Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology

Raymond Kan, Cesare Robotti, and Jay Shanken*

First draft: April 2008
This version: February 2009

*Kan is from the University of Toronto; Robotti is from the Federal Reserve Bank of Atlanta; Shanken is from Emory University and the National Bureau of Economic Research. We thank Wayne Ferson, Nikolay Gospodinov, Ravi Jagannathan, Yaxuan Qi, Guofu Zhou, seminar participants at the Board of Governors of the Federal Reserve System, Concordia University, Federal Reserve Bank of Atlanta, Federal Reserve Bank of New York, and University of Toronto for helpful discussions and comments. Kan gratefully acknowledges financial support from the National Bank Financial of Canada. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Corresponding author: Jay Shanken, Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, Georgia, 30322, USA; telephone: (404)727-4772; fax: (404)727-5238. E-mail: jay_shanken@bus.emory.edu.

Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology

Abstract

Since Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973), the two-pass cross-sectional regression (CSR) methodology has become the most popular approach for estimating and testing asset pricing models. Statistical inference with this method is typically conducted under the assumption that the models are correctly specified, i.e., expected returns are exactly linear in asset betas. This can be a problem in practice since all models are, at best, approximations of reality and are likely to be subject to a certain degree of misspecification. We propose a general methodology for computing misspecification-robust asymptotic standard errors of the risk premia estimates. We also derive the asymptotic distribution of the sample CSR R^2 and develop a test of whether two competing linear beta pricing models have the same population R^2. This provides a formal alternative to the common heuristic of simply comparing the R^2 estimates in evaluating relative model performance. Finally, we provide an empirical application which demonstrates the importance of our new results when applied to a variety of asset pricing models.

I. Introduction

In the empirical asset pricing literature, the popular two-pass cross-sectional regression (CSR) methodology developed by Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) is often used for estimating risk premia and testing pricing models that relate expected security returns to security betas on economic factors (beta pricing models). Although there are many variations of this two-pass methodology, the basic approach always involves two steps. In the first pass, the betas of the test assets are estimated using the usual ordinary least squares (OLS) time series regression of returns on some common factors. In the second pass, the returns on test assets are regressed on the betas estimated from the first pass. By running this second-pass CSR on a period-by-period basis, we obtain time series of the intercept and the slope coefficients. The average values of the intercept and the slope coefficients are then used as estimates of the zero-beta rate and factor risk premia, with standard errors computed from these time series as well.

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. Measurement errors in the betas cause two problems. The first is that the estimated zero-beta rate and risk premia are biased, though Shanken (1992) shows they are consistent as the length of the time series increases to infinity. The second problem is that the usual Fama-MacBeth standard errors for the estimated zero-beta rate and risk premia are inconsistent. Shanken (1992) addresses this by developing an asymptotically valid EIV adjustment of the standard errors. Jagannathan and Wang (1998) extend this asymptotic analysis by relaxing the assumption that returns are homoskedastic conditional on the factors.[1] Finally, the finite sample properties of these two-pass estimators have been studied by Ahn and Gadarowski (2003), Chen and Kan (2003), and Shanken and Zhou (2007).

Standard inference using the two-pass methodology implicitly assumes that expected returns are exactly linear in the betas, i.e., the beta pricing model is correctly specified. It is difficult to justify this assumption when estimating many different models because some (if not all) of the models are bound to be misspecified. Moreover, since asset pricing models are, at best, approximations of reality, it is inevitable that we will often, knowingly or unknowingly (because of limited power), estimate an expected return relation that departs from exact linearity in the betas. The first contribution of this paper is the development of misspecification-robust asymptotic standard errors

[1] Jagannathan, Skoulakis and Wang (2008) provide a synthesis of the two-pass CSR methodology.

for the estimated zero-beta rate and risk premia. Our analysis generalizes the results of Hou and Kimmel (2006) and Shanken and Zhou (2007), which are derived under a normality assumption.

One nice feature of our robust standard errors is that they are applicable whether a model is correctly specified or not. In addition, under a multivariate elliptical assumption, we provide simple expressions for the asymptotic variances of the zero-beta rate and risk premia estimates. In the case of the generalized least squares (GLS) CSR estimators, we prove that the variances are always larger when the model is misspecified. The difference depends on the extent of model misspecification as well as on the correlation between the factors and returns. We show that the misspecification adjustment term can be very large when the underlying factor is poorly mimicked by asset returns, a situation that typically arises when the factors are macroeconomic variables.

Judgement about the empirical success of a beta pricing model is often based on its cross-sectional R^2. A high value is usually considered evidence that the model does a good job of explaining the cross-section of expected returns. Several papers have analyzed the properties of the population values of the cross-sectional R^2 measures. Although there is an exact linear relation between expected returns and betas when a market index (factor portfolio) is mean-variance efficient, Roll and Ross (1994) show that there may be no relation at all, i.e., an OLS R^2 of zero, even if an index is nearly efficient. Kandel and Stambaugh (1995) document related limitations of the OLS R^2 and show that there is a direct relation between the GLS R^2 and the relative efficiency of an index. Lewellen, Nagel and Shanken (2009) provide a multifactor generalization of this result, with mimicking portfolios substituted for non-traded factors. They also argue, as do Jagannathan and Wang (1996), that the OLS R^2 can still be economically meaningful if the objective is to model the expected returns for a particular set of assets.

Jagannathan, Kubota, and Takehara (1998), Kan and Zhang (1999), and Lewellen, Nagel, and Shanken (2009) employ simulation methods to explore sampling issues in estimating the cross-sectional R^2.[2] However, to our knowledge no attempt has been made to derive the asymptotic distribution of the sample cross-sectional R^2. Building on our analysis of parameter estimation under potential model misspecification, the second contribution of this paper is to characterize the asymptotic distribution of the sample R^2, thereby filling a significant gap in the literature.

[2]In contrast to our paper, Jagannathan, Kubota, and Takehara (1998) and Kan and Zhang (1999) examine the sampling errors of the CSR R^2 and risk premia estimates under the assumption that one of the factors is useless (i.e., independent of returns).

