FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Asymptotic Variance Approximations for Invariant Estimators in Uncertain Asset-Pricing Models

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2015-9
October 2015

Abstract: This paper derives explicit expressions for the asymptotic variances of the maximum likelihood and continuously updated GMM estimators under potentially misspecified models. The proposed misspecification-robust variance estimators allow the researcher to conduct valid inference on the model parameters even when the model is rejected by the data. Although the results for the maximum likelihood estimator are only applicable to linear asset-pricing models, the asymptotic distribution of the continuously updated GMM estimator is derived for general, possibly nonlinear, models. The large corrections in the asymptotic variances, which arise from explicitly incorporating model misspecification in the analysis, are illustrated using simulations and an empirical application.

JEL classification: C12; C13; G12

Key words: asset pricing, model misspecification, continuously updated GMM, maximum likelihood, asymptotic approximation, misspecification-robust tests

---

Please address questions regarding content to Nikolay Gospodinov (corresponding author), Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street N.E., Atlanta, GA 30309, nikolay.gospodinov@atl.frb.org; Raymond Kan, Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, kan@chass.utoronto.ca; or Cesare Robotti, Imperial College Business School, South Kensington Campus, London, SW7 2AZ, United Kingdom, c.robotti@imperial.ac.uk.

1 Introduction

Given the complexity of the economic and financial systems, it seems natural to view all economic models only as approximations to the true data generating process (Watson, 1993; White, 1994; Canova, 1994; among others). As argued by Maasoumi (1990), "Misspecification of these models is therefore endemic and inevitable. Omission of relevant variables, inclusion of 'irrelevant variables', incorrect functional forms, incompleteness of systems of relations, and incorrect distributional assumptions are both common and present simultaneously."

Models for which the likelihood function is available are now routinely estimated in a quasi-maximum likelihood framework and the statistical inference is performed using misspecification-robust standard errors (White, 1982, 1994). In contrast, misspecification-robust inference for moment condition models, estimated by the generalized method of moments (GMM), is much less widespread among applied researchers. It is still common practice to use the asymptotic standard errors of Hansen (1982), derived under the assumption of correct model specification, even when the model is rejected by the data. This is unfortunate since most economic models are defined by a set of conditional or unconditional moment restrictions and not allowing for possible (global) misspecification of these moment restrictions would render the GMM inference asymptotically invalid.

Maasoumi and Phillips (1982) and Gallant and White (1988) provide an early analysis of inference in globally misspecified models estimated by instrumental variables and GMM with a fixed weighting matrix, respectively. Hall and Inoue (2003) extended the asymptotic analysis in these studies to the two-step and iterated GMM estimators. They derived the limiting variance of these estimators in the presence of model misspecification and showed that the misspecification adjustment depends on the weighting matrix used in estimation. The consequences of model misspecification for GMM estimation and inference are summarized in Hall (2005). Despite these recent advances in the literature, the use of misspecification-robust standard errors in empirical work with GMM estimators is largely absent.

Misspecification-robust inference proves to be particularly important in evaluating linear asset-pricing models that are often found to be rejected by the data (see Kan and Robotti, 2009, Kan, Robotti, and Shanken, 2013, and Gospodinov, Kan, and Robotti, 2013, 2014, among others). While invariant estimators are believed to posses a number of appealing properties, misspecification-robust inference for these estimators is not yet available in the literature. In this paper, we derive explicit expressions for the asymptotic variances of the ML and the continuously-updated GMM (CU-GMM) estimators (Hansen, 1982; Hansen, Heaton, and Yaron, 1996) in potentially misspecified asset-pricing models.

We focus on the ML and CU-GMM estimators for several reasons. First, the invariance of these estimators to normalizations and transformations of the data is particularly desirable in asset-pricing models (Peñaranda and Sentana, 2015) that could be written in both beta-pricing and stochastic discount factor (SDF) form. Second, the CU-GMM estimator is a member of the class of generalized empirical likelihood (GEL) estimators (Newey and Smith, 2004), which provides an alternative look into the first- and higher-order asymptotic properties of the CU-GMM estimator. In fact, we use the GEL framework to parameterize the degree of model misspecification as the distance of the pseudo-true value of the vector of Lagrange multipliers, associated with the moment conditions, from zero and cast the CU-GMM estimator as a solution to a quasi-likelihood problem. This allows us to work directly with the score function and to sidestep some explicit joint normality assumptions in the approach of Hall and Inoue (2003). Due to the quasi-likelihood interpretation of the estimated augmented parameter vector (the parameters of interest and the Lagrange multipliers), the asymptotic variance of the CU-GMM estimator takes the usual sandwich form as in White (1982, 1994). In this respect, we complement the results in Kitamura (1998) and Schennach (2007), and provide an explicit expression for the asymptotic variance of the CU-GMM estimator in potentially misspecified models. Our results for CU-GMM are derived for linear as well as nonlinear moment condition models.

On the other hand, the maximum likelihood (ML) estimator is developed only for linear beta-pricing models. The usefulness of this estimator is that it can be obtained in a closed form, which facilitates its practical implementation and theoretical analysis. One possibility in deriving the asymptotic distribution of the ML estimator under potentially misspecified models is to extend the two-stage Gaussian quasi-maximum likelihood setting of White (1994), which is robust to distributional assumptions and model misspecification. In contrast, we maintain the normality assumption, which is often imposed in the ML estimation of the beta-pricing model, to obtain a more explicit expression for the asymptotic variance of the estimator. The proposed asymptotic standard errors help us quantify the importance of the model misspecification adjustment when conducting statistical inference. Furthermore, our setup allows us to express the ML estimator as an optimal minimum distance estimator and approximate its limiting behavior under misspecified models using analytical tools for moment condition models as in Hall and Inoue (2003).

Overall, our theoretical and simulation results suggest that the impact of model misspecification on the asymptotic variance of the ML and CU-GMM estimators can be very large and of practical economic significance. It turns out that the size distortions arising from wrongly assuming correct model specification are much larger for these invariant estimators than for the non-invariant estimators studied by Kan and Robotti (2009), Kan, Robotti, and Shanken (2013), and Gospodinov, Kan, and Robotti (2013). For example, the rejection rate of the centered t-test that does not account for model misspecification could be as large as 71% for CU-GMM at the 10% significance level with 300 observations and a degree of model misspecification calibrated to actual data. The proposed misspecification-robust standard errors correct these size distortions and, interestingly, provide substantial improvements even when the model is correctly specified.

The rest of the paper is structured as follows. Sections 2 and 3 derive the limiting distributions of the ML and CU-GMM estimators in misspecified linear asset-pricing models. The asymptotic results for the CU-GMM estimator are also extended to general nonlinear moment condition models. Section 4 provides simulation results on the empirical size and power of t-tests computed with standard errors under correct model specification and misspecification-robust standard errors. Section 5 illustrates the economic significance of the proposed misspecification adjustment using actual data for several popular asset-pricing models. Section 6 concludes.

2 ML Estimation and Misspecification-Robust Inference in the Beta-Pricing Representation

In this section, we discuss the maximum likelihood approach to estimation and statistical inference in unconditional beta-pricing models. Suppose that R_t, the gross returns on N test assets at time t $(t = 1, \ldots, T)$, can be described by the following data generating process:

$$R_t = \alpha + \beta f_t + \epsilon_t, \tag{1}$$

where f_t denotes the realizations of K systematic factors at time t and ϵ_t are the model innovations at time t with $E[\epsilon_t] = 0_N$ and $E[f_t \epsilon_t'] = 0_{K\times N}$. Taking expectations on both sides yields

$$\mu_R = \alpha + \beta\mu_f, \tag{2}$$

where $\mu_f = E[f_t]$ and $\mu_R = E[R_t]$. Under the K-factor asset-pricing model, we have

$$\mu_R = 1_N \gamma_0 + \beta \gamma_1, \tag{3}$$

where 1_N is an $N \times 1$ vector of ones, γ_0 is the zero-beta rate, and γ_1 is the vector of risk premia associated with the K risk factors f_t. Let $\gamma = [\gamma_0,\ \gamma_1']' \in \Gamma$ denote the parameter vector of interest. Comparing (2) with (3), we have the following restrictions on α:

$$\alpha = 1_N \gamma_0 + \beta \phi, \tag{4}$$

where $\phi = \gamma_1 - \mu_f$. The multi-factor model can be written in matrix form as

$$Y = XB + \mathcal{E}, \tag{5}$$

where $B = [\alpha,\ \beta]'$, and the typical rows of X, Y, and $\mathcal{E}$ are $x_t' = [1,\ f_t']$, R_t', and ϵ_t', respectively.

Assumption MLE.A. *Assume that (a) $(f_t,\ \epsilon_t)$ are i.i.d. normally distributed with $V_f = \mathrm{Var}[f_t]$ and $\Sigma = \mathrm{Var}[\epsilon_t]$; (b) the matrix $H = [1_N,\ \beta]$ is of full column rank; and (c) the parameter space Γ is a compact subset of $\mathbb{R}^{K+1}$.*

The ML estimators of μ_f and V_f are

$$\hat{\mu}_f = \frac{1}{T}\sum_{t=1}^{T} f_t, \tag{6}$$

$$\hat{V}_f = \frac{1}{T}\sum_{t=1}^{T}(f_t - \hat{\mu}_f)(f_t - \hat{\mu}_f)'. \tag{7}$$

We partition the parameter vector $\delta = [\mathrm{vec}(B')',\ \mathrm{vech}(\Sigma)',\ \gamma_0,\ \phi']'$ into $\delta = [\delta_1',\ \delta_2']'$, where $\delta_1 = [\mathrm{vec}(B')',\ \mathrm{vech}(\Sigma)']'$ and $\delta_2 = [\gamma_0,\ \phi']'$. Under Assumption MLE.A(a), the log-likelihood function of the unrestricted model (5) is given by

$$\mathcal{L}_T(\delta_1) = -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(R_t - B'x_t)'\Sigma^{-1}(R_t - B'x_t). \tag{8}$$

Then, the unrestricted ML estimators of B and Σ are

$$\hat{B} \equiv [\hat{\alpha},\ \hat{\beta}]' = (X'X)^{-1}(X'Y), \tag{9}$$

$$\hat{\Sigma} = \frac{1}{T}(Y - X\hat{B})'(Y - X\hat{B}), \tag{10}$$

and

$$\mathcal{L}_T(\hat{\delta}_1) = -\frac{T}{2}\log|\hat{\Sigma}| - \frac{NT}{2}[\log(2\pi) + 1]. \tag{11}$$

The concentrated likelihood function is

$$\mathcal{L}_T(\tilde{\delta}_1|\delta_2) = -\frac{T}{2}\log|\tilde{\Sigma}| - \frac{NT}{2}[\log(2\pi)+1], \tag{12}$$

where $\tilde{\Sigma}$ denotes the estimated covariance of the residuals under the constraint (4) that the asset-pricing model holds. Note also that the constraint (4) can be expressed as $\omega'(Q_1 B + Q_2) = 0'_N$, where $\omega = [1,\ -\phi',\ -\gamma_0]'$, $Q_1 = \begin{bmatrix} I_{K+1} \\ 0'_{K+1} \end{bmatrix}$, and $Q_2 = \begin{bmatrix} 0_{(K+1)\times N} \\ 1'_N \end{bmatrix}$. Then, the likelihood ratio statistic of $H_0 : \alpha = 1_N\gamma_0 + \beta\phi$ is given by

$$\mathcal{LR}_T(\delta_2|\hat{\delta}_1) = -T\ \log\left(1 + \frac{\omega'(Q_1\hat{B}+Q_2)\hat{\Sigma}^{-1}(Q_1\hat{B}+Q_2)'\omega}{T\omega' Q_1(X'X)^{-1}Q_1'\omega}\right), \tag{13}$$

using that

$$\mathcal{LR}_T = 2\left[\mathcal{L}_T(\tilde{\delta}_1|\delta_2) - \mathcal{L}_T(\hat{\delta}_1)\right] = -T\ \log\left(\frac{|\tilde{\Sigma}|}{|\hat{\Sigma}|}\right) \tag{14}$$

and (Seber, 1984, p. 410)

$$\tilde{\Sigma} = \hat{\Sigma} + (\omega'(Q_1\hat{B}+Q_2))'[T\omega' Q_1(X'X)^{-1}Q_1'\omega]^{-1}\omega'(Q_1\hat{B}+Q_2). \tag{15}$$

Therefore, the ML estimator of $\delta_2 = [\gamma_0,\ \phi']'$ can be defined as

$$\hat{\delta}_2 = \text{argmin}_{\delta_2} - \mathcal{LR}_T(\delta_2|\hat{\delta}_1). \tag{16}$$

Since this is a ratio of quadratic forms in ω, the minimum is attained when ω is proportional to the eigenvector associated with the largest eigenvalue of

$$[(Q_1\hat{B}+Q_2)\hat{\Sigma}^{-1}(Q_1\hat{B}+Q_2)']^{-1}[TQ_1(X'X)^{-1}Q_1']. \tag{17}$$

Let $p = [p_1, \ldots, p_{K+2}]'$ be the eigenvector associated with the largest eigenvalue of (17). Then, we have

$$\hat{\phi}_i = -p_{i+1}/p_1, \quad i = 1, \ldots, K, \tag{18}$$

$$\hat{\gamma}_0 = -p_{K+2}/p_1, \tag{19}$$

and the ML estimator of γ_1 is simply $\hat{\gamma}_1 = \hat{\phi} + \hat{\mu}_f$.

