Internet Appendix
for
Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology*

Raymond Kan†, Cesare Robotti‡, and Jay Shanken§

*Citation format: Kan, Raymond, Cesare Robotti, and Jay Shanken, YEAR, Internet Appendix to "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," Journal of Finance VOL, PAGES, http://www.afajof.org/IA/YEAR.asp. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

†University of Toronto, Joseph L. Rotman School of Management, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, e-mail: kan@chass.utoronto.ca

‡Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street N.E., Atlanta, GA 30309, e-mail: cesare.robotti@atl.frb.org

§Emory University, Goizueta Business School, 1300 Clifton Road, Atlanta, GA 30322, e-mail: jay_shanken@bus.emory.edu

A Propositions, Lemmas and Proofs

Let f be a K-vector of factors and R a vector of returns on N test assets. We define $Y = [f',\ R']'$ and its mean and covariance matrix as

$$\mu = E[Y] \equiv \begin{bmatrix} \mu_f \\ \mu_R \end{bmatrix}, \tag{A.1}$$

$$V = \mathrm{Var}[Y] \equiv \begin{bmatrix} V_f & V_{fR} \\ V_{Rf} & V_R \end{bmatrix}, \tag{A.2}$$

where V is assumed to be positive definite.[1] The multiple regression betas of the N assets with respect to the K factors are defined as $\beta = V_{Rf}V_f^{-1}$. In addition, we denote the covariance matrix of the regression residuals of the N assets by $\Sigma = V_R - V_{Rf}V_f^{-1}V_{fR}$.

Let $Y_t = [f_t',\ R_t']'$, where f_t is the vector of K proposed factors at time t and R_t is a vector of returns on N test assets at time t. Throughout the various appendices, we assume that the time series $Y_t = [f_t',\ R_t']'$ is jointly stationary and ergodic, with finite fourth moment. Suppose we have T observations on Y_t and denote the sample moments of Y_t by

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_f \\ \hat{\mu}_R \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{A.3}$$

$$\hat{V} = \begin{bmatrix} \hat{V}_f & \hat{V}_{fR} \\ \hat{V}_{Rf} & \hat{V}_R \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{A.4}$$

The estimated multiple regression betas are given by $\hat{\beta} = \hat{V}_{Rf}\hat{V}_f^{-1}$.

Pricing Results

We first present the asymptotic distribution of the risk premium estimates when the weighting matrix W is known.

Proposition A.1. *Let $H = (X'WX)^{-1}$, $A = HX'W$, and $\gamma_t \equiv [\gamma_{0t},\ \gamma_{1t}']' = AR_t$. Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_R$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_{K+1}, V(\hat{\gamma})), \tag{A.5}$$

[1] For most of our analysis, we only need to assume V_f is nonsingular and V_{Rf} is of full column rank. For the case of generalized least squares (GLS) cross-sectional regression (CSR), we also need to assume V_R is nonsingular.

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{A.6}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t, \tag{A.7}$$

$\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_f)']'$, $u_t = e'W(R_t - \mu_R)$, $w_t = \gamma_1' V_f^{-1}(f_t - \mu_f)$, *and* $z_t = [0,\ u_t(f_t - \mu_f)'V_f^{-1}]'$. *When the model is correctly specified, we have:*

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t. \tag{A.8}$$

We do not provide the proof of Proposition A.1 as its proof is similar to that of Proposition A.2 below.

To conduct statistical tests, we need a consistent estimator of $V(\hat{\gamma})$. This can be obtained by replacing h_t with

$$\hat{h}_t = (\hat{\gamma}_t - \hat{\gamma}) - (\hat{\phi}_t - \hat{\phi})\hat{w}_t + \hat{H}\hat{z}_t, \tag{A.9}$$

where $\hat{\gamma}_t \equiv [\hat{\gamma}_{0t},\ \hat{\gamma}_{1t}']' = (\hat{X}'W\hat{X})^{-1}\hat{X}'WR_t$, $\hat{\phi}_t = [\hat{\gamma}_{0t},\ (\hat{\gamma}_{1t} - f_t)']'$, $\hat{\phi} = [\hat{\gamma}_0,\ (\hat{\gamma}_1 - \hat{\mu}_f)']'$, $\hat{u}_t = \hat{e}'W(R_t - \hat{\mu}_R)$ with $\hat{e} = \hat{\mu}_R - \hat{X}\hat{\gamma}$, $\hat{w}_t = \hat{\gamma}_1'\hat{V}_f^{-1}(f_t - \hat{\mu}_f)$, $\hat{H} = (\hat{X}'W\hat{X})^{-1}$ and $\hat{z}_t = [0,\ \hat{u}_t(f_t - \hat{\mu}_f)'\hat{V}_f^{-1}]'$. In particular, if h_t is uncorrelated over time, then we have $V(\hat{\gamma}) = E[h_t h_t']$, and its consistent estimator is given by

$$\hat{V}(\hat{\gamma}) = \frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{h}_t'. \tag{A.10}$$

When h_t is autocorrelated, one can use Newey and West's (1987) method to obtain a consistent estimator of $V(\hat{\gamma})$.

An inspection of (A.7) reveals that there are three sources of asymptotic variance for $\hat{\gamma}$. The first term $\gamma_t - \gamma$ measures the asymptotic variance of $\hat{\gamma}$ when the *true* betas (β) are used in the CSR. For example, if R_t is i.i.d., then γ_t is also i.i.d. and we can use the time series variance of γ_t to compute the standard error of $\hat{\gamma}$. This coincides with the popular Fama and MacBeth (1973) method. Since the betas are estimated with error in the first-pass time series regressions, an errors-in-variables (EIV) problem is introduced in the second-pass CSR. The second term $(\phi_t - \phi)w_t$ is the EIV adjustment term that accounts for the estimation errors in $\hat{\beta}$. The first two terms together give us the $V(\hat{\gamma})$ under the correctly specified model.[2] When the model is misspecified ($e \neq 0_N$), there is a third term Hz_t, which we call the misspecification adjustment term. Traditionally, this term has been ignored by empirical researchers.

We now turn our attention to the asymptotic distribution of $\hat{\gamma}$ when W must be estimated. It is easy to verify that the use of $\hat{W}$ instead of W does not alter the asymptotic distribution of $\hat{\gamma}$ when the model is correctly specified. However, the asymptotic distribution is affected when the model is misspecified. In the following proposition, we present the distribution for the GLS case.

Proposition A.2. *Let $H = (X'V_R^{-1}X)^{-1}$, $A = HX'V_R^{-1}$, and $\gamma_t = [\gamma_{0t},\ \gamma_{1t}']' = AR_t$. Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma} = (\hat{X}'\hat{V}_R^{-1}\hat{X})^{-1}\hat{X}'\hat{V}_R^{-1}\hat{\mu}_R$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_{K+1}, V(\hat{\gamma})), \tag{A.11}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{A.12}$$

with

$$h_t \quad = \quad (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t - (\gamma_t - \gamma)u_t, \tag{A.13}$$

$\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_f)']'$, $u_t = e'V_R^{-1}(R_t - \mu_R)$, $w_t = \gamma_1'V_f^{-1}(f_t - \mu_f)$, $z_t = [0,\ u_t(f_t - \mu_f)'V_f^{-1}]'$. *When the model is correctly specified, we have:*

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t. \tag{A.14}$$

Proof: The proof relies on the fact that $\hat{\gamma}$ is a smooth function of $\hat{\mu}$ and $\hat{V}$. Therefore, once we have the asymptotic distribution of $\hat{\mu}$ and $\hat{V}$, we can use the delta method to obtain the asymptotic distribution of $\hat{\gamma}$. Let

$$\varphi = \begin{bmatrix} \mu \\ \mathrm{vec}(V) \end{bmatrix}, \qquad \hat{\varphi} = \begin{bmatrix} \hat{\mu} \\ \mathrm{vec}(\hat{V}) \end{bmatrix}. \tag{A.15}$$

We first note that $\hat{\mu}$ and $\hat{V}$ can be written as the generalized method of moments (GMM) estimator that uses the moment conditions $E[r_t] = 0_{(N+K)(N+K+1)}$, where

$$r_t = \begin{bmatrix} Y_t - \mu \\ \mathrm{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}. \tag{A.16}$$

[2] It can be verified that this expression coincides with the one given by Jagannathan and Wang (1998) in their Theorem 1, except that our expression is easier to use in practice.

Since this is an exactly identified system of moment conditions, it is straightforward to verify that under the assumption that Y_t is stationary and ergodic with finite fourth moment, we have:[3]

$$\sqrt{T}(\hat{\varphi}-\varphi) \stackrel{A}{\sim} N(0_{(N+K)(N+K+1)}, S_0), \tag{A.17}$$

where

$$S_0 = \sum_{j=-\infty}^{\infty} E[r_t r'_{t+j}]. \tag{A.18}$$

Using the delta method, the asymptotic distribution of $\hat{\gamma}$ under the misspecified model is given by

$$\sqrt{T}(\hat{\gamma}-\gamma) \quad \stackrel{A}{\sim} \quad N\left(0_{K+1}, \left[\frac{\partial \gamma}{\partial \varphi'}\right] S_0 \left[\frac{\partial \gamma}{\partial \varphi'}\right]'\right). \tag{A.19}$$

It is straightforward to obtain:

$$\frac{\partial \gamma}{\partial \mu'_f} = 0_{(K+1)\times K}, \qquad \frac{\partial \gamma}{\partial \mu'_R} = A. \tag{A.20}$$

For the derivative of γ with respect to $\text{vec}(V)$, we first need to obtain $\partial x/\partial \text{vec}(V)'$, where $x = \text{vec}(X)$. In order to prove this identity, we write:

$$V_f = [I_K,\ 0_{K\times N}]V[I_K,\ 0_{K\times N}]', \qquad V_{Rf} = [0_{N\times K},\ I_N]V[I_K,\ 0_{K\times N}]' \tag{A.21}$$

to obtain

$$\frac{\partial \text{vec}(V_f)}{\partial \text{vec}(V)'} = [I_K,\ 0_{K\times N}] \otimes [I_K,\ 0_{K\times N}], \tag{A.22}$$

$$\frac{\partial \text{vec}(V_{Rf})}{\partial \text{vec}(V)'} = [I_K,\ 0_{K\times N}] \otimes [0_{N\times K},\ I_N]. \tag{A.23}$$

With the following identity

$$\begin{aligned}\frac{\partial \text{vec}(V_f^{-1})}{\partial \text{vec}(V)'} &= \frac{\partial \text{vec}(V_f^{-1})}{\partial \text{vec}(V_f)'}\frac{\partial \text{vec}(V_f)}{\partial \text{vec}(V)'} \\ &= -(V_f^{-1}\otimes V_f^{-1})\left([I_K,\ 0_{K\times N}]\otimes[I_K,\ 0_{K\times N}]\right) \\ &= [V_f^{-1},\ 0_{K\times N}]\otimes[-V_f^{-1},\ 0_{K\times N}],\end{aligned} \tag{A.24}$$

we can use the product rule to obtain

$$\begin{aligned}\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(V)'} &= (V_f^{-1}\otimes I_N)\frac{\partial \text{vec}(V_{Rf})}{\partial \text{vec}(V)'} + (I_K \otimes V_{Rf})\frac{\partial \text{vec}(V_f^{-1})}{\partial \text{vec}(V)'} \\ &= [V_f^{-1},\ 0_{K\times N}]\otimes[0_{N\times K},\ I_N] + [V_f^{-1},\ 0_{K\times N}]\otimes[-\beta,\ 0_{N\times N}] \\ &= [V_f^{-1},\ 0_{K\times N}]\otimes[-\beta,\ I_N].\end{aligned} \tag{A.25}$$

[3]Note that S_0 is a singular matrix as $\hat{V}$ is symmetric, so there are redundant elements in $\hat{\varphi}$. We could have written $\hat{\varphi}$ as $[\hat{\mu}',\ \text{vech}(\hat{V})']'$, but the results are the same under both specifications.

Finally, using the identity $\partial x/\partial \text{vec}(\beta)' = [0_K,\ I_K]' \otimes I_N$, we obtain:

$$\frac{\partial x}{\partial \text{vec}(V)'} = \frac{\partial x}{\partial \text{vec}(\beta)'}\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(V)'} = \left([0_K,\ V_f^{-1}]',\ 0_{(K+1)\times N}\right) \otimes [-\beta,\ I_N]. \tag{A.26}$$

Let $K_{m,n}$ be a commutation matrix (see, e.g., Magnus and Neudecker (1999)) such that $K_{m,n}\text{vec}(A) = \text{vec}(A')$ where A is an $m \times n$ matrix. In addition, denote $K_{n,n}$ by K_n. Then, using the product rule, we obtain:

$$\frac{\partial \gamma}{\partial \text{vec}(V)'} = (\mu_R' V_R^{-1} X \otimes I_{K+1})\frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} + (\mu_R' V_R^{-1} \otimes H)\frac{\partial \text{vec}(X')}{\partial \text{vec}(V)'} + (\mu_R' \otimes HX')\frac{\partial \text{vec}(V_R^{-1})}{\partial \text{vec}(V)'}. \tag{A.27}$$

The last two terms are given by

$$\begin{aligned}
(\mu_R' V_R^{-1} \otimes H)\frac{\partial \text{vec}(X')}{\partial \text{vec}(V)'} &= [H\left[0_K,\ V_f^{-1}\right]', 0_{(K+1)\times N}] \otimes [-\mu_R' V_R^{-1}\beta,\ \mu_R' V_R^{-1}], &\text{(A.28)}\\
(\mu_R' \otimes HX')\frac{\partial \text{vec}(V_R^{-1})}{\partial \text{vec}(V)'} &= -[0_K',\ \mu_R' V_R^{-1}] \otimes [0_{(K+1)\times K},\ A]. &\text{(A.29)}
\end{aligned}$$

For the first term, we use the chain rule to obtain

$$\begin{aligned}
& (\mu_R' V_R^{-1} X \otimes I_{K+1})\frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} \\
= \ & (\mu_R' V_R^{-1} X \otimes I_{K+1})\frac{\partial \text{vec}(H)}{\partial \text{vec}(H^{-1})'}\frac{\partial \text{vec}(H^{-1})}{\partial \text{vec}(V)'} \\
= \ & -(\mu_R' V_R^{-1} X \otimes I_{K+1})(H \otimes H)\left[(X'V_R^{-1} \otimes I_{K+1})K_{N,K+1}\frac{\partial x}{\partial \text{vec}(V)'}\right. \\
& \left. + (X' \otimes X')\frac{\partial \text{vec}(V_R^{-1})}{\partial \text{vec}(V)'} + (I_{K+1} \otimes X'V_R^{-1})\frac{\partial x}{\partial \text{vec}(V)'}\right] \\
= \ & -(\gamma' \otimes H)\left\{\left([-X'V_R^{-1}\beta,\ X'V_R^{-1}] \otimes \left[[0_K,\ V_f^{-1}]',\ 0_{(K+1)\times N}\right]\right) K_{N+K}\right. \\
& - [0_{(K+1)\times K},\ X'V_R^{-1}] \otimes [0_{(K+1)\times K},\ X'V_R^{-1}] \\
& \left. + \left[[0_K,\ V_f^{-1}]',\ 0_{(K+1)\times N}\right] \otimes [-X'V_R^{-1}\beta,\ X'V_R^{-1}]\right\} \\
= \ & \left[H[0_K,\ V_f^{-1}]',\ 0_{(K+1)\times N}\right] \otimes [\gamma' X' V_R^{-1}\beta,\ -\gamma' X' V_R^{-1}] \\
& + [0_K',\ \gamma' X' V_R^{-1}] \otimes [0_{(K+1)\times K},\ A] - [\gamma_1' V_f^{-1},\ 0_N'] \otimes [-A\beta,\ A]. \qquad \text{(A.30)}
\end{aligned}$$

Combining the three terms and using the first order condition $\beta' V_R^{-1} e = 0_K$, we have:

$$\begin{aligned}
\frac{\partial \gamma}{\partial \text{vec}(V)'} = \ & \left[H[0_K,\ V_f^{-1}]',\ 0_{(K+1)\times N}\right] \otimes [0_K',\ e'V_R^{-1}] \\
& - \left[\gamma_1' V_f^{-1},\ 0_N'\right] \otimes [-A\beta,\ A] - [0_K',\ e'V_R^{-1}] \otimes \left[0_{(K+1)\times K},\ A\right]. \qquad \text{(A.31)}
\end{aligned}$$

Using the expression for $\partial\gamma/\partial\varphi'$, we can simplify the asymptotic covariance matrix of $\hat{\gamma}$ to

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h'_{t+j}], \tag{A.32}$$

where

$$\begin{aligned}
h_t &= \frac{\partial\gamma}{\partial\varphi'} r_t \\
&= A(R_t-\mu_R) + \mathrm{vec}\left([0'_K,\ e'V_R^{-1}][(Y_t-\mu)(Y_t-\mu)'-V]\begin{bmatrix}[0_K,\ V_f^{-1}]H \\ 0_{N\times(K+1)}\end{bmatrix}\right) \\
&\quad - \mathrm{vec}\left([-A\beta,\ A][(Y_t-\mu)(Y_t-\mu)'-V]\begin{bmatrix}V_f^{-1}\gamma_1 \\ 0_N\end{bmatrix}\right) \\
&\quad - \mathrm{vec}\left([0_{(K+1)\times K},\ A][(Y_t-\mu)(Y_t-\mu)'-V]\begin{bmatrix}0_K \\ V_R^{-1}e\end{bmatrix}\right) \\
&= (\gamma_t-\gamma) + H[0_K,\ V_f^{-1}]'(f_t-\mu_f)u_t - A[(R_t-\mu_R)-\beta(f_t-\mu_f)](f_t-\mu_f)'V_f^{-1}\gamma_1 \\
&\quad - A(R_t-\mu_R)u_t - H[0_K,\ V_f^{-1}]'V_{fR}V_R^{-1}e - A\beta\gamma_1 + A\beta\gamma_1 + Ae \\
&= (\gamma_t-\gamma) + Hz_t - (\phi_t-\phi)w_t - (\gamma_t-\gamma)u_t.
\end{aligned} \tag{A.33}$$

The last equality follows from the first order condition $X'V_R^{-1}e = 0_{K+1}$ (which implies $\beta'V_R^{-1}e = 0_K$ and $Ae = 0_{K+1}$) and the fact that $A\beta = AX[0_K,\ I_K]' = [0_K,\ I_K]'$ gives us

$$A(R_t-\mu_R) - A\beta(f_t-\mu_f) = \gamma_t - \gamma - \begin{bmatrix}0 \\ f_t-\mu_f\end{bmatrix} = \phi_t - \phi. \tag{A.34}$$

Note that when the model is correctly specified, we have $e = 0_N$, $u_t = 0$, and h_t can be simplified to

$$h_t = (\gamma_t-\gamma) - (\phi_t-\phi)w_t. \tag{A.35}$$

This completes the proof.

Comparing (A.13) with the expression for h_t in (A.7), we see that there is an extra term in h_t associated with the use of $\hat{W}$ instead of W. This fourth term vanishes only when the model is correctly specified.

To gain a better understanding of the relative importance of the misspecification adjustment term, in the following lemmas we derive explicit expressions for $V(\hat{\gamma})$ under the assumption that returns and factors are multivariate elliptically distributed, first when W is known, and then for the GLS case.