Finally, although R^2s for competing models are routinely compared in empirical asset pricing studies, no *formal* model comparison test has yet been proposed in this context. This is essential since the R^2 statistics are subject to considerable statistical variation. Consequently, a model with a higher sample R^2 may not truly outperform its competitor. The third contribution of this paper is the introduction of a methodology to formally test whether two beta pricing models have the same population R^2. We find that the asymptotic distribution of the difference in sample R^2s of two models depends on whether the models are correctly specified or not, and on whether the models are nested or non-nested.

After developing the econometric methodology, we provide an in-depth empirical analysis that demonstrates the relevance of our new tests. We examine the performance of a variety of unconditional and conditional beta pricing models that have been proposed as refinements of the static capital asset pricing model (CAPM) and consumption CAPM (CCAPM). We start by investigating whether these models pass a specification test based on the sample cross-sectional R^2 and find that, in many instances, the models are rejected at conventional statistical levels. This provides compelling motivation to explicitly account for model misspecification in the subsequent empirical analysis of R^2s.

Next, we examine whether model misspecification substantially affects the standard errors of the zero-beta rate and risk premia estimates. Consistent with our theoretical results, we find that the t-ratios are about the same under correctly specified and potentially misspecified models when the underlying factors are returns on well diversified portfolios. However, standard errors can differ substantially when the underlying factors are not traded, e.g., macroeconomic factors.

Finally, we analyze whether different beta pricing models have significantly different cross-sectional R^2 measures. It appears that the commonly used returns and factors are sometimes too noisy to conclude that one model clearly outperforms the others. For example, using the commonly employed 25 size and book-to-market ranked portfolios as test assets, there is not much statistical evidence to establish that the five-factor intertemporal capital asset pricing model (ICAPM) of Petkova (2006) outperforms even the simple unconditional CAPM in terms of cross-sectional R^2. However, the advantage of the Fama and French (1993) three-factor model over the CAPM is statistically significant for this metric.

The rest of the paper is organized as follows. Section II presents an asymptotic analysis of the zero-beta rate and risk premia estimates under potentially misspecified models. We also consider an alternative CSR approach that uses covariances with the factors, rather than (multiple regression) betas, as the regressors. In addition, we provide an asymptotic analysis of the sample cross-sectional R^2s under correctly specified and misspecified models. Section III introduces tests of equality of cross-sectional R^2s for two competing models and provides the asymptotic distributions of the test statistics for different scenarios. Section IV presents an empirical application. The final section summarizes our findings and the Appendix contains proofs of all propositions.

II. Asymptotic Analysis under Potentially Misspecified Models

A. Population Measures of Pricing Errors and Cross-Sectional R^2s

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

$$\mu = E[Y] \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \tag{1}$$

$$V = \mathrm{Var}[Y] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{2}$$

where V is assumed to be positive definite. The multiple regression betas of the N assets with respect to the K factors are defined as $\beta = V_{21}V_{11}^{-1}$. These are measures of systematic risk or the sensitivity of returns to the factors. In addition, we denote the covariance matrix of the residuals of the N assets by $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$.

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

$$\mu_2 = X\gamma, \tag{3}$$

where $X = [1_N,\ \beta]$ is assumed to be of full column rank, 1_N is an N-vector of ones, and $\gamma = [\gamma_0,\ \gamma_1']'$ is a vector consisting of the zero-beta rate (γ_0) and risk premia on the K factors (γ_1).[3] When the model is misspecified, the pricing error vector, $\mu_2 - X\gamma$, 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

[3]Note that constant portfolio characteristics can easily be accommodated in the CSR without creating any additional complication. The analysis that includes asset characteristics is available upon request.

an $N \times N$ symmetric positive definite weighting matrix, we define the (pseudo) zero-beta rate and risk premia as the choice of γ that minimizes the quadratic form of pricing errors:

$$\gamma_W \equiv \begin{bmatrix} \gamma_{W,0} \\ \gamma_{W,1} \end{bmatrix} = \text{argmin}_\gamma (\mu_2 - X\gamma)' W (\mu_2 - X\gamma) = (X'WX)^{-1} X'W\mu_2. \tag{4}$$

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

$$e_W = \mu_2 - X\gamma_W. \tag{5}$$

In addition to the aggregate pricing errors, researchers are often interested in a normalized goodness-of-fit measure for a model. A popular measure is the cross-sectional R^2. Following Kandel and Stambaugh (1995), this is defined as

$$\rho_W^2 = 1 - \frac{Q}{Q_0}, \tag{6}$$

where

$$Q_0 = \min_{\gamma_0} (\mu_2 - 1_N\gamma_0)' W (\mu_2 - 1_N\gamma_0) = \mu_2' W \mu_2 - \mu_2' W 1_N (1_N' W 1_N)^{-1} 1_N' W \mu_2, \tag{7}$$

$$Q = e_W' W e_W = \mu_2' W \mu_2 - \mu_2' W X (X'WX)^{-1} X' W \mu_2. \tag{8}$$

In order for ρ_W^2 to be well defined, we need to assume that μ_2 is not proportional to 1_N (the expected returns are not all equal) so that $Q_0 > 0$. Note that $0 \leq \rho_W^2 \leq 1$ and it is a decreasing function of the aggregate pricing errors $Q = e_W' W e_W$. Thus, ρ_W^2 is a natural measure of goodness of fit.

