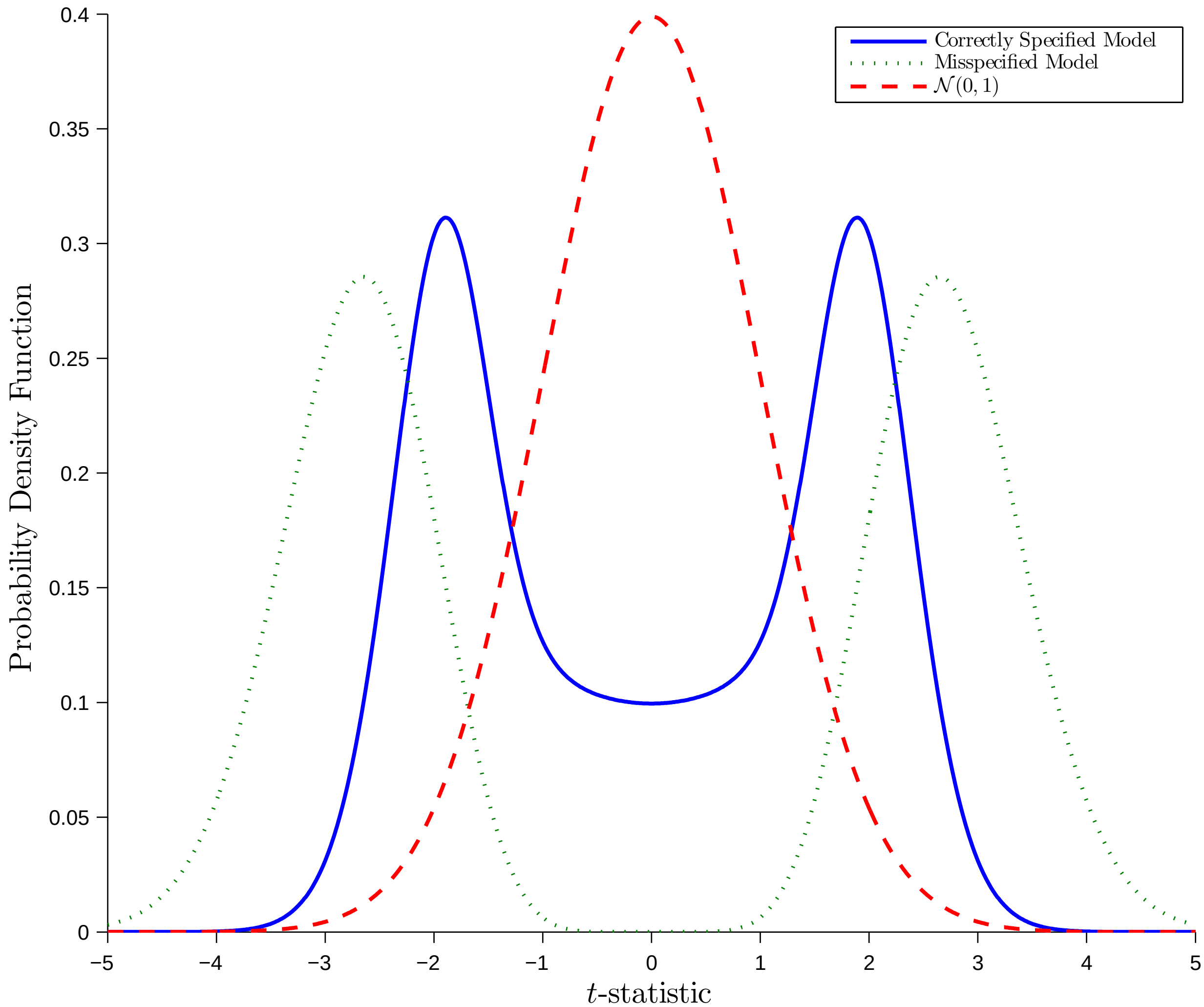


Figure 7. Limiting Distributions of $t(\hat{\lambda}_{1,K-1})$ under Correctly Specified and Misspecified Models. The figure plots the limiting densities of $t(\hat{\lambda}_{1,K-1})$ for correctly specified and misspecified models that contain a spurious factor (for $N - K = 7$), along with the standard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

Online Appendix to "Too Good to Be True? Fallacies in Evaluating Risk Factor Models"

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

This Online Appendix is structured as follows. Appendix A contains proofs of the theorems. Appendix B provides details on the CU-GMM estimation of the beta-pricing model. We refer the readers to the paper for the notation used here.

Appendix A: Proofs of Theorems

A.1 Auxiliary Lemma 1

Auxiliary Lemma 1. *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* $\sigma^2_{f,K-1} = \mathrm{Var}[f_{K-1,t}]$. *Assume that* Y_t *is iid normally distributed. Suppose that the model is misspecified and it contains a useless factor (that is,* $\mathrm{rank}(B) = K-1$*). Then,* $T \to \infty$*, we have (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)* $\frac{\hat{\gamma}^{ML}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} \frac{1}{z_2}$.

Proof. 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

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

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

$$\hat{\psi}_i = -\frac{\hat{v}_i}{\hat{v}_{K+1}}, \quad i = 1, \ldots, K, \tag{A.2}$$

which is asymptotically equivalent to the estimator

$$\tilde{\psi} = (\hat{G}_1'\hat{\Sigma}^{-1}\hat{G}_1)^{-1}(\hat{G}_1'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}). \tag{A.3}$$

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

$$\sqrt{T}\tilde{\psi} \xrightarrow{d} \mathcal{N}(0_K, (G_1'\Sigma^{-1}G_1)^{-1}/\sigma^2_{f,K_1}), \tag{A.4}$$

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

$$\hat{\gamma}_0^{ML} = -\frac{\sqrt{T}\hat{\psi}_1}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} -\frac{z_1}{z_2}, \tag{A.5}$$

$$\hat{\gamma}_{1,i}^{ML} = \hat{\mu}_{f,i} - \frac{\sqrt{T}\hat{\psi}_{i+2}}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} \mu_{f,i} - \frac{z_{i+2}}{z_2}, \quad i = 1, \ldots, K-2, \tag{A.6}$$

$$\frac{\hat{\gamma}^{ML}_{1,K-1}}{\sqrt{T}} = \frac{\hat{\mu}_{f,K-1}}{\sqrt{T}} + \frac{1}{\sqrt{T}\hat{\psi}_2} \xrightarrow{d} \frac{1}{z_2}. \tag{A.7}$$

This completes the proof of the lemma.