Lemma A.1. *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ,*[4] *the asymptotic covariance matrix of $\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_R$ is given by*

$$V(\hat{\gamma}) = \Upsilon_w + \Upsilon_{w1} + \Upsilon_{w1}' + \Upsilon_{w2}, \tag{A.36}$$

where

$$\Upsilon_w = AV_RA' + (1+\kappa)\gamma_1'V_f^{-1}\gamma_1 A\Sigma A', \tag{A.37}$$

$$\Upsilon_{w1} = -(1+\kappa)H[0,\ \gamma_1'V_f^{-1}]'e'WV_RA', \tag{A.38}$$

$$\Upsilon_{w2} = (1+\kappa)e'WV_RWeH\tilde{V}_f^{-1}H, \tag{A.39}$$

with

$$\tilde{V}_f^{-1} = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_f^{-1} \end{bmatrix}. \tag{A.40}$$

Proof: In our proof, we rely on the mixed moments of multivariate elliptical distributions. Lemma 2 of Maruyama and Seo (2003) shows that if (X_i, X_j, X_k, X_l) are jointly multivariate elliptically distributed and with mean zero, we have:

$$E[X_iX_jX_k] = 0, \tag{A.41}$$

$$E[X_iX_jX_kX_l] = (1+\kappa)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), \tag{A.42}$$

where $\sigma_{ij} = \mathrm{Cov}[X_i, X_j]$. We first note that since γ_t, ϕ_t, $V_f^{-1}(f_t - \mu_f)$, w_t, and u_t are all linear functions of R_t and f_t, they are also jointly elliptically distributed. In addition, using (A.34), we have $\phi_t - \phi = A\epsilon_t$, where $\epsilon_t = R_t - \mu_R - \beta(f_t - \mu_f)$, which is uncorrelated with f_t. Using this result and applying (A.41) and (A.42), we can easily show that

$$E[(\gamma_t - \gamma)(\phi_t - \phi)'w_t] = 0_{(K+1)\times(K+1)}, \tag{A.43}$$

$$E[(\gamma_t - \gamma)z_t'] = 0_{(K+1)\times(K+1)}, \tag{A.44}$$

$$E[z_tz_t'] = (1+\kappa)e'WV_RWe\tilde{V}_f^{-1}, \tag{A.45}$$

$$E[(\phi_t - \phi)z_t'w_t] = (1+\kappa)AV_RWe[0,\ \gamma_1'V_f^{-1}], \tag{A.46}$$

$$E[(\phi_t - \phi)(\phi_t - \phi)'w_t^2] = (1+\kappa)\gamma_1'V_f^{-1}\gamma_1A\Sigma A'. \tag{A.47}$$

[4] The kurtosis parameter for an elliptical distribution is defined as $\kappa = \mu_4/(3\sigma^4) - 1$, where σ^2 and μ_4 are its second and fourth central moments, respectively.

Using these results and the i.i.d. assumption, we can now write:

$$\begin{aligned}
V(\hat{\gamma}) &= E[h_t h_t'] \\
&= \mathrm{Var}[\gamma_t] - E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] + E[(\gamma_t-\gamma)z_t']H \\
&\quad + E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] - E[(\phi_t-\phi)(\gamma_t-\gamma)'w_t] - E[(\phi_t-\phi)z_t'w_t]H \\
&\quad + HE[z_t z_t']H + HE[z_t(\gamma_t-\gamma)'] - HE[z_t(\phi_t-\phi)'w_t] \\
&= AV_RA' + (1+\kappa)(\gamma_1'V_f^{-1}\gamma_1)A\Sigma A' + (1+\kappa)e'WV_RWeH\tilde{V}_f^{-1}H \\
&\quad - (1+\kappa)AV_RWe[0,\ \gamma_1'V_f^{-1}]H - (1+\kappa)H[0,\ \gamma_1'V_f^{-1}]'e'WV_RA'. \qquad \text{(A.48)}
\end{aligned}$$

This completes the proof.

Note that when $\kappa = 0$, Lemma A.1 collapses to the expression given by Shanken and Zhou (2007) in their Proposition 1 under normality. For general W, the misspecification adjustment term $\Upsilon_{w1} + \Upsilon_{w1}' + \Upsilon_{w2}$ is not necessarily positive semidefinite. However, for true GLS with $W = V_R^{-1}$ or $W = \Sigma^{-1}$, we have $AV_RWe = Ae = 0_{K+1}$, so Υ_{w1} vanishes, resulting in the following simple expression for $V(\hat{\gamma})$:

$$V(\hat{\gamma}) = H + (1+\kappa)\gamma_1'V_f^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1} + (1+\kappa)QH\tilde{V}_f^{-1}H, \qquad \text{(A.49)}$$

where $H = (X'V_R^{-1}X)^{-1}$ and $Q = e'V_R^{-1}e$. The misspecification adjustment term $(1+\kappa)QH\tilde{V}_f^{-1}H$ is positive semidefinite in this case since $1+\kappa > 0$ (see Bentler and Berkane (1986)) and V_f^{-1} is positive definite. Note that the adjustment term is positively related to the aggregate pricing-error measure Q and the kurtosis parameter κ.

Lemma A.2. *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ, the asymptotic covariance matrix of $\hat{\gamma} = (\hat{X}'\hat{V}_R^{-1}\hat{X})^{-1}\hat{X}'\hat{V}_R^{-1}\hat{\mu}_R$ is given by*

$$V(\hat{\gamma}) = \Upsilon_w + \Upsilon_{w2}, \qquad \text{(A.50)}$$

where

$$\begin{aligned}
\Upsilon_w &= H + (1+\kappa)\gamma_1'V_f^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}, \qquad \text{(A.51)} \\
\Upsilon_{w2} &= (1+\kappa)Q\left[(X'\Sigma^{-1}X)^{-1}\tilde{V}_f^{-1}(X'\Sigma^{-1}X)^{-1} + (X'\Sigma^{-1}X)^{-1}\right], \qquad \text{(A.52)}
\end{aligned}$$

with $H = (X'V_R^{-1}X)^{-1}$, $Q = e'V_R^{-1}e$, and $\tilde{V}_f^{-1} = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_f^{-1} \end{bmatrix}$.

Proof: Under the i.i.d. assumption, the expression for $V(\hat{\gamma})$ is given by

$$
\begin{aligned}
E[h_t h_t'] \;=\;& \mathrm{Var}[\gamma_t] - E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] + E[(\gamma_t-\gamma)z_t']H - E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t] \\
& + E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] - E[(\phi_t-\phi)(\gamma_t-\gamma)'w_t] - E[(\phi_t-\phi)z_t'w_t]H \\
& + E[(\phi_t-\phi)(\gamma_t-\gamma)'w_t u_t] + HE[z_t z_t']H + HE[z_t(\gamma_t-\gamma)'] \\
& - HE[z_t(\phi_t-\phi)'w_t] - HE[z_t(\gamma_t-\gamma)'u_t] + E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t^2] \\
& - E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t] + E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t u_t] - E[(\gamma_t-\gamma)z_t'u_t]H. \quad \text{(A.53)}
\end{aligned}
$$

Following the proof of Lemma A.1, we have:

$$\mathrm{Var}[\gamma_t] = H, \quad \text{(A.54)}$$

$$E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] = 0_{(K+1)\times(K+1)}, \quad \text{(A.55)}$$

$$E[(\gamma_t-\gamma)z_t'] = 0_{(K+1)\times(K+1)}, \quad \text{(A.56)}$$

$$E[z_t z_t'] = (1+\kappa)Q\tilde{V}_f^{-1}, \quad \text{(A.57)}$$

$$E[(\phi_t-\phi)z_t'w_t] = 0_{(K+1)\times(K+1)}, \quad \text{(A.58)}$$

$$E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] = (1+\kappa)\gamma_1'V_f^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}, \quad \text{(A.59)}$$

$$E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t] = 0_{(K+1)\times(K+1)}, \quad \text{(A.60)}$$

$$E[(\phi_t-\phi)(\gamma_t-\gamma)'w_t u_t] = 0_{(K+1)\times(K+1)}, \quad \text{(A.61)}$$

$$E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t^2] = (1+\kappa)QH, \quad \text{(A.62)}$$

$$E[z_t(\gamma_t-\gamma)'u_t] = (1+\kappa)Q\begin{bmatrix} 0 & 0_K' \\ 0_K & I_K \end{bmatrix}. \quad \text{(A.63)}$$

By partitioning H as

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \quad \text{(A.64)}$$

where H_{11} is the $(1,1)$ element of H, and using (A.54)–(A.63), we can write:

$$
\begin{aligned}
E[h_t h_t'] =\;& H + (1+\kappa)\gamma_1'V_f^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1} + (1+\kappa)QH\tilde{V}_f^{-1}H \\
& - (1+\kappa)QH\begin{bmatrix} 0 & 0_K' \\ 0_K & I_K \end{bmatrix} + (1+\kappa)QH - (1+\kappa)Q\begin{bmatrix} 0 & 0_K' \\ 0_K & I_K \end{bmatrix}H \\
=\;& \Upsilon_w + (1+\kappa)Q\left(H\tilde{V}_f^{-1}H + \begin{bmatrix} H_{11} & 0_K' \\ 0_K & -H_{22} \end{bmatrix}\right) \\
=\;& \Upsilon_w + (1+\kappa)Q\begin{bmatrix} H_{12}V_f^{-1}H_{21} + H_{11} & H_{12}V_f^{-1}H_{22} \\ H_{22}V_f^{-1}H_{21} & H_{22}V_f^{-1}H_{22} - H_{22} \end{bmatrix}. \quad \text{(A.65)}
\end{aligned}
$$

By applying the identity $(X'\Sigma^{-1}X)^{-1} = H - \tilde{V}_f$, where $\tilde{V}_f = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_f \end{bmatrix}$, we can verify that the expression of Υ_{w2} in (A.52) is the same as the second term in (A.65) as follows:[5]

$$\begin{aligned}(X'\Sigma^{-1}X)^{-1}\tilde{V}_f^{-1}(X'\Sigma^{-1}X)^{-1} + (X'\Sigma^{-1}X)^{-1} &= (H - \tilde{V}_f)\tilde{V}_f^{-1}(H - \tilde{V}_f) + H - \tilde{V}_f \\ &= H\tilde{V}_f^{-1}H + \begin{bmatrix} H_{11} & 0_K' \\ 0_K & -H_{22} \end{bmatrix}. \end{aligned} \tag{A.66}$$

In particular, the misspecification adjustment term for $V(\hat{\gamma}_1)$ is

$$\begin{aligned} & (1+\kappa)Q(H_{22}V_f^{-1}H_{22} - H_{22}) \\ = \ & (1+\kappa)QH_{22}V_f^{-1}(V_f - V_f H_{22}^{-1}V_f)V_f^{-1}H_{22} \\ = \ & (1+\kappa)QH_{22}V_f^{-1}[V_f - V_{fR}V_R^{-1}V_{Rf} + V_{fR}V_R^{-1}1_N(1_N'V_R^{-1}1_N)^{-1}1_N'V_R^{-1}V_{Rf}]V_f^{-1}H_{22}, \end{aligned} \tag{A.67}$$

where the last equality is obtained by writing H_{22}^{-1} as

$$H_{22}^{-1} = \beta'V_R^{-1}\beta - \beta'V_R^{-1}1_N(1_N'V_R^{-1}1_N)^{-1}1_N'V_R^{-1}\beta. \tag{A.68}$$

This completes the proof.

Note that the term $V_f - V_{fR}V_R^{-1}V_{Rf}$ in (A.67) is the variance of the residuals from projecting the factors on the returns. For factors that have very low correlations with the returns (e.g., macroeconomic factors), the impact of this term and hence of the misspecification adjustment on the asymptotic variance of $\hat{\gamma}_1$ can be very large.

In the following proposition, we present the asymptotic distribution of $\hat{\lambda}$, the estimated parameters in the covariance-based model, for various cases. Since the derivation is very similar to the derivation for $\hat{\gamma}$, we do not provide the proof.

Proposition A.3. *Under a potentially misspecified model, the asymptotic distribution of $\hat{\lambda}$ is given by*

$$\sqrt{T}(\hat{\lambda} - \lambda) \stackrel{A}{\sim} N(0_{K+1}, V(\hat{\lambda})), \tag{A.69}$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t\tilde{h}_{t+j}']. \tag{A.70}$$

[5] By comparing $V(\hat{\gamma})$ for the estimated GLS case with the $V(\hat{\gamma})$ for the true GLS case in (A.49), it is easy to see that the use of $\hat{V}_R^{-1}$ instead of V_R^{-1} as weighting matrix increases the asymptotic variance of $\hat{\gamma}_0$ but reduces the asymptotic variance of $\hat{\gamma}_1$.

To simplify the expressions for $\tilde{h}_t$, we denote the last K elements of λ by λ_1 and define $\tilde{G}_t = (R_t - \mu_R)(f_t - \mu_f)' - V_{Rf}$, $\tilde{z}_t = [0,\ u_t(f_t - \mu_f)']'$, $\tilde{H} = (C'WC)^{-1}$, $\tilde{A} = \tilde{H}C'W$, $\lambda_t = \tilde{A}R_t$, and $u_t = e'W(R_t - \mu_R)$.

(1) With a known weighting matrix W, $\hat{\lambda} = (\hat{C}'W\hat{C})^{-1}\hat{C}'W\hat{\mu}_R$ and

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A}\tilde{G}_t\lambda_1 + \tilde{H}\tilde{z}_t. \tag{A.71}$$

(2) For estimated GLS, $\hat{\lambda} = (\hat{C}'\hat{V}_R^{-1}\hat{C})^{-1}\hat{C}'\hat{V}_R^{-1}\hat{\mu}_R$ and

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A}\tilde{G}_t\lambda_1 + \tilde{H}\tilde{z}_t - (\lambda_t - \lambda)u_t. \tag{A.72}$$

When the model is correctly specified, we have:

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A}\tilde{G}_t\lambda_1. \tag{A.73}$$

Results for the Sample R^2

We characterize the asymptotic distribution of $\hat{\rho}^2$ in the following proposition.

Proposition A.4. *In the following, we set W to be V_R^{-1} for the GLS case.*

(1) When $\rho^2 = 1$,

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \overset{A}{\sim} -\sum_{j=1}^{N-K-1} \frac{\xi_j}{Q_0} x_j, \tag{A.74}$$

where the x_j's are independent χ^2_1 random variables, and the ξ_j's are the eigenvalues of

$$P'W^{\frac{1}{2}}SW^{\frac{1}{2}}P, \tag{A.75}$$

where P is an $N \times (N - K - 1)$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}}C$, S is the asymptotic covariance matrix of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \epsilon_t y_t$, $\epsilon_t = R_t - \mu_R - \beta(f_t - \mu_f)$, and $y_t = 1 - \lambda_1'(f_t - \mu_f)$ is the normalized stochastic discount factor (SDF).

(2) When $0 < \rho^2 < 1$,

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{A.76}$$

where

$$n_t = 2\left[-u_t y_t + (1-\rho^2)v_t\right]/Q_0 \quad \text{for known } W, \tag{A.77}$$

$$n_t = \left[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)\right]/Q_0 \quad \text{for } \hat{W} = \hat{V}_R^{-1}, \tag{A.78}$$

with $e_0 = [I_N - 1_N(1_N' W 1_N)^{-1} 1_N' W]\mu_R$, $u_t = e' W(R_t - \mu_R)$, *and* $v_t = e_0' W(R_t - \mu_R)$.

(3) When $\rho^2 = 0$,

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \tag{A.79}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$[\beta' W\beta - \beta' W 1_N (1_N' W 1_N)^{-1} 1_N' W\beta] V(\hat{\gamma}_1), \tag{A.80}$$

where $V(\hat{\gamma}_1)$ *is given in Proposition A.1 (for known weighting matrix* W*) or Proposition A.2 (for estimated GLS).*[6]

Proof: (1) $\rho^2 = 1$: We first derive the asymptotic distribution of

$$T\hat{Q} = T[\hat{\mu}_R' \hat{W}\hat{\mu}_R - \hat{\mu}_R' \hat{W}\hat{X}(\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W}\hat{\mu}_R] \tag{A.81}$$

under $H_0 : \rho^2 = 1$, where $\hat{W} \xrightarrow{\text{a.s.}} W$ (this includes the known weighting matrix case as a special case). This can be accomplished by using the GMM results of Hansen (1982). Let $\theta = (\theta_1', \ \theta_2')'$, where $\theta_1 = (\alpha', \ \text{vec}(\beta)')'$ and $\theta_2 = \gamma$. Define

$$g_t(\theta) \equiv \begin{bmatrix} g_{1t}(\theta_1) \\ g_{2t}(\theta) \end{bmatrix} = \begin{bmatrix} l_t \otimes \epsilon_t \\ R_t - X\gamma \end{bmatrix}, \tag{A.82}$$

where $l_t = [1, \ f_t']'$ and $\epsilon_t = R_t - \alpha - \beta f_t$. When the model is correctly specified, we have $E[g_t(\theta)] = 0_{p+N}$, where $p = N(K+1)$. The sample moments of $g_t(\theta)$ are given by

$$\bar{g}_T(\theta) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T g_{1t}(\theta_1) \\ \frac{1}{T}\sum_{t=1}^T g_{2t}(\theta) \end{bmatrix}. \tag{A.83}$$

Let $\hat{\theta} = (\hat{\theta}_1', \ \hat{\theta}_2')'$, where $\hat{\theta}_1 = (\hat{\alpha}', \ \text{vec}(\hat{\beta})')'$ is the ordinary least squares (OLS) estimator of α and β, and

$$\hat{\theta}_2 = \hat{\gamma} = (\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W}\hat{\mu}_R \tag{A.84}$$

[6] In the proof of this proposition, we show that $\rho^2 = 0$ if and only if $\gamma_1 = 0_K$. Therefore, another way to test $H_0 : \rho^2 = 0$ is to test the equivalent hypothesis $H_0 : \gamma_1 = 0_K$, which can be easily performed by using a Wald test. When computing $V(\hat{\gamma}_1)$ for the test of $H_0 : \rho^2 = 0$, one could also impose the null hypothesis $H_0 : \gamma_1 = 0_K$ and drop the EIV term $(\phi_t - \phi)w_t$ in the expressions for h_t in Propositions A.1 and A.2.

is the second-pass CSR estimator of γ. Note that $\hat{\theta}$ is the solution to the following first order condition

$$B_T \bar{g}_T(\theta) = 0_{p+K+1}, \tag{A.85}$$

where

$$B_T = \begin{bmatrix} I_p & 0_{p\times N} \\ 0_{(K+1)\times p} & \hat{X}'\hat{W} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} I_p & 0_{p\times N} \\ 0_{(K+1)\times p} & X'W \end{bmatrix} \equiv B. \tag{A.86}$$

Writing

$$l_t \otimes \epsilon_t = \text{vec}(\epsilon_t l_t') = (l_t \otimes I_N)\text{vec}(\epsilon_t), \tag{A.87}$$

$$\epsilon_t = R_t - \alpha - \beta f_t = R_t - (l_t' \otimes I_N)\theta_1, \tag{A.88}$$

$$\beta\gamma_1 = (\gamma_1' \otimes I_N)\text{vec}(\beta), \tag{A.89}$$

we have:

$$\frac{\partial g_{1t}(\theta_1)}{\partial \theta_1'} = -l_t l_t' \otimes I_N, \tag{A.90}$$

$$\frac{\partial g_{1t}(\theta_1)}{\partial \theta_2'} = 0_{p\times(K+1)}, \tag{A.91}$$

$$\frac{\partial g_{2t}(\theta)}{\partial \theta_1'} = [0,\ -\gamma_1'] \otimes I_N, \tag{A.92}$$

$$\frac{\partial g_{2t}(\theta)}{\partial \theta_2'} = -X. \tag{A.93}$$

Let

$$\begin{aligned} D_T &= \frac{\partial \bar{g}_T(\theta)}{\partial \theta'} \\ &= \begin{bmatrix} -\left(\frac{1}{T}\sum_{t=1}^T l_t l_t'\right) \otimes I_N & 0_{p\times(K+1)} \\ [0,\ -\gamma_1'] \otimes I_N & -X \end{bmatrix} \\ &\xrightarrow{\text{a.s.}} \begin{bmatrix} -E[l_t l_t'] \otimes I_N & 0_{p\times(K+1)} \\ [0,\ -\gamma_1'] \otimes I_N & -X \end{bmatrix} \equiv D. \end{aligned} \tag{A.94}$$