White (1994, Theorem 6.11) provides the asymptotic distribution of $\hat{\delta}_2$ under potential model misspecification and non-normality of ϵ_t. To obtain explicit expressions for the asymptotic covariance of $\hat{\gamma} = [\hat{\gamma}_0,\ \hat{\gamma}_1']'$ in globally misspecified models, in the following we deviate from White (1994) and maintain the joint normality assumption in MLE.A. This allows us to isolate and quantify the impact of model misspecification on the asymptotic covariance of $\hat{\gamma}$.

Note that the ML estimator of γ can also be expressed as

$$\hat{\gamma} = \text{argmin}_{\gamma} \frac{(\hat{\mu}_R - \hat{H}\gamma)'\hat{\Sigma}^{-1}(\hat{\mu}_R - \hat{H}\gamma)}{1 + \gamma_1'\hat{V}_f^{-1}\gamma_1}, \tag{20}$$

where $\hat{\mu}_R = \frac{1}{T}\sum_{t=1}^{T} R_t$ and $\hat{H} = [1_N,\ \hat{\beta}]$. Define the pseudo-true values of γ as

$$\gamma_* \equiv \begin{bmatrix} \gamma_{0*} \\ \gamma_{1*} \end{bmatrix} = \text{argmin}_{\gamma} \frac{(\mu_R - H\gamma)'\Sigma^{-1}(\mu_R - H\gamma)}{1 + \gamma_1'V_f^{-1}\gamma_1}, \tag{21}$$

and let $M = \left[1_N,\ \beta + \frac{(\mu_R - H\gamma_*)\gamma_{1*}'V_f^{-1}}{1+\gamma_{1*}'V_f^{-1}\gamma_{1*}}\right]$, $S_* = (\mu_R - H\gamma_*)'\Sigma^{-1}(\mu_R - H\gamma_*)$, $c_* = 1 + \gamma_{1*}'V_f^{-1}\gamma_{1*}$, $C_1 = 2M'\Sigma^{-1}M - H'\Sigma^{-1}H$, $C = H'\Sigma^{-1}H - \frac{S_*}{c_*}\tilde{V}_f^{-1}$, $c_* = 1 + \gamma_{1*}'V_f^{-1}\gamma_{1*}$,

$$\tilde{V}_f = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_f \end{bmatrix} \tag{22}$$

and

$$\tilde{V}_f^{-1} = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_f^{-1} \end{bmatrix}. \tag{23}$$

Theorem 1 below derives the asymptotic distribution of $\hat{\gamma}$ for globally misspecified models.

Theorem 1. *Suppose that Assumption MLE.A is satisfied and $\mu_R \neq H\gamma$, that is, the model is misspecified. Then, we have*

$$\sqrt{T}(\hat{\gamma} - \gamma_*) \xrightarrow{d} \mathcal{N}\left(0_{K+1}, \Omega_m\right), \tag{24}$$

where $\Omega_m = C^{-1}\left\{c_*C_1 + C_1\tilde{V}_fC_1 + S_*\left[\left(1 - \frac{1}{c_*^2}\right)C_1 + \left(1 + \frac{S_*(c_*-1)}{c_*^2}\right)\tilde{V}_f^{-1} + \frac{1}{c_*^2}H'\Sigma^{-1}H\right]\right\}C^{-1}$.

Proof. See Appendix.

Note that when the model is correctly specified, we have $S_* = 0$, $M = H$, and $C_1 = C = H'\Sigma^{-1}H$. In this case,

$$\sqrt{T}(\hat{\gamma} - \gamma_*) \xrightarrow{d} \mathcal{N}\left(0_{K+1}, \Omega_c\right), \tag{25}$$

where $\Omega_c = c_*(H'\Sigma^{-1}H)^{-1} + \tilde{V}_f$.

3 CU-GMM Estimation and Misspecification-Robust Inference in the SDF Representation

The $N \times 1$ vector of pricing errors (moment conditions) of the linear asset-pricing model at time t are given by

$$e_t(\lambda) = R_t \tilde{f}_t' \lambda - 1_N, \tag{26}$$

where $\tilde{f}_t = [1,\ f_t']'$ and $\lambda = [\lambda_0,\ \lambda_1']' \in \Lambda$ is a $(K+1) \times 1$ parameter vector. A model is globally misspecified if for all values of λ we have

$$E[e_t(\lambda)] \equiv e(\lambda) = G\lambda - 1_N \neq 0_N, \tag{27}$$

where $G = E[R_t \tilde{f}_t']$. Let $V(\lambda) = \lim_{T\to\infty} \text{Var}\left(T^{-1/2} \sum_{t=1}^{T} (e_t(\lambda) - e(\lambda))\right)$ be a positive definite matrix and λ_* denote the pseudo-true value of λ, which is defined as

$$\lambda_* \equiv \begin{bmatrix} \lambda_{0*} \\ \lambda_{1*} \end{bmatrix} = \text{argmin}_\lambda e(\lambda)' V(\lambda)^{-1} e(\lambda). \tag{28}$$

In the case of correctly specified models, $e(\lambda_*) = 0_N$ and λ_* is the true value of λ.

Assumption GMM.A. *Assume that (a) $Y_t \equiv [f_t',\ R_t']'$ is a jointly stationary and ergodic process; (b) $e_t(\lambda_*) - e(\lambda_*)$ forms a martingale difference sequence with variance matrix $V(\lambda_*)$; (c) $E[(e_t(\lambda) - e(\lambda))(e_t(\lambda) - e(\lambda))']$ is non-singular in some neighborhood of λ_*; and (d) the parameter space Λ is a compact subset of R^{K+1}.*

Assumption GMM.A imposes some restrictions on the dynamic behavior of the data and the moment conditions. The martingale difference sequence assumption in GMM.A(b) can be relaxed by modifying the structure of the estimation problem along the lines suggested by Smith (2011). Let $g_t = R_t \tilde{f}_t'$, $G_T = \frac{1}{T}\sum_{t=1}^{T} g_t = \frac{1}{T}\sum_{t=1}^{T} R_t \tilde{f}_t'$, and $\bar{e}_T(\lambda) = \frac{1}{T}\sum_{t=1}^{T} e_t(\lambda) = G_T\lambda - 1_N$ is an $N \times 1$ vector of sample pricing errors with a sample variance (given Assumption GMM.A(b))

$$V_T(\lambda) = \frac{1}{T}\sum_{t=1}^{T} [e_t(\lambda) - \bar{e}_T(\lambda)][e_t(\lambda) - \bar{e}_T(\lambda)]'. \tag{29}$$

Then, the CU-GMM estimator of λ is defined as[1]

$$\hat{\lambda} = \left[\hat{\lambda}_0,\ \hat{\lambda}_1'\right]' = \text{argmin}_\lambda \bar{e}_T(\lambda)' V_T(\lambda)^{-1} \bar{e}_T(\lambda). \tag{30}$$

[1] Newey and Smith (2004, footnote 2) establish the equivalence of this CU-GMM estimator and the CU-GMM estimator based on $V_T(\lambda) = \frac{1}{T}\sum_{t=1}^{T} e_t(\lambda) e_t(\lambda)'$.

In deriving the asymptotic variance for the CU-GMM estimator in (30) under model misspecification, we follow an approach that allows us to write the estimator of an augmented parameter vector as a solution to the score function of a just-identified problem. The point of departure is the observation that the CU-GMM estimator can be defined equivalently as a solution to a nonparametric likelihood problem that minimizes the Euclidean distance between a probability measure P_T that satisfies exactly the moment conditions, that is, $E\left[e(\lambda)|P_T\right] = \int e(\lambda)dP_T = 0_N$, and the empirical probability measure (see Antoine, Bonnal, and Renault, 2007, and Newey and Smith, 2004, among others). This primal problem can be recast conveniently as a dual (saddle-point) problem, where the duality parameter $\rho(\lambda)$ is an $N \times 1$ vector of Lagrange multipliers associated with the moment conditions $e(\lambda) = 0_N$. Let $\rho_* \equiv \rho_*(\lambda)$ denote the pseudo-true value of ρ and $\theta = [\rho', \lambda']' \in \Theta$ be an augmented $(N + K + 1)$ parameter vector with a pseudo-true value $\theta_* = [\rho_*', \lambda_*']'$. For correctly specified models, we have $\rho_* = 0_N$ while for misspecified models, $\|\rho_*(\lambda)\| > 0$ for all $\lambda \in \Lambda$.

Let $\hat{\theta} = \left[\hat{\rho}', \hat{\lambda}'\right]'$. The first-order conditions of this nonparametric likelihood problem are given by (Antoine, Bonnal, and Renault, 2007)

$$\bar{s}_T(\hat{\theta}) \equiv \frac{1}{T}\sum_{t=1}^{T} s_t(\theta)\bigg|_{\theta=\hat{\theta}} = 0_{N+K+1}, \tag{31}$$

where

$$s_t(\theta) = -\left[\begin{array}{c} \left[1 + \rho'\left(e_t(\lambda) - e(\lambda)\right)\right] e_t(\lambda) \\ \left[1 + \rho'\left(e_t(\lambda) - e(\lambda)\right)\right] g_t'\rho \end{array}\right]. \tag{32}$$

The $(N+K+1)$ vector $s_t(\theta)$ can be interpreted as the score function of a quasi-likelihood problem. As argued above, we augment the first-order conditions for the parameter vector of interest λ with the parameter vector of Lagrange multipliers ρ in order to make the model misspecification, which is reflected in ρ, explicit in deriving the limiting distribution. Note also that from the first N equations in (31), we have $\hat{\rho} = -V_T(\hat{\lambda})^{-1}\bar{e}_T(\hat{\lambda})$.

Let $w_t(\theta_*) = [1 + \rho_*'\left(e_t(\lambda_*) - e(\lambda_*)\right)]$, $B = E[w_t(\theta_*)g_t] + E[\left(e_t(\lambda_*) - e(\lambda_*)\right)\rho_*'(g_t - G)]$, $C = E[(g_t - G)'\rho_*\rho_*'(g_t - G)]$, and $V = V(\lambda_*)$. Next, we state the limiting distribution of the CU-GMM estimator in misspecified models.

Theorem 2. *Suppose that Assumption GMM.A holds, G is of full column rank, and Y_t has finite eighth moments. Then, it follows that*

$$\sqrt{T}(\hat{\theta} - \theta_*) \xrightarrow{d} \mathcal{N}(0_{N+K+1}, \Xi), \tag{33}$$

where $\Xi \equiv E[l_t l_t']$, $l_t \equiv [l_{1t}', l_{2t}']'$, *and*

$$l_{1t} = V^{-1}\left[w_t(\theta_*)e_t(\lambda_*) - Bl_{2t}\right], \quad (34)$$

$$l_{2t} = (C - B'V^{-1}B)^{-1}w_t(\theta_*)\left[g_t'\rho_* - B'V^{-1}e_t(\lambda_*)\right]. \quad (35)$$

Proof. See Appendix.

The variance matrix Ξ in Theorem 2 can be consistently estimated using the sample analogs of (34) and (35). Importantly, the result in Theorem 2 can be easily extended to nonlinear moment condition models. Let $g_t^{(2)}(\lambda) = (\partial/\partial\lambda')\mathrm{vec}(g_t(\lambda))$, where $g_t(\lambda) = \partial e_t(\lambda)/\partial\lambda'$ is now a function of λ, and $\tilde{C} = (I_{K+1} \otimes \rho_*')E[g_t^{(2)}(\lambda_*)] + E[(g_t(\lambda_*) - G(\lambda_*))'\rho_*\rho_*'(g_t(\lambda_*) - G(\lambda_*))]$. The following theorem states the result for possibly misspecified nonlinear models.