A.2 Proof of Theorem 1

part (a): Let $\sigma_i^2 = \text{Var}[z_i]$, $\sigma_{ij} \equiv \text{Cov}[z_i, z_j]$, $\rho_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$, $G_2 = [1_N,\ \beta_1, \ldots, \beta_{K-2}]$, $\hat{G}_2 = [1_N,\ \hat{\beta}_1, \ldots, \hat{\beta}_{K-2}]$, 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 $i = 1, \ldots, K-1$. We start with the squared t-ratio of the useless factor, $t^2(\hat{\gamma}_{1,K-1}^{ML})$. Using the formula for the inverse of a partitioned matrix, we obtain

$$\begin{aligned}
s^2(\hat{\gamma}_{1,K-1}^{ML}) &= (1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML})\left(\hat{\beta}_{K-1}'[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_{K-1}\right)^{-1} + \hat{\sigma}_{f,K-1}^2 \\
&= \left(\frac{\hat{\gamma}_{1,K-1}^{ML}}{\hat{\sigma}_{f,K-1}}\right)^2 \left(\hat{\beta}_{K-1}'[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}]\hat{\beta}_{K-1}\right)^{-1} + O_p(T^{\frac{1}{2}}) \qquad \text{(A.8)}
\end{aligned}$$

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:

$$\sqrt{T}\hat{\sigma}_{f,K-1}\hat{\Sigma}^{-\frac{1}{2}}\hat{\beta}_{K-1} \xrightarrow{d} u \sim \mathcal{N}(0_N, I_N), \qquad \text{(A.9)}$$

we obtain

$$\begin{aligned}
t^2(\hat{\gamma}_{1,K-1}^{ML}) &= \frac{T(\hat{\gamma}_{1,K-1}^{ML})^2\hat{\beta}_{K-1}'[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2\hat{\Sigma}^{-1}]\hat{\beta}_{K-1}}{(\hat{\gamma}_{1,K-1}^{ML}/\hat{\sigma}_{f,K-1})^2} + O_p(T^{-\frac{1}{2}}) \\
&= u'[I_N - \hat{\Sigma}^{-\frac{1}{2}}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-\frac{1}{2}}]u + O_p(T^{-\frac{1}{2}}) \\
&\xrightarrow{d} u'[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u \sim \chi^2_{N-K+1}. \qquad \text{(A.10)}
\end{aligned}$$

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

$$\begin{aligned}
&(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1} + \frac{(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}}{\hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{\beta}_{K-1} - \hat{\beta}_{K-1}'\hat{\Sigma}^{-1}\hat{G}_2(\hat{G}_2'\hat{\Sigma}^{-1}\hat{G}_2)^{-1}\hat{G}_2'\hat{\Sigma}^{-1}\hat{\beta}_{K-1}} \\
&= (G_2'\Sigma^{-1}G_2)^{-1} + \frac{(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}uu'\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}}{u'[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u} + O_p(T^{-\frac{1}{2}}). \qquad \text{(A.11)}
\end{aligned}$$

Note that we can write

$$I_N - \Sigma^{-\frac{1}{2}}G_1(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}} = I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}} - hh', \qquad \text{(A.12)}$$

where

$$h = \frac{[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha}{\left(\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha\right)^{\frac{1}{2}}}. \qquad \text{(A.13)}$$

With this expression, we can write

$$u'[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u = u'[I_N - \Sigma^{-\frac{1}{2}}G_1(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}}]u + (h'u)^2$$
$$= x + \tilde{z}_2^2, \tag{A.14}$$

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 $\sigma_{i,j} \equiv \mathrm{Cov}[z_i, z_j] = \boldsymbol{\iota}_{K,i}'(G_1'\Sigma^{-1}G_1)^{-1}\boldsymbol{\iota}_{K,j}/\sigma^2_{f,K-1}$. Using the formula for the inverse of a partitioned matrix, we obtain

$$\begin{aligned} z_2 &= \frac{1}{\sigma_{f,K-1}}\boldsymbol{\iota}_{K,2}'(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-\frac{1}{2}}u \\ &= \frac{1}{\sigma_{f,K-1}}\frac{\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]u}{\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha}. \end{aligned} \tag{A.15}$$

It follows that

$$\sigma_2^2 = \frac{1}{\sigma^2_{f,K-1}\alpha'\Sigma^{-\frac{1}{2}}[I_N - \Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}]\Sigma^{-\frac{1}{2}}\alpha} \tag{A.16}$$

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.11), we have

$$\begin{aligned} w_i &\stackrel{d}{\to} \boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i} + \frac{\boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}uu'\Sigma^{-\frac{1}{2}}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}}{x + \tilde{z}_2^2} \\ &= \boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}\left(1 + \frac{q_i^2}{x + \tilde{z}_2^2}\right), \end{aligned} \tag{A.17}$$

where

$$q_i = \frac{\boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}G_2'\Sigma^{-\frac{1}{2}}u}{[\boldsymbol{\iota}_{K-1,i}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i}]^{\frac{1}{2}}} \sim \mathcal{N}(0,1). \tag{A.18}$$