Hansen (1982, Lemma 4.1) shows that when the model is correctly specified,[7] we have:

$$\sqrt{T}\bar{g}_T(\hat{\theta}) \stackrel{A}{\sim} N(0_{p+N}, [I_{p+N} - D(BD)^{-1}B]S_g[I_{p+N} - D(BD)^{-1}B]'), \tag{A.95}$$

[7] Although it is possible that some of the GMM sample moment conditions are not asymptotically normally distributed (see Gospodinov, Kan and Robotti (2010) for details), our results on the asymptotic distribution of $T(\hat{\rho}^2 - 1)$ are not affected by this problem.

where

$$S_g = \sum_{j=-\infty}^{\infty} E[g_t(\theta)g_{t+j}(\theta)']. \tag{A.96}$$

Using the partitioned matrix inverse formula, it is easy to verify that

$$E[l_t l_t']^{-1} = \begin{bmatrix} 1+\mu_f' V_f^{-1}\mu_f & -\mu_f' V_f^{-1} \\ -V_f^{-1}\mu_f & V_f^{-1} \end{bmatrix}. \tag{A.97}$$

It follows that

$$\begin{aligned}
BD &= \begin{bmatrix} -E[l_t l_t'] \otimes I_N & 0_{p\times(K+1)} \\ [0,\ -\gamma_1'] \otimes X'W & -H^{-1} \end{bmatrix}, && \text{(A.98)}\\
(BD)^{-1} &= \begin{bmatrix} -E[l_t l_t']^{-1} \otimes I_N & 0_{p\times(K+1)} \\ [-\gamma_1' V_f^{-1}\mu_f,\ \gamma_1' V_f^{-1}] \otimes A & -H \end{bmatrix}, && \text{(A.99)}\\
D(BD)^{-1}B &= \begin{bmatrix} I_p & 0_{p\times N} \\ [-\gamma_1' V_f^{-1}\mu_f,\ \gamma_1' V_f^{-1}] \otimes (I_N - XA) & XA \end{bmatrix}, && \text{(A.100)}\\
I_{p+N} - D(BD)^{-1}B &= \begin{bmatrix} 0_{p\times p} & 0_{p\times N} \\ [\gamma_1' V_f^{-1}\mu_f,\ -\gamma_1' V_f^{-1}] \otimes (I_N - XA) & I_N - XA \end{bmatrix}. && \text{(A.101)}
\end{aligned}$$

We now provide a simplification of the asymptotic distribution of $\bar{g}_{2T}(\hat{\theta})$. From (A.95), we have:

$$\sqrt{T}\bar{g}_{2T}(\hat{\theta}) \overset{A}{\sim} N(0_N, V_q), \tag{A.102}$$

where

$$V_q = \sum_{j=-\infty}^{\infty} E[q_t(\theta)q_{t+j}(\theta)'], \tag{A.103}$$

and

$$\begin{aligned}
q_t(\theta) &= [0_{N\times p},\ I_N][I_{p+N} - D(BD)^{-1}B]g_t(\theta) \\
&= -(I_N - XA)\epsilon_t \gamma_1' V_f^{-1}(f_t - \mu_f) + (I_N - XA)(R_t - X\gamma) \\
&= (I_N - XA)[R_t - \epsilon_t \gamma_1' V_f^{-1}(f_t - \mu_f)] \\
&= (I_N - XA)\epsilon_t y_t \\
&= [I_N - X(X'WX)^{-1}X'W]\epsilon_t y_t \\
&= W^{-\frac{1}{2}}[I_N - W^{\frac{1}{2}}X(X'WX)^{-1}X'W^{\frac{1}{2}}]W^{\frac{1}{2}}\epsilon_t y_t \\
&= W^{-\frac{1}{2}}[I_N - W^{\frac{1}{2}}C(C'WC)^{-1}C'W^{\frac{1}{2}}]W^{\frac{1}{2}}\epsilon_t y_t \\
&= W^{-\frac{1}{2}}PP'W^{\frac{1}{2}}\epsilon_t y_t, && \text{(A.104)}
\end{aligned}$$

where $y_t = 1 - \lambda_1'(f_t - \mu_f) = 1 - \gamma_1' V_f^{-1}(f_t - \mu_f)$. The fourth equality follows from the fact that, under $H_0 : \rho^2 = 1$, $(I_N - XA)R_t = (I_N - XA)\epsilon_t$. With this expression of $q_t(\theta)$, we can write V_q as

$$V_q = W^{-\frac{1}{2}} PP' W^{\frac{1}{2}} S W^{\frac{1}{2}} PP' W^{-\frac{1}{2}}. \tag{A.105}$$

Having derived the asymptotic distribution of $\bar{g}_{2T}(\hat{\theta})$, the asymptotic distribution of $\hat{Q}$ is given by

$$T\hat{Q} = T\bar{g}_{2T}(\hat{\theta})'\hat{W}\bar{g}_{2T}(\theta) \stackrel{A}{\sim} \sum_{j=1}^{N-K-1} \xi_j x_j, \tag{A.106}$$

where the x_j's are independent χ_1^2 random variables, and the ξ_j's are the $N-K-1$ nonzero eigenvalues of

$$W^{\frac{1}{2}} V_q W^{\frac{1}{2}} = PP' W^{\frac{1}{2}} S W^{\frac{1}{2}} PP'. \tag{A.107}$$

Equivalently, the ξ_j's are the eigenvalues of $P'W^{\frac{1}{2}} S W^{\frac{1}{2}} P$. Since $\hat{Q}_0 \xrightarrow{\text{a.s.}} Q_0 > 0$, we have:

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \stackrel{A}{\sim} -\sum_{j=1}^{N-K-1} \frac{\xi_j}{Q_0} x_j. \tag{A.108}$$

(2) $0 < \rho^2 < 1$: The proof uses the same notation and delta method employed in the proof of Proposition A.2 to obtain the asymptotic distribution of $\hat{\rho}^2$ as

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \stackrel{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{A.109}$$

where

$$n_t = \frac{\partial \rho^2}{\partial \varphi'} r_t. \tag{A.110}$$

Obtaining an explicit expression for n_t requires computing $\partial \rho^2 / \partial \varphi'$. For both the known weighting matrix case and the estimated GLS case, we have:

$$\frac{\partial \rho^2}{\partial \mu_f} = 0_K, \tag{A.111}$$

$$\frac{\partial \rho^2}{\partial \mu_R} = 2Q_0^{-1} W[(1-\rho^2)e_0 - e]. \tag{A.112}$$

Equation (A.111) follows because ρ^2 does not depend on μ_f. For (A.112), using the first order conditions $1_N' W e_0 = 0$ and $X'We = 0_{K+1}$, we have:

$$\frac{\partial Q_0}{\partial \mu_R} = 2We_0, \qquad \frac{\partial Q}{\partial \mu_R} = 2We. \tag{A.113}$$

It follows that

$$\frac{\partial \rho^2}{\partial \mu_R} = -Q_0^{-1}\frac{\partial Q}{\partial \mu_R} + Q_0^{-2}Q\frac{\partial Q_0}{\partial \mu_R} = -2Q_0^{-1}We + 2QQ_0^{-2}We_0 = 2Q_0^{-1}W[(1-\rho^2)e_0 - e]. \quad \text{(A.114)}$$

The expression for $\partial\rho^2/\partial\text{vec}(V)'$, however, depends on whether we use a known W or an estimate of W, say $\hat{W}$, as the weighting matrix. We start with the known weighting matrix W case. Differentiating $Q = e'We$ with respect to $\text{vec}(V)$, we obtain:

$$\frac{\partial Q}{\partial \text{vec}(V)'} = 2e'W\frac{\partial(\mu_R - X\gamma)}{\partial \text{vec}(V)'} = -2e'W\left[(\gamma' \otimes I_N)\frac{\partial x}{\partial \text{vec}(V)'} + X\frac{\partial \gamma}{\partial \text{vec}(V)'}\right]. \quad \text{(A.115)}$$

Note that the second term vanishes because of the first order condition $X'We = 0_{K+1}$. Using (A.26) for the first term and the fact that $\beta'We = 0_K$ gives

$$\frac{\partial Q}{\partial \text{vec}(V)'} = -2e'W\left([\gamma_1'V_f^{-1},\ 0_N'] \otimes [-\beta,\ I_N]\right) = -2\left([\gamma_1'V_f^{-1},\ 0_N'] \otimes [0_K',\ e'W]\right). \quad \text{(A.116)}$$

Since $Q_0 = e_0'We_0$ does not depend on V, we have:

$$\frac{\partial \rho^2}{\partial \text{vec}(V)'} = -Q_0^{-1}\frac{\partial Q}{\partial \text{vec}(V)'} = 2Q_0^{-1}[\gamma_1'V_f^{-1},\ 0_N'] \otimes [0_K',\ e'W]. \quad \text{(A.117)}$$

Therefore, for the known weighting matrix W case, n_t is given by

$$\begin{aligned} n_t &= \frac{\partial \rho^2}{\partial \varphi'}r_t \\ &= 2Q_0^{-1}[(1-\rho^2)e_0' - e']W(R_t - \mu_R) + 2Q_0^{-1}e'W(R_t - \mu_R)(f_t - \mu_f)'V_f^{-1}\gamma_1 \\ &= 2Q_0^{-1}[-u_t y_t + (1-\rho^2)v_t]. \end{aligned} \quad \text{(A.118)}$$

We now turn to the $\hat{W} = \hat{V}_R^{-1}$ case. Differentiating $Q = e'V_R^{-1}e$ with respect to $\text{vec}(V)$, we obtain:

$$\begin{aligned} \frac{\partial Q}{\partial \text{vec}(V)'} &= 2e'V_R^{-1}\frac{\partial(\mu_R - X\gamma)}{\partial \text{vec}(V)'} + (e' \otimes e')\frac{\partial \text{vec}(V_R^{-1})}{\partial \text{vec}(V)'} \\ &= -2\left([\gamma_1'V_f^{-1},\ 0_N'] \otimes [0_K',\ e'V_R^{-1}]\right) - (e' \otimes e')\left([0_{N\times K},\ V_R^{-1}] \otimes [0_{N\times K},\ V_R^{-1}]\right) \\ &= -[2\gamma_1'V_f^{-1},\ e'V_R^{-1}] \otimes [0_K',\ e'V_R^{-1}]. \end{aligned} \quad \text{(A.119)}$$

Similarly, we have:

$$\frac{\partial Q_0}{\partial \text{vec}(V)'} = -[0_K',\ e_0'V_R^{-1}] \otimes [0_K',\ e_0'V_R^{-1}]. \quad \text{(A.120)}$$

It follows that for the GLS case

$$\begin{aligned}\frac{\partial\rho^2}{\partial\text{vec}(V)'} &= -Q_0^{-1}\frac{\partial Q}{\partial\text{vec}(V)'} + Q_0^{-2}Q\frac{\partial Q_0}{\partial\text{vec}(V)'} \\ &= Q_0^{-1}\left[2\gamma_1'V_f^{-1},\ e'V_R^{-1}\right]\otimes\left[0_K',\ e'V_R^{-1}\right] \\ &\quad - Q_0^{-1}(1-\rho^2)\left[0_K',\ e_0'V_R^{-1}\right]\otimes\left[0_K',\ e_0'V_R^{-1}\right]. \end{aligned} \tag{A.121}$$

Therefore, we have for the GLS case:

$$\begin{aligned} n_t &= \frac{\partial\rho^2}{\partial\varphi'}r_t \\ &= 2Q_0^{-1}[(1-\rho^2)e_0' - e']V_R^{-1}(R_t-\mu_R) + Q_0^{-1}e'V_R^{-1}(R_t-\mu_R)[2\gamma_1'V_f^{-1}(f_t-\mu_f) \\ &\quad + e'V_R^{-1}(R_t-\mu_R)] - Q_0^{-1}(1-\rho^2)[e_0'V_R^{-1}(R_t-\mu_R)]^2 - Q_0^{-1}Q + Q_0^{-1}(1-\rho^2)Q_0 \\ &= Q_0^{-1}[u_t^2 - 2u_ty_t + (1-\rho^2)(2v_t - v_t^2)]. \end{aligned} \tag{A.122}$$

(3) $\rho^2 = 0$: We start by rewriting $Q_0 - Q$ as

$$\begin{aligned} Q_0 - Q &= \mu_R'WX(X'WX)^{-1}X'W\mu_R - \mu_R'W1_N(1_N'W1_N)^{-1}1_N'W\mu_R \\ &= \mu_R'WX(X'WX)^{-1}X'W\mu_R - \mu_R'WX\begin{bmatrix}(1_N'W1_N)^{-1} & 0_K' \\ 0_K & 0_{K\times K}\end{bmatrix}X'W\mu_R \\ &= \gamma'(X'WX)\gamma - \gamma'(X'WX)\begin{bmatrix}(1_N'W1_N)^{-1} & 0_K' \\ 0_K & 0_{K\times K}\end{bmatrix}(X'WX)\gamma \\ &= \gamma'(X'WX)\gamma - \gamma'\begin{bmatrix}1_N'W1_N & 1_N'W\beta \\ \beta'W1_N & \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta\end{bmatrix}\gamma \\ &= \gamma_1'[\beta'W\beta - \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta]\gamma_1. \end{aligned} \tag{A.123}$$

The matrix in the middle is positive definite because X is assumed to be of full column rank, so the necessary and sufficient condition for $Q_0 = Q$ (i.e., $\rho^2 = 0$) is $\gamma_1 = 0_K$. Note that (A.123) also holds for its sample counterpart, so we can write $\hat{\rho}^2$ as

$$\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0} = \frac{\hat{Q}_0 - \hat{Q}}{\hat{Q}_0} = \frac{\hat{\gamma}_1'[\hat{\beta}'\hat{W}\hat{\beta} - \hat{\beta}'\hat{W}1_N(1_N'\hat{W}1_N)^{-1}1_N'\hat{W}\hat{\beta}]\hat{\gamma}_1}{\hat{Q}_0}. \tag{A.124}$$

Under the null hypothesis $H_0 : \gamma_1 = 0_K$, we have:

$$\sqrt{T}\hat{\gamma}_1 \stackrel{A}{\sim} N(0_K, V(\hat{\gamma}_1)), \tag{A.125}$$

where $V(\hat{\gamma}_1)$ is the asymptotic covariance matrix of $\hat{\gamma}_1$ obtained under the misspecified model. As $\hat{Q}_0 \xrightarrow{\text{a.s.}} Q_0 > 0$ and

$$\hat{\beta}'\hat{W}\hat{\beta} - \hat{\beta}'\hat{W}1_N(1_N'\hat{W}1_N)^{-1}1_N'\hat{W}\hat{\beta} \xrightarrow{\text{a.s.}} \beta'W\beta - \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta, \tag{A.126}$$

it follows that

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \tag{A.127}$$

where the x_j's are independent χ^2_1 random variables and the ξ_j's are the eigenvalues of

$$[\beta' W\beta - \beta' W 1_N (1_N' W 1_N)^{-1} 1_N' W\beta] V(\hat{\gamma}_1). \tag{A.128}$$

This completes the proof.

Model Comparison Tests

Nested Models

Lemma A.3. $\rho_A^2 = \rho_B^2$ *if and only if* $\lambda_{A,2} = 0_{K_2}$.

Proof: Partition $C_A = [C_{Aa},\ C_{Ab}]$, where C_{Aa} is the first $K_1 + 1$ columns of C_A and C_{Ab} is the last K_2 columns of C_A. Using the fact that $C_{Aa} = C_B$, we can write the difference between Q_B and Q_A as

$$\begin{aligned}
Q_B - Q_A &= \mu_R' W C_A (C_A' W C_A)^{-1} C_A' W \mu_R - \mu_R' W C_B (C_B' W C_B)^{-1} C_B' W \mu_R \\
&= \mu_R' W C_A (C_A' W C_A)^{-1} C_A' W \mu_R - \mu_R' W C_A \begin{bmatrix} (C_{Aa}' W C_{Aa})^{-1} & 0_{(K_1+1)\times K_2} \\ 0_{K_2\times(K_1+1)} & 0_{K_2\times K_2} \end{bmatrix} C_A' W \mu_R \\
&= \lambda_A' (C_A' W C_A) \lambda_A - \lambda_A' (C_A' W C_A) \begin{bmatrix} (C_{Aa}' W C_{Aa})^{-1} & 0_{(K_1+1)\times K_2} \\ 0_{K_2\times(K_1+1)} & 0_{K_2\times K_2} \end{bmatrix} (C_A' W C_A) \lambda_A \\
&= \lambda_{A,2}' [C_{Ab}' W C_{Ab} - C_{Ab}' W C_{Aa} (C_{Aa}' W C_{Aa})^{-1} (C_{Aa}' W C_{Ab})] \lambda_{A,2} \\
&= \lambda_{A,2}' \tilde{H}_{A,22}^{-1} \lambda_{A,2},
\end{aligned} \tag{A.129}$$

where $\tilde{H}_{A,22}$ is the lower right $K_2 \times K_2$ submatrix of $\tilde{H}_A = (C_A' W C_A)^{-1}$. Since C_A is assumed to be of full column rank, $\tilde{H}_{A,22}^{-1}$ is a positive definite matrix. It follows that $Q_A = Q_B$ if and only if $\lambda_{A,2} = 0_{K_2}$. This completes the proof.

By this lemma, to test whether the models have the same ρ^2, one can simply perform a test of $H_0 : \lambda_{A,2} = 0_{K_2}$. Let $\hat{V}(\hat{\lambda}_{A,2})$ be a consistent estimator of the asymptotic covariance matrix of $\sqrt{T}(\hat{\lambda}_{A,2} - \lambda_{A,2})$. Then, under the null hypothesis,

$$T\hat{\lambda}_{A,2}' \hat{V}(\hat{\lambda}_{A,2})^{-1} \hat{\lambda}_{A,2} \overset{A}{\sim} \chi^2_{K_2}, \tag{A.130}$$

and this statistic can be used to test $H_0 : \rho_A^2 = \rho_B^2$. If $K_2 = 1$, we can also use the t-ratio associated with $\hat{\lambda}_{A,2}$ to perform the test. However, it is important to note that, in general, we cannot conduct this test using the usual standard error of $\hat{\lambda}_{A,2}$, which assumes that model A is correctly specified. Instead, we need to rely on the misspecification-robust standard error of $\hat{\lambda}$ given in Proposition A.3.

In the next proposition, we derive the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ and use this statistic to test $H_0 : \rho_A^2 = \rho_B^2$.

Proposition A.5. *Partition* $\tilde{H}_A = (C_A' W C_A)^{-1}$ *as*

$$\tilde{H}_A = \begin{bmatrix} \tilde{H}_{A,11} & \tilde{H}_{A,12} \\ \tilde{H}_{A,21} & \tilde{H}_{A,22} \end{bmatrix}, \tag{A.131}$$

where $\tilde{H}_{A,22}$ *is* $K_2 \times K_2$. *Under the null hypothesis* $H_0 : \rho_A^2 = \rho_B^2$,

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \stackrel{A}{\sim} \sum_{j=1}^{K_2} \frac{\xi_j}{Q_0} x_j, \tag{A.132}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of* $\tilde{H}_{A,22}^{-1} V(\hat{\lambda}_{A,2})$.

We do not provide the proof of Proposition A.5 since this proposition is a special case of Proposition A.6 below when $K_3 = 0$.