While the betas are typically used as the regressors in the second-pass CSR, there is a potential issue with the use of multiple regression betas when $K > 1$: in general, the beta of an asset with respect to a particular factor depends on what other factors are included in the first-pass time-series OLS regression. As a consequence, the interpretation of the risk premia γ_1 in the context of model selection becomes problematic (more discussion on this issue later in Section III.A). To overcome this problem, we propose an alternative second-pass CSR that uses the covariances V_{21} as the regressors. Let $C = [1_N,\ V_{21}]$ and λ_W be the choice of coefficients that minimizes the quadratic form of pricing errors:

$$\lambda_W \equiv \begin{bmatrix} \lambda_{W,0} \\ \lambda_{W,1} \end{bmatrix} = \text{argmin}_\lambda (\mu_2 - C\lambda)' W (\mu_2 - C\lambda) = (C'WC)^{-1} C'W\mu_2. \tag{9}$$

Given (4) and (9), there is a one-to-one correspondence between γ_W and λ_W:

$$\lambda_{W,0} = \gamma_{W,0}, \qquad \lambda_{W,1} = V_{11}^{-1}\gamma_{W,1}. \tag{10}$$

It is easy to see that the pricing errors from this alternative second-pass CSR, $e_W = \mu_2 - C\lambda_W$, are the same as those in (5). It follows that the ρ_W^2 for these two CSRs are also identical. However, it is important to note that unless V_{11} is a diagonal matrix, $\lambda_{W,1i} = 0$ does not imply $\gamma_{W,1i} = 0$, and vice versa.[4]

It should be emphasized that unless the model is correctly specified, γ_W, λ_W, e_W, and ρ_W^2 depend on the choice of W. Popular choices of W in the literature are $W = I_N$ (OLS CSR), $W = V_{22}^{-1}$ (GLS CSR),[5] and $W = \Sigma_d^{-1}$ (weighted least squares (WLS) CSR), where Σ_d is a diagonal matrix containing the diagonal elements of Σ. To simplify the notation, we suppress the subscript W from γ_W, λ_W, e_W, and ρ_W^2 when the choice of W is clear from the context.

B. Sample Measures of Pricing Errors and Cross-Sectional R^2s

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 paper, we assume the time series Y_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

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{11}$$

$$\hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{12}$$

The popular two-pass method first estimates the betas of the N assets by running the following multivariate regression:

$$R_t = \alpha + \beta f_t + \epsilon_t, \quad t = 1, \ldots, T. \tag{13}$$

[4]See Jagannathan and Wang (1998) and Cochrane (2005, Chapter 13.4) for a discussion of this issue. Another solution to this problem is to use simple regression betas as the regressors in the second-pass CSR, as in Chen, Roll, and Ross (1986) and Jagannathan and Wang (1996, 1998). Kan and Robotti (2009) provide asymptotic results for the CSR with simple regression betas under potentially misspecified models.

[5]As pointed out by Lewellen, Nagel, and Shanken (2009), γ_W and $e_W' W e_W$ are the same regardless of whether we use $W = V_{22}^{-1}$ or $W = \Sigma^{-1}$. However, it should be noted that the ρ_W^2 are different for $W = V_{22}^{-1}$ and $W = \Sigma^{-1}$. For the purpose of model comparison, it makes sense to use a common W across models, so we prefer to use $W = V_{22}^{-1}$ for the case of GLS CSR.

The estimated betas from this first-pass time-series regression are given by the matrix $\hat{\beta} = \hat{V}_{21}\hat{V}_{11}^{-1}$. We then run a single CSR of $\hat{\mu}_2$ on $\hat{X} = [1_N,\ \hat{\beta}]$ to estimate γ_W in the second pass.[6] When the weighting matrix W is known (say OLS CSR), we can estimate γ_W in (4) by

$$\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_2. \tag{14}$$

Similarly, letting $\hat{C} = [1_N,\ \hat{V}_{21}]$, we estimate λ_W in (9) by

$$\hat{\lambda} = (\hat{C}'W\hat{C})^{-1}\hat{C}'W\hat{\mu}_2. \tag{15}$$

In the GLS and WLS cases, the weighting matrix W involves unknown parameters and, therefore, we need to substitute a consistent estimate of W, say $\hat{W}$, in (14) and (15). This is typically the corresponding matrix of sample moments, for example, $\hat{W} = \hat{V}_{22}^{-1}$ for GLS and $\hat{W} = \hat{\Sigma}_d^{-1}$ for WLS.

The sample measure of ρ^2 is similarly defined as

$$\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0}, \tag{16}$$

where $\hat{Q}_0$ and $\hat{Q}$ are consistent estimators of Q_0 and Q in (7) and (8), respectively. When W is known, we estimate Q_0 and Q using

$$\hat{Q}_0 = \hat{\mu}_2'W\hat{\mu}_2 - \hat{\mu}_2'W1_N(1_N'W1_N)^{-1}1_N'W\hat{\mu}_2, \tag{17}$$

$$\hat{Q} = \hat{\mu}_2'W\hat{\mu}_2 - \hat{\mu}_2'W\hat{X}(\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_2. \tag{18}$$

When W is not known, we replace W with $\hat{W}$ in the formulas above.

C. *Asymptotic Distribution of $\hat{\gamma}$ under Potentially Misspecified Models*

When computing the standard error of $\hat{\gamma}$, researchers typically rely on the asymptotic distribution of $\hat{\gamma}$ under the assumption that the model is correctly specified. Shanken (1992) presents the asymptotic distribution of $\hat{\gamma}$ under the conditional homoskedasticity assumption on the residuals. Jagannathan and Wang (1998) extend Shanken's results by allowing for conditional heteroskedasticity as well as autocorrelated errors.

Two recent papers have investigated the asymptotic distribution of $\hat{\gamma}$ under potentially misspecified models. Hou and Kimmel (2006) derive the asymptotic distribution of $\hat{\gamma}$ for the case of

[6]Some studies allow $\hat{\beta}$ to change throughout the sample period. For example, in the original Fama and MacBeth (1973) study, the betas used in the CSR for month t were estimated from data prior to that month. We do not study this case here mainly because the estimator of γ from this alternative procedure is generally not consistent.

GLS CSR with a known value of γ_0, and Shanken and Zhou (2007) present asymptotic results for the OLS, WLS, and GLS cases with γ_0 unknown. However, both analyses are somewhat restrictive, as they rely on the i.i.d. normality assumption. We now relax this assumption.[7]

We first present the asymptotic distribution of $\hat{\gamma}$ when W is known.