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

$$(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-1}G_2 = [\boldsymbol{\iota}_{K,1},\ \boldsymbol{\iota}_{K,3}, \ldots, \boldsymbol{\iota}_{K,K}], \tag{A.19}$$

it is straightforward to show that

$$\begin{aligned} \mathrm{Cov}[z_1, q_1] &= \frac{\boldsymbol{\iota}_{K,1}'(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-1}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}}{\sigma_{f,K-1}[\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} \\ &= [\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma^2_{f,K-1}]^{\frac{1}{2}}, \end{aligned} \tag{A.20}$$

$$\mathrm{Cov}[z_2, q_1] = \frac{\boldsymbol{\iota}_{K,2}'(G_1'\Sigma^{-1}G_1)^{-1}G_1'\Sigma^{-1}G_2(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}}{\sigma_{f,K-1}[\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}]^{\frac{1}{2}}} = 0. \tag{A.21}$$

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

$$\frac{1}{\sigma_{f,K-1}^2}\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1} = \sigma_1^2 - \frac{\sigma_{1,2}^2}{\sigma_2^2} = \sigma_1^2(1-\rho_{1,2}^2). \tag{A.22}$$

It follows that

$$\text{Cov}\left[z_1 - \frac{\sigma_{12}}{\sigma_2^2}z_2, q_1\right] = [\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma_{f,K-1}^2]^{\frac{1}{2}} = \sigma_1\sqrt{1-\rho_{1,2}^2}. \tag{A.23}$$

Therefore, $z_1 - (\sigma_{1,2}/\sigma_2^2)z_2$ is perfectly correlated with q_1 and we can write

$$z_1 = \frac{\sigma_{1,2}}{\sigma_2^2}z_2 + \sqrt{1-\rho_{1,2}^2}\sigma_1 q_1 = \sigma_1\left(\rho_{1,2}\tilde{z}_2 + \sqrt{1-\rho_{1,2}^2}q_1\right). \tag{A.24}$$

Similarly,

$$z_{i+1} = \frac{\sigma_{i+1,2}}{\sigma_2^2}z_2 + \sqrt{1-\rho_{i+1,2}^2}\sigma_{i+1}q_i = \sigma_{i+1}\left(\rho_{i+1,2}\tilde{z}_2 + \sqrt{1-\rho_{i+1,2}^2}q_i\right), \quad i = 2,\dots,K-1. \tag{A.25}$$

Let

$$b_i = \frac{x+\tilde{z}_2^2}{x+\tilde{z}_2^2+q_i^2}, \quad i = 1,\dots,K-1. \tag{A.26}$$

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

$$\begin{aligned} t(\hat{\gamma}_0^{ML}) &\xrightarrow{d} -\frac{z_1|z_2|b_1^{\frac{1}{2}}}{z_2[\boldsymbol{\iota}_{K-1,1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,1}/\sigma_{f,K-1}^2]^{\frac{1}{2}}} \\ &= -\left(\frac{\rho_{1,2}|\tilde{z}_2|}{\sqrt{1-\rho_{1,2}^2}} + q_1\right)b_1^{\frac{1}{2}}, \end{aligned} \tag{A.27}$$

$$\begin{aligned} t(\hat{\gamma}_{1,i}^{ML}) &\xrightarrow{d} \frac{\left(\mu_{f,i} - \frac{z_{i+2}}{z_2}\right)|z_2|b_{i+1}^{\frac{1}{2}}}{[\boldsymbol{\iota}_{K-1,i+1}'(G_2'\Sigma^{-1}G_2)^{-1}\boldsymbol{\iota}_{K-1,i+1}/\sigma_{f,K-1}^2]^{\frac{1}{2}}} \\ &= \left(\frac{\frac{\mu_{f,i}\sigma_2}{\sigma_{i+2}} - \rho_{i+2,2}}{\sqrt{1-\rho_{i+2,2}^2}}|\tilde{z}_2| - q_{i+1}\right)b_{i+1}^{\frac{1}{2}}, \quad i = 1,\dots,K-2. \end{aligned} \tag{A.28}$$

Defining $\bar{Z}_0 = -\left(\frac{\rho_{1,2}|\tilde{z}_2|}{\sqrt{1-\rho_{1,2}^2}} + q_1\right)b_1^{\frac{1}{2}}$ and $\bar{Z}_i = \left(\frac{\frac{\mu_{f,i}\sigma_2}{\sigma_{i+2}} - \rho_{i+2,2}}{\sqrt{1-\rho_{i+2,2}^2}}|\tilde{z}_2| - q_{i+1}\right)b_{i+1}^{\frac{1}{2}}$ for $i = 1,\dots,K-2$, delivers the desired result. This completes the proof of part (a).

part (b): 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

$$
\begin{aligned}
\hat{\mu}_R^{ML} &= 1_N\hat{\gamma}_0^{ML} + \hat{\beta}^{ML}\hat{\gamma}_1^{ML} \\
&= 1_N\hat{\gamma}_0^{ML} + \hat{\beta}\hat{\gamma}_1^{ML} + \hat{e}\frac{\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}{1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}} \\
&= \hat{\mu}_R - \hat{e} + \hat{e}\frac{\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}{1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}} \\
&= \hat{\mu}_R - \hat{e}\frac{1}{1+\hat{\gamma}_1^{ML\prime}\hat{V}_f^{-1}\hat{\gamma}_1^{ML}}.
\end{aligned}
\tag{A.29}
$$

Using the result from Auxiliary Lemma 1 that $\hat{\gamma}_{1,i}^{ML} = O_p(1)$ for $i = 1,\dots,K-2$ 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

$$
R^2_{ML} = \mathrm{Corr}(\hat{\mu}_R^{ML}, \hat{\mu}_R)^2 \xrightarrow{p} 1 \tag{A.30}
$$