Again, we emphasize that the misspecification-robust version of $V(\hat{\lambda}_{A,2})$ should be used to test $H_0 : \rho_A^2 = \rho_B^2$. Model misspecification tends to create additional sampling variation in $\hat{\rho}_A^2 - \hat{\rho}_B^2$. Without taking this into account, one might mistakenly reject the null hypothesis when it is true. In actual testing, we replace ξ_j with its sample counterpart $\hat{\xi}_j$, where the $\hat{\xi}_j$'s are the eigenvalues of $\hat{\tilde{H}}_{A,22}^{-1} \hat{V}(\hat{\lambda}_{A,2})$, and $\hat{\tilde{H}}_{A,22}$ and $\hat{V}(\hat{\lambda}_{A,2})$ are consistent estimators of $\tilde{H}_{A,22}$ and $V(\hat{\lambda}_{A,2})$, respectively.[8]

Non-Nested Models

Testing $H_0 : \rho_A^2 = \rho_B^2$ is more complicated for non-nested models. The reason is that under H_0, there are three possible asymptotic distributions for $\hat{\rho}_A^2 - \hat{\rho}_B^2$, depending on why the two models have the same cross-sectional R^2. To see this, we first define the normalized SDFs for models A and B as

$$y_A = 1 - (f_1 - E[f_1])'\lambda_{A,1} - (f_2 - E[f_2])'\lambda_{A,2}, \qquad y_B = 1 - (f_1 - E[f_1])'\lambda_{B,1} - (f_3 - E[f_3])'\lambda_{B,3}. \tag{A.133}$$

[8] In the empirical application in the paper, we use the weighted chi-squared test in Proposition A.5 for nested models. Results for the Wald test of $\lambda_{A,2} = 0_{K_2}$ based on Lemma A.3 are consistent with those shown in Table IV.

At first sight, it may appear that $y_A = y_B$ is equivalent to the joint restriction $\lambda_{A,1} = \lambda_{B,1}$, $\lambda_{A,2} = 0_{K_2}$ and $\lambda_{B,3} = 0_{K_3}$. The following lemma shows that the first equality is redundant, however, since it is implied by the other two.

Lemma A.4. *For non-nested models, $y_A = y_B$ if and only if $\lambda_{A,2} = 0_{K_2}$ and $\lambda_{B,3} = 0_{K_3}$.*

Proof: Given that $y_A = y_B$ if and only if $\lambda_{A,1} = \lambda_{B,1}$, $\lambda_{A,2} = 0_{K_2}$, and $\lambda_{B,3} = 0_{K_3}$, it suffices to show that $\lambda_{A,2} = 0_{K_2}$ and $\lambda_{B,3} = 0_{K_3}$ imply $\lambda_{A,1} = \lambda_{B,1}$. Premultiplying both sides of $\lambda_A = (C_A' W C_A)^{-1} C_A' W \mu_R$ by $C_A' W C_A$, we obtain:

$$\begin{bmatrix} C_{Aa}' W C_{Aa} & C_{Aa}' W C_{Ab} \\ C_{Ab}' W C_{Aa} & C_{Ab}' W C_{Ab} \end{bmatrix} \begin{bmatrix} \lambda_{A,0} \\ \lambda_{A,1} \\ \lambda_{A,2} \end{bmatrix} = \begin{bmatrix} C_{Aa}' W \mu_R \\ C_{Ab}' W \mu_R \end{bmatrix}, \tag{A.134}$$

where C_{Aa} is the first K_1+1 columns of C_A and C_{Ab} is the last K_2 columns of C_A. When $\lambda_{A,2} = 0_{K_2}$, the first block of this equation gives us

$$\begin{bmatrix} \lambda_{A,0} \\ \lambda_{A,1} \end{bmatrix} = (C_{Aa}' W C_{Aa})^{-1} C_{Aa}' W \mu_R. \tag{A.135}$$

Similarly for model B, when $\lambda_{B,3} = 0_{K_3}$, we have:

$$\begin{bmatrix} \lambda_{B,0} \\ \lambda_{B,1} \end{bmatrix} = (C_{Ba}' W C_{Ba})^{-1} C_{Ba}' W \mu_R, \tag{A.136}$$

where C_{Ba} is the first K_1+1 columns of C_B. Since C_{Aa} and C_{Ba} are both equal to $[1_N,\ \mathrm{Cov}[R_t, f_{1t}']]$, we have $\lambda_{A,0} = \lambda_{B,0}$ and $\lambda_{A,1} = \lambda_{B,1}$. This completes the proof.

Lemma A.4 shows that $y_A = y_B$ implies that the two models have the same pricing errors ($e_A = e_B$) and cross-sectional R^2 ($\rho_A^2 = \rho_B^2$). Note that this lemma is applicable even when the models are misspecified. It implies that we can test $H_0 : y_A = y_B$ by testing the joint hypothesis $H_0 : \lambda_{A,2} = 0_{K_2},\ \lambda_{B,3} = 0_{K_3}$. Let $\psi = [\lambda_{A,2}',\ \lambda_{B,3}']'$ and $\hat{\psi} = [\hat{\lambda}_{A,2}',\ \hat{\lambda}_{B,3}']'$. It can be easily established that under $H_0 : y_A = y_B$, the asymptotic distribution of $\hat{\psi}$ is given by

$$\sqrt{T}(\hat{\psi} - \psi) \stackrel{A}{\sim} N(0_{K_2+K_3}, V(\hat{\psi})), \tag{A.137}$$

where

$$V(\hat{\psi}) = \sum_{j=-\infty}^{\infty} E[\tilde{q}_t \tilde{q}_{t+j}'], \tag{A.138}$$

and $\tilde{q}_t$ is a K_2+K_3 vector obtained by stacking up the last K_2 and K_3 elements of $\tilde{h}_t$ for models A and B, respectively, where $\tilde{h}_t$ is given in Proposition A.3.

Let $\hat{V}(\hat{\psi})$ be a consistent estimator of $V(\hat{\psi})$. Then, under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$,

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \overset{A}{\sim} \chi^2_{K_2+K_3}, \tag{A.139}$$

and this statistic can be used to test $H_0 : y_A = y_B$. As in the nested models case, it is important to conduct this test using the misspecification-robust standard error of $\hat{\psi}$.

The following proposition gives the asymptotic distribution of $\hat{\rho}^2_A - \hat{\rho}^2_B$ given $H_0 : y_A = y_B$.

Proposition A.6. *Let $\tilde{H}_A = (C_A'WC_A)^{-1}$ and $\tilde{H}_B = (C_B'WC_B)^{-1}$, and partition them as*

$$\tilde{H}_A = \begin{bmatrix} \tilde{H}_{A,11} & \tilde{H}_{A,12} \\ \tilde{H}_{A,21} & \tilde{H}_{A,22} \end{bmatrix}, \qquad \tilde{H}_B = \begin{bmatrix} \tilde{H}_{B,11} & \tilde{H}_{B,13} \\ \tilde{H}_{B,31} & \tilde{H}_{B,33} \end{bmatrix}, \tag{A.140}$$

where $\tilde{H}_{A,11}$ and $\tilde{H}_{B,11}$ are $(K_1+1)\times(K_1+1)$. Under the null hypothesis $H_0 : y_A = y_B$,

$$T(\hat{\rho}^2_A - \hat{\rho}^2_B) \overset{A}{\sim} \sum_{j=1}^{K_2+K_3} \frac{\xi_j}{Q_0} x_j, \tag{A.141}$$

where the x_j's are independent χ^2_1 random variables and the ξ_j's are the eigenvalues of

$$\begin{bmatrix} \tilde{H}^{-1}_{A,22} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}^{-1}_{B,33} \end{bmatrix} V(\hat{\psi}). \tag{A.142}$$

Proof: We first derive a simplified expression for $Q_B - Q_A$. The aggregate pricing-error measure for model A is given by

$$Q_A = e_A'We_A = \mu_R'W\mu_R - \mu_R'WC_A(C_A'WC_A)^{-1}C_A'W\mu_R. \tag{A.143}$$

We now introduce a model M that uses only f_1 as factors. The aggregate pricing-error measure for model M is given by

$$Q_M = e_M'We_M = \mu_R'W\mu_R - \mu_R'WC_M(C_M'WC_M)^{-1}C_M'W\mu_R, \tag{A.144}$$

where $C_M = [1_N,\ \mathrm{Cov}[R, f_1']]$. Using the fact that the $C_{Aa} = C_{Ba} = C_M$ and (A.129), we can write the difference between Q_M and Q_A as

$$Q_M - Q_A = \lambda_{A,2}'\tilde{H}^{-1}_{A,22}\lambda_{A,2}. \tag{A.145}$$

Similarly, we have:

$$Q_M - Q_B = \lambda'_{B,3}\tilde{H}^{-1}_{B,33}\lambda_{B,3}. \tag{A.146}$$

Subtracting (A.146) from (A.145), we obtain:

$$Q_B - Q_A = \lambda'_{A,2}\tilde{H}^{-1}_{A,22}\lambda_{A,2} - \lambda'_{B,3}\tilde{H}^{-1}_{B,33}\lambda_{B,3} = \psi' \left[\begin{array}{cc} \tilde{H}^{-1}_{A,22} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}^{-1}_{B,33} \end{array} \right] \psi, \tag{A.147}$$

where $\psi = [\lambda'_{A,2},\ \lambda'_{B,3}]'$. This equation also holds for its sample counterpart, and under the null hypothesis $H_0 : \psi = 0_{K_2+K_3}$, we have $\sqrt{T}V(\hat{\psi})^{-\frac{1}{2}}\hat{\psi} \stackrel{A}{\sim} N(0_{K_2+K_3}, I_{K_2+K_3})$. It follows that

$$T(\hat{Q}_B - \hat{Q}_A) \stackrel{A}{\sim} \sum_{j=1}^{K_2+K_3} \xi_j x_j, \tag{A.148}$$

where the x_j's are independent χ^2_1 random variables and the ξ_j's are the eigenvalues of

$$\left[\begin{array}{cc} \tilde{H}^{-1}_{A,22} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}^{-1}_{B,33} \end{array} \right] V(\hat{\psi}). \tag{A.149}$$

Since $\hat{\rho}^2_A - \hat{\rho}^2_B = (\hat{Q}_B - \hat{Q}_A)/\hat{Q}_0$ and $\hat{Q}_0 \stackrel{\text{a.s.}}{\longrightarrow} Q_0 > 0$, we have

$$T(\hat{\rho}^2_A - \hat{\rho}^2_B) \stackrel{A}{\sim} \sum_{j=1}^{K_2+K_3} \frac{\xi_j}{Q_0} x_j. \tag{A.150}$$

This completes the proof.

Note that we can think of the earlier nested models scenario as a special case of testing $H_0 : y_A = y_B$ with $K_3 = 0$. The only difference is that the ξ_j's in Proposition A.5 are all positive whereas some of the ξ_j's in Proposition A.6 are negative. As a result, we need to perform a two-sided test based on $\hat{\rho}^2_A - \hat{\rho}^2_B$ in the non-nested models case.[9]

When $y_A \neq y_B$, the asymptotic distribution of $\hat{\rho}^2_A - \hat{\rho}^2_B$ given $H_0 : \rho^2_A = \rho^2_B$ depends on whether the models are correctly specified or not. The following proposition presents a simple chi-squared statistic for testing whether models A and B are both correctly specified.

Proposition A.7. *Let $n_A = N - K_1 - K_2 - 1$ and $n_B = N - K_1 - K_3 - 1$. Also let P_A be an $N \times n_A$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}}C_A$ and P_B be an $N \times n_B$ orthonormal*

[9] Following Davidson and MacKinnon (2003, p.174), the p-value of a two-sided test associated with a realized statistic $\hat{\tau}$ that has a possibly asymmetric distribution is computed as $p = 2\min[F(\hat{\tau}), 1 - F(\hat{\tau})]$, where $F(\hat{\tau})$ is the cumulative density function of the statistic $\hat{\tau}$.

matrix with columns orthogonal to $W^{\frac{1}{2}}C_B$*. Let* ϵ_{At} *and* ϵ_{Bt} *be the residuals of models* A *and* B*, respectively, and define*

$$g_t(\theta) = \begin{bmatrix} g_{At}(\lambda_A) \\ g_{Bt}(\lambda_B) \end{bmatrix} = \begin{bmatrix} \epsilon_{At} y_{At} \\ \epsilon_{Bt} y_{Bt} \end{bmatrix}, \tag{A.151}$$

where $\theta = (\lambda_A',\ \lambda_B')'$*, and*

$$S \equiv \begin{bmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)']. \tag{A.152}$$

If $y_A \neq y_B$ *and the null hypothesis* $H_0 : \rho_A^2 = \rho_B^2 = 1$ *holds, then*

$$T \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix}' \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AB} \hat{W}^{\frac{1}{2}} \hat{P}_B \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BB} \hat{W}^{\frac{1}{2}} \hat{P}_B \end{bmatrix}^{-1} \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix} \stackrel{A}{\sim} \chi^2_{n_A+n_B}, \tag{A.153}$$

where $\hat{e}_A$ *and* $\hat{e}_B$ *are the sample pricing errors of models* A *and* B*, and* $\hat{P}_A$*,* $\hat{P}_B$*, and* $\hat{S}$ *are consistent estimators of* P_A*,* P_B*, and* S*, respectively.*

Proof: See the proof of Proposition A.8.

An alternative specification test makes use of the cross-sectional R^2s. The relevant asymptotic distribution is given in the following proposition.

Proposition A.8. *Using the notation in Proposition A.7, if* $y_A \neq y_B$ *and the null hypothesis* $H_0 : \rho_A^2 = \rho_B^2 = 1$ *holds, then*

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \stackrel{A}{\sim} \sum_{j=1}^{n_A+n_B} \frac{\xi_j}{Q_0} x_j, \tag{A.154}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$\begin{bmatrix} -P_A' W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A & -P_A' W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B \\ P_B' W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A & P_B' W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B \end{bmatrix}. \tag{A.155}$$

Proof: In the proof of Proposition A.4, we show that when model A is correctly specified,

$$\sqrt{T} \hat{e}_A \stackrel{A}{\sim} N(0_N, V_{q_A}), \tag{A.156}$$

where

$$V_{q_A} = \sum_{j=-\infty}^{\infty} E[q_{At} q_{A,t+j}'], \tag{A.157}$$

with

$$q_{At} = W^{-\frac{1}{2}} P_A P_A' W^{\frac{1}{2}} \epsilon_{At} y_{At} = W^{-\frac{1}{2}} P_A P_A' W^{\frac{1}{2}} g_{At}. \tag{A.158}$$

A similar result holds for model B. Stacking up the pricing errors of the two models, we have:

$$\sqrt{T} \begin{bmatrix} \hat{e}_A \\ \hat{e}_B \end{bmatrix} \overset{A}{\sim} N(0_{2N}, V_q), \tag{A.159}$$

where

$$V_q = \sum_{j=-\infty}^{\infty} E[q_t q_{t+j}'], \tag{A.160}$$

and

$$q_t = \begin{bmatrix} q_{At} \\ q_{Bt} \end{bmatrix} = \begin{bmatrix} W^{-\frac{1}{2}} P_A P_A' W^{\frac{1}{2}} g_{At} \\ W^{-\frac{1}{2}} P_B P_B' W^{\frac{1}{2}} g_{Bt} \end{bmatrix}. \tag{A.161}$$

We can then write V_q as

$$V_q = \begin{bmatrix} W^{-\frac{1}{2}} P_A P_A' W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A P_A' W^{-\frac{1}{2}} & W^{-\frac{1}{2}} P_A P_A' W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B P_B' W^{-\frac{1}{2}} \\ W^{-\frac{1}{2}} P_B P_B' W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A P_A' W^{-\frac{1}{2}} & W^{-\frac{1}{2}} P_B P_B' W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B P_B' W^{-\frac{1}{2}} \end{bmatrix}. \tag{A.162}$$

It follows that

$$z = \sqrt{T} \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix} \overset{A}{\sim} N(0_{n_A+n_B}, V_z), \tag{A.163}$$

where

$$V_z = \begin{bmatrix} P_A' W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A & P_A' W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B \\ P_B' W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A & P_B' W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B \end{bmatrix}. \tag{A.164}$$

Then, we have:

$$z' \hat{V}_z^{-1} z \overset{A}{\sim} \chi^2_{n_A+n_B}. \tag{A.165}$$

This completes the proof of Proposition A.7.

Using the first order condition $\hat{C}_A' \hat{W}' \hat{e}_A = 0_{K_1+K_2+1}$, we can write:

$$\begin{aligned} T\hat{Q}_A &= T\hat{e}_A' \hat{W}^{\frac{1}{2}} [\hat{P}_A \hat{P}_A' + \hat{W}^{\frac{1}{2}} \hat{C}_A (\hat{C}_A' \hat{W} \hat{C}_A)^{-1} \hat{C}_A' \hat{W}^{\frac{1}{2}}] \hat{W}^{\frac{1}{2}} \hat{e}_A \\ &= T\hat{e}_A' \hat{W}^{\frac{1}{2}} \hat{P}_A \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ &= z_A' z_A, \end{aligned} \tag{A.166}$$

where z_A is the first n_A elements of z. Similarly, $T\hat{Q}_B = z_B' z_B$, where z_B is the last n_B elements of z. Let $Q\Xi Q'$ be the eigenvalue decomposition of

$$V_z^{\frac{1}{2}} \begin{bmatrix} -I_{n_A} & 0_{n_A \times n_B} \\ 0_{n_B \times n_A} & I_{n_B} \end{bmatrix} V_z^{\frac{1}{2}}, \tag{A.167}$$

where $\Xi = \text{Diag}(\xi_1, \ldots, \xi_{n_A+n_B})$ is a diagonal matrix of the eigenvalues of (A.167) or, equivalently, of the eigenvalues of (A.155). Writing $\tilde{z} = Q'V_z^{-\frac{1}{2}}z \overset{A}{\sim} N(0_{n_A+n_B}, I_{n_A+n_B})$, we have:

$$T(\hat{Q}_B - \hat{Q}_A) = z' \begin{bmatrix} -I_{n_A} & 0_{n_A \times n_B} \\ 0_{n_B \times n_A} & I_{n_B} \end{bmatrix} z = z'V_z^{-\frac{1}{2}} Q \Xi Q' V_z^{-\frac{1}{2}} z = \tilde{z}'\Xi\tilde{z} = \sum_{j=1}^{n_A+n_B} \xi_j x_j, \quad \text{(A.168)}$$

where $x_j = \tilde{z}_j^2 \overset{A}{\sim} \chi^2_1$, $j = 1, \ldots, n_A + n_B$, and they are asymptotically independent of each other.

Since $\hat{\rho}^2_A - \hat{\rho}^2_B = (\hat{Q}_B - \hat{Q}_A)/\hat{Q}_0$ and $\hat{Q}_0 \xrightarrow{\text{a.s.}} Q_0 > 0$, we have:

$$T(\hat{\rho}^2_A - \hat{\rho}^2_B) \overset{A}{\sim} \sum_{j=1}^{n_A+n_B} \frac{\xi_j}{Q_0} x_j. \quad \text{(A.169)}$$

This completes the proof of Proposition A.8.

Note that the ξ_j's are not all positive because $\hat{\rho}^2_A - \hat{\rho}^2_B$ can be negative. Thus, again, we need to perform a two-sided test of $H_0 : \rho^2_A = \rho^2_B$.

The asymptotic distribution of $\hat{\rho}^2_A - \hat{\rho}^2_B$ changes when the models are misspecified and the next proposition presents the appropriate distribution for this case.

Proposition A.9. *Suppose $y_A \neq y_B$ and $0 < \rho^2_A = \rho^2_B < 1$.*[10] *We have:*

$$\sqrt{T}(\hat{\rho}^2_A - \hat{\rho}^2_B) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]\right). \quad \text{(A.170)}$$

When the weighting matrix W is known,

$$d_t = 2Q_0^{-1}\left[u_{Bt}y_{Bt} - u_{At}y_{At} - (\rho^2_A - \rho^2_B)v_t\right], \quad \text{(A.171)}$$

where $u_{At} = e'_A W(R_t - \mu_R)$, $u_{Bt} = e'_B W(R_t - \mu_R)$, and v_t is defined in Proposition A.4. For estimated GLS,

$$d_t = Q_0^{-1}\left[u^2_{At} - 2u_{At}y_{At} - u^2_{Bt} + 2u_{Bt}y_{Bt} - (\rho^2_A - \rho^2_B)(2v_t - v^2_t)\right], \quad \text{(A.172)}$$

where $u_{At} = e'_A V_R^{-1}(R_t - \mu_R)$ and $u_{Bt} = e'_B V_R^{-1}(R_t - \mu_R)$.[11]

[10] Since $\rho^2_A = \rho^2_B = 0$ implies $y_A = y_B = 1$, this case is already covered by the test based on Lemma A.4.