Theorem 3. *In addition to Assumption GMM.A, assume that (a) the pseudo-true values* λ_* *and* ρ_* *are unique and* λ_* *is in the interior of* Λ*; (b)* $e_t(\lambda)$ *is twice continuously differentiable in* λ *and* $E\left[\sup_{\lambda\in\Lambda}|e_t(\lambda)|\right] < \infty$*; (c)* $E\left[\sup_{\theta\in\mathcal{N}(\theta_*)}\left\|\frac{\partial}{\partial\theta'}s_t(\theta)\right\|\right] < \infty$ *for some neighborhood* N *of* θ_**; (d)* $E\left\|s_t(\theta_*)s_t(\theta_*)'\right\|$ *exists and is finite; (h)* $E\left[\frac{\partial}{\partial\theta'}s_t(\theta_*)\right]$ *is of full rank. Then, it follows that*

$$\sqrt{T}(\hat{\theta} - \theta_*) \xrightarrow{d} N(0_{N+K+1}, \tilde{\Xi}), \quad (36)$$

where $\tilde{\Xi} \equiv E[\tilde{l}_t\tilde{l}_t']$, $\tilde{l}_t \equiv [\tilde{l}_{1t}', \tilde{l}_{2t}']$ *and*

$$\tilde{l}_{1t} = V^{-1}\left[w_t(\theta_*)e_t(\lambda_*) - B\tilde{l}_{2t}\right], \quad (37)$$

$$\tilde{l}_{2t} = (\tilde{C} - B'V^{-1}B)^{-1}w_t(\theta_*)\left[g_t(\lambda_*)'\rho_* - B'V^{-1}e_t(\lambda_*)\right]. \quad (38)$$

Proof. See Appendix.

Note that for linear models, $g_t^{(2)}(\lambda_*)$ is a null matrix and $\tilde{C} = C = E[(g_t - G)'\rho_*\rho_*'(g_t - G)]$. Thus, the result in Theorem 3 reduces to the asymptotic distribution in Theorem 2. Furthermore, for correctly specified models, the limiting distribution in Theorem 3 specializes to the result in Theorem 3.2 of Newey and Smith (2004). More specifically, for correctly specified models, we have $\rho_* = 0_N$, $w_t(\theta_*) = 1$, $B = G$, $C = 0_{(K+1)\times(K+1)}$, $(C - B'V^{-1}B)^{-1} = -(G'V^{-1}G)^{-1}$, and

$$l_{1t} = V^{-1}\left[e_t(\lambda_*) - Gl_{2t}\right], \quad (39)$$

$$l_{2t} = (G'V^{-1}G)^{-1}G'V^{-1}e_t(\lambda_*). \quad (40)$$

Peñaranda and Sentana (2015) show the equivalence between the CU-GMM estimation of the SDF and beta-pricing frameworks. Let[2]

$$w_t(\hat{\lambda}) = \frac{1 - (e_t(\hat{\lambda}) - \bar{e}(\hat{\lambda}))'\hat{W}_e(\hat{\lambda})^{-1}\bar{e}(\hat{\lambda})}{T}. \tag{41}$$

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

4 Monte Carlo Simulations

In this section, we evaluate the performance of the proposed variance estimators by reporting the empirical size and power of t-tests that are constructed using standard errors under correct model specification and misspecification-robust standard errors. To facilitate the power comparisons, we report size-adjusted power in all tables. In our simulations, we consider the popular linear model of Fama and French (FF3, 1993) with a constant term and three risk factors ($\tilde{f}_t = [1,\ mkt_t,\ smb_t,\ hml_t]'$), where mkt denotes the excess return (in excess of the one-month T-bill rate) on the value-weighted stock market index (NYSE-AMEX-NASDAQ), smb is the return difference between portfolios of stocks with small and large market capitalizations, and hml is the return difference between portfolios of stocks with high and low book-to-market ratios ("value" and "growth" stocks, respectively). The asset-pricing model can either be correctly specified or misspecified.

In our baseline simulations, the returns on the test assets and the risk factors f_t are drawn from a multivariate normal distribution. In addition, we analyze the impact of non-normality and finite moment requirements on our variance approximations by drawing the returns and the factors from a multivariate t-distribution with eight degrees of freedom.[3] The variance matrix of the simulated test asset returns, R_t, is set equal to the estimated variance matrix from the 1963:7–2015:7 sample of monthly returns on the 25 Fama-French size and book-to-market ranked portfolios and the 10

[2] Newey and Smith (2004) and Antoine, Bonnal, and Renault (2007) show that $w_t(\hat{\lambda})$, $t = 1, \ldots, T$, in (41) represent the implied probability weights associated with the CU-GMM estimator.

[3] In our empirical application, the degree-of-freedom parameter of the multivariate t-distribution is estimated to be 8.1.

industry portfolios ($N = 35$). For misspecified models, the means of the simulated returns are set equal to the means of the actual returns. Then, for example, one can use the Hansen and Jagannathan distance (HJD, 1997) to quantify the degree of model misspecification. The resulting HJD for FF3 is 0.3996, which is in line with the HJD values commonly reported in empirical applications with monthly data. For correctly specified models, the means of the simulated returns are set such that the asset-pricing model restrictions are satisfied (that is, the pricing errors are zero). The means and variances of the simulated factors are calibrated to those of the observed factors during the 1963:7–2015:7 sample period.[4] The variance matrix of the risk factors and the returns is set equal to the variance matrix estimated from the data. The time-series sample sizes are $T = 300$, 600, 1200, and 3600. The number of Monte Carlo replications is set equal to 100,000.

For the beta-pricing model, the vector of risk premium parameters γ is estimated by the ML estimator $\hat{\gamma}$. The estimator $\hat{\gamma}$ is used to construct a consistent estimate of the variance matrix

$$\hat{\Omega}_c = \hat{c}(\hat{H}'\hat{\Sigma}^{-1}\hat{H})^{-1} + \hat{\tilde{V}}_f, \tag{42}$$

under the assumption of a correctly specified model, and the variance matrix

$$\hat{\Omega}_m = \hat{C}^{-1}\left\{\hat{c}\hat{C}_1 + \hat{C}_1\widehat{\tilde{V}}_f\hat{C}_1 + \hat{S}\left[\left(1 - \frac{1}{\hat{c}^2}\right)\hat{C}_1 + \left(1 + \frac{\hat{S}(\hat{c}-1)}{\hat{c}^2}\right)\widehat{\tilde{V}}_f^{-1} + \frac{1}{\hat{c}^2}\hat{H}'\hat{\Sigma}^{-1}\hat{H}\right]\right\}\hat{C}^{-1}, \tag{43}$$

under the assumption of a misspecified model, where $\hat{c} = 1 + \hat{\gamma}_1'\hat{V}_f^{-1}\hat{\gamma}_1$, $\hat{S} = (\hat{\mu}_R - \hat{H}\hat{\gamma})'\hat{\Sigma}^{-1}(\hat{\mu}_R - \hat{H}\hat{\gamma})$, $\hat{M} = \left[1_N,\ \hat{\beta} + \frac{(\hat{\mu}_R - \hat{H}\hat{\gamma})\hat{\gamma}_1'\hat{V}_f^{-1}}{1+\hat{\gamma}_1'\hat{V}_f^{-1}\hat{\gamma}_1}\right]$, $\hat{C}_1 = 2\hat{M}'\hat{\Sigma}^{-1}\hat{M} - \hat{H}'\hat{\Sigma}^{-1}\hat{H}$, $\hat{C} = \hat{H}'\hat{\Sigma}^{-1}\hat{H} - \frac{\hat{S}}{\hat{c}}\widehat{\tilde{V}}_f^{-1}$,

$$\widehat{\tilde{V}}_f = \begin{bmatrix} 0 & 0_K' \\ 0_K & \hat{V}_f \end{bmatrix}, \tag{44}$$

and

$$\widehat{\tilde{V}}_f^{-1} = \begin{bmatrix} 0 & 0_K' \\ 0_K & \hat{V}_f^{-1} \end{bmatrix}. \tag{45}$$

The square roots of the diagonal elements of $\hat{\Omega}_c$ and $\hat{\Omega}_m$ are then used to obtain the t-tests under correct model specification, denoted by $t_c(\hat{\gamma})$, and the misspecification-robust t-tests, denoted by $t_m(\hat{\gamma})$.

Tables I and II report the actual probabilities of rejection for the MLE t-tests ($t_c(\hat{\gamma})$ and $t_m(\hat{\gamma})$) of $H_0 : \gamma_{1,i} = \gamma_{*1,i}$ and $H_0 : \gamma_{1,i} = 0$ ($i = 1, \ldots, K$) using standard normal critical values. The

[4] This choice of sample period is dictated by factor data availability. The test asset return and the factor data are obtained from Kenneth French's website.

factors and the returns are assumed to be multivariate normally distributed. The (pseudo-) true values γ_* (reported in the table legends) are set equal to their ML estimates from the actual data with μ_R being dependent on whether the model is correctly specified or misspecified, as described above.

Tables I and II about here

Table I presents the results for the FF3 specification when the model is correctly specified. Table II is for the misspecified model. Although the model is correctly specified, the centered t-test under correct specification, t_c, tends to slightly overreject in small samples. Interestingly, the centered misspecification-robust t-test, t_m, corrects these size distortions and provides improvements despite the fact that the true misspecification adjustment is zero in this case. When the model is misspecified, the t-tests t_c are no longer valid, and this is reflected in the fairly significant overrejections. In contrast, the centered misspecification-robust t-tests t_m are almost perfectly sized even in small samples. For example, for $T = 600$ and a 10% significance level, the centered t_c statistic for mkt rejects the null hypothesis 21.6% of the time under model misspecification ($t_c(\hat{\gamma}_{1,1})$ in Panel A of Table II). In contrast, the centered misspecification-robust t_m statistic rejects the null hypothesis 9.7% of the time under model misspecification ($t_m(\hat{\gamma}_{1,1})$ in Panel B of Table II). As for power, both tests behave very similarly. It should be noted that power can be low at times. This depends on, among other things, how far from zero the (pseudo-) true parameters are.

We explore departures from the normality assumption in Tables III and IV. In these tables, the returns and the factors are multivariate t-distributed with eight degrees of freedom. Note that this distribution (i) generates fat tails and conditional heteroskedasticity in returns, and (ii) makes the MLE inference invalid since the normality assumption is violated.

Tables III and IV about here

When the model is correctly specified (Table III), the impact of non-normality on t_c and t_m is negligible, and the size and power properties of the two tests are very similar to the ones under normality in Table I. When the model is misspecified, the centered misspecification-robust t-test tends to slightly overreject the null in very large samples but is almost perfectly sized in small samples. For example, for $T = 3600$ and a 10% significance level, the centered t_m statistic for mkt

rejects the null hypothesis 11.1% of the time ($t_m(\hat{\gamma}_{1,1})$ in Panel B of Table IV). The centered t_c statistic continues to be theoretically invalid since the model is misspecified, and it exhibits slightly bigger overrejections compared to the normal case. As for power, both tests behave similarly, with power being about the same as under normality. Overall, t_m enjoys very nice size and power properties and seems to be little affected by the presence of heavy tails in financial data.

For the SDF representation of the asset-pricing model, the parameter vector $\theta = [\rho', \lambda']'$ is estimated using the CU-GMM estimator $\hat{\theta} = [\hat{\rho}', \hat{\lambda}']'$. Let $w_t(\hat{\theta}) = 1 + \hat{\rho}'[e_t(\hat{\lambda}) - \bar{e}_T(\hat{\lambda})]$, $\hat{B} = \frac{1}{T}\sum_{t=1}^T w_t(\hat{\theta})g_t + \frac{1}{T}\sum_{t=1}^T [e_t(\hat{\lambda}) - \bar{e}_T(\hat{\lambda})]\hat{\rho}'(g_t - G_T)$, and $\hat{C} = \frac{1}{T}\sum_{t=1}^T (g_t - G_T)'\hat{\rho}\hat{\rho}'(g_t - G_T)$. Then,

$$\hat{l}_{1t} = V_T(\hat{\lambda})^{-1}\left[w_t(\hat{\theta})e_t(\hat{\lambda}) - \hat{B}\hat{l}_{2t}\right], \quad (46)$$

$$\hat{l}_{2t} = (\hat{C} - \hat{B}'V_T(\hat{\lambda})^{-1}\hat{B})^{-1}w_t(\hat{\theta})\left[g_t'\hat{\rho} - \hat{B}'V_T(\hat{\lambda})^{-1}e_t(\hat{\lambda})\right], \quad (47)$$

which are used to construct a consistent estimator $\hat{\Xi}$ of the asymptotic variance matrix of $\hat{\theta}$ in Theorem 2. The square roots of the last $K+1$ diagonal elements of $\hat{\Xi}$ are used to construct the misspecification-robust t-tests, denoted by $t_m(\hat{\lambda})$. The variance estimator of $\hat{\theta}$ under correct model specification is obtained from

$$\hat{l}_{1t} = V_T(\hat{\lambda})^{-1}\left[e_t(\hat{\lambda}) - G_T\hat{l}_{2t}\right], \quad (48)$$

$$\hat{l}_{2t} = (G_T'V_T(\hat{\lambda})^{-1}G_T)^{-1}G_T'V_T(\hat{\lambda})^{-1}e_t(\hat{\lambda}), \quad (49)$$

and the square roots of the last $K+1$ diagonal elements are used to construct the t-tests under correct model specification, denoted by $t_c(\hat{\lambda})$.