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

A.3 Auxiliary Lemma 2

Auxiliary Lemma 2. *Let $z = [z_1, z_2, \dots, z_K]' \sim \mathcal{N}(0_K, \sigma^2_{f,K-1}(H_1'U^{-1}H_1)^{-1})$, where $H_1 = [1_N,\ D_1]$ and $D_1 = [d_1,\ d_2, \dots, d_{K-1}]$ with d_i $(i = 1,\dots,K)$ being the i-th column of D. Assume that Y_t is a jointly stationary and ergodic process with a finite fourth moment, $\{\mathrm{vec}(D_t - D) : t \geq 1\}$ is a martingale difference sequence. Suppose that the model is misspecified and it contains a useless factor (that is, $\mathrm{rank}(D) = K-1$). Then, $T \to \infty$, we have (i) $\hat{\lambda}_0 \xrightarrow{d} -\frac{z_2}{z_1}$ if $\mu_{f,K-1} = 0$ or $\frac{\hat{\lambda}_0}{\sqrt{T}} \xrightarrow{d} \frac{\mu_{f,K-1}}{z_1}$ if $\mu_{f,K-1} \neq 0$; (ii) $\hat{\lambda}_{1,i} \xrightarrow{d} -\frac{z_{i+2}}{z_1}$ for $i = 1,\dots,K-2$; and (iii) $\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{z_1}$.*

Proof. We first perform the following parameterization of the problem. Let

$$
g_t(v) = H_t \begin{bmatrix} v \\ 1 \end{bmatrix}. \tag{A.31}
$$

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

$$
v^* = \begin{bmatrix} 0 \\ -\mu_{f,K-1} \\ 0_{K-2} \end{bmatrix}. \tag{A.32}
$$

Consider the CU-GMM estimator of v^*:

$$
\hat{v} = \mathrm{argmin}_v \bar{g}(v)'\hat{W}_g(v)^{-1}\bar{g}(v), \tag{A.33}
$$

where $\bar{g}(v) = \sum_{t=1}^{T} g_t(v)/T$ and $\hat{W}_g(v) = \frac{1}{T}\sum_{t=1}^{T}[g_t(v) - \bar{g}(v)][g_t(v) - \bar{g}(v)]'$. The asymptotic distribution of $\hat{v}$ is given by

$$\sqrt{T}(\hat{v} - v^*) \xrightarrow{d} \mathcal{N}\left(0_K, (H_1' S_g^{-1} H_1)^{-1}\right), \tag{A.34}$$

where

$$S_g = E[g_t(v^*)g_t(v^*)'] = E[R_t R_t'(f_{K-1,t} - \mu_{f,K-1})^2] = U\sigma_{f,K-1}^2, \tag{A.35}$$

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

$$\check{v} = (\hat{H}_1'\hat{U}^{-1}\hat{H}_1)^{-1}\hat{H}_1'\hat{U}^{-1}\hat{d}_K. \tag{A.36}$$

Let

$$z \sim \mathcal{N}(0_K, \sigma_{f,K-1}^2(H_1'U^{-1}H_1)^{-1}). \tag{A.37}$$

Then, we have

$$\sqrt{T}\hat{v}_1 \xrightarrow{d} z_1, \tag{A.38}$$

$$\sqrt{T}(\hat{v}_2 + \mu_{f,K-1}) \xrightarrow{d} z_2, \tag{A.39}$$

$$\sqrt{T}\hat{v}_i \xrightarrow{d} z_i, \quad i = 3, \dots, K. \tag{A.40}$$

Due to the invariance property of CU-GMM, we know that $[-1,\ \hat{\lambda}']$ is proportional to $[\hat{v}',\ 1]$. Then, we have $\hat{\lambda}_0 = -\frac{\hat{v}_2}{\hat{v}_1}$, $\hat{\lambda}_{1,i} = -\frac{\hat{v}_{i+2}}{\hat{v}_1}$ for $i = 1, \dots, K-2$, and $\hat{\lambda}_{1,K-1} = -\frac{1}{\hat{v}_1}$. Therefore, the limiting distributions of the $K-1$ elements of $\hat{\lambda}_1$ are given by

$$\hat{\lambda}_{1,i} \xrightarrow{d} -\frac{z_{i+2}}{z_1}, \quad i = 1, \dots, K-2, \tag{A.41}$$

$$\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{z_1}. \tag{A.42}$$

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 the lemma.

A.4 Proof of Theorem 2

part (a): It is easy to show that

$$\sigma_1^2 = \frac{\sigma_{f,K-1}^2}{1_N'[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]1_N} = \frac{\sigma_{f,K-1}^2}{\delta^2}, \tag{A.43}$$

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

$$\frac{\hat{\lambda}_{1,K-1}}{\sqrt{T}} \xrightarrow{d} -\frac{1}{\sigma_1\tilde{z}_1} = -\frac{\delta}{\sigma_{f,K-1}\tilde{z}_1}, \tag{A.44}$$

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

$$\frac{e_t(\hat{\lambda})}{\sqrt{T}} = -\frac{R_t(f_{K-1,t} - \mu_{f,K-1})}{z_1} + O_p(T^{-\frac{1}{2}}), \tag{A.45}$$

we can show that

$$\frac{\hat{W}_e(\hat{\lambda})}{T} = \frac{\sigma^2_{f,K-1}}{z_1^2} U + o_p(1). \tag{A.46}$$

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

$$\begin{aligned} t^2(\hat{\lambda}_{1,K-1}) &= \frac{T\hat{\lambda}^2_{1,K-1}}{\iota'_{K,K}(\hat{D}'\hat{W}_e(\hat{\lambda})^{-1}\hat{D})^{-1}\iota_{K,K}} \\ &= \frac{T\hat{d}'_K[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]\hat{d}_K}{\sigma^2_{f,K-1}} + o_p(1), \end{aligned} \tag{A.47}$$

where $\iota_{K,K}$ is a generic $K \times 1$ selector vector with one for its K-th element and zero otherwise. Let

$$\tilde{d}_K = \hat{d}_K - \hat{\mu}_R \mu_{f,K-1} = \frac{1}{T}\sum_{t=1}^{T} R_t(f_{K-1,t} - \mu_{f,K-1}). \tag{A.48}$$