[11] One could impose $H_0 : \rho^2_A = \rho^2_B$ in (A.171) and (A.172) and the v_t terms would drop out of these expressions. However, our simulation results indicate that not imposing $H_0 : \rho^2_A = \rho^2_B$ in the computation of the standard errors leads to improved finite-sample properties of the normal test. Similarly, we obtain better finite-sample performance when, in the GLS case, we multiply u_t and v_t by $(T - N - 2)/T$.

Proof: We start from the known weighting matrix case. Using the results of Proposition A.4, we obtain the following expressions for models A and B:

$$n_{At} = \left[\frac{\partial \rho_A^2}{\partial \varphi}\right]' r_t = 2Q_0^{-1}[-u_{At}y_{At} + (1-\rho_A^2)v_t], \tag{A.173}$$

$$n_{Bt} = \left[\frac{\partial \rho_B^2}{\partial \varphi}\right]' r_t = 2Q_0^{-1}[-u_{Bt}y_{Bt} + (1-\rho_B^2)v_t]. \tag{A.174}$$

Now, using the delta method and equations (A.15)–(A.18), the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ when both models are misspecified is given by

$$\sqrt{T}(\hat{\rho}_A^2 - \hat{\rho}_B^2 - (\rho_A^2 - \rho_B^2)) \stackrel{A}{\sim} N\left(0, \left[\frac{\partial(\rho_A^2 - \rho_B^2)}{\partial \varphi}\right]' S_0 \left[\frac{\partial(\rho_A^2 - \rho_B^2)}{\partial \varphi}\right]\right). \tag{A.175}$$

With the analytical expressions of n_{At} and n_{Bt}, the asymptotic variance of $\sqrt{T}(\hat{\rho}_A^2 - \hat{\rho}_B^2)$ can be written as

$$\sum_{j=-\infty}^{\infty} E[d_t d_{t+j}], \tag{A.176}$$

where

$$d_t = \left(\frac{\partial \rho_A^2}{\partial \varphi} - \frac{\partial \rho_B^2}{\partial \varphi}\right)' r_t = n_{At} - n_{Bt}. \tag{A.177}$$

Therefore, we have:

$$d_t = 2Q_0^{-1}\left[u_{Bt}y_{Bt} - u_{At}y_{At} - (\rho_A^2 - \rho_B^2)v_t\right]. \tag{A.178}$$

Using the same type of proof for the GLS case with $\hat{W} = \hat{V}_R^{-1}$, we obtain:

$$d_t = Q_0^{-1}\left[u_{At}^2 - 2u_{At}y_{At} - u_{Bt}^2 + 2u_{Bt}y_{Bt} - (\rho_A^2 - \rho_B^2)(2v_t - v_t^2)\right]. \tag{A.179}$$

This completes the proof.

Note that if $y_{At} = y_{Bt}$, then $\rho_A^2 = \rho_B^2$, $u_{At} = u_{Bt}$, and hence $d_t = 0$. Or, if $y_{At} \neq y_{Bt}$, but both models are correctly specified (i.e., $u_{At} = u_{Bt} = 0$ and $\rho_A^2 = \rho_B^2 = 1$), then again $d_t = 0$. Thus, the normal test cannot be used in these cases, consistent with the maintained assumptions in the proposition.

Discussion of the Sequential Test

Given the three distinct cases described above, testing $H_0 : \rho_A^2 = \rho_B^2$ for non-nested models entails a sequential test, as suggested by Vuong (1989). In our context, this involves first testing $H_0 : y_A = y_B$ using (A.139) or (A.141). If we reject $H_0 : y_A = y_B$, then we use (A.153) or (A.154) to test

$H_0 : \rho_A^2 = \rho_B^2 = 1$. This second test can be viewed as a generalization of the cross-sectional regression test (CSRT) of Shanken (1985) and later multivariate tests of the validity of the expected return relation for a single pricing model. Finally, if the hypothesis that both models are correctly specified is also rejected, we proceed to evaluate $H_0 : 0 < \rho_A^2 = \rho_B^2 < 1$ using the normal test in Proposition A.9. Let α_1, α_2 and α_3 be the significance levels employed in these three tests. Then the sequential test has an asymptotic significance level that is bounded above by $\max[\alpha_1, \alpha_2, \alpha_3]$. Thus, if $\alpha_1 = \alpha_2 = \alpha_3 = 0.05$, the significance level of this procedure for testing $H_0 : \rho_A^2 = \rho_B^2$ is asymptotically no larger than 5%.[12]

In our empirical application in the paper, we implement the sequential test by using (A.141), (A.154), and the normal test in Proposition A.9.

[12] Note that for the sequential test to reject $\rho_A^2 = \rho_B^2$, all three tests must reject. Consider the first scenario, $y_A = y_B$. $\mathrm{P}(\text{reject } \rho_A^2 = \rho_B^2 \mid y_A = y_B) \leq \mathrm{P}(\text{test 1 rejects} \mid y_A = y_B) = \alpha_1$. Similarly, the probability that the sequential test rejects under the second and third scenarios cannot exceed α_2 and α_3, respectively. Under $H_0 : \rho_A^2 = \rho_B^2$, one of the three scenarios must hold, so the true probability of rejection cannot exceed the maximum.

B Analysis with Portfolio Characteristics

We show how to accommodate portfolio characteristics in the CSR. In particular, we derive the asymptotic distributions of the estimated parameters, sample cross-sectional R^2s, and model comparison tests when both portfolio characteristics and estimated betas (or covariances) are used in the CSR. The proofs of the various lemmas and propositions are omitted since they are similar to the ones of Appendix A.

We are interested in determining whether the unconditional betas with respect to K factors and L portfolio characteristics help explain the unconditional expected returns on N test assets. Let Z_t be an $N \times L$ matrix of L portfolio characteristics associated with the N test assets at the beginning of period t. The proposed model states that unconditional expected returns are linear in $\beta = V_{Rf}V_f^{-1}$ and $\mu_Z = E[Z_t]$:

$$\mu_R = X\gamma, \tag{B.1}$$

where $X = [1_N,\ \beta,\ \mu_Z]$. In reality, the proposed model could be misspecified. In this case, the vector of pseudo-true parameters γ is defined as

$$\gamma = (X'WX)^{-1}(X'W\mu_R), \tag{B.2}$$

where W is an $N \times N$ positive definite weighting matrix. We partition the $(K + L + 1)$-vector γ as $\gamma = [\gamma_0,\ \gamma_1',\ \gamma_2']'$, where γ_0 is the zero-beta rate, γ_1 is a K-vector of parameters associated with the K systematic factors, and γ_2 is an L-vector of parameters associated with the L portfolio characteristics.

Since β and μ_Z are not observable, we need to use their sample estimates

$$\hat{\beta} = \left[\frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu}_R)(f_t - \hat{\mu}_f)'\right]\left[\frac{1}{T}\sum_{t=1}^{T}(f_t - \hat{\mu}_f)(f_t - \hat{\mu}_f)'\right]^{-1}, \qquad \hat{\mu}_Z = \frac{1}{T}\sum_{t=1}^{T} Z_t, \tag{B.3}$$

in the second-pass CSR. Let $\hat{X} = [1_N,\ \hat{\beta},\ \hat{\mu}_Z]$, the sample estimate of γ is given by

$$\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}(\hat{X}'W\hat{\mu}_R). \tag{B.4}$$

Note that this setup coincides with the one proposed by Jagannathan and Wang (1996) except that we (1) take into account the estimation error in $\hat{\mu}_Z$, and (2) allow for potential model misspecification.

Pricing Results

In the following proposition, we present the asymptotic distribution of $\hat{\gamma}$ when the weighting matrix W is known.

Proposition B.1. *Let $H = (X'WX)^{-1}$, $A = HX'W$, and $\gamma_t \equiv [\gamma_{0t},\ \gamma_{1t}',\ \gamma_{2t}']' = AR_t$. Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma}$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \stackrel{A}{\sim} N(0_{K+L+1}, V(\hat{\gamma})), \tag{B.5}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{B.6}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t - A(Z_t - \mu_Z)\gamma_2 + Hz_t, \tag{B.7}$$

$\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)',\ \gamma_{2t}']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_f)',\ \gamma_2']'$, $u_t = e'W(R_t - \mu_R)$, $w_t = \gamma_1' V_f^{-1}(f_t - \mu_f)$, *and* $z_t = [0,\ u_t(f_t - \mu_f)' V_f^{-1},\ e'WZ_t]'$. *When the model is correctly specified, we have:*

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t - A(Z_t - \mu_Z)\gamma_2. \tag{B.8}$$

The first term $(\gamma_t - \gamma)$ is the Fama-MacBeth term, which ignores the estimation errors in $\hat{\beta}$ and $\hat{\mu}_Z$. The second term $(\phi_t - \phi)w_t$ is the EIV adjustment term for $\hat{\beta}$. The third term $A(Z_t - \mu_Z)\gamma_2$ is the EIV adjustment term for $\hat{\mu}_Z$. The final term Hz_t is the misspecification adjustment term due to model misspecification.

We now turn our attention to the asymptotic distribution of $\hat{\gamma}$ when W must be estimated. In the following proposition, we present the distribution for the GLS case

Proposition B.2. *Let $H = (X'V_R^{-1}X)^{-1}$, $A = HX'V_R^{-1}$, and $\gamma_t = [\gamma_{0t},\ \gamma_{1t}',\ \gamma_{2t}']' = AR_t$. Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma} = (\hat{X}'\hat{V}_R^{-1}\hat{X})^{-1}\hat{X}'\hat{V}_R^{-1}\hat{\mu}_R$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \stackrel{A}{\sim} N(0_{K+L+1}, V(\hat{\gamma})), \tag{B.9}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{B.10}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t - A(Z_t - \mu_Z)\gamma_2 + Hz_t - (\gamma_t - \gamma)u_t, \quad \text{(B.11)}$$

$\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)',\ \gamma_{2t}']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_f)',\ \gamma_2']'$, $u_t = e'V_R^{-1}(R_t - \mu_R)$, $w_t = \gamma_1' V_f^{-1}(f_t - \mu_f)$, *and* $z_t = [0,\ u_t(f_t - \mu_f)'V_f^{-1},\ e'V_R^{-1}Z_t]'$. *When the model is correctly specified, we have:*

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t - A(Z_t - \mu_Z)\gamma_2. \quad \text{(B.12)}$$

Note that when the model is correctly specified, the estimation error in the weighting matrix does not affect the asymptotic distribution of $\hat{\gamma}$.

If we replace $\hat{\beta}$ by $\hat{V}_{Rf}$ in the second-pass CSR, we have

$$\hat{\lambda} = (\hat{C}'W\hat{C})^{-1}\hat{C}'W\hat{\mu}_R, \quad \text{(B.13)}$$

where

$$\hat{C} = [1_N,\ \hat{V}_{Rf},\ \hat{\mu}_Z]. \quad \text{(B.14)}$$

Also, define the population counterpart of $\hat{\lambda}$ as

$$\lambda = (C'WC)^{-1}C'W\mu_R, \quad \text{(B.15)}$$

where

$$C = [1_N,\ V_{Rf},\ \mu_Z]. \quad \text{(B.16)}$$

We denote the K-vector of parameters associated with the K risk factors by λ_1 and the L-vector of parameters associated with the L portfolio characteristics by λ_2. It is easy to see that there is a one-to-one mapping between γ and λ, which is given by

$$\lambda_0 = \gamma_0, \qquad \lambda_1 = V_f^{-1}\gamma_1, \qquad \lambda_2 = \gamma_2. \quad \text{(B.17)}$$

The next proposition derives the asymptotic distribution of $\hat{\lambda}$ under potentially misspecified models.

Proposition B.3. *Under a potentially misspecified model, the asymptotic distribution of* $\hat{\lambda}$ *is given by*

$$\sqrt{T}(\hat{\lambda} - \lambda) \overset{A}{\sim} N(0_{K+L+1}, V(\hat{\lambda})), \quad \text{(B.18)}$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t \tilde{h}_{t+j}']. \tag{B.19}$$

To simplify the expressions for $\tilde{h}_t$*, we define* $\tilde{G}_t = (R_t - \mu_R)(f_t - \mu_f)' - V_{Rf}$*,* $\tilde{z}_t = [0,\ u_t(f_t - \mu_f)',\ e'WZ_t]'$*,* $\tilde{H} = (C'WC)^{-1}$*,* $\tilde{A} = \tilde{H}C'W$*,* $\lambda_t = \tilde{A}R_t$*, and* $u_t = e'W(R_t - \mu_R)$*.*

(1) With a known weighting matrix W*,* $\hat{\lambda} = (\hat{C}'W\hat{C})^{-1}\hat{C}'W\hat{\mu}_R$ *and*

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A}\tilde{G}_t\lambda_1 - \tilde{A}(Z_t - \mu_Z)\lambda_2 + \tilde{H}\tilde{z}_t. \tag{B.20}$$

(2) For estimated GLS, $\hat{\lambda} = (\hat{C}'\hat{V}_R^{-1}\hat{C})^{-1}\hat{C}'\hat{V}_R^{-1}\hat{\mu}_R$ *and*

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A}\tilde{G}_t\lambda_1 - \tilde{A}(Z_t - \mu_Z)\lambda_2 + \tilde{H}\tilde{z}_t - (\lambda_t - \lambda)u_t. \tag{B.21}$$

When the model is correctly specified, we have:

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A}\tilde{G}_t\lambda_1 - \tilde{A}(Z_t - \mu_Z)\lambda_2. \tag{B.22}$$

Results for the Sample R^2

We characterize the asymptotic distribution of $\hat{\rho}^2$ in the following proposition.

Proposition B.4. *In the following, we set* W *to be* V_R^{-1} *for the GLS case.*

(1) When $\rho^2 = 1$*,*

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \overset{A}{\sim} -\sum_{j=1}^{N-K-L-1} \frac{\xi_j}{Q_0} x_j, \tag{B.23}$$

where the x_j*'s are independent* χ_1^2 *random variables, and the* ξ_j*'s are the eigenvalues of*

$$P'W^{\frac{1}{2}}SW^{\frac{1}{2}}P, \tag{B.24}$$

where P *is an* $N \times (N-K-L-1)$ *orthonormal matrix with columns orthogonal to* $W^{\frac{1}{2}}C$*,* S *is the asymptotic covariance matrix of* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\epsilon_t y_t - (Z_t - \mu_Z)\lambda_2]$*,* $\epsilon_t = R_t - \mu_R - \beta(f_t - \mu_f)$*, and* $y_t = 1 - \lambda_1'(f_t - \mu_f)$ *is the normalized SDF.*

(2) When $0 < \rho^2 < 1$*,*

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{B.25}$$

where

$$n_t = 2\left[-u_t + (1-\rho^2)v_t + \gamma' z_t\right]/Q_0 \qquad \text{for known } W, \tag{B.26}$$

$$n_t = \left[u_t^2 - 2u_t + (1-\rho^2)(2v_t - v_t^2) + 2\gamma' z_t\right]/Q_0 \qquad \text{for } \hat{W} = \hat{V}_R^{-1}, \tag{B.27}$$

with $e_0 = [I_N - 1_N(1_N' W 1_N)^{-1} 1_N' W]\mu_R$, $u_t = e'W(R_t - \mu_R)$, $v_t = e_0' W(R_t - \mu_R)$, *and* $z_t = [0,\ u_t(f_t - \mu_f)' V_f^{-1},\ e'WZ_t]'$.

(3) When $\rho^2 = 0$,

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K+L} \frac{\xi_j}{Q_0} x_j, \tag{B.28}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$[X_1' W X_1 - X_1' W 1_N (1_N' W 1_N)^{-1} 1_N' W X_1]([0_{K+L},\ I_{K+L}] V(\hat{\gamma}) [0_{K+L},\ I_{K+L}]'), \tag{B.29}$$

where $X_1 = [\beta,\ \mu_Z]$ *and* $V(\hat{\gamma})$ *is given in Proposition B.1 (for known weighting matrix* W*) or Proposition B.2 (for estimated GLS).*

Model Comparison Tests

Consider models A and B. Let f_1, f_2, and f_3 be three sets of distinct factors, where f_i is of dimension $K_i \times 1$, $i = 1, 2, 3$. Similarly, let Z_1, Z_2, and Z_3 be three sets of distinct portfolio characteristics, where Z_i is of dimension $N \times L_i$, $i = 1, 2, 3$. Assume that model A uses factors f_1 and f_2 and portfolio characteristics Z_1 and Z_2 while model B uses factors f_1 and f_3 and portfolio characteristics Z_1 and Z_3. Therefore, model A specifies that the expected returns on the test assets are linear in the betas (or covariances) with respect to f_1 and f_2 and the means of Z_1 and Z_2, i.e.,

$$\mu_R = 1_N \lambda_{A,0} + \text{Cov}[R, f_1']\lambda_{A,1} + \mu_{Z_1}\lambda_{A,2} + \text{Cov}[R, f_2']\lambda_{A,3} + \mu_{Z_2}\lambda_{A,4} = C_A \lambda_A, \tag{B.30}$$

where $C_A = [1_N,\ \text{Cov}[R, f_1'],\ \mu_{Z_1},\ \text{Cov}[R, f_2'],\ \mu_{Z_2}]$ and $\lambda_A = [\lambda_{A,0},\ \lambda_{A,1}',\ \lambda_{A,2}',\ \lambda_{A,3}',\ \lambda_{A,4}']'$. Similarly, model B specifies that expected returns are linear in the betas (or covariances) with respect to f_1 and f_3 and the means of Z_1 and Z_3, i.e.,

$$\mu_R = 1_N \lambda_{B,0} + \text{Cov}[R, f_1']\lambda_{B,1} + \mu_{Z_1}\lambda_{B,2} + \text{Cov}[R, f_3']\lambda_{B,3} + \mu_{Z_3}\lambda_{B,4} = C_B \lambda_B, \tag{B.31}$$

where $C_B = [1_N,\ \text{Cov}[R, f_1'],\ \mu_{Z_1},\ \text{Cov}[R, f_3'],\ \mu_{Z_3}]$ and $\lambda_B = [\lambda_{B,0},\ \lambda_{B,1}',\ \lambda_{B,2}',\ \lambda_{B,3}',\ \lambda_{B,4}']'$.

Nested Models

Without loss of generality, assume $K_3 = 0$ and $L_3 = 0$, so that model A nests model B. In addition, assume $K_2 + L_2 > 0$.

Lemma B.1. *$\rho_A^2 = \rho_B^2$ if and only if $\lambda_{A,3} = 0_{K_2}$ and $\lambda_{A,4} = 0_{L_2}$.*

By the lemma, to test whether two nested models have the same R^2, one can simply perform a test of $H_0 : \lambda_{A,3} = 0_{K_2},\ \lambda_{A,4} = 0_{L_2}$ using a Wald test. Let $\hat{V}([\hat{\lambda}'_{A,3},\ \hat{\lambda}'_{A,4}]')$ be a consistent estimator of $V([\hat{\lambda}'_{A,3}, \hat{\lambda}'_{A,4}]')$, the asymptotic covariance matrix of $\sqrt{T}([\hat{\lambda}'_{A,3},\ \hat{\lambda}'_{A,4}]' - [\lambda'_{A,3},\ \lambda'_{A,4}]')$. Then, under the null hypothesis,

$$T[\hat{\lambda}'_{A,3},\ \hat{\lambda}'_{A,4}]\hat{V}([\hat{\lambda}'_{A,3},\ \hat{\lambda}'_{A,4}]')^{-1}[\hat{\lambda}'_{A,3},\ \hat{\lambda}'_{A,4}]' \stackrel{A}{\sim} \chi^2_{K_2+L_2}, \tag{B.32}$$

and this statistic can be used to test $H_0 : \rho_A^2 = \rho_B^2$.