Tables V and VI report the actual probabilities of rejection for the CU-GMM t-tests ($t_c(\hat{\lambda})$ and $t_m(\hat{\lambda})$) of $H_0 : \lambda_{1,i} = \lambda_{*1,i}$ and $H_0 : \lambda_{1,i} = 0$ ($i = 1, \ldots, K$) using standard normal critical values. The factors and the returns are multivariate normally distributed. The (pseudo-) true values λ_* need to be computed under the joint normality assumption. For this purpose, partition

$$\text{Var}\begin{bmatrix} f_t \\ R_t \end{bmatrix} = \begin{bmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{bmatrix}. \quad (50)$$

It is easy to show that under the *i.i.d.* multivariate elliptical distributional assumption on the factors and the returns, the optimal weighting matrix (the variance matrix of the moment conditions) is given by

$$V(\lambda) = [(\lambda_0 + \mu_f'\lambda_1)^2 + (1+\kappa)\lambda_1'V_f\lambda_1]V_R + (\lambda_0 + \mu_f'\lambda_1)(\mu_R\lambda_1'V_{fR} + V_{Rf}\lambda_1\mu_R') + (\lambda_1'V_f\lambda_1)\mu_R\mu_R' + (1+2\kappa)V_{Rf}\lambda_1\lambda_1'V_{fR}, \quad (51)$$

where κ is the multivariate excess kurtosis of the factors and the returns. The weighting matrix under the normality assumption[5] is obtained by setting $\kappa = 0$ and the (pseudo-) true values are set equal to the CU-GMM estimates from the actual data using this form of the weighting matrix and the value of μ_R corresponding to correctly specified or misspecified models.

Tables V and VI about here

While the pattern of results is somewhat similar to those for the MLE, the CU-GMM estimator appears to be much more sensitive to model misspecification. This is partly due to the numerical instability of the CU-GMM estimator, especially when N is large, which leads to poorer asymptotic approximations and more pronounced size distortions. For example, in the correctly specified FF3 model with $T = 600$, the centered t_c test rejects the null for the market factor 16.7% of the time at the 10% significance level while the centered t_m test rejects the null 8.9% of the time (Panels A and B of Table V). For the misspecified FF3 model with $T = 600$, the corresponding rejection rates for the centered t_c and t_m tests are 60% and 10.8% (Panels A and B of Table VI), respectively. In fact, the rejection rates for the centered t_c test can be as large as 27.3% (Panel A of Table V) for correctly specified models and 68.2% (Panel A of Table VI) for misspecified models at the 10% significance level.

This should serve as a warning signal to applied researchers who routinely use standard errors constructed under the assumption of a correctly specified model in evaluating the statistical significance of the SDF parameters. It suggests that the researcher will conclude erroneously (with very high probability) that the risk factor is important for the pricing of the test assets. While the centered misspecification-robust t-tests also exhibit some slight size distortions for small sample sizes,[6] their empirical size approaches quickly the nominal level when T increases. Importantly, the misspecification-robust t-tests provide statistically large size corrections not only for the case of misspecified models but also for correctly specified models where the t_c tests are theoretically valid. Moreover, as Tables V and VI illustrate, the effective size correction that the misspecification-robust t-tests perform does not reflect negatively on the power of the tests neither in correctly specified nor in misspecified models.

[5] We keep the notation general (with possibly nonzero κ) because in Tables VII and VIII we also generate data from a multivariate t-distribution with eight degrees of freedom.

[6] These size distortions are somewhat expected for a small T and a relatively large N given the small number of time series observations per moment condition.

Finally, in Tables VII and VIII, we conducted simulations with data drawn from a multivariate t-distribution with eight degrees of freedom. In this case, the inference based on t_m is borderline valid since the degrees of freedom need to be at least eight for our variance approximation under model misspecification to work.

Tables VII and VIII about here

Overall, the simulations suggest that our proposed method continues to work well under this more extreme scenario. While there are some overrejections for the centered t_m test for small sample sizes in misspecified models, they appear to be due primarily to the large number of test assets (moment restrictions) used in our analysis. In simulations that are not reported to conserve space ($N = 10$ and $N = 25$), these size distortions largely disappear. As in the previous tables, the size-adjusted power is similar for the t_c and t_m.

5 Empirical Application

We use our methodology to estimate the parameters γ and λ of three asset-pricing models. The first model is the simple static CAPM with $\tilde{f}_t = [1,\ mkt_t]'$, where mkt is the excess return on the value-weighted stock market index that was defined in the previous section. The CAPM performed well in early tests, but has fared poorly since. The second model is the three-factor specification of Fama and French (FF3, 1993) with $\tilde{f}_t = [1,\ mkt_t,\ smb_t,\ hml_t]'$ that is described in the simulation part of the paper. Finally, we consider the five-factor model of Fama and French (FF5, 2015), an empirical specification that is becoming increasingly popular in the asset-pricing literature. For this model, $\tilde{f}_t = [1,\ mkt_t,\ smb_t,\ hml_t,\ rmw_t,\ cma_t]'$, where rmw (profitability factor) is the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios, and cma (investment factor) is the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios. The test asset returns R_t are (as in the simulation section of the paper) the monthly returns on the value-weighted 25 Fama-French size and book-to-market ranked portfolios and the 10 industry portfolios ($N = 35$) for the period July 1963 – July 2015. As argued in Lewellen, Nagel, and Shanken (2010), the 25 Fama-French portfolios appear to be characterized by a strong factor structure, and the inclusion of the industry portfolios presents a greater challenge to the various asset-pricing models.

Kan and Zhou (2006) argue that the monthly portfolio returns on the 25 Fama-French benchmark portfolios and the three factor portfolios of Fama and French (1993) are well described by a multivariate t distribution with eight degrees of freedom. When we apply the ML methods described in Section 2.1 of Kan and Zhou (2006) to our dataset of 40 financial time series (that is, 35 benchmark portfolios and five factors), we obtain 8.1 as an estimate of the degrees of freedom parameter of the multivariate t-distribution. Additional tests based on Mardia's (1970) measures of multivariate skewness and kurtosis (see Section 1.2 of Kan and Zhou, 2006) also indicate that the number of degrees of freedom of the multivariate t distribution is at least eight in our dataset. Given the outcome of these tests, our regularity assumption of finite eighth moments for CU-GMM does not appear to be at odds with the financial data used in our empirical analysis.

In addition to the invariant ML and CU-GMM estimators, we also present results for the non-invariant generalized least squares (GLS) cross-sectional regression (CSR) and HJD estimators in the beta-pricing and SDF representations, respectively.[7] While inefficient compared to the invariant estimators, CSR and HJD provide useful benchmarks given their numerical stability and popularity in empirical work.

To quantify the degree of misspecification of these models, we performed a model specification test using each of the four estimators. For all models and estimators, the null of correct model specification is strongly rejected with p-values equal to 0.000. To determine whether the models are well identified, we also applied the Cragg and Donald (1997) rank test to the beta-pricing and SDF representations of the models. The results from the rank test suggest that the models are well identified as the test rejects the null of a reduced rank with p-values of 0.000. In summary, these pre-tests provide convincing evidence that the models are misspecified but properly identified.

Hence, to ensure valid statistical inference, the standard errors for the estimated parameters need to be adjusted to account for the additional uncertainty arising from model misspecification. However, it is common practice in empirical work to employ the traditional standard errors derived under the assumption of correct model specification, even when the null of correct model specification is rejected by the data. For this reason, in Table IX, we report t-statistics constructed under the assumption of a correctly specified model (t_c) in addition to the misspecification-robust t-statistics (t_m).

[7] For the GLS CSR estimator and related misspecification-robust t-tests, we refer the reader to Kan, Robotti, and Shanken (2013). For the HJD estimator and related misspecification-robust t-tests, we refer the reader to Kan and Robotti (2009) and Gospodinov, Kan, and Robotti (2013).

Table IX about here

For the beta-pricing model, the ML and CSR estimators (Panel A of Table IX) deliver similar results. In addition, the differences between the t_c and t_m tests are generally small and rarely lead to different conclusions regarding the statistical significance of the individual parameters (the only noticeable exception is the investment factor in FF5 estimated by ML). This is likely due to the fact that all factors are traded and the model misspecification adjustment is typically not large in this scenario (see Kan, Robotti, and Shanken, 2013). It also appears that the misspecification adjustment for the ML standard errors is larger than the corresponding adjustment for the CSR estimator although part of the reason may arise from the asymptotic efficiency (larger t_c statistics) of the MLE under correct model specification.

The model misspecification adjustment is much more pronounced for the CU-GMM estimator in the SDF representation of the model (Panel B of Table IX). For example, consider FF5. When using standard errors constructed under correct model specification, one would conclude that, except for *mkt*, all factors are priced at the 5% significance level. In contrast, incorporating model misspecification in the analysis produces standard errors that are much larger than those constructed under correct model specification. In particular, the new profitability and investment factors of Fama and French (2015) do not appear to be priced at the 5% significance level. The inference based on misspecification-robust standard errors suggests that only *smb* is priced (albeit with much smaller t-statistics) at the 5% significance level. The SDF parameter estimates on all the other risk factors are statistically insignificant. The evidence of pricing in CAPM and FF3 is also much weaker once the uncertainty associated with potential model misspecification is incorporated in the inference procedure. As for the beta-pricing representation, the non-invariant estimator (HJD) in the SDF setup exhibits less sensitivity to model misspecification (see Gospodinov, Kan, and Robotti, 2013), although the evidence of pricing for *mkt*, *hml*, *rmw*, and *cma* in FF5 is even weaker than for CU-GMM.

6 Conclusions

This paper derives the asymptotic variance of the ML and CU-GMM estimators in potentially misspecified models, represented either in beta-pricing or SDF form. This fills an important gap in the literature given the increasing popularity of invariant estimators and the widespread belief that economic models are inherently misspecified. The new expressions for the asymptotic variances of the ML and CU-GMM estimators are explicit and easy-to-use in practice.

We illustrate the importance of using misspecification-robust standard errors of the parameter estimates in the context of various linear asset-pricing models. While, as expected, the misspecification-robust tests deliver impressive improvements when the true model is misspecified, these tests also tend to provide substantial small-sample corrections when the model is correctly specified, especially for CU-GMM. All these size corrections are achieved at no apparent cost associated with loss of power. As a result, the main recommendation that emerges from our analysis is that the proposed misspecification-robust standard errors should always be used in applied work regardless of whether the model is believed (based, for example, on the outcome of a pre-test of overidentifying restrictions) to be correctly specified or misspecified.