Proposition 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 $\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_2$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_{K+1}, V(\hat{\gamma})), \tag{19}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{20}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t u_t, \tag{21}$$

$\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_1)']'$, $u_t = e'W(R_t - \mu_2)$, $w_t = \gamma_1' V_{11}^{-1}(f_t - \mu_1)$, and $z_t = [0,\ (f_t - \mu_1)'V_{11}^{-1}]'$. When the model is correctly specified, we have:

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t. \tag{22}$$

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

$$\hat{h}_t = (\hat{\gamma}_t - \hat{\gamma}) - (\hat{\phi}_t - \hat{\phi})\hat{w}_t + \hat{H}\hat{z}_t\hat{u}_t, \tag{23}$$

where $\hat{\gamma}_t \equiv [\hat{\gamma}_{0t},\ \hat{\gamma}_{1t}']' = (\hat{X}'W\hat{X})^{-1}\hat{X}'WR_t$, $\hat{\phi}_t = [\hat{\gamma}_{0t},\ (\hat{\gamma}_{1t} - f_t)']'$, $\hat{\phi} = [\hat{\gamma}_0,\ (\hat{\gamma}_1 - \hat{\mu}_1)']'$, $\hat{u}_t = \hat{e}'W(R_t - \hat{\mu}_2)$ with $\hat{e} = \hat{\mu}_2 - \hat{X}\hat{\gamma}$, $\hat{w}_t = \hat{\gamma}_1'\hat{V}_{11}^{-1}(f_t - \hat{\mu}_1)$, $\hat{H} = (\hat{X}'W\hat{X})^{-1}$ and $\hat{z}_t = [0,\ (f_t - \hat{\mu}_1)'\hat{V}_{11}^{-1}]'$. 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

$$\hat{V}(\hat{\gamma}) = \frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{h}_t'. \tag{24}$$

[7] For the case of misspecified GMM, White (1994) and Hall and Inoue (2003) provide an asymptotic analysis of the parameter estimates. However, the two-pass CSR is not a standard GMM procedure that estimates β and γ jointly. Instead, the two-pass CSR can be interpreted as a sequential GMM that first estimates β from one set of moment conditions and then estimates γ using a different set of moment conditions by plugging in the estimated β (see Pagan (1984) for an analysis of regressions with generated regressors and Newey (1984) for a discussion of sequential GMM). As a result, the asymptotic analyses of White (1994) and Hall and Inoue (2003) cannot be directly applied to the two-pass CSR estimator of γ.

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 (21) 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. However, since $\hat{\beta}$ is used in place of β in the actual second-pass CSR, there is an EIV problem. 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.[8] When the model is misspecified ($e \neq 0_N$), there is a third term $Hz_t u_t$, which we call the misspecification adjustment term. Traditionally, this term has been ignored by empirical researchers.

To gain a better understanding of the relative importance of the misspecification adjustment term, in the following lemma we derive an explicit expression for $V(\hat{\gamma})$ under the assumption that returns and factors are multivariate elliptically distributed.

Lemma 1. *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ,[9] the asymptotic variance of $\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_2$ is given by*

$$V(\hat{\gamma}) = \Upsilon_w + \Upsilon_{w1} + \Upsilon_{w1}' + \Upsilon_{w2}, \tag{25}$$

where

$$\begin{aligned} \Upsilon_w &= AV_{22}A' + (1+\kappa)\gamma_1' V_{11}^{-1}\gamma_1 A\Sigma A', &(26)\\ \Upsilon_{w1} &= -(1+\kappa)H[0,\ \gamma_1' V_{11}^{-1}]' e' W V_{22} A' &(27)\\ \Upsilon_{w2} &= (1+\kappa) e' W V_{22} W e H \tilde{V}_{11}^{-1} H, &(28)\end{aligned}$$

with $\tilde{V}_{11}^{-1} = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11}^{-1} \end{bmatrix}$.

Note that when $\kappa = 0$, Lemma 1 collapses to the expression given by Shanken and Zhou (2007) in their Proposition 1 under normality. For general W, the misspecification adjustment term

[8] 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.

[9] 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.

$\Upsilon_{w1} + \Upsilon_{w1}' + \Upsilon_{w2}$ is not necessarily positive semidefinite. However, for true GLS with $W = V_{22}^{-1}$ or $W = \Sigma^{-1}$, we have $AV_{22}We = Ae = 0_{K+1}$, so Υ_{w1} vanishes, resulting in the following simple expression for $V(\hat{\gamma})$:

$$V(\hat{\gamma}) = H + (1+\kappa)\gamma_1' V_{11}^{-1}\gamma_1 (X'\Sigma^{-1}X)^{-1} + (1+\kappa)QH\tilde{V}_{11}^{-1}H, \tag{29}$$

where $H = (X'V_{22}^{-1}X)^{-1}$ and $Q = e'V_{22}^{-1}e$. The misspecification adjustment term $(1+\kappa)QH\tilde{V}_{11}^{-1}H$ is positive semidefinite in this case since $1+\kappa > 0$ (see Bentler and Berkane (1986)) and V_{11}^{-1} is positive definite. Note that the adjustment term is positively related to the aggregate pricing errors Q and the kurtosis parameter κ.

We now turn our attention to the asymptotic distribution of $\hat{\gamma}$ when W must be estimated. Under a correctly specified model, the use of $\hat{W}$ instead of W does not alter the asymptotic distribution of $\hat{\gamma}$ (proof is available upon request). However, the asymptotic distribution is affected when the model is misspecified. In the following proposition, we present the distribution for the GLS case.[10]

Proposition 2. *Let $H = (X'V_{22}^{-1}X)^{-1}$, $A = HX'V_{22}^{-1}$, and $\gamma_t = [\gamma_{0t},\ \gamma_{1t}']' = AR_t$. Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma} = (\hat{X}'\hat{V}_{22}^{-1}\hat{X})^{-1}\hat{X}'\hat{V}_{22}^{-1}\hat{\mu}_2$ is given by*

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_{K+1}, V(\hat{\gamma})), \tag{30}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h_{t+j}'], \tag{31}$$

with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + Hz_t u_t - (\gamma_t - \gamma)u_t, \tag{32}$$

$\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_1)']'$, $u_t = e'V_{22}^{-1}(R_t - \mu_2)$, $w_t = \gamma_1' V_{11}^{-1}(f_t - \mu_1)$, $z_t = [0,\ (f_t - \mu_1)'V_{11}^{-1}]'$. *When the model is correctly specified, we have:*

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t. \tag{33}$$

[10] Various results for the WLS case are available upon request.