Alternatively, it is possible to derive the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ and use this statistic to test $H_0 : \rho_A^2 = \rho_B^2$.

Proposition B.5. *Define $\tilde{H}_{A,22}$ as the lower right $(K_2 + L_2) \times (K_2 + L_2)$ submatrix of $\tilde{H}_A = (C'_A W C_A)^{-1}$. Under the null hypothesis $H_0 : \rho_A^2 = \rho_B^2$,*

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \stackrel{A}{\sim} \sum_{j=1}^{K_2+L_2} \frac{\xi_j}{Q_0} x_j, \tag{B.33}$$

where the x_j's are independent χ_1^2 random variables, the ξ_j's are the eigenvalues of the matrix $\tilde{H}_{A,22}^{-1} V([\hat{\lambda}'_{A,3},\ \hat{\lambda}'_{A,4}]')$.

Non-Nested Models

Testing $H_0 : \rho_A^2 = \rho_B^2$ is more complicated for non-nested models. The reason is that under H_0, there are three possible asymptotic distributions of $\hat{\rho}_A^2 - \hat{\rho}_B^2$, depending on why the two models have the same cross-sectional R^2.

We first provide a lemma which will be useful for deriving the first asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$.

Lemma B.2. *The conditions* $\lambda'_{A,1}f_{1,t} + \lambda'_{A,3}f_{2,t} = \lambda'_{B,1}f_{1,t} + \lambda'_{B,3}f_{3,t}$ *and* $Z_{1t}\lambda_{A,2} + Z_{2t}\lambda_{A,4} = Z_{1t}\lambda_{B,2} + Z_{3t}\lambda_{B,4}$ *hold if and only if*

$$\lambda_{A,3} = 0_{K_2}, \qquad \lambda_{B,3} = 0_{K_3}, \qquad \lambda_{A,4} = 0_{L_2}, \qquad \lambda_{B,4} = 0_{L_3}. \tag{B.34}$$

The above lemma implies that when (B.34) holds, the pricing errors of the two models are the same ($e_A = e_B$) and the two models have the same cross-sectional R^2 ($\rho_A^2 = \rho_B^2$).

A pre-test of (B.34) can be obtained in two ways. We can perform a Wald test of $H_0 : \psi = 0_{K_2+L_2+K_3+L_3}$, where $\psi = [\lambda'_{A,3},\ \lambda'_{A,4},\ \lambda'_{B,3},\ \lambda'_{B,4}]'$. Alternatively, we can derive the asymptotic distribution of $T(\hat{\rho}_A^2 - \hat{\rho}_B^2)$.

Proposition B.6. *Under the conditions in (B.34),*

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{K_2+K_3+L_2+L_3} \frac{\xi_j}{Q_0} x_j, \tag{B.35}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$\begin{bmatrix} \tilde{H}_{A,22}^{-1} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}_{B,22}^{-1} \end{bmatrix} V(\hat{\psi}), \tag{B.36}$$

where $\tilde{H}_{A,22}$ *is the lower right* $(K_2 + L_2) \times (K_2 + L_2)$ *submatrix of* $\tilde{H}_A = (C'_A W C_A)^{-1}$ *and* $\tilde{H}_{B,22}$ *is the lower right* $(K_3 + L_3) \times (K_3 + L_3)$ *submatrix of* $\tilde{H}_B = (C'_B W C_B)^{-1}$.

Models A and B can also be both correctly specified and the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ is different in this case. Below, we provide two different pre-tests of $H_0 : \rho_A^2 = \rho_B^2 = 1$. The first test is a chi-squared test of $e_A = e_B = 0_N$, which is given in the following proposition:

Proposition B.7. *Let* $n_A = N - K_1 - K_2 - L_1 - L_2 - 1$ *and* $n_B = N - K_1 - K_3 - L_1 - L_3 - 1$. *Also let* P_A *be an* $N \times n_A$ *orthonormal matrix with columns orthogonal to* $W^{\frac{1}{2}}C_A$ *and* P_B *be an* $N \times n_B$ *orthonormal matrix with columns orthogonal to* $W^{\frac{1}{2}}C_B$. *Let* ϵ_{At} *and* ϵ_{Bt} *be the regression residuals of the* N *test assets in models* A *and* B*, respectively, and define*

$$g_t(\theta) = \begin{bmatrix} g_{At}(\lambda_A) \\ g_{Bt}(\lambda_B) \end{bmatrix} = \begin{bmatrix} \epsilon_{At}y_{At} - (Z_{1,t} - \mu_{Z_1})\lambda_{A,2} - (Z_{2,t} - \mu_{Z_2})\lambda_{A,4} \\ \epsilon_{Bt}y_{Bt} - (Z_{1,t} - \mu_{Z_1})\lambda_{B,2} - (Z_{3,t} - \mu_{Z_3})\lambda_{B,4} \end{bmatrix}, \tag{B.37}$$

where $\theta = (\lambda_A',\ \lambda_B')'$, $y_{At} = 1 - \lambda_{A,1}' f_{1,t} - \lambda_{A,3}' f_{2,t}$, $y_{Bt} = 1 - \lambda_{B,1}' f_{1,t} - \lambda_{B,3}' f_{3,t}$, *and*

$$S \equiv \begin{bmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)']. \tag{B.38}$$

If (B.34) does not hold and the null hypothesis $H_0 : \rho_A^2 = \rho_B^2 = 1$ *is satisfied, then*

$$T \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix}' \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AB} \hat{W}^{\frac{1}{2}} \hat{P}_B \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BB} \hat{W}^{\frac{1}{2}} \hat{P}_B \end{bmatrix}^{-1} \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix} \overset{A}{\sim} \chi^2_{n_A + n_B}, \tag{B.39}$$

where $\hat{e}_A$ *and* $\hat{e}_B$ *are the sample pricing errors of models* A *and* B*, and* $\hat{P}_A$*,* $\hat{P}_B$*, and* $\hat{S}$ *are consistent estimators of* P_A*,* P_B*, and* S*, respectively.*

The second pre-test of $H_0 : \rho_A^2 = \rho_B^2 = 1$ is a weighted chi-squared test based on the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$, which is given in the following proposition:

Proposition B.8. *Assuming (B.34) does not hold and* $H_0 : \rho_A^2 = \rho_B^2 = 1$ *is satisfied, then*

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{n_A + n_B} \frac{\xi_j}{Q_0} x_j, \tag{B.40}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$\begin{bmatrix} -P_A' W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A & -P_A' W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B \\ P_B' W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A & P_B' W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B \end{bmatrix}. \tag{B.41}$$

Finally, if (B.34) does not hold and both models are misspecified, we can test $H_0 : \rho_A^2 - \rho_B^2$ using the normal test provided in the next proposition.

Proposition B.9. *Suppose (B.34) does not hold and* $0 < \rho_A^2 = \rho_B^2 < 1$*. We have:*

$$\sqrt{T}(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]\right). \tag{B.42}$$

When the weighting matrix W *is known,*

$$d_t = 2Q_0^{-1}\left[u_{Bt} - u_{At} - (\rho_A^2 - \rho_B^2) v_t + (\gamma_A' z_{At} - \gamma_B' z_{Bt})\right], \tag{B.43}$$

where $u_{At} = e_A' W (R_t - \mu_R)$*,* $u_{Bt} = e_B' W (R_t - \mu_R)$*,* v_t *is defined in Proposition B.4,* γ_A *and* γ_B *are the* γ*'s for models* A *and* B*, respectively, and* z_{At} *and* z_{Bt} *are the* z_t*'s for models* A *and* B*, respectively. For estimated GLS,*

$$d_t = Q_0^{-1}\left[u_{At}^2 - 2(u_{At} - u_{Bt}) - u_{Bt}^2 - (\rho_A^2 - \rho_B^2)(2v_t - v_t^2) + 2(\gamma_A' z_{At} - \gamma_B' z_{Bt})\right], \tag{B.44}$$

where $u_{At} = e'_A V_R^{-1}(R_t - \mu_R)$ *and* $u_{Bt} = e'_B V_R^{-1}(R_t - \mu_R)$.[13]

The normal test in Proposition B.9 will break down when $d_t = 0$. There are two different scenarios for $d_t = 0$. The first case occurs when $\lambda'_{A,1} f_{1,t} + \lambda'_{A,3} f_{2,t} = \lambda'_{B,1} f_{1,t} + \lambda'_{B,3} f_{3,t}$ and $Z_{1t}\lambda_{A,2} + Z_{2t}\lambda_{A,4} = Z_{1t}\lambda_{B,2} + Z_{3t}\lambda_{B,4}$. The second case occurs when $\rho_A^2 = \rho_B^2 = 1$.

[13] One could impose $H_0 : \rho_A^2 = \rho_B^2$ in (B.43) and (B.44) and the v_t terms would drop out of these expressions.

C The Price of Covariance Risk

As mentioned in the paper (see Section II.A), there are some subtle differences between the prices of beta risk and the prices of covariance risk when the risk factors are correlated. Let $\gamma = [\gamma_0,\ \gamma_1',\ \gamma_2']'$ be the zero-beta rate and risk premia for two sets of factors, f_1 and f_2. The standard relation between multiple regression betas and covariances then implies that there is a one-to-one correspondence between γ and λ; the zero-beta rates are identical and the usual risk premia are obtained by multiplying the prices of covariance risk by the factor covariance matrix:

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} \text{Var}[f_1] & \text{Cov}[f_1, f_2'] \\ \text{Cov}[f_2, f_1'] & \text{Var}[f_2] \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \tag{C.1}$$

Hence, when $\lambda_2 = 0_{K_2}$, the risk premia associated with f_2 are $\gamma_2 = \text{Cov}[f_2, f_1']\lambda_1$. Clearly, γ_2 can still be nonzero unless f_1 and f_2 are uncorrelated.[14] Similarly, we can show that $\gamma_2 = 0_{K_2}$ does not imply $\lambda_2 = 0_{K_2}$ unless f_1 and f_2 are uncorrelated.

Here, we provide some numerical illustrations of these points. In the first example, we consider two factors with

$$V_f = \begin{bmatrix} 15 & -10 \\ -10 & 15 \end{bmatrix}. \tag{C.2}$$

Suppose there are four assets and their expected returns and covariances with the two factors are

$$\mu_R = [2,\ 3,\ 4,\ 5]', \qquad V_{fR} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 5 & 2 & 1 \end{bmatrix}. \tag{C.3}$$

It is clear that the covariances of the four assets with respect to the first factor alone can fully explain μ_R because μ_R is exactly linear in the first row of V_{fR}. As a result, the second factor is irrelevant from a cross-sectional expected return perspective. However, when we compute the (multiple regression) beta matrix with respect to the two factors, we obtain:

$$\beta = V_{Rf} V_f^{-1} = \begin{bmatrix} 0.36 & 0.64 & 0.52 & 0.56 \\ 0.44 & 0.76 & 0.48 & 0.44 \end{bmatrix}'. \tag{C.4}$$

Simple calculations give $\gamma = [1,\ 15,\ -10]'$ and γ_2 is nonzero even though f_2 is irrelevant.[15]

[14] When $\lambda_2 = 0_{K_2}$, we see that $\gamma_1 = \text{Var}[f_1]\lambda_1$. Consequently, the risk premia for f_1 stay the same when we add f_2 to the model.

[15] This suggests that when the CAPM is true, it does not imply that the betas with respect to the other two Fama-French factors should not be priced. See Grauer and Janmaat (2009) for a discussion of this point.

In the second example, we change μ_R to $[10,\ 17,\ 14,\ 15]'$. In this case, the covariances with respect to f_1 alone do not fully explain μ_R (in fact, the OLS R^2 for the model with just f_1 is only 28%). However, it is easy to see that μ_R is linear in the first column of the beta matrix, implying that the R^2 of the full model is 100%. Simple calculations give us $\gamma = [1,\ 25,\ 0]'$ and $\gamma_2 = 0$ even though f_2 is needed in the factor model, along with f_1, to explain μ_R.

D Excess Returns Analysis

We provide the necessary tools for implementing the excess returns analysis described in the paper. The proofs of the various lemmas and propositions are omitted since they are similar to the ones of Appendix A.

Let f be a K-vector of factors and R a vector of excess returns (i.e., returns on zero investment portfolios) on N test assets. In many applications, R is a vector of returns on N assets in excess of the risk-free rate. The multiple regression betas of the N assets with respect to the K factors are defined as $\beta = V_{Rf}V_f^{-1}$.

The proposed K-factor beta pricing model specifies that asset expected excess returns are linear in the betas, i.e.,

$$\mu_R = \beta\gamma, \tag{D.1}$$

where γ is a vector of risk premia on the K factors. When the model is misspecified, the pricing-error vector, $\mu_R - \beta\gamma$, will be nonzero for all values of γ. In that case, it makes sense to choose γ to minimize some aggregation of pricing errors. Denoting by W an $N \times N$ symmetric positive definite weighting matrix, we define the (pseudo-true) risk premia as

$$\gamma = (\beta'W\beta)^{-1}\beta'W\mu_R. \tag{D.2}$$

The corresponding pricing errors on the N assets are then given by

$$e = \mu_R - \beta\gamma \tag{D.3}$$

and the cross-sectional R^2 is defined as

$$\rho^2 = 1 - \frac{Q}{Q_0}, \tag{D.4}$$

where

$$Q_0 = \mu_R'W\mu_R, \tag{D.5}$$

$$Q = e'We = \mu_R'W\mu_R - \mu_R'W\beta(\beta'W\beta)^{-1}\beta'W\mu_R. \tag{D.6}$$

The estimated betas from the first-pass time-series regression are given by the matrix $\hat{\beta} = \hat{V}_{Rf}\hat{V}_f^{-1}$. We then run a single CSR of $\hat{\mu}_R$ on $\hat{\beta}$ to estimate γ in the second pass. When the weighting matrix

W is known (say OLS CSR), we can estimate γ in (D.2) by

$$\hat{\gamma} = (\hat{\beta}' W \hat{\beta})^{-1} \hat{\beta}' W \hat{\mu}_R. \tag{D.7}$$

Instead of using $\hat{\beta}$, we can use $\hat{V}_{Rf}$ in the second-pass CSR. The pseudo-true parameters of this alternative second-pass CSR are given by

$$\lambda = (V_{fR} W V_{Rf})^{-1} V_{fR} W \mu_R. \tag{D.8}$$

Similarly, we can estimate λ in (D.8) by

$$\hat{\lambda} = (\hat{V}_{fR} W \hat{V}_{Rf})^{-1} \hat{V}_{fR} W \hat{\mu}_R. \tag{D.9}$$

In the GLS case, the weighting matrix W involves unknown parameters and, therefore, we need to substitute a consistent estimate of W, $\hat{W} = \hat{V}_R^{-1}$, in (D.7) and (D.9).

The sample measure of ρ^2 is similarly defined as

$$\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0}, \tag{D.10}$$

where $\hat{Q}_0$ and $\hat{Q}$ are consistent estimators of Q_0 and Q in (D.5) and (D.6), respectively.

Pricing Results

Proposition D.1. *Let $H = (\beta' W \beta)^{-1}$, $A = H\beta' W$, and $\gamma_t = AR_t$. Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma} = (\hat{\beta}' W \hat{\beta})^{-1} \hat{\beta}' W \hat{\mu}_R$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_K, V(\hat{\gamma})), \tag{D.11}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{D.12}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi) w_t + H z_t, \tag{D.13}$$

$\phi_t = \gamma_t - f_t$, $\phi = \gamma - \mu_f$, $u_t = e' W (R_t - \mu_R)$, $w_t = \gamma' V_f^{-1} (f_t - \mu_f)$, and $z_t = V_f^{-1} (f_t - \mu_f) u_t$. When the model is correctly specified, we have:

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi) w_t. \tag{D.14}$$

Proposition D.2. *Let* $H = (\beta' V_R^{-1}\beta)^{-1}$, $A = H\beta' V_R^{-1}$, *and* $\gamma_t = AR_t$. *Under a potentially misspecified model, the asymptotic distribution of* $\hat{\gamma} = (\hat{\beta}'\hat{V}_R^{-1}\hat{\beta})^{-1}\hat{\beta}'\hat{V}_R^{-1}\hat{\mu}_R$ *is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_{K+1}, V(\hat{\gamma})), \tag{D.15}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{D.16}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t - (\gamma_t - \gamma)u_t, \tag{D.17}$$

$\phi_t = \gamma_t - f_t$, $\phi = \gamma - \mu_f$, $u_t = e'V_R^{-1}(R_t - \mu_R)$, $w_t = \gamma' V_f^{-1}(f_t - \mu_f)$, $z_t = V_f^{-1}(f_t - \mu_f)u_t$. *When the model is correctly specified, we have:*

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t. \tag{D.18}$$

Lemma D.1. *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter* κ, *the asymptotic covariance matrix of* $\hat{\gamma} = (\hat{\beta}' W\hat{\beta})^{-1}\hat{\beta}' W\hat{\mu}_R$ *is given by*

$$V(\hat{\gamma}) = \Upsilon_w + \Upsilon_{w1} + \Upsilon_{w1}' + \Upsilon_{w2}, \tag{D.19}$$

where

$$\Upsilon_w = AV_R A' + (1+\kappa)\gamma' V_f^{-1}\gamma A\Sigma A', \tag{D.20}$$

$$\Upsilon_{w1} = -(1+\kappa)HV_f^{-1}\gamma e' W V_R A', \tag{D.21}$$

$$\Upsilon_{w2} = (1+\kappa)e' W V_R W e H V_f^{-1} H. \tag{D.22}$$

Lemma D.2. *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter* κ, *the asymptotic covariance matrix of* $\hat{\gamma} = (\hat{\beta}'\hat{V}_R^{-1}\hat{\beta})^{-1}\hat{\beta}'\hat{V}_R^{-1}\hat{\mu}_R$ *is given by*

$$V(\hat{\gamma}) = \Upsilon_w + \Upsilon_{w2}, \tag{D.23}$$

where

$$\Upsilon_w = H + (1+\kappa)\gamma' V_f^{-1}\gamma(\beta'\Sigma^{-1}\beta)^{-1}, \tag{D.24}$$

$$\Upsilon_{w2} = (1+\kappa)Q[HV_f^{-1}H - H], \tag{D.25}$$

with $H = (\beta' V_R^{-1}\beta)^{-1}$ *and* $Q = e' V_R^{-1} e$.

Proposition D.3. *Under a potentially misspecified model, the asymptotic distribution of $\hat{\lambda}$ is given by*

$$\sqrt{T}(\hat{\lambda} - \lambda) \stackrel{A}{\sim} N(0_K, V(\hat{\lambda})), \tag{D.26}$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t \tilde{h}'_{t+j}]. \tag{D.27}$$

To simplify the expressions for $\tilde{h}_t$, we define $\tilde{G}_t = (R_t - \mu_R)(f_t - \mu_f)' - V_{Rf}$, $\tilde{H} = (V_{fR} W V_{Rf})^{-1}$, $\tilde{A} = \tilde{H} V_{fR} W$, $\lambda_t = \tilde{A} R_t$, $u_t = e' W (R_t - \mu_R)$, and $\tilde{z}_t = (f_t - \mu_f) u_t$.