Appendix A: Proofs of Lemmas and Theorems

A.1 Preliminary Lemma A.1

LEMMA A.1. *The matrix* $C = H'\Sigma^{-1}H - \frac{S_*}{c_*}\tilde{V}_f^{-1}$ *is a positive definite matrix.*

Proof: Let η be a $(K+2)$-vector, $\tilde{A} = [\mu_R,\ H]'\Sigma^{-1}[\mu_R,\ H]$ and $\tilde{B} = \begin{bmatrix} 1 & 0 & 0_K' \\ 0 & 0 & 0_K' \\ 0_K & 0_K & V_f^{-1} \end{bmatrix}$. Then, we can write the minimization problem in (21) as

$$\min_{\eta} \frac{\eta'\tilde{A}\eta}{\eta'\tilde{B}\eta}. \tag{A.1}$$

By restricting $\eta = [0,\ \gamma']'$, it is easy to see that

$$\min_{\eta} \frac{\eta'\tilde{A}\eta}{\eta'\tilde{B}\eta} < \min_{\eta:\eta=[0,\ \gamma']'} \frac{\eta'\tilde{A}\eta}{\eta'\tilde{B}\eta} = \min_{\gamma} \frac{\gamma' H'\Sigma^{-1}H\gamma}{\gamma'\tilde{V}_f^{-1}\gamma}. \tag{A.2}$$

Note that it is a strict inequality because when the model is identified, the optimal η on the left hand side is chosen such that the first element is normalized to one (that is, nonzero). Since the left hand side is equal to S_*/c_*, the largest eigenvalue of $(H'\Sigma^{-1}H)^{-1}\tilde{V}_f^{-1}$ is less than c_*/S_*, which in turn implies that $H'\Sigma^{-1}H - (S_*/c_*)\tilde{V}_f^{-1}$ is a positive definite matrix. This completes the proof.■

A.2 Proof of Theorem 1

Let

$$\hat{M} = \left[1_N,\ \hat{\beta} + \frac{\hat{m}(\hat{\gamma})\hat{\gamma}_1'\hat{V}_f^{-1}}{1 + \hat{\gamma}_1'\hat{V}_f^{-1}\hat{\gamma}_1}\right], \tag{A.3}$$

where $\hat{m}(\gamma) = \hat{\mu}_R - \hat{H}\gamma$. The first order conditions of (20) and (21) are given by

$$\hat{M}'\hat{\Sigma}^{-1}\hat{m}(\hat{\gamma}) = 0_{K+1}, \tag{A.4}$$

$$M'\Sigma^{-1}m_* = 0_{K+1}, \tag{A.5}$$

where $m_* \equiv m(\gamma_*) = \mu_R - H\gamma_*$. Using a Taylor series expansion, we can write

$$\sqrt{T}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*)] = -\sqrt{T}H(\hat{\gamma} - \gamma_*) + O_p(T^{-\frac{1}{2}}), \tag{A.6}$$

and in addition, using the fact that $\hat{\Sigma} \xrightarrow{p} \Sigma$, $\hat{M} \xrightarrow{p} M$, and $\hat{m}(\gamma_*) \xrightarrow{p} m_*$, we have

$$\begin{aligned}
& \sqrt{T}M'\Sigma^{-1}[\hat{m}(\gamma_*) - m_*] \\
= \; & \sqrt{T}M'\Sigma^{-1}\hat{m}(\gamma_*) \\
= \; & -\sqrt{T}(\hat{M} - M)'\Sigma^{-1}\hat{m}(\gamma_*) - \sqrt{T}\hat{M}'(\hat{\Sigma}^{-1} - \Sigma^{-1})\hat{m}(\gamma_*) - \sqrt{T}\hat{M}'\hat{\Sigma}^{-1}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*)] \\
= \; & -\sqrt{T}(\hat{M} - M)'\Sigma^{-1}m_* - \sqrt{T}M'(\hat{\Sigma}^{-1} - \Sigma^{-1})m_* - \sqrt{T}M'\Sigma^{-1}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*)] + O_p(T^{-\frac{1}{2}}).
\end{aligned} \tag{A.7}$$

Under the normality assumption,

$$\sqrt{T}\text{vec}(\hat{\Sigma}^{-1} - \Sigma^{-1}) \xrightarrow{d} \mathcal{N}\left(0_{N^2}, (\Sigma^{-1} \otimes \Sigma^{-1})(I_{N^2} + K_N)\right), \tag{A.8}$$

where K_N is an $N^2 \times N^2$ commutation matrix. Then, defining $S_* = m_*'\Sigma^{-1}m_*$ and using the fact that $M'\Sigma^{-1}m_* = 0_{K+1}$, we can obtain the limiting distribution of the second term in (A.7) as

$$\sqrt{T}M'(\hat{\Sigma}^{-1} - \Sigma^{-1})m_* \xrightarrow{d} \mathcal{N}(0_{K+1}, S_*M'\Sigma^{-1}M), \tag{A.9}$$

and it is asymptotically independent of $\hat{m}(\gamma_*)$.

For the third term in (A.7), we have

$$-\sqrt{T}M'\Sigma^{-1}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*)] = \sqrt{T}M'\Sigma^{-1}H(\hat{\gamma} - \gamma_*) + O_p(T^{-\frac{1}{2}}) = \sqrt{T}M'\Sigma^{-1}M(\hat{\gamma} - \gamma_*) + O_p(T^{-\frac{1}{2}}), \tag{A.10}$$

where the last equality follows from the fact that $M'\Sigma^{-1}M = M'\Sigma^{-1}H$ because of (A.5).

It remains to expand the first term in (A.7). Writing

$$\begin{aligned}
& \sqrt{T}(\hat{M} - M)'\Sigma^{-1}m_* \\
= \; & \begin{bmatrix} 0 \\ \sqrt{T}(\hat{\beta} - \beta)'\Sigma^{-1}m_* + \left[\frac{\sqrt{T}\hat{V}_f^{-1}\hat{\gamma}_1\hat{m}(\hat{\gamma})'}{1+\hat{\gamma}_1'\hat{V}_f^{-1}\hat{\gamma}_1} - \frac{\sqrt{T}V_f^{-1}\gamma_{1*}m_*'}{1+\gamma_{1*}'V_f^{-1}\gamma_{1*}}\right]\Sigma^{-1}m_* \end{bmatrix} \\
= \; & \sqrt{T}(\hat{H} - H)'\Sigma^{-1}m_* + \begin{bmatrix} 0 \\ \left[\frac{\sqrt{T}\hat{V}_f^{-1}\hat{\gamma}_1\hat{m}(\hat{\gamma})'}{1+\hat{\gamma}_1'\hat{V}_f^{-1}\hat{\gamma}_1} - \frac{\sqrt{T}V_f^{-1}\gamma_{1*}m_*'}{1+\gamma_{1*}'V_f^{-1}\gamma_{1*}}\right]\Sigma^{-1}m_* \end{bmatrix}.
\end{aligned} \tag{A.11}$$

The second term in (A.11) has three sources of randomness. Using the delta method and letting

$c_* = 1 + \gamma_{1*}' V_f^{-1} \gamma_{1*}$, we can approximate the second term in (A.11) as

$$\begin{aligned}
&\left[\frac{\sqrt{T}\hat{V}_f^{-1}\hat{\gamma}_1 \hat{m}(\hat{\gamma})'}{1+\hat{\gamma}_1'\hat{V}_f^{-1}\hat{\gamma}_1} - \frac{\sqrt{T}V_f^{-1}\gamma_{1*}m_*'}{1+\gamma_{1*}'V_f^{-1}\gamma_{1*}}\right]\Sigma^{-1}m_* \\
= \;& \frac{\sqrt{T}V_f^{-1}\gamma_{1*}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*) + \hat{m}(\gamma_*) - m_*]'\Sigma^{-1}m_*}{c_*} \\
&+ \frac{\sqrt{T}(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*}S_*}{c_*} - \frac{V_f^{-1}\gamma_{1*}S_*}{c_*^2}\sqrt{T}\gamma_{1*}'(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*} \\
&+ \frac{\sqrt{T}V_f^{-1}(\hat{\gamma}_1 - \gamma_{1*})S_*}{c_*} - \frac{V_f^{-1}\gamma_{1*}S_*}{c_*^2}2\gamma_{1*}'V_f^{-1}\sqrt{T}(\hat{\gamma}_1 - \gamma_{1*}) + O_p(T^{-\frac{1}{2}}). \qquad \text{(A.12)}
\end{aligned}$$

Combining the second and the third terms in (A.12), we have

$$\frac{\sqrt{T}(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*}S_*}{c_*} - \frac{V_f^{-1}\gamma_{1*}S_*}{c_*^2}\sqrt{T}\gamma_{1*}'(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*} = \frac{S_*}{c_*}A\sqrt{T}(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*}, \qquad \text{(A.13)}$$

where

$$A = I_K - \frac{V_f^{-1}\gamma_{1*}\gamma_{1*}'}{c_*}. \qquad \text{(A.14)}$$

It can be readily shown that

$$\frac{S_*}{c_*}A\sqrt{T}(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*} \xrightarrow{d} \mathcal{N}\left(0_K, \frac{S_*^2}{c_*^2}\left[(c_* - 1)V_f^{-1} + \left(\frac{2}{c_*^2} - 1\right)V_f^{-1}\gamma_{1*}\gamma_{1*}'V_f^{-1}\right]\right), \qquad \text{(A.15)}$$

and this random variable is independent of $\hat{\Sigma}$, $\hat{\mu}_R$, and $\hat{\beta}$. Combining the last two terms in (A.12), we have

$$\left[\begin{array}{c} 0 \\ \frac{\sqrt{T}V_f^{-1}(\hat{\gamma}_1 - \gamma_{1*})S_*}{c_*} - \frac{V_f^{-1}\gamma_{1*}S_*}{c_*^2}2\gamma_{1*}'V_f^{-1}\sqrt{T}(\hat{\gamma}_1 - \gamma_{1*}) \end{array}\right] = \frac{S_*}{c_*}B\sqrt{T}(\hat{\gamma} - \gamma_*), \qquad \text{(A.16)}$$

where

$$B = \left[\begin{array}{cc} 0 & 0_K' \\ 0_K & V_f^{-1} - \frac{2V_f^{-1}\gamma_{1*}\gamma_{1*}'V_f^{-1}}{c_*} \end{array}\right]. \qquad \text{(A.17)}$$

Collecting all these terms, we obtain

$$\begin{aligned}
&\sqrt{T}M'\Sigma^{-1}[\hat{m}(\gamma_*) - m_*] + \sqrt{T}(\hat{H} - H)'\Sigma^{-1}m_* + \sqrt{T}(M - H)'\Sigma^{-1}[\hat{m}(\gamma_*) - m_*] \\
&+ \left[\begin{array}{c} 0 \\ \frac{\sqrt{T}A(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*}S_*}{c_*} \end{array}\right] + \sqrt{T}M'(\hat{\Sigma}^{-1} - \Sigma^{-1})m_* \\
= \;& -\sqrt{T}(M - H)'\Sigma^{-1}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*)] - \frac{\sqrt{T}B(\hat{\gamma} - \gamma_*)S_*}{c_*} - \sqrt{T}M'\Sigma^{-1}[\hat{m}(\hat{\gamma}) - \hat{m}(\gamma_*)] \\
\Rightarrow \;& \sqrt{T}(2M - H)'\Sigma^{-1}[\hat{m}(\gamma_*) - m_*] + \sqrt{T}(\hat{H} - H)'\Sigma^{-1}m_* \\
&+ \left[\begin{array}{c} 0 \\ \frac{\sqrt{T}A(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*}S_*}{c_*} \end{array}\right] + \sqrt{T}M'(\hat{\Sigma}^{-1} - \Sigma^{-1})m_* \\
= \;& \left[(2M - H)'\Sigma^{-1}H - \frac{S_*}{c_*}B\right]\sqrt{T}(\hat{\gamma} - \gamma_*). \qquad \text{(A.18)}
\end{aligned}$$

Using the fact that

$$C = (2M - H)'\Sigma^{-1}H - \frac{S_*}{c_*}B = 2M'\Sigma^{-1}M - H'\Sigma^{-1}H - \frac{S_*}{c_*}B, \tag{A.19}$$

we can then write

$$\begin{aligned} \sqrt{T}(\hat{\gamma} - \gamma_*) \quad \xrightarrow{d} \quad & C^{-1}(2M - H)'\Sigma^{-1}\sqrt{T}[\hat{m}(\gamma_*) - m_*] + \sqrt{T}C^{-1}(\hat{H} - H)'\Sigma^{-1}m_* \\ & + C^{-1}\left[\begin{array}{c} 0 \\ \frac{\sqrt{T}A(\hat{V}_f^{-1} - V_f^{-1})\gamma_{1*}S_*}{c_*} \end{array}\right] + C^{-1}M'\sqrt{T}(\hat{\Sigma}^{-1} - \Sigma^{-1})m_*. \end{aligned} \tag{A.20}$$

The last two terms in (A.20) are independent of each other and also independent of the first two terms, and their variances are given by

$$\frac{S_*^2}{c_*^2}C^{-1}\left[\begin{array}{cc} 0 & 0_K' \\ 0_K & (\gamma_{1*}'V_f^{-1}\gamma_{1*})V_f^{-1} + \left(\frac{2}{c_*^2} - 1\right)V_f^{-1}\gamma_{1*}\gamma_{1*}'V_f^{-1} \end{array}\right]C^{-1} + S_*C^{-1}M'\Sigma^{-1}MC^{-1}. \tag{A.21}$$