Comparing (32) with the expression for h_t in (21), 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.

In order to gain a more concrete understanding of the misspecification adjustment term, in the following lemma we derive an explicit expression for $V(\hat{\gamma})$ in the GLS case under the multivariate elliptical assumption.

Lemma 2. *When the factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ, the asymptotic variance of $\hat{\gamma} = (\hat{X}'\hat{V}_{22}^{-1}\hat{X})^{-1}\hat{X}'\hat{V}_{22}^{-1}\hat{\mu}_2$ is given by*

$$V(\hat{\gamma}) = \Upsilon_w + \Upsilon_{w2}, \tag{34}$$

where

$$\Upsilon_w = H + (1+\kappa)\gamma_1'V_{11}^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}, \tag{35}$$

$$\Upsilon_{w2} = (1+\kappa)Q\left[(X'\Sigma^{-1}X)^{-1}\tilde{V}_{11}^{-1}(X'\Sigma^{-1}X)^{-1} + (X'\Sigma^{-1}X)^{-1}\right], \tag{36}$$

with $H = (X'V_{22}^{-1}X)^{-1}$, $Q = e'V_{22}^{-1}e$, and $\tilde{V}_{11}^{-1} = \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11}^{-1} \end{bmatrix}$.

When $Q > 0$, the misspecification adjustment term Υ_{w2} is positive definite since it is the sum of two matrices, the first positive semidefinite and the second positive definite. In the proof of Lemma 2, we show that the misspecification adjustment term crucially depends on the variance of the residuals from projecting the factors on the returns. For factors that have very low correlation with returns (e.g., macroeconomic factors), therefore, the impact of misspecification on the asymptotic variance of $\hat{\gamma}_1$ can be very large.

D. Asymptotic Distribution of $\hat{\lambda}$ under Potentially Misspecified Models

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.[11]

[11] A proof of this proposition is available upon request. The asymptotic distribution of $\hat{\lambda}$ under the i.i.d. multivariate elliptical distributional assumption is also available upon request.

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

$$\sqrt{T}(\hat{\lambda}-\lambda) \stackrel{A}{\sim} N(0_{K+1}, V(\hat{\lambda})), \tag{37}$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t \tilde{h}'_{t+j}]. \tag{38}$$

To simplify the expressions for $\tilde{h}_t$, we define $\tilde{G}_t = V_{21} - (R_t - \mu_2)(f_t - \mu_1)'$, $\tilde{z}_t = [0,\ (f_t - \mu_1)']'$, $\tilde{H} = (C'WC)^{-1}$, $\tilde{A} = \tilde{H}C'W$, $\lambda_t = \tilde{A}R_t$, and $u_t = e'W(R_t - \mu_2)$.

(1) With a known weighting matrix W, $\hat{\lambda} = (\hat{C}'W\hat{C})^{-1}\hat{C}'W\hat{\mu}_2$ and

$$\tilde{h}_t = (\lambda_t - \lambda) + \tilde{A}\tilde{G}_t\lambda_1 + \tilde{H}\tilde{z}_t u_t. \tag{39}$$

(2) For GLS, $\hat{\lambda} = (\hat{C}'\hat{V}_{22}^{-1}\hat{C})^{-1}\hat{C}'\hat{V}_{22}^{-1}\hat{\mu}_2$ and

$$\tilde{h}_t = (\lambda_t - \lambda) + \tilde{A}\tilde{G}_t\lambda_1 + \tilde{H}\tilde{z}_t u_t - (\lambda_t - \lambda)u_t. \tag{40}$$

When the model is correctly specified, we have:

$$\tilde{h}_t = (\lambda_t - \lambda) + \tilde{A}\tilde{G}_t\lambda_1. \tag{41}$$

E. Asymptotic Distribution of the Sample Cross-Sectional R^2

The sample R^2 ($\hat{\rho}^2$) in the second-pass CSR is a popular measure of goodness of fit for a model. A high $\hat{\rho}^2$ is viewed as evidence that the model under study does a good job of explaining the cross-section of expected returns. Lewellen, Nagel, and Shanken (2009) point out several pitfalls in this approach and explore simulation techniques to obtain approximate confidence intervals for ρ^2. In this subsection, we provide the first formal statistical analysis of $\hat{\rho}^2$.

In the following proposition, we show that the asymptotic distribution of $\hat{\rho}^2$ crucially depends on whether (1) the population ρ^2 is 1 (i.e., a correctly specified model), (2) $0 < \rho^2 < 1$ (a misspecified model that provides some explanatory power for the expected returns on the test assets), or (3) $\rho^2 = 0$ (a misspecified model that does not explain any of the cross-sectional variation in expected returns for the test assets).

Proposition 4. *In the following, we set W to be V_{22}^{-1} for the GLS case. (1) When $\rho^2 = 1$,*

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \overset{A}{\sim} -\sum_{j=1}^{N-K-1} \frac{\xi_j}{Q_0} x_j, \tag{42}$$

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

$$P' W^{\frac{1}{2}} S W^{\frac{1}{2}} P, \tag{43}$$

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 $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t y_t$, and $y_t = 1 - \lambda_1'(f_t - \mu_1)$ is the normalized stochastic discount factor (SDF).

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

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{44}$$

where

$$n_t = 2\left[-u_t y_t + (1-\rho^2) v_t\right] / Q_0 \qquad \text{for known } W, \tag{45}$$

$$n_t = \left[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)\right] / Q_0 \qquad \text{for } \hat{W} = \hat{V}_{22}^{-1}, \tag{46}$$

with $e_0 = [I_N - 1_N(1_N' W 1_N)^{-1} 1_N' W]\mu_2$, $u_t = e' W (R_t - \mu_2)$, and $v_t = e_0' W (R_t - \mu_2)$.