(1) With a known weighting matrix W, $\hat{\lambda} = (\hat{V}_{fR} W \hat{V}_{Rf})^{-1} \hat{V}_{fR} W \hat{\mu}_R$ and

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A} \tilde{G}_t \lambda + \tilde{H} \tilde{z}_t. \tag{D.28}$$

(2) For estimated GLS, $\hat{\lambda} = (\hat{V}_{fR} \hat{V}_R^{-1} \hat{V}_{Rf})^{-1} \hat{V}_{fR} \hat{V}_R^{-1} \hat{\mu}_R$ and

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A} \tilde{G}_t \lambda + \tilde{H} \tilde{z}_t - (\lambda_t - \lambda) u_t. \tag{D.29}$$

When the model is correctly specified, we have:

$$\tilde{h}_t = (\lambda_t - \lambda) - \tilde{A} \tilde{G}_t \lambda. \tag{D.30}$$

Results for the Sample R^2

Proposition D.4. *In the following, we set W to be V_R^{-1} for the GLS case.*

(1) When $\rho^2 = 1$,

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \stackrel{A}{\sim} -\sum_{j=1}^{N-K} \frac{\xi_j}{Q_0} x_j, \tag{D.31}$$

where the x_j's are independent χ^2_1 random variables, and the ξ_j's are the eigenvalues of

$$P' W^{\frac{1}{2}} S W^{\frac{1}{2}} P, \tag{D.32}$$

where P is an $N \times (N - K)$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}} V_{Rf}$, S is the asymptotic covariance matrix of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t y_t$, $\epsilon_t = R_t - \mu_R - \beta(f_t - \mu_f)$, and $y_t = 1 - \lambda'(f_t - \mu_f)$ is the normalized SDF.

(2) When $0 < \rho^2 < 1$,

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{D.33}$$

where

$$n_t = 2\left[-u_t y_t + (1-\rho^2)v_t\right]/Q_0 \qquad \text{for known } W, \tag{D.34}$$

$$n_t = \left[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)\right]/Q_0 \qquad \text{for } \hat{W} = \hat{V}_R^{-1}, \tag{D.35}$$

with $u_t = e'W(R_t - \mu_R)$ *and* $v_t = \mu_R' W(R_t - \mu_R)$.

(3) When $\rho^2 = 0$,

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \tag{D.36}$$

where the x_j*'s are independent* χ_1^2 *random variables and the* ξ_j*'s are the eigenvalues of*

$$(\beta' W \beta) V(\hat{\gamma}), \tag{D.37}$$

where $V(\hat{\gamma})$ *is given in Proposition D.1 (for known weighting matrix* W*) or Proposition D.2 (for estimated GLS).*

Model Comparison Tests

Consider two competing beta pricing models. Let f_1, f_2, and f_3 be three sets of distinct factors, where f_i is of dimension $K_i \times 1$, $i = 1, 2, 3$. Assume that model A uses f_1 and f_2, while Model B uses f_1 and f_3 as factors. Therefore, model A requires that the expected returns on the test assets are linear in the betas or covariances with respect to f_1 and f_2, i.e.,

$$\mu_R = \text{Cov}[R, f_1']\lambda_{A,1} + \text{Cov}[R, f_2']\lambda_{A,2} = C_A \lambda_A, \tag{D.38}$$

where $C_A = [\text{Cov}[R, f_1'],\ \text{Cov}[R, f_2']]$ and $\lambda_A = [\lambda_{A,1}',\ \lambda_{A,2}']'$. Model B requires that expected returns are linear in the betas or covariances with respect to f_1 and f_3, i.e.,

$$\mu_R = \text{Cov}[R, f_1']\lambda_{B,1} + \text{Cov}[R, f_3']\lambda_{B,3} = C_B \lambda_B, \tag{D.39}$$

where $C_B = [\text{Cov}[R, f_1'],\ \text{Cov}[R, f_3']]$ and $\lambda_B = [\lambda_{B,1}',\ \lambda_{B,3}']'$.

Given a weighting matrix W, the λ_i that maximizes the ρ^2 of model i is given by

$$\lambda_i \quad = \quad (C_i'WC_i)^{-1}C_i'W\mu_R, \tag{D.40}$$

where C_i is assumed to have full column rank, $i = A, B$. For each model, the pricing-error vector e_i, the aggregate pricing-error measure Q_i, and the corresponding goodness-of-fit measure ρ_i^2 are all defined at the beginning of Appendix D.

Nested Models

Lemma D.3. *$\rho_A^2 = \rho_B^2$ if and only if $\lambda_{A,2} = 0_{K_2}$.*

Proposition D.5. *Partition $\tilde{H}_A = (C_A'WC_A)^{-1}$ as*

$$\tilde{H}_A = \begin{bmatrix} \tilde{H}_{A,11} & \tilde{H}_{A,12} \\ \tilde{H}_{A,21} & \tilde{H}_{A,22} \end{bmatrix}, \tag{D.41}$$

where $\tilde{H}_{A,22}$ is $K_2 \times K_2$. Under the null hypothesis $H_0 : \rho_A^2 = \rho_B^2$,

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{K_2} \frac{\xi_j}{Q_0} x_j, \tag{D.42}$$

where the x_j's are independent χ_1^2 random variables and the ξ_j's are the eigenvalues of $\tilde{H}_{A,22}^{-1}V(\hat{\lambda}_{A,2})$.

Non-Nested Models

Define the normalized SDFs for models A and B as

$$y_A = 1 - (f_1 - E[f_1])'\lambda_{A,1} - (f_2 - E[f_2])'\lambda_{A,2}, \qquad y_B = 1 - (f_1 - E[f_1])'\lambda_{B,1} - (f_3 - E[f_3])'\lambda_{B,3}. \tag{D.43}$$

Lemma D.4. *For non-nested models, $y_A = y_B$ if and only if $\lambda_{A,2} = 0_{K_2}$ and $\lambda_{B,3} = 0_{K_3}$.*

Proposition D.6. *Let $\tilde{H}_A = (C_A'WC_A)^{-1}$ and $\tilde{H}_B = (C_B'WC_B)^{-1}$, and partition them as*

$$\tilde{H}_A = \begin{bmatrix} \tilde{H}_{A,11} & \tilde{H}_{A,12} \\ \tilde{H}_{A,21} & \tilde{H}_{A,22} \end{bmatrix}, \qquad \tilde{H}_B = \begin{bmatrix} \tilde{H}_{B,11} & \tilde{H}_{B,13} \\ \tilde{H}_{B,31} & \tilde{H}_{B,33} \end{bmatrix}, \tag{D.44}$$

where $\tilde{H}_{A,11}$ and $\tilde{H}_{B,11}$ are $K_1 \times K_1$. Under the null hypothesis $H_0 : y_A = y_B$,

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{K_2+K_3} \frac{\xi_j}{Q_0} x_j, \tag{D.45}$$

where the x_j's are independent χ_1^2 random variables and the ξ_j's are the eigenvalues of

$$\begin{bmatrix} \tilde{H}_{A,22}^{-1} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & -\tilde{H}_{B,33}^{-1} \end{bmatrix} V(\hat{\psi}). \tag{D.46}$$

Proposition D.7. *Let $n_A = N - K_1 - K_2$ and $n_B = N - K_1 - K_3$. Also let P_A be an $N \times n_A$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}} C_A$ and P_B be an $N \times n_B$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}} C_B$. Let ϵ_{At} and ϵ_{Bt} be the residuals of models A and B, respectively, and define*

$$g_t(\theta) = \begin{bmatrix} g_{At}(\lambda_A) \\ g_{Bt}(\lambda_B) \end{bmatrix} = \begin{bmatrix} \epsilon_{At} y_{At} \\ \epsilon_{Bt} y_{Bt} \end{bmatrix}, \tag{D.47}$$

where $\theta = (\lambda_A', \lambda_B')'$, and

$$S \equiv \begin{bmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)']. \tag{D.48}$$

If $y_A \neq y_B$ and the null hypothesis $H_0 : \rho_A^2 = \rho_B^2 = 1$ holds, then

$$T \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix}' \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AB} \hat{W}^{\frac{1}{2}} \hat{P}_B \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BB} \hat{W}^{\frac{1}{2}} \hat{P}_B \end{bmatrix}^{-1} \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix} \overset{A}{\sim} \chi^2_{n_A+n_B}, \tag{D.49}$$

where $\hat{e}_A$ and $\hat{e}_B$ are the sample pricing errors of models A and B, and $\hat{P}_A$, $\hat{P}_B$, and $\hat{S}$ are consistent estimators of P_A, P_B, and S, respectively.

Proposition D.8. *Using the notation in Proposition D.7, if $y_A \neq y_B$ and the null hypothesis $H_0 : \rho_A^2 = \rho_B^2 = 1$ holds, then*

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{n_A+n_B} \frac{\xi_j}{Q_0} x_j, \tag{D.50}$$

where the x_j's are independent χ_1^2 random variables and the ξ_j's are the eigenvalues of

$$\begin{bmatrix} -P_A' W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A & -P_A' W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B \\ P_B' W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A & P_B' W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B \end{bmatrix}. \tag{D.51}$$

Proposition D.9. *Suppose $y_A \neq y_B$ and $0 < \rho_A^2 = \rho_B^2 < 1$. We have:*

$$\sqrt{T}(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]\right). \tag{D.52}$$

When the weighting matrix W is known,

$$d_t = 2Q_0^{-1}\left[u_{Bt}y_{Bt} - u_{At}y_{At} - (\rho_A^2 - \rho_B^2)v_t\right], \tag{D.53}$$

where $u_{At} = e_A' W(R_t - \mu_R)$, $u_{Bt} = e_B' W(R_t - \mu_R)$, and v_t is defined in Proposition D.4 in Appendix D. With the GLS weighting matrix $\hat{W} = \hat{V}_R^{-1}$,

$$d_t = Q_0^{-1}\left[u_{At}^2 - 2u_{At}y_{At} - u_{Bt}^2 + 2u_{Bt}y_{Bt} - (\rho_A^2 - \rho_B^2)(2v_t - v_t^2)\right], \tag{D.54}$$

where $u_{At} = e_A' V_R^{-1}(R_t - \mu_R)$ and $u_{Bt} = e_B' V_R^{-1}(R_t - \mu_R)$.

In the following, we report additional estimation results for the excess returns case. These results complement the ones in Table V in the paper.

Table VII

Estimates and *t*-ratios of Risk Premia with a Constrained Zero-Beta Rate

The table presents the estimation results of eight beta pricing models. The models include the CAPM, the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the Fama and French (1993) three-factor model (FF3), the intertemporal CAPM (ICAPM) specification of Petkova (2006), the consumption CAPM (CCAPM), the conditional consumption CAPM (CC-CAY) of Lettau and Ludvigson (2001), the ultimate consumption CAPM (U-CCAPM) of Parker and Julliard (2005), and the durable consumption CAPM (D-CCAPM) of Yogo (2006). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios and five industry portfolios. The data are from February 1959 to July 2007 (582 observations). We report parameter estimates $\hat{\gamma}$ (multiplied by 100), the Fama and MacBeth (1973) t-ratio under correctly specified models (t-ratio$_{fm}$), the Shanken (1992) and the Jagannathan and Wang (1998) t-ratios under correctly specified models that account for the EIV problem (t-ratio$_s$ and t-ratio$_{jw}$, respectively), and our model misspecification-robust t-ratios (t-ratio$_{pm}$).

Panel A: OLS

	CAPM	C-LAB			FF3		
	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{lab}$	$\hat{\gamma}_{prem}$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{smb}$	$\hat{\gamma}_{hml}$
Estimate	0.63	0.57	−0.20	0.40	0.50	0.16	0.39
t-ratio$_{fm}$	3.33	3.15	−1.46	3.13	2.75	1.24	3.21
t-ratio$_s$	3.32	3.11	−0.97	2.09	2.75	1.24	3.21
t-ratio$_{jw}$	3.30	3.14	−0.82	2.13	2.74	1.24	3.19
t-ratio$_{pm}$	3.31	2.94	−0.69	1.43	2.74	1.23	3.15

	ICAPM					CCAPM
	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$	$\hat{\gamma}_{cg}$
Estimate	0.53	0.31	−0.10	−0.06	−0.59	0.67
t-ratio$_{fm}$	2.90	3.86	−1.55	−5.27	−4.17	3.38
t-ratio$_s$	2.81	2.19	−0.88	−3.35	−2.38	2.58
t-ratio$_{jw}$	2.85	2.03	−0.79	−3.40	−2.15	2.47
t-ratio$_{pm}$	2.82	1.99	−0.79	−3.26	−2.20	2.48

	CC-CAY			U-CCAPM	D-CCAPM		
	$\hat{\gamma}_{cay}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cg\cdot cay}$	$\hat{\gamma}_{cg36}$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cgdur}$
Estimate	0.67	0.35	0.01	4.67	0.55	0.93	0.00
t-ratio$_{fm}$	1.63	2.19	2.31	3.63	3.06	3.70	0.00
t-ratio$_s$	1.19	1.60	1.69	2.11	3.00	2.36	0.00
t-ratio$_{jw}$	1.25	1.51	1.52	2.19	3.02	2.25	0.00
t-ratio$_{pm}$	0.29	0.32	0.26	2.20	2.91	0.96	0.00

Table VII (Continued)
Estimates and t-ratios of Risk Premia with a Constrained Zero-Beta Rate

Panel B: GLS

	CAPM	C-LAB			FF3		
	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{lab}$	$\hat{\gamma}_{prem}$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{smb}$	$\hat{\gamma}_{hml}$
Estimate	0.50	0.51	−0.12	0.02	0.51	0.23	0.41
t-ratio$_{fm}$	2.81	2.82	−1.84	0.22	2.82	1.80	3.51
t-ratio$_{s}$	2.81	2.81	−1.75	0.21	2.82	1.80	3.50
t-ratio$_{jw}$	2.81	2.82	−1.77	0.21	2.82	1.79	3.49
t-ratio$_{pm}$	2.80	2.82	−0.76	0.09	2.82	1.79	3.49

	ICAPM					CCAPM
	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$	$\hat{\gamma}_{cg}$
Estimate	0.52	0.24	−0.07	−0.04	−0.42	0.26
t-ratio$_{fm}$	2.91	5.15	−1.94	−4.72	−4.36	2.44
t-ratio$_{s}$	2.89	3.56	−1.36	−3.70	−3.03	2.33
t-ratio$_{jw}$	2.90	3.52	−1.19	−3.69	−2.74	2.26
t-ratio$_{pm}$	2.89	2.44	−0.94	−3.02	−2.29	1.24

	CC-CAY			U-CCAPM	D-CCAPM		
	$\hat{\gamma}_{cay}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cg \cdot cay}$	$\hat{\gamma}_{cg36}$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cgdur}$
Estimate	0.73	0.27	0.00	1.95	0.50	0.16	0.62
t-ratio$_{fm}$	2.88	2.37	0.50	3.86	2.80	1.31	1.60
t-ratio$_{s}$	2.50	2.06	0.44	3.36	2.79	1.27	1.54
t-ratio$_{jw}$	2.46	2.08	0.42	3.72	2.81	1.28	1.56
t-ratio$_{pm}$	1.38	1.08	0.17	2.16	2.80	0.63	1.00

Table VIII
Estimates and t-ratios of Prices of Covariance Risk with a Constrained Zero-Beta Rate (OLS Case)

The table presents the estimation results of eight beta pricing models. The models include the CAPM, the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the Fama and French (1993) three-factor model (FF3), the intertemporal CAPM (ICAPM) specification of Petkova (2006), the consumption CAPM (CCAPM), the conditional consumption CAPM (CC-CAY) of Lettau and Ludvigson (2001), the ultimate consumption CAPM (U-CCAPM) of Parker and Julliard (2005), and the durable consumption CAPM (D-CCAPM) of Yogo (2006). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios and five industry portfolios. The data are from February 1959 to July 2007 (582 observations). We report parameter estimates $\hat{\lambda}$ and the model misspecification-robust t-ratio (t-ratio$_{pm}$).

	CAPM	C-LAB			FF3		
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{lab}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$
Estimate	336.31	20.28	−153.11	240.85	439.97	1.87	8.15
t-ratio$_{pm}$	2.99	0.10	−0.81	1.54	3.56	1.20	4.70

	ICAPM					CCAPM
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{term}$	$\hat{\lambda}_{def}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rf}$	$\hat{\lambda}_{cg}$
Estimate	−2161.60	288.94	−271.79	−802.35	−107.68	10980.68
t-ratio$_{pm}$	−2.09	1.09	−1.26	−2.10	−1.31	2.42

	CC-CAY			U-CCAPM	D-CCAPM		
	$\hat{\lambda}_{cay}$	$\hat{\lambda}_{cg}$	$\hat{\lambda}_{cg\cdot cay}$	$\hat{\lambda}_{cg36}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{cg}$	$\hat{\lambda}_{cgdur}$
Estimate	2337.11	63.90	6707.55	4214.10	−126.03	160.84	−10.54
t-ratio$_{pm}$	0.29	0.42	0.28	2.18	−0.24	0.90	−0.43

Table IX
Tests of Equality of Cross-Sectional R^2s with a Constrained Zero-Beta Rate

The table presents pairwise tests of equality of the OLS and GLS cross-sectional R^2s of eight beta pricing models. The models include the CAPM, the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the Fama and French (1993) three-factor model (FF3), the intertemporal CAPM (ICAPM) specification of Petkova (2006), the consumption CAPM (CCAPM), the conditional consumption CAPM (CC-CAY) of Lettau and Ludvigson (2001), the ultimate consumption CAPM (U-CCAPM) of Parker and Julliard (2005), and the durable consumption CAPM (D-CCAPM) of Yogo (2006). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios and five industry portfolios. The data are from February 1959 to July 2007 (582 observations). We report the difference between the sample cross-sectional R^2s of the models in row i and column j, $\hat{\rho}_i^2 - \hat{\rho}_j^2$, and the associated p-value (in parenthesis) for the test of $H_0 : \rho_i^2 = \rho_j^2$. The p-values are computed under the assumption that the models are potentially misspecified.

Panel A: OLS

	C-LAB	FF3	ICAPM	CCAPM	CC-CAY	U-CCAPM	D-CCAPM
CAPM	−0.035	−0.100	−0.115	−0.022	−0.028	−0.088	−0.025
	(0.292)	(**0.001**)	(0.062)	(0.594)	(0.481)	(0.108)	(0.525)
C-LAB		−0.065	−0.080	0.013	0.007	−0.053	0.009
		(0.280)	(0.222)	(0.809)	(0.888)	(0.331)	(0.877)
FF3			−0.015	0.078	0.072	0.012	0.074
			(0.310)	(0.183)	(0.266)	(0.639)	(0.220)
ICAPM				0.093	0.087	0.026	0.089
				(0.148)	(0.216)	(0.393)	(0.174)
CCAPM					−0.006	−0.066	−0.004
					(0.913)	(0.206)	(0.913)
CC-CAY						−0.060	0.002
						(0.270)	(0.962)
U-CCAPM							0.063
							(0.274)

Panel B: GLS

	C-LAB	FF3	ICAPM	CCAPM	CC-CAY	U-CCAPM	D-CCAPM
CAPM	−0.032	−0.216	−0.281	0.014	−0.047	−0.052	−0.025
	(0.635)	(**0.000**)	(0.071)	(0.843)	(0.614)	(0.597)	(0.618)
C-LAB		−0.184	−0.248	0.046	−0.015	−0.020	0.007
		(0.054)	(0.139)	(0.618)	(0.883)	(0.857)	(0.923)
FF3			−0.065	0.230	0.169	0.164	0.191
			(0.681)	(**0.009**)	(0.148)	(0.134)	(**0.008**)
ICAPM				0.295	0.234	0.229	0.256
				(0.095)	(0.210)	(0.233)	(0.127)
CCAPM					−0.061	−0.066	−0.039
					(0.491)	(0.516)	(0.342)
CC-CAY						−0.005	0.022
						(0.963)	(0.796)
U-CCAPM							0.027
							(0.787)

Table X
Multiple Model Comparison Tests with a Constrained Zero-Beta Rate

The table presents multiple model comparison tests of the OLS and GLS cross-sectional R^2s of eight beta pricing models. The models include the CAPM, the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the Fama and French (1993) three-factor model (FF3), the intertemporal CAPM (ICAPM) specification of Petkova (2006), the consumption CAPM (CCAPM), the conditional consumption CAPM (CC-CAY) of Lettau and Ludvigson (2001), the ultimate consumption CAPM (U-CCAPM) of Parker and Julliard (2005), and the durable consumption CAPM (D-CCAPM) of Yogo (2006). The models are estimated using monthly excess returns on the 25 Fama-French size and book-to-market ranked portfolios and five industry portfolios. The data are from February 1959 to July 2007 (582 observations). We report the benchmark models in column 1 and their sample R^2s in column 2. r in column 3 denotes the number of alternative models in each multiple non-nested model comparison. LR in column 4 is the value of the likelihood ratio statistic with p-value given in column 5. s in column 6 denotes the number of models that nest the benchmark model. Finally, $\hat{\rho}^2_M - \hat{\rho}^2$ in column 7 denotes the difference between the sample R^2 of the expanded model (M) and the sample R^2 of the benchmark model with p-value given in column 8.