Since

$$H'\Sigma^{-1}H - M'\Sigma^{-1}M = \frac{S_*}{c_*^2}\left[\begin{array}{cc} 0 & 0_K' \\ 0_K & V_f^{-1}\gamma_{1*}\gamma_{1*}'V_f^{-1} \end{array}\right], \tag{A.22}$$

we can write

$$C = H'\Sigma^{-1}H - \frac{S_*}{c_*}\tilde{V}_f^{-1}. \tag{A.23}$$

Given that

$$\hat{m}(\gamma_*) - m_* = \hat{\alpha} - \alpha - (\hat{\beta} - \beta)\phi_* + \hat{\beta}(\hat{\mu}_f - \mu_f), \tag{A.24}$$

where $\phi_* = \gamma_{1*} - \mu_f$, we obtain

$$\sqrt{T}[\hat{m}(\gamma_*) - m_*] \xrightarrow{d} \mathcal{N}\left(0_N, (1 + \gamma_{1*}'V_f^{-1}\gamma_{1*})\Sigma + H\tilde{V}_fH'\right). \tag{A.25}$$

Hence, the asymptotic variance of the first term in (A.20) is

$$\begin{aligned} & c_*C^{-1}(2M - H)'\Sigma^{-1}(2M - H)C^{-1} + C^{-1}(2M - H)'\Sigma^{-1}H\tilde{V}_fH'\Sigma^{-1}(2M - H)C^{-1} \\ & = c_*C^{-1}H'\Sigma^{-1}HC^{-1} + C^{-1}(2M - H)'\Sigma^{-1}H\tilde{V}_fH'\Sigma^{-1}(2M - H)C^{-1}, \end{aligned} \tag{A.26}$$

where the invertibility of C follows from Lemma A.1. Using that under Assumption MLE.A,

$$\sqrt{T}\text{vec}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0_{NK}, V_f^{-1} \otimes \Sigma\right), \tag{A.27}$$

we obtain the asymptotic variance of the second term in (A.20) as

$$S_*C^{-1}\tilde{V}_f^{-1}C^{-1}. \tag{A.28}$$

Let $\tilde{B} = \sqrt{T}[\hat{\alpha} - \alpha,\ \hat{\beta} - \beta]$ and $\tilde{b} = \text{vec}(\tilde{B})$. We have

$$\tilde{b} \xrightarrow{d} \mathcal{N}\left(0_{N(K+1)}, \left[\begin{array}{cc} 1 + \mu_f' V_f^{-1}\mu_f & -\mu_f' V_f^{-1} \\ -V_f^{-1}\mu_f & V_f^{-1} \end{array}\right] \otimes \Sigma\right). \tag{A.29}$$

Then, using

$$\begin{aligned} E[\sqrt{T}[\hat{m}(\gamma_*) - m_*]m_*'\Sigma^{-1}\sqrt{T}(\hat{\beta} - \beta)] &= E[\sqrt{T}[\hat{\alpha} - \alpha - (\hat{\beta} - \beta)\phi_*]m_*'\Sigma^{-1}\sqrt{T}(\hat{\beta} - \beta)] \\ &= E\left[\tilde{B}\left[\begin{array}{c} 1 \\ -\phi_* \end{array}\right] m_*'\Sigma^{-1}\tilde{B}\left[\begin{array}{c} 0_K' \\ I_K \end{array}\right]\right] \\ &= E\left[([1,\ -\phi_*'] \otimes I_N)\tilde{b}\tilde{b}'\left(\left[\begin{array}{c} 0_K' \\ I_K \end{array}\right] \otimes \Sigma^{-1}m_*\right)\right] \\ &= -\gamma_{1*}' V_f^{-1} \otimes m_* \\ &= -m_*\gamma_{1*}' V_f^{-1}, \end{aligned} \tag{A.30}$$

we obtain the asymptotic variance between the first and second terms in (A.20) as

$$\begin{aligned} C^{-1}(2M - H)'\Sigma^{-1}m_*[0,\ -\gamma_{1*}' V_f^{-1}]C^{-1} &= C^{-1}H'\Sigma^{-1}m_*[0,\ \gamma_{1*}' V_f^{-1}]C^{-1} \\ &= c_* C^{-1}H'\Sigma^{-1}(M - H)C^{-1}. \end{aligned} \tag{A.31}$$

Combining all the results, we obtain

$$\sqrt{T}(\hat{\gamma} - \gamma_*) \xrightarrow{d} \mathcal{N}(0_{K+1}, \Omega), \tag{A.32}$$

where

$$\begin{aligned} \Omega &= c_* C^{-1}H'\Sigma^{-1}HC^{-1} \\ &\quad + C^{-1}(2M - H)'\Sigma^{-1}H\tilde{V}_f H'\Sigma^{-1}(2M - H)C^{-1} \\ &\quad + S_* C^{-1}\tilde{V}_f^{-1}C^{-1} + c_* C^{-1}H'\Sigma^{-1}(M - H)C^{-1} \\ &\quad + c_* C^{-1}(M - H)'\Sigma^{-1}HC^{-1} \\ &\quad + \frac{S_*^2}{c_*^2}C^{-1}\left[\begin{array}{cc} 0 & 0_K' \\ 0_K & (\gamma_{1*}' V_f^{-1}\gamma_{1*})V_f^{-1} + \left(\frac{2}{c_*^2} - 1\right)V_f^{-1}\gamma_{1*}\gamma_{1*}' V_f^{-1} \end{array}\right]C^{-1} \\ &\quad + S_* C^{-1}M'\Sigma^{-1}MC^{-1}. \end{aligned} \tag{A.33}$$

Let $C_1 = 2M'\Sigma^{-1}M - H'\Sigma^{-1}H$. Then, we can write

$$\begin{aligned} \Omega &= c_* C^{-1}C_1 C^{-1} + C^{-1}C_1\tilde{V}_f C_1 C^{-1} + S_* C^{-1}\tilde{V}_f^{-1}C^{-1} + S_* C^{-1}M'\Sigma^{-1}MC^{-1} \\ &\quad + \frac{S_*^2}{c_*^2}C^{-1}\left[\begin{array}{cc} 0 & 0_K' \\ 0_K & (\gamma_{1*}' V_f^{-1}\gamma_{1*})V_f^{-1} + \left(\frac{2}{c_*^2} - 1\right)V_f^{-1}\gamma_{1*}\gamma_{1*}' V_f^{-1} \end{array}\right]C^{-1}. \end{aligned} \tag{A.34}$$

Using the identities

$$M'\Sigma^{-1}M = C_1 + \frac{S_*}{c_*^2}\begin{bmatrix} 0 & 0_K' \\ 0_K & V_f^{-1}\gamma_{1*}\gamma_{1*}'V_f^{-1} \end{bmatrix}, \tag{A.35}$$

$$H'\Sigma^{-1}H - C_1 = \frac{2S_*}{c_*^2}\begin{bmatrix} 0 & 0_K' \\ 0_K & V_f^{-1}\gamma_{1*}\gamma_{1*}'V_f^{-1} \end{bmatrix}, \tag{A.36}$$

we can write Ω as

$$\Omega = C^{-1}\left\{c_*C_1 + C_1\tilde{V}_fC_1 + S_*\left[\left(1-\frac{1}{c_*^2}\right)C_1 + \left(1+\frac{S_*(c_*-1)}{c_*^2}\right)\tilde{V}_f^{-1} + \frac{1}{c_*^2}H'\Sigma^{-1}H\right]\right\}C^{-1}. \tag{A.37}$$

This completes the proof.■

A.3 Proof of Theorem 2

A mean value expansion of $\bar{s}_T(\hat{\theta})$ about θ_* yields

$$0_{N+K+1} = \bar{s}_T(\theta_*) + H_T(\tilde{\theta})(\hat{\theta}-\theta_*) \tag{A.38}$$

or

$$\sqrt{T}(\hat{\theta}-\theta_*) = -\left[H_T\left(\tilde{\theta}\right)\right]^{-1}\sqrt{T}\bar{s}_T\left(\theta_*\right), \tag{A.39}$$

where $H_T(\theta) = \frac{1}{T}\sum_{t=1}^T h_t(\theta)$ with $h_t(\theta) = (\partial/\partial\theta')s_t(\theta)$, and $\tilde{\theta}$ is an intermediate point on the line segment joining $\hat{\theta}$ and θ_*. More specifically,

$$h_t(\theta) = -\begin{bmatrix} (e_t(\lambda)-e(\lambda))(e_t(\lambda)-e(\lambda))' & w_t(\theta)g_t + (e_t(\lambda)-e(\lambda))\rho'(g_t-G) \\ w_t(\theta)g_t' + (g_t-G)'\rho(e_t(\lambda)-e(\lambda))' & (g_t-G)'\rho\rho'(g_t-G) \end{bmatrix}, \tag{A.40}$$

where $w_t(\theta) = [1+\rho'(e_t(\lambda)-e(\lambda))]$. Our regularity conditions ensure that

$$\sqrt{T}\bar{s}_T\left(\theta_*\right) \xrightarrow{d} \mathcal{N}(0_{N+K+1}, S), \tag{A.41}$$

and

$$\sqrt{T}(\hat{\theta}-\theta_*) \xrightarrow{d} \mathcal{N}(0_{N+K+1}, H^{-1}S(H')^{-1}), \tag{A.42}$$

where $S = E[s_t(\theta_*)s_t(\theta_*)']$,

$$H \equiv E[H_T(\theta_*)] = \begin{bmatrix} V & B \\ B' & C \end{bmatrix}, \tag{A.43}$$

and V, B, and C are defined in the text.

To derive the explicit expression for the asymptotic variance matrix of $\hat{\theta}$ in Theorem 2, we write

$$H^{-1}S(H')^{-1} = E[l_tl_t'], \tag{A.44}$$

where

$$l_t \equiv \begin{bmatrix} l_{1t} \\ l_{2t} \end{bmatrix} = H^{-1} s_t(\theta_*). \tag{A.45}$$

From the definition of H in (A.43), we can use the formula for the inverse of a partitioned matrix to obtain

$$H^{-1} = \begin{bmatrix} V^{-1}(I_N + B\tilde{H}B'V^{-1}) & -V^{-1}B\tilde{H} \\ -\tilde{H}'B'V^{-1} & \tilde{H} \end{bmatrix}, \tag{A.46}$$

where $\tilde{H} = (C - B'V^{-1}B)^{-1}$. Observe that $C - B'V^{-1}B$ is the Schur complement of V in H and its invertibility follows from our assumptions and the properties of Schur complements. Using (A.46) and (32), we can express l_{1t} and l_{2t} as

$$l_{1t} = V^{-1}\left[w_t(\theta_*)e_t(\lambda_*) - Bl_{2t}\right], \tag{A.47}$$

$$l_{2t} = \tilde{H}w_t(\theta_*)\left[g_t'\rho_* - B'V^{-1}e_t(\lambda_*)\right]. \tag{A.48}$$

This delivers the desired result.■

A.4 Proof of Theorem 3

Note that in the case of nonlinear moment conditions, the upper-left, upper-right, lower-left, and lower-right blocks of the $h_t(\theta)$ matrix are given by

$$-(e_t(\lambda) - e(\lambda))(e_t(\lambda) - e(\lambda))', \tag{A.49}$$

$$-\left[w_t(\theta)g_t(\lambda) + (e_t(\lambda) - e(\lambda))\rho'(g_t(\lambda) - G(\lambda))\right], \tag{A.50}$$

$$-\left[w_t(\theta)g_t(\lambda)' + (g_t(\lambda) - G(\lambda))\rho(e_t(\lambda) - e(\lambda))'\right], \tag{A.51}$$

$$-\left[w_t(\theta)(I_{K+1} \otimes \rho')g_t^{(2)}(\lambda) + (g_t(\lambda) - G(\lambda))'\rho\rho'(g_t(\lambda) - G(\lambda))\right], \tag{A.52}$$

respectively. The rest of the proof follows similar arguments as those in the proof of Theorem 2.■

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

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

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

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

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

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 highly 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 SDF setups.[8] By augmenting $\bar{e}(\lambda)$ in the SDF representation with additional (just-identified) moment conditions for μ_f, V_f, and β, the CU-GMM estimate of the augmented parameter vector $\theta = [\lambda_0,\ \lambda_1',\ \beta_1',\ldots,\beta_K',\ \mu_f',\ \text{vech}(V_f)']'$ becomes numerically identical to the CU-GMM estimate of ϕ in the beta-pricing model. However, the estimation of θ can be performed in a sequential manner which offers substantial computational advantages. The following theorem presents a general result for this sequential estimation.

Theorem 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

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

[8] Shanken and Zhou (2007) show that under some particular Kronecker structure for the weighting matrix $\hat{W}_g$, the GMM estimator of the beta-pricing model is numerically identical to the MLE.