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

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \tag{47}$$

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

$$[\beta' W \beta - \beta' W 1_N (1_N' W 1_N)^{-1} 1_N' W \beta] V(\hat{\gamma}_1), \tag{48}$$

where $V(\hat{\gamma}_1)$ is given in Proposition 1 (for known weighting matrix W) or Proposition 2 (for GLS).[12]

The first asymptotic distribution in (42) allows us to perform a specification test of the beta pricing model. This is an alternative to the various multivariate asset pricing tests that have

[12]In the proof of Proposition 4, 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 $H_0 : \rho^2 = 0$, we can 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 1 and 2.

been developed in the literature.[13] Whereas the earlier tests focus on an aggregate measure of pricing errors, the R^2-based test examines aggregate pricing errors in relation to the cross-sectional variation in expected returns. In contrast, the asymptotic distribution in (47) permits a test of whether the model has *any* explanatory power for expected returns, i.e., whether we can reject $H_0 : \rho^2 = 0$.

When $0 < \rho^2 < 1$, the primary case of interest, Proposition 4 shows that asymptotically, $\hat{\rho}^2$ is normally distributed around its true value. From the results of the proposition, we see that n_t approaches zero when $\rho^2 \to 0$ or $\rho^2 \to 1$. Consequently, se($\hat{\rho}^2$) tends to be lowest when ρ^2 is close to zero or one, and se($\hat{\rho}^2$) is not monotonic in ρ^2. Note that the asymptotic normal distribution of $\hat{\rho}^2$ breaks down for the two extreme cases ($\rho^2 = 0$ or 1).[14] Intuitively, the normal distribution fails because, by construction, $\hat{\rho}^2$ will always be above zero (even when $\rho^2 = 0$) and below one (even when $\rho^2 = 1$).[15]

III. Tests for Comparing Two Competing Models

One way to think about model comparison and selection is to ask whether two competing beta pricing models have the same population cross-sectional R^2. In this section, we derive the asymptotic distribution of the difference between the sample R^2s of two models. We show that this distribution depends on whether the two models are nested or non-nested and whether the models are correctly specified or not. For model comparison, we focus on the R^2 of the CSR with known weighting matrix W and on the R^2 of the GLS CSR that uses $\hat{W} = \hat{V}_{22}^{-1}$ as the weighting matrix.

Our analysis in this section is related to the model selection tests of Kan and Robotti (2008) and Li, Xu, and Zhang (2009), which are based on the earlier work of Vuong (1989), Rivers and Vuong (2002), and Golden (2003). Whereas Kan and Robotti (2008) and Li, Xu, and Zhang (2009) conduct tests of equality of the Hansen-Jagannathan (1997) distances of two competing asset pricing models, our objective is to test for equality of the cross-sectional R^2s of two models. The asymptotic distributions of the model selection tests developed here are derived under general

[13]See Campbell, Lo, and MacKinlay (1997) or Cochrane (2005) and the included references.

[14]This is because when $\rho^2 = 1$, we have $e = 0_N$ and $u_t = 0$, so n_t in (45) and (46) becomes zero. Similarly, when $\rho^2 = 0$, we have $y_t = 1$, $e = e_0$ and $u_t = v_t$, so again n_t in (45) and (46) vanishes.

[15]As a result, we need to use a weighted sum of independent chi-squared random variables with one degree of freedom to characterize the sampling variation of $\hat{\rho}^2$ for these two extreme cases. The asymptotic distribution of $\hat{\rho}^2$ under the i.i.d. multivariate elliptical distributional assumption is available upon request.

distributional assumptions.[16]

We 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.,

$$\mu_2 = 1_N \lambda_{A,0} + \mathrm{Cov}[R, f_1'] \lambda_{A,1} + \mathrm{Cov}[R, f_2'] \lambda_{A,2} = C_A \lambda_A, \tag{49}$$

where $C_A = [1_N,\ \mathrm{Cov}[R, f_1'],\ \mathrm{Cov}[R, f_2']]$ and $\lambda_A = [\lambda_{A,0},\ \lambda_{A,1}',\ \lambda_{A,2}']'$. Model B requires that expected returns are linear in the betas or covariances with respect to f_1 and f_3, i.e.,

$$\mu_2 = 1_N \lambda_{B,0} + \mathrm{Cov}[R, f_1'] \lambda_{B,1} + \mathrm{Cov}[R, f_3'] \lambda_{B,3} = C_B \lambda_B, \tag{50}$$

where $C_B = [1_N,\ \mathrm{Cov}[R, f_1'],\ \mathrm{Cov}[R, f_3']]$ and $\lambda_B = [\lambda_{B,0},\ \lambda_{B,1}',\ \lambda_{B,3}']'$.

In general, both models can be misspecified. Following the development in Section II.A, given a weighting matrix W, the λ_i that maximizes the ρ^2 of model i is given by

$$\lambda_i \quad = \quad (C_i' W C_i)^{-1} C_i' W \mu_2, \tag{51}$$

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 errors Q_i, the time t multivariate regression residual vector ϵ_{it}, the time t normalized SDF y_{it}, and the corresponding goodness-of-fit measure ρ_i^2 are all defined as in Section II.

When $K_2 = 0$, model B nests model A as a special case. Similarly, when $K_3 = 0$, model A nests model B. When both $K_2 > 0$ and $K_3 > 0$, the two models are non-nested. We study the nested models case in the next subsection and deal with non-nested models in Section III.B.

A. Nested Models

Without loss of generality, we assume $K_3 = 0$, so that model A nests model B. In this case, the following lemma shows that $\rho_A^2 = \rho_B^2$ is equivalent to a restriction on the parameters of model A.

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

[16] The asymptotic distributions of our model selection tests under the multivariate elliptical distributional assumption are available upon request.

Note that Lemma 3 is applicable even when the models are misspecified. 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 variance of $\sqrt{T}(\hat{\lambda}_{A,2} - \lambda_{A,2})$. Then, under the null hypothesis,

$$T\hat{\lambda}'_{A,2}\hat{V}(\hat{\lambda}_{A,2})^{-1}\hat{\lambda}_{A,2} \overset{A}{\sim} \chi^2_{K_2}, \tag{52}$$

and this statistic can be used to test $H_0 : \rho^2_A = \rho^2_B$. 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 $\hat{\lambda}$, 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 3.