Panel A: OLS

Benchmark	$\hat{\rho}^2$	r	LR	p-value	s	$\hat{\rho}^2_M - \hat{\rho}^2$	p-value
CAPM	0.858	2	2.592	0.106	4	0.121	0.155
C-LAB	0.893	5	1.491	0.282			
FF3	0.958	5	1.029	0.535			
ICAPM	0.972	5	0.000	0.810			
CCAPM	0.880	4	2.089	0.165	2	0.019	0.952
CC-CAY	0.886	5	1.534	0.289			
U-CCAPM	0.946	5	0.730	0.575			
D-CCAPM	0.883	5	1.852	0.228			

Panel B: GLS

Benchmark	$\hat{\rho}^2$	r	LR	p-value	s	$\hat{\rho}^2_M - \hat{\rho}^2$	p-value
CAPM	0.058	2	0.381	0.421	4	0.354	0.296
C-LAB	0.091	5	4.351	0.102			
FF3	0.274	5	0.169	0.738			
ICAPM	0.339	5	0.000	0.680			
CCAPM	0.044	4	7.137	0.023	2	0.077	0.655
CC-CAY	0.105	5	2.412	0.210			
U-CCAPM	0.110	5	2.418	0.197			
D-CCAPM	0.083	5	7.594	0.032			

E Multiple Model Comparison

We discuss the details of the multiple model comparison test and provide a numerically efficient procedure for computing its p-value.

Our multiple model comparison test is based on the multivariate inequality test of Wolak (1989). Let $\delta = (\delta_2, \ldots, \delta_p)$ and $\hat{\delta} = (\hat{\delta}_2, \ldots, \hat{\delta}_p)$, where $\delta_i = \rho_1^2 - \rho_i^2$ and $\hat{\delta}_i = \hat{\rho}_1^2 - \hat{\rho}_i^2$ for $i = 2, \ldots, p$. We are interested in testing

$$H_0 : \delta \geq 0_r \quad \text{vs.} \quad H_1 : \delta \in \Re^r, \tag{E.1}$$

where $r = p - 1$ is the number of non-negativity restrictions. Under the null hypothesis, model 1 (the benchmark) performs at least as well as models 2 to p (the competing models).

We assume that

$$\sqrt{T}(\hat{\delta} - \delta) \stackrel{A}{\sim} N(0_r, \Sigma_{\hat{\delta}}). \tag{E.2}$$

Sufficient conditions for this assumption to hold are i) $0 < \rho_i^2 < 1$, and ii) the implied SDFs of the different models are distinct (see Appendix A).

The test statistic is constructed by first solving the following quadratic programming problem

$$\min_{\delta} (\hat{\delta} - \delta)' \hat{\Sigma}_{\hat{\delta}}^{-1} (\hat{\delta} - \delta) \quad \text{s.t.} \quad \delta \geq 0_r, \tag{E.3}$$

where $\hat{\Sigma}_{\hat{\delta}}$ is a consistent estimator of $\Sigma_{\hat{\delta}}$. Let $\tilde{\delta}$ be the optimal solution of the problem in (E.3). The likelihood ratio test of the null hypothesis is given by

$$LR = T(\hat{\delta} - \tilde{\delta})' \hat{\Sigma}_{\hat{\delta}}^{-1} (\hat{\delta} - \tilde{\delta}). \tag{E.4}$$

For computational purposes, it is convenient to consider the dual problem

$$\min_{\lambda} \lambda' \hat{\delta} + \frac{1}{2} \lambda' \hat{\Sigma}_{\hat{\delta}} \lambda \quad \text{s.t.} \quad \lambda \geq 0_r. \tag{E.5}$$

Let $\tilde{\lambda}$ be the optimal solution of the problem in (E.5). The Kuhn-Tucker test of the null hypothesis is given by

$$KT = T \tilde{\lambda}' \hat{\Sigma}_{\hat{\delta}} \tilde{\lambda}. \tag{E.6}$$

It can be readily shown that $LR = KT$.

To conduct statistical inference, we need to derive the asymptotic distribution of LR. Wolak (1989) shows that under $H_0 : \delta = 0_r$ (i.e., the least favorable value of δ under the null hypothesis), LR has a weighted chi-squared distribution

$$LR \overset{A}{\sim} \sum_{i=0}^{r} w_i(\Sigma_{\hat{\delta}}^{-1})X_i = \sum_{i=0}^{r} w_{r-i}(\Sigma_{\hat{\delta}})X_i, \tag{E.7}$$

where the X_i's are independent χ^2 random variables with i degrees of freedom, $\chi_0^2 \equiv 0$, and the weights w_i sum up to one. To compute the p-value of LR, we replace $\Sigma_{\hat{\delta}}^{-1}$ with $\hat{\Sigma}_{\hat{\delta}}^{-1}$ in the weight functions.

The biggest hurdle in determining the p-value of this multivariate inequality test is the computation of the weights. For a given $r \times r$ covariance matrix $\Sigma = (\sigma_{ij})$, the expressions for the weights $w_i(\Sigma)$, $i = 0, \ldots, r$, are given in Kudo (1963). The weights depend on Σ through the correlation coefficients $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$. When $r = 1$, $w_0 = w_1 = 1/2$. When $r = 2$,

$$w_0 = \frac{1}{2} - w_2, \tag{E.8}$$

$$w_1 = \frac{1}{2}, \tag{E.9}$$

$$w_2 = \frac{1}{4} + \frac{\arcsin(\rho_{12})}{2\pi}. \tag{E.10}$$

When $r = 3$,

$$w_0 = \frac{1}{2} - w_2, \tag{E.11}$$

$$w_1 = \frac{1}{2} - w_3, \tag{E.12}$$

$$w_2 = \frac{3}{8} + \frac{\arcsin(\rho_{12\cdot3}) + \arcsin(\rho_{13\cdot2}) + \arcsin(\rho_{23\cdot1})}{4\pi}, \tag{E.13}$$

$$w_3 = \frac{1}{8} + \frac{\arcsin(\rho_{12}) + \arcsin(\rho_{13}) + \arcsin(\rho_{23})}{4\pi}, \tag{E.14}$$

where

$$\rho_{ij\cdot k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{[(1-\rho_{ik}^2)(1-\rho_{jk}^2)]^{\frac{1}{2}}}. \tag{E.15}$$

For $r > 3$, the computation of the weights is more complicated. Following Kudo (1963), we let $P = \{1, \ldots, r\}$. There are 2^r subsets of P, which are indexed by M. Let $n(M)$ be the number of elements in M and M' be the complement of M relative to P. Define Σ_M as the submatrix that consists of the rows and columns in the set M, $\Sigma_{M'}$ as the submatrix that consists of the rows and columns in the set M', $\Sigma_{M,M'}$ the submatrix with rows corresponding to the elements in M and columns corresponding to the elements in M' ($\Sigma_{M',M}$ is similarly defined), and $\Sigma_{M\cdot M'} = \Sigma_M - \Sigma_{M,M'}\Sigma_{M'}^{-1}\Sigma_{M',M}$. Kudo (1963) shows that

$$w_i(\Sigma) = \sum_{M:\, n(M)=i} P(\Sigma_{M'}^{-1})P(\Sigma_{M\cdot M'}), \tag{E.16}$$

where $P(A)$ is the probability for a multivariate normal distribution with zero mean and covariance matrix A to have all positive elements. In the above equation, we use the convention that $P[\Sigma_{\emptyset\cdot P}] = 1$ and $P[\Sigma_{\emptyset}^{-1}] = 1$. Using (E.16), we have $w_0(\Sigma) = P(\Sigma^{-1})$ and $w_r(\Sigma) = P(\Sigma)$.

Researchers have typically used a Monte Carlo approach to compute the positive orthant probability $P(A)$. However, the Monte Carlo approach is not efficient because it requires a large number of simulations to achieve the accuracy of a few digits, even when r is relatively small.

We overcome this problem by using a formula for the positive orthant probability due to Childs (1967) and Sun (1988a). Let $R = (r_{ij})$ be the correlation matrix corresponding to A. Childs (1967) and Sun (1988a) show that

$$\begin{aligned} P_{2k}(A) &= \frac{1}{2^{2k}} + \frac{1}{2^{2k-1}\pi} \sum_{1\le i<j\le 2k} \arcsin(r_{ij}) \\ &\quad + \sum_{j=2}^{k} \frac{1}{2^{2k-j}\pi^j} \sum_{1\le i_1<\cdots<i_{2j}\le 2k} I_{2j}\left(R_{(i_1,\ldots,i_{2j})}\right), \end{aligned} \tag{E.17}$$

$$\begin{aligned} P_{2k+1}(A) &= \frac{1}{2^{2k+1}} + \frac{1}{2^{2k}\pi} \sum_{1\le i<j\le 2k+1} \arcsin(r_{ij}) \\ &\quad + \sum_{j=2}^{k} \frac{1}{2^{2k+1-j}\pi^j} \sum_{1\le i_1<\cdots<i_{2j}\le 2k+1} I_{2j}\left(R_{(i_1,\ldots,i_{2j})}\right), \end{aligned} \tag{E.18}$$

where $R_{(i_1,\ldots,i_{2j})}$ denotes the submatrix consisting of the $(i_1,\ldots,i_{2j})$-th rows and columns of R, and

$$I_{2j}(\Lambda) = \frac{(-1)^j}{(2\pi)^j} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{2j} \frac{1}{\omega_i}\right) \exp\left(-\frac{\omega'\Lambda\omega}{2}\right) \mathrm{d}\omega_1 \cdots \mathrm{d}\omega_{2j}, \tag{E.19}$$

where Λ is a $2j \times 2j$ covariance matrix and $\omega = (\omega_1,\ldots,\omega_{2j})'$. Sun (1988a) provides a recursive relation for $I_{2j}(\Lambda)$ that allows us to obtain I_{2j} starting from I_2. Sun's formula enables us to compute the $2j$-th order multivariate integral I_{2j} using a $(j-1)$-th order multivariate integral, which can be obtained numerically using the Gauss-Legendre quadrature method. Sun (1988b) provides a Fortran subroutine to compute $P(A)$ for $r \le 9$. We improve on Sun's program and are able to accurately compute $P(A)$ and hence $w_i(\Sigma)$ for $r \le 11$.

F Simulation Designs

We provide a detailed description of the various simulation designs. In all of our simulations, the factors and the returns on the test assets are drawn from a multivariate normal distribution. We incorporate the pricing-model restrictions for the different scenarios by changing the mean return vector μ_R. The covariance matrix of the factors and returns, V, is chosen based on the covariance matrix estimated from the data, i.e., $V = \hat{V}$. Since the distribution of $\hat{\rho}^2$ is independent of μ_f, without loss of generality we set $\mu_f = 0_K$ in all simulation designs.

Single R^2s

We start with the specification tests — the R^2 test based on Proposition A.4 and the approximate F-test. To evaluate the size properties of these tests, we simulate data from a world in which FF3 is exactly true. The corresponding mean return vector is set to be

$$\mu_R = \hat{X}\hat{\gamma}, \tag{F.1}$$

where $\hat{X}$ and $\hat{\gamma}$ are the sample estimates of X and γ. Here, and in the calibration of other simulation parameters below, we refer to the estimates obtained using the actual data. To analyze the power of the specification tests, we set $\mu_R = \hat{\mu}_R$, which implies that the population R^2s for FF3 are 0.747 (OLS) and 0.298 (GLS), the sample values reported in Table I in the paper.

Turning to the size properties of the test of $H_0 : \rho^2 = 0$, we simulate a world in which FF3 has no explanatory power, i.e., we set

$$\mu_R = \hat{\gamma}_0 1_N + \hat{e}, \tag{F.2}$$

where $\hat{\gamma}_0$ and $\hat{e}$ are the estimated zero-beta rate and sample pricing errors from FF3. To study the power of the test of $H_0 : \rho^2 = 0$, we set $\mu_R = \hat{\mu}_R$.

Pairwise Tests of Equality of Cross-Sectional R^2s

For nested models, we consider CAPM (model B), which is nested by FF3 (model A). To evaluate the size of the weighted chi-squared test described in Proposition A.5, we choose μ_R such that $0 < \rho_A^2 = \rho_B^2 < 1$. Specifically, we set

$$\mu_R = \hat{C}_B\hat{\lambda}_B + \hat{e}_A, \tag{F.3}$$

where $\hat{C}_B$ and $\hat{\lambda}_B$ are the sample estimates of C_B and λ_B obtained from CAPM and $\hat{e}_A$ are the sample pricing errors obtained from FF3. This will guarantee that $\lambda_{A,2} = 0_{K_2}$ and $0 < \rho_A^2 = \rho_B^2 < 1$. This simulation design yields population R^2s of 0.313 (OLS) and 0.132 (GLS). To evaluate the power of the test, we set $\mu_R = \hat{\mu}_R$, which implies that the population R^2s for FF3 and CAPM are 0.747 and 0.115 for OLS and 0.298 and 0.107 for GLS, the sample values reported in Table I in the paper.

For the non-nested models case, it is more complicated to generate μ_R such that $\rho_A^2 = \rho_B^2$. Since we focus on the normal test (Proposition A.9), we need to generate μ_R such that $y_A \neq y_B$ and also both models are misspecified. We define

$$\mu_R = (\hat{C}_A\hat{\lambda}_A + \hat{C}_B\hat{\lambda}_B)/2 + a\hat{e}_A + b\hat{e}_B, \tag{F.4}$$

where a and b are chosen such that

$$\mu_R'\hat{W}\hat{C}_A(\hat{C}_A'\hat{W}\hat{C}_A)^{-1}\hat{C}_A'\hat{W}\mu_R = \mu_R'\hat{W}\hat{C}_B(\hat{C}_B'\hat{W}\hat{C}_B)^{-1}\hat{C}_B'\hat{W}\mu_R, \tag{F.5}$$

i.e., $\rho_A^2 = \rho_B^2 = \rho^2$, and ρ^2 is set to be as close as possible to $(\hat{\rho}_A^2 + \hat{\rho}_B^2)/2$. With our choice of a and b, ρ^2 is the same for FF3 and C-LAB: 0.647 for OLS and 0.203 for the GLS case. These are the averages of the sample R^2s reported in Table I in the paper. To evaluate the power of the test, we set $\mu_R = \hat{\mu}_R$, which implies that the population R^2s for FF3 and C-LAB are set equal to their sample values in Table I in the paper.

Multiple Tests of Equality of Cross-Sectional R^2s

Finally, we examine the multiple-comparison inequality test for non-nested models. To evaluate the size of the test, we consider the case in which all models have the same ρ^2 value, so as to maximize the likelihood of rejection under the null. We simulate six different single-factor models corresponding to the factors vw, smb, $cg36$, lab, $prem$, and rf and implement the likelihood ratio test with $r = 5$. We now explain how we can set μ_R such that the cross-sectional R^2 for each single-factor model is the same.

Let $V_{Rf,i} = \text{Cov}[R_t, f_{it}]$ for $i = 1, \ldots, K$. Suppose W is the weighting matrix. Let

$$M = I_N - \eta(\eta'\eta)^{-1}\eta', \tag{F.6}$$

where $\eta = W^{\frac{1}{2}}1_N$.

The cross-sectional R^2 of the model with factor i is given by

$$\rho_i^2 = \frac{(V_{Rf,i}' W^{\frac{1}{2}} M W^{\frac{1}{2}} \mu_R)^2}{(V_{Rf,i}' W^{\frac{1}{2}} M W^{\frac{1}{2}} V_{Rf,i})(\mu_R' W^{\frac{1}{2}} M W^{\frac{1}{2}} \mu_R)}. \tag{F.7}$$

Let

$$V_{Rf,i}^n = \frac{V_{Rf,i}}{(V_{Rf,i}' W^{\frac{1}{2}} M W^{\frac{1}{2}} V_{Rf,i})^{\frac{1}{2}}}, \tag{F.8}$$

we can then write

$$\rho_i^2 = \frac{(V_{Rf,i}^n{}' W^{\frac{1}{2}} M W^{\frac{1}{2}} \mu_R)^2}{\mu_R' W^{\frac{1}{2}} M W^{\frac{1}{2}} \mu_R}. \tag{F.9}$$

To ensure that all models have the same ρ^2, a sufficient condition is

$$V_{Rf,i}^n{}' W^{\frac{1}{2}} M W^{\frac{1}{2}} \mu_R = c, \tag{F.10}$$

where c is a constant. Let $V_{Rf}^n = [V_{Rf,1}^n, \ldots, V_{Rf,K}^n]$, we have

$$V_{Rf}^n{}' W^{\frac{1}{2}} M W^{\frac{1}{2}} \mu_R = c 1_K. \tag{F.11}$$

If we set $\mu_R = V_{Rf} \lambda_1$, then

$$\lambda_1 = c (V_{Rf}^n{}' W^{\frac{1}{2}} M W^{\frac{1}{2}} V_{Rf})^{-1} 1_K, \tag{F.12}$$

and we can choose μ_R to be

$$\mu_R = \hat{c} \hat{V}_{Rf} (\hat{V}_{Rf}^n{}' \hat{W}^{\frac{1}{2}} \hat{M} \hat{W}^{\frac{1}{2}} \hat{V}_{Rf})^{-1} 1_K, \tag{F.13}$$

where $\hat{M}$ is a consistent estimator of M and $\hat{V}_{Rf}^n$ is a consistent estimator of V_{Rf}^n. In our simulations, we choose $\hat{c} = \hat{V}_{Rf}^n{}' \hat{W}^{\frac{1}{2}} \hat{M} \hat{W}^{\frac{1}{2}} \hat{\mu}_R$ when the factor is the value-weighted market return. The common ρ^2 for the various models is 0.306 for OLS and 0.235 for the GLS case.

To examine the power of the test, we set $\mu_R = \hat{\mu}_R$ and simulate five of our original models (CCAPM, U-CCAPM, C-LAB, FF3, and ICAPM), so that the population R^2 of each model is set equal to its sample R^2 in Table I in the paper.

G Additional References

Bentler, Peter M., and Maia Berkane, 1986, Greatest lower bound to the elliptical theory kurtosis parameter, *Biometrika* 73, 240–241.

Childs, Donald R., 1967, Reduction of the multivariate normal integral to characteristic form, *Biometrika* 54, 293–300.

Davidson, Russell, and James D. MacKinnon, 2003. *Econometric Theory and Methods* (Oxford University Press, New York).

Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2010, Further results on the limiting distribution of GMM sample moment conditions, Working paper, Federal Reserve Bank of Atlanta.

Grauer, Robert R., and Johannus A. Janmaat, 2009, On the power of cross-sectional and multivariate tests of the CAPM, *Journal of Banking and Finance* 33, 775–787.

Kudo, Akio, 1963, A multivariate analogue of the one-sided test, *Biometrika* 50, 403–418.

Magnus, Jan R., and Heinz Neudecker, 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Wiley, New York).

Maruyama, Yosihito, and Takashi Seo, 2003, Estimation of moment parameter in elliptical distributions, *Journal of the Japan Statistical Society* 33, 215–229.

Sun, Hong-Jie, 1988a, A general reduction method for n-variate normal orthant probability, *Communications in Statistics – Theory and Methods* 11, 3913–3921.

Sun, Hong-Jie, 1988b, A Fortran subroutine for computing normal orthant probabilities of dimensions up to nine, *Communications in Statistics – Simulation and Computation* 17, 1097–1111.