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

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

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

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

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 Theorem B.1.■

Theorem 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 Theorem 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 highly desirable when performing numerical optimization. The following lemma demonstrates how to solve for $\hat{\theta}_2$ after $\tilde{\theta}_1$ is obtained from the smaller system.

Lemma B.1. *Let*

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

and

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

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

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

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,*

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

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.1.■

Lemma B.1 shows that when g_{2t} is linear in θ_2, $\hat{\theta}_2$ has a closed-form solution. When $h_{2t}(\theta_1) = I_{K_2}$, which is the case of the asset-pricing models considered in this paper, we have

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

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.

References

[1] Antoine, B., H. Bonnal, and E. Renault (2007), On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood, *Journal of Econometrics* 138, 461–487.

[2] Canova, F. (1994), Statistical inference in calibrated models, *Journal of Applied Econometrics* 9 (S1), S123–S144.

[3] Cragg, J. G., and S. G. Donald (1997), Inferring the rank of a matrix, *Journal of Econometrics* 76, 223–250.

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

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

[6] Gallant, A. R., and H. White (1988), *A Unified Theory of Estimation and Inference in Nonlinear Models*, Basil Blackwell, Oxford.

[7] Gospodinov, N., R. Kan, and C. Robotti (2013), Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics* 173, 108–125.

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

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

[10] Hall, A. R., and A. Inoue (2003), The large sample behaviour of the generalized method of moments estimator in misspecified models, *Journal of Econometrics* 114, 361–394.

[11] Hansen, L. P. (1982), Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

[12] Hansen, L. P., J. Heaton, and A. Yaron (1996), Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.

[13] Hansen, L. P., and R. Jagannathan (1997), Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

[14] Kan, R., and C. Robotti (2009), Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies* 22, 3449–3490.

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

[16] Kan, R., and G. Zhou (2006), Modeling non-normality using multivariate t: Implications for asset pricing, mimeo.

[17] Kitamura, Y. (1998), Comparing misspecified dynamic econometric models using nonparametric likelihood, mimeo.

[18] Lewellen, J. W., S. Nagel, and J. Shanken (2010), A skeptical appraisal of asset pricing tests, *Journal of Financial Economics* 96, 175–194.

[19] Maasoumi, E. (1990), How to live with misspecification if you must, *Journal of Econometrics* 44, 67–86.

[20] Maasoumi, E., and P. C. B. Phillips (1982), On the behavior of inconsistent instrumental variable estimators, *Journal of Econometrics* 19, 183–201.

[21] Mardia, K. V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika* 57, 519–530.

[22] Newey, W. K., and R. J. Smith (2004), Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* 72, 219–255.

[23] Peñaranda, F., and E. Sentana (2015) A unifying approach to the empirical evaluation of asset pricing models, *Review of Economics and Statistics* 97, 412–435.

[24] Schennach, S. M. (2007), Point estimation with exponentially tilted empirical likelihood, *Annals of Statistics* 35, 634–672.

[25] Seber, G. A. F. (1984), *Multivariate Observations*, John Wiley & Sons, New York.

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

[27] Smith, R. J. (2011), GEL criteria for moment condition models, *Econometric Theory* 27, 1192–1235.

[28] Watson, M. W. (1993), Measures of fit for calibrated models, *Journal of Political Economy* 101, 1011–1041.

[29] White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1–25.

[30] White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University. Press, New York.

Table I
Size and power properties of MLE t-tests under normality: Correctly specified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \gamma_{1,i} = \gamma_{*1,i}$ and $H_0 : \gamma_{1,i} = 0$ ($i = 1, \ldots, K$) for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The true values of the risk premium parameters are $\gamma_{*1,1} = -0.2829$, $\gamma_{*1,2} = 0.2196$, and $\gamma_{*1,3} = 0.2801$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate normally distributed.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
				Panel A: t_c			
$t_c(\hat{\gamma}_{1,1})$	300	0.137	0.077	0.020	0.199	0.119	0.035
	600	0.119	0.063	0.015	0.296	0.198	0.071
	1200	0.109	0.056	0.012	0.480	0.356	0.160
	3600	0.102	0.052	0.010	0.873	0.795	0.592
$t_c(\hat{\gamma}_{1,2})$	300	0.102	0.051	0.010	0.339	0.234	0.092
	600	0.102	0.052	0.011	0.535	0.406	0.195
	1200	0.102	0.051	0.010	0.789	0.689	0.449
	3600	0.102	0.051	0.010	0.996	0.989	0.954
$t_c(\hat{\gamma}_{1,3})$	300	0.103	0.051	0.011	0.509	0.387	0.176
	600	0.102	0.052	0.011	0.762	0.652	0.410
	1200	0.101	0.050	0.011	0.955	0.917	0.774
	3600	0.100	0.050	0.010	1.000	1.000	0.999
				Panel B: t_m			
$t_m(\hat{\gamma}_{1,1})$	300	0.089	0.043	0.008	0.199	0.119	0.035
	600	0.095	0.046	0.009	0.296	0.199	0.072
	1200	0.096	0.048	0.009	0.480	0.356	0.160
	3600	0.098	0.049	0.009	0.873	0.795	0.592
$t_m(\hat{\gamma}_{1,2})$	300	0.102	0.051	0.010	0.338	0.234	0.093
	600	0.101	0.052	0.011	0.535	0.406	0.195
	1200	0.101	0.050	0.010	0.789	0.689	0.449
	3600	0.102	0.051	0.010	0.996	0.989	0.954
$t_m(\hat{\gamma}_{1,3})$	300	0.101	0.051	0.011	0.509	0.386	0.176
	600	0.102	0.052	0.011	0.762	0.652	0.410
	1200	0.100	0.050	0.011	0.955	0.917	0.774
	3600	0.100	0.050	0.010	1.000	1.000	0.999

Table II
Size and power properties of MLE t-tests under normality: Misspecified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \gamma_{1,i} = \gamma_{*1,i}$ and $H_0 : \gamma_{1,i} = 0$ $(i = 1, \ldots, K)$ for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The pseudo-true values of the risk premium parameters are $\gamma_{*1,1} = -0.7463$, $\gamma_{*1,2} = 0.2525$, and $\gamma_{*1,3} = 0.3307$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate normally distributed.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
		Panel A: t_c					
$t_c(\hat{\gamma}_{1,1})$	300	0.242	0.164	0.068	0.475	0.349	0.153
	600	0.216	0.141	0.052	0.748	0.636	0.398
	1200	0.203	0.129	0.045	0.954	0.916	0.776
	3600	0.196	0.120	0.041	1.000	1.000	1.000
$t_c(\hat{\gamma}_{1,2})$	300	0.103	0.052	0.010	0.406	0.291	0.125
	600	0.103	0.053	0.011	0.635	0.507	0.271
	1200	0.103	0.052	0.011	0.880	0.802	0.590
	3600	0.103	0.051	0.010	1.000	0.999	0.990
$t_c(\hat{\gamma}_{1,3})$	300	0.105	0.053	0.012	0.621	0.497	0.262
	600	0.105	0.054	0.011	0.868	0.787	0.568
	1200	0.103	0.052	0.011	0.988	0.975	0.908
	3600	0.102	0.052	0.011	1.000	1.000	1.000
		Panel B: t_m					
$t_m(\hat{\gamma}_{1,1})$	300	0.095	0.047	0.009	0.476	0.354	0.160
	600	0.097	0.048	0.009	0.744	0.633	0.398
	1200	0.098	0.048	0.009	0.953	0.913	0.772
	3600	0.097	0.048	0.010	1.000	1.000	0.999
$t_m(\hat{\gamma}_{1,2})$	300	0.102	0.051	0.010	0.406	0.291	0.125
	600	0.101	0.052	0.011	0.635	0.507	0.271
	1200	0.101	0.051	0.010	0.880	0.802	0.590
	3600	0.102	0.051	0.010	1.000	0.999	0.990
$t_m(\hat{\gamma}_{1,3})$	300	0.101	0.051	0.011	0.621	0.498	0.261
	600	0.102	0.052	0.011	0.868	0.787	0.568
	1200	0.101	0.050	0.010	0.988	0.975	0.908
	3600	0.100	0.051	0.010	1.000	1.000	1.000

Table III
Size and power properties of MLE t-tests under non-normality: Correctly specified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \gamma_{1,i} = \gamma_{*1,i}$ and $H_0 : \gamma_{1,i} = 0$ ($i = 1, \ldots, K$) for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The true values of the risk premium parameters are $\gamma_{*1,1} = -0.2829$, $\gamma_{*1,2} = 0.2196$, and $\gamma_{*1,3} = 0.2801$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate t-distributed. The number of degrees of freedom of the t-distribution is set equal to eight.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
				Panel A: t_c			
$t_c(\hat{\gamma}_{1,1})$	300	0.135	0.074	0.019	0.196	0.117	0.035
	600	0.117	0.063	0.015	0.295	0.195	0.068
	1200	0.111	0.056	0.012	0.470	0.352	0.158
	3600	0.103	0.053	0.011	0.871	0.791	0.575
$t_c(\hat{\gamma}_{1,2})$	300	0.102	0.052	0.011	0.343	0.235	0.090
	600	0.101	0.050	0.010	0.541	0.417	0.203
	1200	0.100	0.050	0.010	0.797	0.696	0.453
	3600	0.100	0.050	0.010	0.996	0.989	0.956
$t_c(\hat{\gamma}_{1,3})$	300	0.103	0.052	0.011	0.510	0.386	0.181
	600	0.103	0.051	0.010	0.763	0.656	0.416
	1200	0.100	0.050	0.011	0.956	0.918	0.774
	3600	0.100	0.050	0.010	1.000	1.000	0.999
				Panel B: t_m			
$t_m(\hat{\gamma}_{1,1})$	300	0.089	0.044	0.008	0.196	0.117	0.035
	600	0.095	0.046	0.009	0.296	0.195	0.069
	1200	0.099	0.048	0.009	0.470	0.352	0.158
	3600	0.099	0.050	0.010	0.871	0.791	0.574
$t_m(\hat{\gamma}_{1,2})$	300	0.102	0.051	0.010	0.343	0.235	0.090
	600	0.100	0.050	0.010	0.541	0.417	0.203
	1200	0.099	0.050	0.010	0.797	0.696	0.453
	3600	0.100	0.050	0.010	0.996	0.989	0.956
$t_m(\hat{\gamma}_{1,3})$	300	0.102	0.051	0.011	0.510	0.386	0.181
	600	0.102	0.051	0.010	0.763	0.656	0.416
	1200	0.100	0.050	0.011	0.956	0.918	0.774
	3600	0.100	0.050	0.010	1.000	1.000	0.999

Table IV
Size and power properties of MLE t-tests under non-normality: Misspecified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \gamma_{1,i} = \gamma_{*1,i}$ and $H_0 : \gamma_{1,i} = 0$ $(i = 1, \ldots, K)$ for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The pseudo-true values of the risk premium parameters are $\gamma_{*1,1} = -0.7463$, $\gamma_{*1,2} = 0.2525$, and $\gamma_{*1,3} = 0.3307$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate t-distributed. The number of degrees of freedom of the t-distribution is set equal to eight.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
		Panel A: t_c					
$t_c(\hat{\gamma}_{1,1})$	300	0.248	0.169	0.070	0.455	0.332	0.146
	600	0.226	0.150	0.058	0.723	0.609	0.364
	1200	0.216	0.139	0.052	0.943	0.895	0.733
	3600	0.209	0.135	0.049	1.000	1.000	0.999
$t_c(\hat{\gamma}_{1,2})$	300	0.104	0.052	0.011	0.408	0.292	0.124
	600	0.103	0.051	0.011	0.638	0.516	0.282
	1200	0.101	0.051	0.011	0.882	0.808	0.594
	3600	0.101	0.050	0.010	0.999	0.998	0.990
$t_c(\hat{\gamma}_{1,3})$	300	0.106	0.054	0.012	0.621	0.498	0.267
	600	0.105	0.053	0.011	0.868	0.790	0.575
	1200	0.102	0.052	0.011	0.988	0.974	0.909
	3600	0.102	0.052	0.010	1.000	1.000	1.000
		Panel B: t_m					
$t_m(\hat{\gamma}_{1,1})$	300	0.103	0.052	0.011	0.456	0.335	0.148
	600	0.106	0.054	0.011	0.720	0.608	0.367
	1200	0.108	0.056	0.012	0.941	0.893	0.731
	3600	0.111	0.057	0.013	1.000	1.000	0.999
$t_m(\hat{\gamma}_{1,2})$	300	0.102	0.051	0.010	0.408	0.292	0.123
	600	0.101	0.050	0.010	0.638	0.516	0.282
	1200	0.100	0.050	0.010	0.882	0.808	0.594
	3600	0.100	0.050	0.010	0.999	0.998	0.990
$t_m(\hat{\gamma}_{1,3})$	300	0.102	0.051	0.011	0.621	0.498	0.267
	600	0.102	0.051	0.010	0.868	0.791	0.574
	1200	0.100	0.051	0.011	0.988	0.974	0.909
	3600	0.101	0.050	0.010	1.000	1.000	1.000