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

$$\begin{aligned} &T\hat{d}'_K[U^{-1} - U^{-1}\hat{D}_1(\hat{D}_1'U^{-1}\hat{D}_1)^{-1}\hat{D}_1'U^{-1}]\hat{d}_K = T\tilde{d}'_K[U^{-1} - U^{-1}\hat{D}_1(\hat{D}_1'U^{-1}\hat{D}_1)^{-1}\hat{D}_1'U^{-1}]\tilde{d}_K \\ &= T\tilde{d}'_K[U^{-1} - U^{-1}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}]\tilde{d}_K + o_p(1). \end{aligned} \tag{A.49}$$

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

$$\frac{1}{\sigma_{f,K-1}}\sqrt{T}P_U'U^{-\frac{1}{2}}\tilde{d}_K \xrightarrow{d} \mathcal{N}(0_{N-K+1}, I_{N-K+1}) \tag{A.50}$$

and

$$t^2(\hat{\lambda}_{1,K-1}) \xrightarrow{d} \chi^2_{N-K+1}. \tag{A.51}$$

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

$$I_N - U^{-\frac{1}{2}}H_1(H_1'U^{-1}H_1)^{-1}H_1'U^{-\frac{1}{2}} = I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}} - hh', \tag{A.52}$$

where

$$h = \frac{[I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}1_N}{\delta} = \frac{P_UP_U'U^{-\frac{1}{2}}1_N}{\delta} \tag{A.53}$$

and $h'h = 1$. Note that

$$\sqrt{T}h'U^{-\frac{1}{2}}\tilde{d}_K/\sigma_{f,K-1} = \sqrt{T}1_N'U^{-\frac{1}{2}}P_UP_U'U^{-\frac{1}{2}}\tilde{d}_K/(\sigma_{f,K-1}\delta) \xrightarrow{d} \tilde{z}_1 \sim \mathcal{N}(0,1), \tag{A.54}$$

$$T\tilde{d}'_K[I_N - U^{-\frac{1}{2}}H_1(H_1'U^{-1}H_1)^{-1}H_1'U^{-\frac{1}{2}}]\tilde{d}_K/\sigma^2_{f,K-1} \xrightarrow{d} x \sim \chi^2_{N-K}, \tag{A.55}$$

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

$$\sigma_{f,K-1}^2 \iota_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\iota_{K-1,i} = \sigma_{i+1}^2 - \frac{\sigma_{1,i+1}}{\sigma_1^2} = \sigma_{i+1}^2(1-\rho_{1,i+1}^2), \tag{A.56}$$

where $\sigma_i^2 = \mathrm{Var}[z_i]$, $\sigma_{i,j} \equiv \mathrm{Cov}[z_i, z_j]$, and $\rho_{i,j} = \sigma_{i,j}/(\sigma_i\sigma_j)$. In addition, we can easily show that for $i = 2, \ldots, K-1$

$$\frac{\sqrt{T}\iota_{K-1,i}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K}{\sigma_{f,K-1}[\iota_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\iota_{K-1,i}]^{\frac{1}{2}}} \stackrel{d}{\to} q_{i+1} \sim \mathcal{N}(0,1). \tag{A.57}$$

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

$$\frac{\sqrt{T}\iota_{K-1,1}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K}{\sigma_{f,K-1}[\iota_{K-1,1}'(D_1'U^{-1}D_1)^{-1}\iota_{K-1,1}]^{\frac{1}{2}}} \stackrel{d}{\to} q_2 \sim \mathcal{N}(0,1). \tag{A.58}$$

Note that the q_i's are independent of $\tilde{z}_1$ and x. If $\mu_{f,K-1} \neq 0$, we have $\hat{d}_K \stackrel{p}{\to} \mu_R\mu_{f,K-1} = D_1[\mu_{f,K-1},\ 0_{K-2}']'$ and hence

$$\begin{aligned}\iota_{K-1,1}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K/\sigma_{f,K-1} &\stackrel{p}{\to} \iota_{K-1,1}'(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}D_1[\mu_{f,K-1},\ 0_{K-2}']'/\sigma_{f,K-1} \\ &= \mu_{f,K-1}/\sigma_{f,K-1}. \end{aligned} \tag{A.59}$$

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

$$\frac{\sigma_{f,K-1}^2}{z_1^2}\left[(D_1'U^{-1}D_1)^{-1} + \frac{(D_1'U^{-1}D_1)^{-1}D_1'U^{-1}\hat{d}_K\hat{d}_K'U^{-1}D_1(D_1'U^{-1}D_1)^{-1}}{\tilde{d}_K'U^{-\frac{1}{2}}[I_N - U^{-\frac{1}{2}}D_1(D_1'U^{-1}D_1)^{-1}D_1'U^{-\frac{1}{2}}]U^{-\frac{1}{2}}\tilde{d}_K}\right]. \tag{A.60}$$

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

$$\frac{\sigma_{f,K-1}^2\iota_{K-1,i}'(D_1'U^{-1}D_1)^{-1}\iota_{K-1,i}}{z_1^2}\left(1 + \frac{q_{i+1}^2}{x+\tilde{z}_1^2}\right) = \frac{\sigma_{i+1}^2(1-\rho_{1,i+1}^2)}{z_1^2}\left(1 + \frac{q_{i+1}^2}{x+\tilde{z}_1^2}\right). \tag{A.61}$$

