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

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

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

Appendix A: Proofs of Theorems

A.1 Auxiliary Lemma 1

Auxiliary Lemma 1. *Let* $z = [z_1, z_2, \ldots, z_K]' \sim \mathcal{N}(0_K, (G_1'\Sigma^{-1}G_1)^{-1}/\sigma^2_{f,K-1})$*, where* $G_1 = [1_N,\ \alpha,\ \beta_1, \ldots, \beta_{K-2}]$ *and* $\sigma^2_{f,K-1} = \mathrm{Var}[f_{K-1,t}]$*. Assume that* Y_t *is iid normally distributed. Suppose that the model is misspecified and it contains a useless factor (that is,* $\mathrm{rank}(B) = K-1$*). Then,* $T \to \infty$*, we have (i)* $\hat{\gamma}_0^{ML} \xrightarrow{d} -\frac{z_1}{z_2}$*; (ii)* $\hat{\gamma}_{1,i}^{ML} \xrightarrow{d} \mu_{f,i} - \frac{z_{i+2}}{z_2}$ *for* $i = 1, \ldots, K-2$*; and (iii)* $\frac{\hat{\gamma}_{1,K-1}^{ML}}{\sqrt{T}} \xrightarrow{d} \frac{1}{z_2}$.

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

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

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

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

which is asymptotically equivalent to the estimator

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

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

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

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

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

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

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

This completes the proof of the lemma.

A.2 Proof of Theorem 1

part (a): Let $\sigma_i^2 = \text{Var}[z_i]$, $\sigma_{ij} \equiv \text{Cov}[z_i, z_j]$, $\rho_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$, $G_2 = [1_N,\ \beta_1, \dots, \beta_{K-2}]$, $\hat{G}_2 = [1_N,\ \hat{\beta}_1, \dots, \hat{\beta}_{K-2}]$, and define the random variables $\tilde{z}_2 \equiv z_2/\sigma_2 \sim \mathcal{N}(0,1)$, $x \sim \chi^2_{N-K}$, $q_i \sim \mathcal{N}(0,1)$, where x and q_i are independent of $\tilde{z}_2$, and $b_i = (x+\tilde{z}_2^2)/(x+\tilde{z}_2^2+q_i^2)$ for $i = 1, \dots, K-1$. We start with the squared t-ratio of the useless factor, $t^2(\hat{\gamma}_{1,K-1}^{ML})$. Using the formula for the inverse of a partitioned matrix, we obtain

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

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

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

we obtain

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

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

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

Note that we can write

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

where

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

With this expression, we can write

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

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

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

It follows that

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

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

Denote by w_i the i-th diagonal element of $(\hat{B}_1'\hat{\Sigma}^{-1}\hat{B}_1)^{-1}$, $i = 1, \ldots, K-1$. Using (A.11), we have

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

where

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

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

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

it is straightforward to show that

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

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

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

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

It follows that

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

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

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

Similarly,

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

Let

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

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

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

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

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

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

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

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

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

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

A.3 Auxiliary Lemma 2

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

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

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

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

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

Consider the CU-GMM estimator of v^*:

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

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

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

where

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

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

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

Let

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

Then, we have

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

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

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

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

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

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

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

A.4 Proof of Theorem 2

part (a): It is easy to show that

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

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

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

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

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

we can show that

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

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

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

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

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

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

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

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

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

and

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

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

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

where

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

and $h'h = 1$. Note that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Let

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

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

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

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

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

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

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

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

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

This completes the proof of part (a).

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

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

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

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

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

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

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

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

The problem with implementing this CU-GMM estimator is that the parameter vector ϕ is high dimensional especially when the number of test assets N is large. Peñaranda and Sentana (2015) show that CU-GMM delivers numerically identical estimates in the beta-pricing and linear SDF setups.[1] By augmenting $E[e_t(\lambda)]$ in the SDF representation with additional (just-identified) moment conditions for μ_f, V_f, and β, the CU-GMM estimate of the augmented parameter vector $\theta = [\lambda_0,\ \lambda_1',\ \beta_1', \ldots, \beta_K',\ \mu_f',\ \text{vech}(V_f)']'$ becomes numerically identical to the CU-GMM estimate of ϕ in the beta-pricing model. However, the estimation of θ can be performed in a sequential manner which offers substantial computational advantages. The following lemma presents a general result for this sequential estimation.

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

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

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

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

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

where $\bar{g}_1(\theta_1) = \frac{1}{T}\sum_{t=1}^T g_{1t}(\theta_1)$, $\hat{W}_{11}(\theta_1) = \frac{1}{T}\sum_{t=1}^T (g_{1t}(\theta_1) - \bar{g}_1(\theta_1))(g_{1t}(\theta_1) - \bar{g}_1(\theta_1))'$, $\bar{g}(\theta) = \frac{1}{T}\sum_{t=1}^T g_t(\theta)$, and $\hat{W}(\theta) = \frac{1}{T}\sum_{t=1}^T (g_t(\theta) - \bar{g}(\theta))(g_t(\theta) - \bar{g}(\theta))'$. Then, $\tilde{\theta}_1 = \hat{\theta}_1$.

Proof. Let

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

where

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

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

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

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

Similarly, we define

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

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

$$\hat{W}(\theta)^{-1} = \left[\begin{array}{cc} \hat{W}^{11}(\theta) & \hat{W}^{12}(\theta) \\ \hat{W}^{21}(\theta) & \hat{W}^{22}(\theta) \end{array}\right]. \tag{B.13}$$

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

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

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

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

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

Lemma B.1 establishes that for CU-GMM, adding a new set of just-identified moment conditions to the original system does not alter the estimates of the original parameters. This numerical equivalence can also be shown for the corresponding tests for over-identifying restrictions. The result in Lemma B.1 has implications for speeding up the optimization problem in the CU-GMM estimation. The key is to discard the subset of moment conditions that are exactly identified and only perform the over-identifying restriction test on the remaining smaller set of moment conditions. This will lead to fewer moment conditions and parameters in the system, which is desirable when performing numerical optimization. The following lemma demonstrates how to solve for $\hat{\theta}_2$ after $\tilde{\theta}_1$ is obtained from the smaller system.

Lemma B.2. *Let*

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

and

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

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

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

and $r_t(\hat{\theta}) = r_{1t}(\tilde{\theta}_1)$. Furthermore, if g_{2t}, conditional on θ_1, is linear in θ_2, that is,

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

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

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

Proof. Using the formula for the inverse of a partitioned matrix, we have $-(\hat{W}^{22})^{-1}\hat{W}^{21} = \hat{W}_{21}\hat{W}_{11}^{-1}$. Plugging this in (B.15) and noting that $\hat{\theta}_1 = \tilde{\theta}_1$, we obtain

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

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

(B.24) as

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

For the larger system, we have

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

where the third equality follows from (B.15), the fourth equality follows from the formula for the inverse of a partitioned matrix, and the last equality follows because $\hat{\theta}_1 = \tilde{\theta}_1$. The expression for $\hat{\theta}_2$ can be obtained by plugging $g_{2t}(\theta_1, \theta_2) = h_{1t}(\theta_1) - h_{2t}(\theta_1)\theta_2$ into (B.25) and solving for $\hat{\theta}_2$. This completes the proof of Lemma B.2.

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

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

Adding an extra set of just-identified moment conditions proves to be straightforward since $r_t(\hat{\theta}) = r_{1t}(\tilde{\theta}_1)$ and r_t does not need to be recomputed for the larger system.

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

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

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

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

Let

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

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