Alternatively, in keeping with the common practice of comparing cross-sectional R^2s, we can derive the asymptotic distribution of $\hat{\rho}^2_A - \hat{\rho}^2_B$ and use this statistic to test $H_0 : \rho^2_A = \rho^2_B$. The next proposition presents the distribution.

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

$$\tilde{H}_A = \begin{bmatrix} \tilde{H}_{A,11} & \tilde{H}_{A,12} \\ \tilde{H}_{A,21} & \tilde{H}_{A,22} \end{bmatrix}, \tag{53}$$

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

$$T(\hat{\rho}^2_A - \hat{\rho}^2_B) \overset{A}{\sim} \sum_{j=1}^{K_2} \frac{\xi_j}{Q_0} x_j, \tag{54}$$

where the x_j's are independent χ^2_1 random variables and the ξ_j's are the eigenvalues of $\tilde{H}^{-1}_{A,22} V(\hat{\lambda}_{A,2})$.

Again, we emphasize that the misspecification-robust version of $V(\hat{\lambda}_{A,2})$ should be used to test $H_0 : \rho^2_A = \rho^2_B$. Model misspecification tends to create additional sampling variation in $\hat{\rho}^2_A - \hat{\rho}^2_B$. 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 $\hat{\tilde{H}}^{-1}_{A,22}\hat{V}(\hat{\lambda}_{A,2})$, computed from the consistent estimators $\hat{\tilde{H}}_{A,22}$ and $\hat{V}(\hat{\lambda}_{A,2})$.

Before considering the more complicated case of non-nested models, it is worth clarifying a point about risk premia, which we suspect is not widely understood. Lemma 3 implies that whether the extra factors f_2 improve the cross-sectional R^2 depends on whether any of the prices of covariance risk associated with f_2 are nonzero. However, $\lambda_2 = 0_{K_2}$ does not mean that the usual risk premia (coefficients on the multiple-regression betas) associated with f_2 are zero. To see this, let $\gamma = [\gamma_0,\ \gamma_1',\ \gamma_2']'$ be the zero-beta rate and the risk premia for f_1 and f_2. Then, using the one-to-one correspondence between λ and γ in (10), we have:

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Var}[f_1] & \mathrm{Cov}[f_1, f_2'] \\ \mathrm{Cov}[f_2, f_1'] & \mathrm{Var}[f_2] \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \tag{55}$$

Hence, the risk premia associated with f_2 are $\gamma_2 = \mathrm{Cov}[f_2, f_1']\lambda_1$ when λ_2 is zero. As we see, γ_2 can still be nonzero in this case unless f_1 and f_2 are uncorrelated.[17] 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.

In other words, finding a significant t-ratio on a factor risk premium — the case of a so-called "priced" factor — need not imply that inclusion of that factor will add to the cross-sectional explanatory power of a model. Similarly, finding that a factor is not "priced" in the usual sense need not imply that the factor is unimportant in explaining cross-sectional differences in expected returns. However, by Lemma 3, the corresponding implications do hold if the explanatory variables are *simple* regression betas or covariances with the factors. We provide some examples to illustrate these points.

In the first example, we consider two factors with

$$V_{11} = \begin{bmatrix} 15 & -10 \\ -10 & 15 \end{bmatrix}. \tag{56}$$

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

$$\mu_2 = [2,\ 3,\ 4,\ 5]', \qquad V_{12} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 5 & 2 & 1 \end{bmatrix}. \tag{57}$$

It is clear that the covariances (or simple regression betas) of the four assets with respect to the first factor alone can fully explain μ_2 because μ_2 is exactly linear in the first row of V_{12}. 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:

$$\beta = V_{21}V_{11}^{-1} = \begin{bmatrix} 0.36 & 0.64 & 0.52 & 0.56 \\ 0.44 & 0.76 & 0.48 & 0.44 \end{bmatrix}'. \tag{58}$$

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

Simple calculations give $\gamma = [1,\ 15,\ -10]'$ and γ_2 is nonzero even though f_2 is irrelevant.[18]

In the second example, we change μ_2 to $[10,\ 17,\ 14,\ 15]'$. In this case, the covariances (or simple regression betas) with respect to f_1 alone do not fully explain μ_2 (in fact, the OLS ρ^2 for the model with just f_1 is only 28%). However, it is easy to see that μ_2 is linear in the first column of the beta matrix, implying that the ρ^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_2.

B. Non-Nested Models

The test of $H_0 : \rho_A^2 = \rho_B^2$ 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, first let us define the normalized SDFs for models A and B as

$$y_A = 1-(f_1-E[f_1])'\lambda_{A,1}-(f_2-E[f_2])'\lambda_{A,2}, \qquad y_B = 1-(f_1-E[f_1])'\lambda_{B,1}-(f_3-E[f_3])'\lambda_{B,3}. \quad (59)$$

In the Appendix, we show that $y_A = y_B$ implies that the two models have the same pricing errors and hence $\rho_A^2 = \rho_B^2$. If $y_A \neq y_B$, there are additional cases in which $\rho_A^2 = \rho_B^2$. A second possibility is that both models are correctly specified (i.e., $\rho_A^2 = \rho_B^2 = 1$). This occurs, for example, if model A is correctly specified and the factors f_3 in model B are given by $f_3 = f_2 + \epsilon$, where ϵ is pure "noise" — a vector of measurement errors with mean zero, independent of returns. In this case, we have $C_A = C_B$ and both models produce zero pricing errors. A third possibility is that the two models produce different pricing errors but the same overall goodness of fit. Intuitively, one model might do a good job of pricing some assets that the other prices poorly and vice versa, such that the aggregate pricing errors are the same ($\rho_A^2 = \rho_B^2 < 1$). As it turns out, each of these three scenarios results in a different asymptotic distribution for $\hat{\rho}_A^2 - \hat{\rho}_B^2$.

1. $y_A = y_B$ Case

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 since

[18] 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.

it is implied by the other two.