Let

$$b_{i+1} = \frac{x+\tilde{z}_1^2}{q_{i+1}^2 + x + \tilde{z}_1^2} \tag{A.62}$$

and note that, using the same arguments as in the proof of part (b) in Theorem 1, we can write

$$z_{i+1} = \sigma_{i+1}\left(\rho_{1,i+1}\tilde{z}_1 + \sqrt{1-\rho_{1,i+1}^2}q_{i+1}\right). \tag{A.63}$$

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

$$t(\hat{\lambda}_{1,i}) \xrightarrow{d} -\frac{z_{i+2}/z_1}{\sigma_{i+2}\sqrt{1-\rho_{1,i+2}^2}|z_1|b_{i+1}^{-\frac{1}{2}}} = -\frac{|\tilde{z}_1|}{\tilde{z}_1}\frac{\left(\rho_{1,i+2}\tilde{z}_1 + \sqrt{1-\rho_{1,i+2}^2}q_{i+2}\right)b_{i+1}^{\frac{1}{2}}}{\sqrt{1-\rho_{1,i+2}^2}}$$
$$= -\left(\frac{\rho_{1,i+2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,i+2}^2}} + q_{i+2}\right)b_{i+1}^{\frac{1}{2}}. \tag{A.64}$$

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

$$t(\hat{\lambda}_0) \xrightarrow{d} -\left(\frac{\rho_{1,2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,2}^2}} + q_2\right)b_1^{\frac{1}{2}}. \tag{A.65}$$

Defining $\tilde{Z}_i = -\left(\frac{\rho_{1,i+2}|\tilde{z}_1|}{\sqrt{1-\rho_{1,i+2}^2}} + q_{i+2}\right)b_{i+1}^{\frac{1}{2}}$ for $i = 0, \ldots, K-2$, delivers the desired result. If $\mu_{f,K-1} \neq 0$, we have

$$t^2(\hat{\lambda}_0) \xrightarrow{d} \frac{\frac{\mu_{f,K-1}^2}{z_1^2}}{\frac{\sigma_{f,K-1}^2}{z_1^2}\left[\frac{\mu_{f,K-1}^2}{\sigma_{f,K-1}^2(x+\tilde{z}_1^2)}\right]} = x + \tilde{z}_1^2 \sim \chi_{N-K+1}^2. \tag{A.66}$$

This completes the proof of part (a).

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

Appendix B: CU-GMM Estimation of the Beta-Pricing Model

Let $\phi = [\gamma_0,\ \gamma_1',\ \beta_1', \ldots, \beta_K',\ \mu_f',\ \mathrm{vech}(V_f)']'$ denote the vector of parameters of interest, where β_i is an $N \times 1$ vector. In addition, let

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

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

$$\hat{W}_g(\phi) = \frac{1}{T}\sum_{t=1}^T (g_t(\phi) - \bar{g}(\phi))(g_t(\phi) - \bar{g}(\phi))'. \tag{B.2}$$

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

$$\hat{\phi} = \text{argmin}_{\phi}\ \bar{g}(\phi)'\hat{W}_g(\phi)^{-1}\bar{g}(\phi). \tag{B.3}$$

The problem with implementing this CU-GMM estimator is that the parameter vector ϕ is high dimensional especially when the number of test assets N is large. Peñaranda and Sentana (2015) show that CU-GMM delivers numerically identical estimates in the beta-pricing and linear SDF setups.[1] By augmenting $E[e_t(\lambda)]$ 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 $\theta = [\lambda_0,\ \lambda_1',\ \beta_1', \ldots, \beta_K',\ \mu_f',\ \text{vech}(V_f)']'$ 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 lemma presents a general result for this sequential estimation.

Lemma B.1. *Let $\theta = [\theta_1',\ \theta_2']'$, where θ_1 is $K_1 \times 1$ and θ_2 is $K_2 \times 1$, and*

$$E[g_t(\theta)] = \begin{bmatrix} E[g_{1t}(\theta_1)] \\ E[g_{2t}(\theta)] \end{bmatrix} = \begin{bmatrix} 0_{N_1} \\ 0_{N_2} \end{bmatrix}, \tag{B.4}$$

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

$$\tilde{\theta}_1 = \text{argmin}_{\theta_1}\ \bar{g}_1(\theta_1)'\hat{W}_{11}(\theta_1)^{-1}\bar{g}_1(\theta_1), \tag{B.5}$$

$$\hat{\theta} \equiv \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \text{argmin}_{\theta}\ \bar{g}(\theta)'\hat{W}(\theta)^{-1}\bar{g}(\theta), \tag{B.6}$$

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 $\hat{W}(\theta) = \frac{1}{T}\sum_{t=1}^{T}(g_t(\theta) - \bar{g}(\theta))(g_t(\theta) - \bar{g}(\theta))'$. Then, $\tilde{\theta}_1 = \hat{\theta}_1$.

Proof. Let

$$\tilde{D}_{11}(\theta_1) = \frac{1}{T}\sum_{t=1}^{T} \tilde{w}_t(\theta_1)\frac{\partial g_{1t}(\theta_1)}{\partial \theta_1'}, \tag{B.7}$$

where

$$\tilde{w}_t(\theta_1) = 1 - \bar{g}_1(\theta_1)'\hat{W}_{11}(\theta_1)^{-1}[g_{1t}(\theta_1) - \bar{g}_1(\theta_1)]. \tag{B.8}$$

The first-order conditions for the smaller system are given by

$$\tilde{D}_{11}(\tilde{\theta}_1)'\hat{W}_{11}(\tilde{\theta}_1)^{-1}\bar{g}_1(\tilde{\theta}_1) = 0_{N_1}. \tag{B.9}$$

[1] 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 ML estimator.