Table V
Size and power properties of CU-GMM *t*-tests under normality: Correctly specified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \lambda_{1,i} = \lambda_{*1,i}$ and $H_0 : \lambda_{1,i} = 0$ $(i = 1, \ldots, K)$ for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The true values of the SDF parameters are $\lambda_{*1,1} = 1.4497$, $\lambda_{*1,2} = -3.2283$, and $\lambda_{*1,3} = -3.1090$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate normally distributed.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
		Panel A: t_c					
$t_c(\hat{\lambda}_{1,1})$	300	0.273	0.192	0.088	0.161	0.092	0.023
	600	0.167	0.100	0.031	0.255	0.163	0.054
	1200	0.130	0.071	0.017	0.414	0.296	0.124
	3600	0.108	0.055	0.012	0.806	0.708	0.469
$t_c(\hat{\lambda}_{1,2})$	300	0.241	0.163	0.067	0.406	0.279	0.100
	600	0.159	0.093	0.027	0.696	0.576	0.325
	1200	0.127	0.068	0.016	0.932	0.880	0.715
	3600	0.109	0.057	0.012	1.000	1.000	0.998
$t_c(\hat{\lambda}_{1,3})$	300	0.240	0.162	0.067	0.341	0.226	0.080
	600	0.157	0.091	0.026	0.602	0.474	0.247
	1200	0.125	0.067	0.016	0.868	0.788	0.571
	3600	0.107	0.055	0.012	0.999	0.998	0.988
		Panel B: t_m					
$t_m(\hat{\lambda}_{1,1})$	300	0.076	0.035	0.005	0.166	0.095	0.027
	600	0.089	0.042	0.007	0.257	0.164	0.055
	1200	0.095	0.046	0.009	0.414	0.298	0.125
	3600	0.098	0.048	0.010	0.806	0.709	0.469
$t_m(\hat{\lambda}_{1,2})$	300	0.090	0.042	0.008	0.417	0.295	0.116
	600	0.103	0.051	0.010	0.700	0.581	0.337
	1200	0.103	0.051	0.010	0.932	0.881	0.718
	3600	0.102	0.052	0.011	1.000	1.000	0.998
$t_m(\hat{\lambda}_{1,3})$	300	0.089	0.041	0.006	0.351	0.243	0.093
	600	0.100	0.050	0.009	0.607	0.479	0.253
	1200	0.102	0.050	0.010	0.868	0.789	0.575
	3600	0.100	0.050	0.010	0.999	0.998	0.988

Table VI
Size and power properties of CU-GMM *t*-tests under normality: Misspecified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \lambda_{1,i} = \lambda_{*1,i}$ and $H_0 : \lambda_{1,i} = 0$ ($i = 1, \ldots, K$) for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The pseudo-true values of the SDF parameters are $\lambda_{*1,1} = 7.3017$, $\lambda_{*1,2} = -7.3402$, and $\lambda_{*1,3} = -3.5071$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate normally distributed.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
		Panel A: t_c					
$t_c(\hat{\lambda}_{1,1})$	300	0.682	0.622	0.509	0.346	0.180	0.010
	600	0.600	0.528	0.401	0.631	0.480	0.143
	1200	0.536	0.461	0.330	0.905	0.836	0.604
	3600	0.490	0.412	0.279	1.000	0.999	0.996
$t_c(\hat{\lambda}_{1,2})$	300	0.531	0.454	0.323	0.466	0.328	0.113
	600	0.442	0.359	0.227	0.829	0.721	0.441
	1200	0.386	0.302	0.176	0.990	0.976	0.903
	3600	0.346	0.262	0.141	1.000	1.000	1.000
$t_c(\hat{\lambda}_{1,3})$	300	0.551	0.478	0.352	0.163	0.097	0.028
	600	0.456	0.376	0.249	0.246	0.152	0.044
	1200	0.395	0.314	0.187	0.443	0.316	0.127
	3600	0.355	0.271	0.148	0.876	0.800	0.592
		Panel B: t_m					
$t_m(\hat{\lambda}_{1,1})$	300	0.143	0.076	0.018	0.355	0.242	0.080
	600	0.108	0.053	0.011	0.607	0.489	0.258
	1200	0.095	0.045	0.008	0.875	0.804	0.608
	3600	0.093	0.045	0.008	0.999	0.998	0.992
$t_m(\hat{\lambda}_{1,2})$	300	0.099	0.047	0.008	0.482	0.361	0.169
	600	0.088	0.040	0.006	0.819	0.730	0.508
	1200	0.091	0.041	0.006	0.987	0.973	0.912
	3600	0.095	0.046	0.008	1.000	1.000	1.000
$t_m(\hat{\lambda}_{1,3})$	300	0.116	0.060	0.012	0.177	0.109	0.034
	600	0.093	0.045	0.008	0.268	0.181	0.071
	1200	0.090	0.042	0.007	0.456	0.347	0.171
	3600	0.096	0.047	0.009	0.873	0.799	0.600

Table VII
Size and power properties of CU-GMM t-tests under non-normality: Correctly specified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \lambda_{1,i} = \lambda_{*1,i}$ and $H_0 : \lambda_{1,i} = 0$ ($i = 1, \ldots, K$) for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The true values of the SDF parameters are $\lambda_{*1,1} = 1.4497$, $\lambda_{*1,2} = -3.2283$, and $\lambda_{*1,3} = -3.1090$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate t-distributed. The number of degrees of freedom of the t-distribution is set equal to eight.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
			Panel A: t_c				
$t_c(\hat{\lambda}_{1,1})$	300	0.348	0.265	0.144	0.156	0.085	0.021
	600	0.204	0.132	0.048	0.248	0.157	0.050
	1200	0.149	0.085	0.023	0.410	0.291	0.120
	3600	0.116	0.061	0.013	0.800	0.700	0.469
$t_c(\hat{\lambda}_{1,2})$	300	0.314	0.230	0.115	0.371	0.248	0.089
	600	0.200	0.127	0.044	0.677	0.548	0.294
	1200	0.149	0.085	0.024	0.926	0.868	0.686
	3600	0.117	0.062	0.014	1.000	1.000	0.998
$t_c(\hat{\lambda}_{1,3})$	300	0.311	0.229	0.115	0.312	0.200	0.066
	600	0.195	0.124	0.043	0.581	0.447	0.217
	1200	0.145	0.083	0.023	0.861	0.773	0.539
	3600	0.115	0.061	0.014	0.999	0.998	0.987
			Panel B: t_m				
$t_m(\hat{\lambda}_{1,1})$	300	0.097	0.050	0.011	0.157	0.089	0.023
	600	0.089	0.043	0.008	0.247	0.156	0.051
	1200	0.095	0.046	0.009	0.411	0.294	0.124
	3600	0.100	0.050	0.009	0.800	0.700	0.469
$t_m(\hat{\lambda}_{1,2})$	300	0.109	0.057	0.012	0.373	0.252	0.095
	600	0.110	0.056	0.012	0.680	0.552	0.303
	1200	0.110	0.057	0.012	0.928	0.869	0.689
	3600	0.106	0.054	0.012	1.000	1.000	0.998
$t_m(\hat{\lambda}_{1,3})$	300	0.106	0.056	0.012	0.317	0.207	0.073
	600	0.107	0.054	0.011	0.584	0.454	0.225
	1200	0.107	0.055	0.012	0.862	0.774	0.543
	3600	0.104	0.053	0.011	0.999	0.998	0.987

Table VIII
Size and power properties of CU-GMM *t*-tests under non-normality: Misspecified model

The table presents the actual probabilities of rejection for the t-tests of $H_0 : \lambda_{1,i} = \lambda_{*1,i}$ and $H_0 : \lambda_{1,i} = 0$ ($i = 1, \ldots, K$) for different levels of significance. The model includes a constant term and three risk factors (FF3 specification). The pseudo-true values of the SDF parameters are $\lambda_{*1,1} = 10.5705$, $\lambda_{*1,2} = -9.2722$, and $\lambda_{*1,3} = -3.1037$. Panel A presents the empirical size and power for t-tests that are constructed assuming that the model is correctly specified (t_c). Panel B reports the empirical size and power for misspecification-robust t-tests (t_m). The factors and the returns are multivariate t-distributed. The number of degrees of freedom of the t-distribution is set equal to eight.

		Size			Power		
		Level of Significance			Level of Significance		
	T	10%	5%	1%	10%	5%	1%
		Panel A: t_c					
$t_c(\hat{\lambda}_{1,1})$	300	0.713	0.660	0.558	0.269	0.077	0.001
	600	0.673	0.616	0.508	0.540	0.328	0.015
	1200	0.660	0.599	0.491	0.809	0.660	0.229
	3600	0.673	0.616	0.511	0.995	0.987	0.908
$t_c(\hat{\lambda}_{1,2})$	300	0.588	0.517	0.391	0.427	0.276	0.066
	600	0.527	0.451	0.323	0.733	0.581	0.247
	1200	0.502	0.424	0.293	0.957	0.907	0.660
	3600	0.515	0.438	0.310	1.000	1.000	0.997
$t_c(\hat{\lambda}_{1,3})$	300	0.620	0.555	0.438	0.150	0.087	0.024
	600	0.565	0.493	0.371	0.174	0.098	0.027
	1200	0.528	0.453	0.330	0.241	0.139	0.033
	3600	0.519	0.444	0.318	0.483	0.333	0.091
		Panel B: t_m					
$t_m(\hat{\lambda}_{1,1})$	300	0.187	0.123	0.057	0.286	0.138	0.010
	600	0.151	0.095	0.042	0.495	0.317	0.049
	1200	0.136	0.084	0.034	0.728	0.575	0.194
	3600	0.135	0.083	0.031	0.965	0.925	0.708
$t_m(\hat{\lambda}_{1,2})$	300	0.125	0.069	0.019	0.422	0.297	0.101
	600	0.106	0.057	0.016	0.701	0.573	0.289
	1200	0.104	0.056	0.016	0.918	0.858	0.623
	3600	0.113	0.064	0.020	0.996	0.992	0.969
$t_m(\hat{\lambda}_{1,3})$	300	0.173	0.104	0.031	0.157	0.093	0.026
	600	0.139	0.080	0.022	0.202	0.127	0.040
	1200	0.120	0.066	0.017	0.288	0.196	0.073
	3600	0.113	0.058	0.014	0.526	0.421	0.224

Table IX
Test statistics for various asset-pricing models

The table reports test statistics for the three asset-pricing models (CAPM, FF3, and FF5) described in Section 5. CSR and HJD denote the GLS cross-sectional regression and Hansen-Jagannathan distance estimators, respectively. $t(x)$ 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 (t_c) and model misspecification (t_m).

	t_c			t_m		
	CAPM	FF3	FF5	CAPM	FF3	FF5
			Panel A: Beta-Pricing Representation			
MLE						
$t(mkt)$	−2.92	−3.05	−1.34	−2.38	−2.43	−0.75
$t(smb)$		2.04	1.93		2.04	1.90
$t(hml)$		2.85	2.54		2.84	2.45
$t(rmw)$			−0.85			−0.44
$t(cma)$			5.09			1.63
CSR						
$t(mkt)$	−2.53	−2.61	−1.99	−2.37	−2.39	−1.74
$t(smb)$		2.04	2.02		2.04	2.03
$t(hml)$		2.86	2.72		2.86	2.70
$t(rmw)$			0.08			0.06
$t(cma)$			3.05			2.39
			Panel B: SDF Representation			
CU-GMM						
$t(mkt)$	4.00	4.84	−1.74	2.07	1.68	−0.84
$t(smb)$		−4.97	−4.92		−1.53	−2.10
$t(hml)$		−3.51	5.14		−1.25	1.62
$t(rmw)$			−5.68			−1.46
$t(cma)$			−7.15			−1.86
HJD						
$t(mkt)$	2.72	2.57	0.87	2.49	2.33	0.71
$t(smb)$		−3.03	−2.90		−2.98	−2.70
$t(hml)$		−1.85	0.78		−1.86	0.58
$t(rmw)$			−1.15			−1.02
$t(cma)$			−1.80			−1.30