Similarly, we define

$$\hat{D}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \hat{w}_t(\theta)\frac{\partial g_t(\theta)}{\partial \theta'} \equiv \begin{bmatrix} \hat{D}_{11}(\theta) & 0_{N_1\times N_2} \\ \hat{D}_{21}(\theta) & \hat{D}_{22}(\theta) \end{bmatrix}, \tag{B.10}$$

where

$$\hat{w}_t(\theta) = 1 - \bar{g}(\theta)'\hat{W}(\theta)^{-1}[g_t(\theta) - \bar{g}(\theta)]. \tag{B.11}$$

The first-order conditions for the larger system are given by

$$\hat{D}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\bar{g}(\hat{\theta}) = 0_{N_1+N_2}. \tag{B.12}$$

Let

$$\hat{W}(\theta)^{-1} = \begin{bmatrix} \hat{W}^{11}(\theta) & \hat{W}^{12}(\theta) \\ \hat{W}^{21}(\theta) & \hat{W}^{22}(\theta) \end{bmatrix}. \tag{B.13}$$

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

$$\begin{aligned} 0_{N_1+N_2} &= \hat{D}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\bar{g}(\hat{\theta}) \\ &= \begin{bmatrix} (\hat{D}_{11}'\hat{W}^{11} + \hat{D}_{21}'\hat{W}^{21})\bar{g}_1(\hat{\theta}_1) + (\hat{D}_{11}'\hat{W}^{12} + \hat{D}_{21}'\hat{W}^{22})\bar{g}_2(\hat{\theta}) \\ \hat{D}_{22}'\hat{W}^{21}\bar{g}_1(\hat{\theta}_1) + \hat{D}_{22}'\hat{W}^{22}\bar{g}_2(\hat{\theta}) \end{bmatrix}. \end{aligned} \tag{B.14}$$

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

$$\bar{g}_2(\hat{\theta}) = -(\hat{W}^{22})^{-1}\hat{W}^{21}\bar{g}_1(\hat{\theta}_1). \tag{B.15}$$

Plugging this equation into the first subset of first-order conditions, we obtain

$$\begin{aligned} 0_{N_1} &= (\hat{D}_{11}'\hat{W}^{11} + \hat{D}_{21}'\hat{W}^{21})\bar{g}_1(\hat{\theta}_1) - (\hat{D}_{11}'\hat{W}^{12} + \hat{D}_{21}'\hat{W}^{22})(\hat{W}^{22})^{-1}\hat{W}^{21}\bar{g}_1(\hat{\theta}_1) \\ &= \hat{D}_{11}(\hat{\theta}_1)'\hat{W}_{11}(\hat{\theta}_1)^{-1}\bar{g}_1(\hat{\theta}_1), \end{aligned} \tag{B.16}$$

where the last identity is obtained by using the partitioned matrix inverse formula, which implies that

$$\hat{W}_{11}(\theta_1)^{-1} = \hat{W}^{11}(\theta) - \hat{W}^{12}(\theta)\hat{W}^{22}(\theta)^{-1}\hat{W}^{21}(\theta). \tag{B.17}$$

In addition, defining $\bar{g}_2(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\theta)$ and using (B.15), we have

$$\begin{aligned} \hat{w}_t(\hat{\theta}) &= 1 - \bar{g}'(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}\begin{bmatrix} g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta}) - \bar{g}_2(\hat{\theta}) \end{bmatrix} \\ &= 1 - [\bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{21}(\hat{\theta}),\ \ \bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta}) + \bar{g}_2(\hat{\theta})'\hat{W}^{22}(\hat{\theta})]\begin{bmatrix} g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta}) - \bar{g}_2(\hat{\theta}) \end{bmatrix} \\ &= 1 - \bar{g}_1(\hat{\theta}_1)'\hat{W}_{11}(\hat{\theta}_1)^{-1}[g_{1t}(\hat{\theta}_1) - \bar{g}_1(\hat{\theta}_1)] \\ &= \tilde{w}_t(\hat{\theta}_1), \end{aligned} \tag{B.18}$$

which only depends on $\hat{\theta}_1$. Therefore, we have $\hat{D}_{11}(\hat{\theta}_1) = \tilde{D}_{11}(\hat{\theta}_1)$ and (B.16) is identical to the first-order conditions for the smaller system. It follows that $\hat{\theta}_1 = \tilde{\theta}_1$. This completes the proof of Lemma B.1.

Lemma B.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. This numerical equivalence can also be shown for the corresponding tests for over-identifying restrictions. The result in Lemma B.1 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 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 B.2. *Let*

$$r_t(\hat{\theta}) = \bar{g}(\hat{\theta})'\hat{W}(\hat{\theta})^{-1}[g_t(\hat{\theta}) - \bar{g}(\hat{\theta})] \tag{B.19}$$

and

$$r_{1t}(\tilde{\theta}_1) = \bar{g}_1(\tilde{\theta}_1)'\hat{W}_{11}(\tilde{\theta}_1)^{-1}[g_{1t}(\tilde{\theta}_1) - \bar{g}_1(\tilde{\theta}_1)]. \tag{B.20}$$

The estimate $\hat{\theta}_2$ is given by the solution to

$$\frac{1}{T}\sum_{t=1}^{T} g_{2t}(\tilde{\theta}_1, \hat{\theta}_2)[1 - r_{1t}(\tilde{\theta}_1)] = 0_{K_2} \tag{B.21}$$

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,

$$g_{2t}(\theta_1, \theta_2) = h_{1t}(\theta_1) - h_{2t}(\theta_1)\theta_2, \tag{B.22}$$

where h_{1t} and h_{2t} are functions of the data and θ_1, then

$$\hat{\theta}_2 = \left(\sum_{t=1}^{T} h_{2t}(\tilde{\theta}_1)[1 - r_{1t}(\tilde{\theta}_1)]\right)^{-1} \sum_{t=1}^{T} h_{1t}(\tilde{\theta}_1)[1 - r_{1t}(\tilde{\theta}_1)]. \tag{B.23}$$

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 (B.15) and noting that $\hat{\theta}_1 = \tilde{\theta}_1$, we obtain

$$\bar{g}_2(\tilde{\theta}_1, \hat{\theta}_2) = \hat{W}_{21}(\tilde{\theta}_1, \hat{\theta}_2)\hat{W}_{11}(\tilde{\theta}_1)^{-1}\bar{g}_1(\tilde{\theta}_1). \tag{B.24}$$

This is a system of K_2 equations with K_2 unknowns. Using the expression for $r_{1t}(\tilde{\theta}_1)$, we can write

(B.24) as

$$\begin{aligned}\bar{g}_2(\tilde{\theta}_1,\hat{\theta}_2) &= \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\tilde{\theta}_1,\hat{\theta}_2)r_{1t}(\tilde{\theta}_1) \\ \Rightarrow\ 0_{K_2} &= \frac{1}{T}\sum_{t=1}^{T} g_{2t}(\tilde{\theta}_1,\hat{\theta}_2)[1-r_{1t}(\tilde{\theta}_1)]. \end{aligned} \tag{B.25}$$

For the larger system, we have

$$\begin{aligned} r_t(\hat{\theta}) &= \begin{bmatrix} \bar{g}_1(\hat{\theta}_1) \\ \bar{g}_2(\hat{\theta}) \end{bmatrix}' \begin{bmatrix} \hat{W}^{11}(\hat{\theta}) & \hat{W}^{12}(\hat{\theta}) \\ \hat{W}^{21}(\hat{\theta}) & \hat{W}^{22}(\hat{\theta}) \end{bmatrix} \begin{bmatrix} g_{1t}(\hat{\theta}_1)-\bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta})-\bar{g}_2(\hat{\theta}) \end{bmatrix} \\ &= \left[\bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta})+\bar{g}_2(\hat{\theta})'\hat{W}^{21}(\hat{\theta}),\ \bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta})+\bar{g}_2(\hat{\theta})'\hat{W}^{22}(\hat{\theta})\right] \begin{bmatrix} g_{1t}(\hat{\theta}_1)-\bar{g}_1(\hat{\theta}_1) \\ g_{2t}(\hat{\theta})-\bar{g}_2(\hat{\theta}) \end{bmatrix} \\ &= \bar{g}_1(\hat{\theta}_1)'\hat{W}^{11}(\hat{\theta})[g_{1t}(\hat{\theta}_1)-\bar{g}_1(\hat{\theta}_1)] - \bar{g}_1(\hat{\theta}_1)'\hat{W}^{12}(\hat{\theta})(\hat{W}^{22}(\hat{\theta}))^{-1}\hat{W}^{21}(\hat{\theta})[g_{1t}(\hat{\theta}_1)-\bar{g}_1(\hat{\theta}_1)] \\ &= \bar{g}_1(\hat{\theta}_1)'\hat{W}_{11}^{-1}(\hat{\theta}_1)[g_{1t}(\hat{\theta}_1)-\bar{g}_1(\hat{\theta}_1)] \\ &= r_{1t}(\tilde{\theta}_1), \end{aligned} \tag{B.26}$$

where the third equality follows from (B.15), 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 (B.25) and solving for $\hat{\theta}_2$. This completes the proof of Lemma B.2.

Lemma B.2 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

$$\hat{\theta}_2 = \frac{\sum_{t=1}^{T} h_{1t}(\tilde{\theta}_1)[1-r_{1t}(\tilde{\theta}_1)]}{\sum_{t=1}^{T}[1-r_{1t}(\tilde{\theta}_1)]}. \tag{B.27}$$

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.

Lemma B.1 and Lemma B.2 allow us to efficiently implement the CU-GMM estimation of the beta-pricing model. Let $g_t(\lambda) = R_t x_t'\lambda - 1_N = D_t\lambda - 1_N$, $\bar{g}(\lambda) = \frac{1}{T}\sum_{t=1}^{T} g_t(\lambda) = \hat{D}\lambda - 1_N$, and

$$\hat{W}_g(\lambda) = \frac{1}{T}\sum_{t=1}^{T}[g_t(\lambda)-\bar{g}(\lambda)][g_t(\lambda)-\bar{g}(\lambda)]'. \tag{B.28}$$

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

$$\hat{\lambda} = [\hat{\lambda}_0,\ \hat{\lambda}_1']' = \mathrm{argmin}_{\lambda}\bar{g}(\lambda)'\hat{W}_g(\lambda)^{-1}\bar{g}(\lambda). \tag{B.29}$$

Let

$$w_t(\hat{\lambda}) = \frac{1-(g_t(\hat{\lambda})-\bar{g}(\hat{\lambda}))'\hat{W}_g(\hat{\lambda})^{-1}\bar{g}(\hat{\lambda})}{T}. \tag{B.30}$$

The CU-GMM estimates of the parameters μ_f, V_f, and β can be obtained as $\hat{\mu}_f^{CU} = \sum_{t=1}^T w_t(\hat{\lambda}) f_t$, $\hat{V}_f^{CU} = \sum_{t=1}^T w_t(\hat{\lambda}) f_t(f_t - \hat{\mu}_f^{CU})'$, and $\hat{\beta}^{CU} = \sum_{t=1}^T w_t(\hat{\lambda}) R_t(f_t - \hat{\mu}_f^{CU})'(\hat{V}_f^{CU})^{-1}$. These estimates can be used to construct estimates of the zero-beta rate and risk premium parameters, $\hat{\gamma}_0 = \frac{1}{\hat{\lambda}_0 + \hat{\mu}_f^{CU\prime}\hat{\lambda}_1}$ and $\hat{\gamma}_1 = -\frac{\hat{V}_f^{CU}\hat{\lambda}_1}{\hat{\lambda}_0 + \hat{\mu}_f^{CU\prime}\hat{\lambda}_1}$, respectively. The asymptotic variances of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ can then be obtained by the delta method.