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

Note that Lemma 4 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 $\hat{\psi} = [\hat{\lambda}'_{A,2},\ \hat{\lambda}'_{B,3}]'$. Arguing, as in the proof of Proposition 3, we can establish that under $H_0 : y_A = y_B$, the asymptotic distribution of $\hat{\psi}$ is

$$\sqrt{T}(\hat{\psi} - \psi) \overset{A}{\sim} N(0_{K_2+K_3}, V(\hat{\psi})), \tag{60}$$

where

$$V(\hat{\psi}) = \sum_{j=-\infty}^{\infty} E[\tilde{q}_t \tilde{q}'_{t+j}], \tag{61}$$

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 3.

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

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \overset{A}{\sim} \chi^2_{K_2+K_3}, \tag{62}$$

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 $\hat{\psi}$.

Alternatively, we can perform a test based on the popular ρ^2 metric, the main focus of this paper. The following proposition gives the asymptotic distribution of $\hat{\rho}^2_A - \hat{\rho}^2_B$ given $H_0 : y_A = y_B$.

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

$$\tilde{H}_A = \begin{bmatrix} \tilde{H}_{A,11} & \tilde{H}_{A,12} \\ \tilde{H}_{A,21} & \tilde{H}_{A,22} \end{bmatrix}, \qquad \tilde{H}_B = \begin{bmatrix} \tilde{H}_{B,11} & \tilde{H}_{B,13} \\ \tilde{H}_{B,31} & \tilde{H}_{B,33} \end{bmatrix}, \tag{63}$$

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

$$T(\hat{\rho}^2_A - \hat{\rho}^2_B) \overset{A}{\sim} \sum_{j=1}^{K_2+K_3} \frac{\xi_j}{Q_0} x_j, \tag{64}$$

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

$$\begin{bmatrix} \tilde{H}^{-1}_{A,22} & 0_{K_2 \times K_3} \\ 0_{K_3 \times K_2} & -\tilde{H}^{-1}_{B,33} \end{bmatrix} V(\hat{\psi}). \tag{65}$$

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 5 are all positive whereas some of the ξ_j's in Proposition 6 are negative. As a result, we need to perform a two-sided test based on $\hat{\rho}_A^2 - \hat{\rho}_B^2$ in the non-nested model case.

If we fail to reject $H_0 : y_A = y_B$, we are finished since equality of ρ_A^2 and ρ_B^2 is implied by this hypothesis. Otherwise, we need to consider the case $y_A \neq y_B$.

2. $y_A \neq y_B$ Case

As noted earlier, when $y_A \neq y_B$, the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ given $H_0 : \rho_A^2 = \rho_B^2$ 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. As this *joint* specification test focuses on the pricing errors, it can be viewed as a generalization of the cross-sectional regression test (CSRT) of Shanken (1985), which tests the validity of the expected return relation for a single pricing model.

Proposition 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 matrix with columns orthogonal to $W^{\frac{1}{2}} C_B$. Define*

$$g_t(\theta) = \begin{bmatrix} g_{At}(\lambda_A) \\ g_{Bt}(\lambda_B) \end{bmatrix} = \begin{bmatrix} \epsilon_{At} y_{At} \\ \epsilon_{Bt} y_{Bt} \end{bmatrix}, \tag{66}$$

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

$$S \equiv \begin{bmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)']. \tag{67}$$

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

$$T \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix}' \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{S}_{AB} \hat{W}^{\frac{1}{2}} \hat{P}_B \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BA} \hat{W}^{\frac{1}{2}} \hat{P}_A & \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{S}_{BB} \hat{W}^{\frac{1}{2}} \hat{P}_B \end{bmatrix} \begin{bmatrix} \hat{P}_A' \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}_B' \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix} \stackrel{A}{\sim} \chi^2_{n_A + n_B}, \tag{68}$$

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.

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

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

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{n_A+n_B} \frac{\xi_j}{Q_0} x_j, \tag{69}$$

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

$$\begin{bmatrix} -P_A' W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A & -P_A' W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B \\ P_B' W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A & P_B' W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B \end{bmatrix}. \tag{70}$$

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

If the hypothesis that both models are correctly specified is not rejected, we are finished, as the data are consistent with $H_0 : \rho_A^2 = \rho_B^2 = 1$. Otherwise, we need to determine whether $\rho_A^2 = \rho_B^2$ for some value less than one. As in our earlier analysis for $\hat{\rho}^2$, the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ changes when the models are misspecified. Proposition 9 presents the appropriate distribution for this case.

Proposition 9. *Suppose $y_A \neq y_B$ and $0 < \rho_A^2 = \rho_B^2 < 1$.*[19] *We have:*

$$\sqrt{T}(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]\right). \tag{71}$$

When the weighting matrix W is known,

$$d_t = 2Q_0^{-1}(u_{Bt} y_{Bt} - u_{At} y_{At}), \tag{72}$$

where $u_{At} = e_A' W(R_t - \mu_2)$ and $u_{Bt} = e_B' W(R_t - \mu_2)$. With the GLS weighting matrix $\hat{W} = \hat{V}_{22}^{-1}$,

$$d_t = Q_0^{-1}(u_{At}^2 - 2u_{At} y_{At} - u_{Bt}^2 + 2u_{Bt} y_{Bt}), \tag{73}$$

where $u_{At} = e_A' V_{22}^{-1}(R_t - \mu_2)$ and $u_{Bt} = e_B' V_{22}^{-1}(R_t - \mu_2)$.

Note that if $y_{At} = y_{Bt}$, then $u_{At} = u_{Bt}$, and hence $d_t = 0$. Or, if $y_{At} \neq y_{Bt}$, but both models are correctly specified (i.e., $u_{At} = u_{Bt} = 0$), then again $d_t = 0$. Thus, the normal test cannot be used in these cases, consistent with the maintained assumptions in the proposition.[20]

[19] Since $\rho_A^2 = \rho_B^2 = 0$ implies $y_A = y_B = 1$, this case is already covered by the test based on Lemma 4.

[20] Note that, depending on whether the asymptotic distribution is normal (Proposition 9) or a linear combination of independent chi-squared random variables (Propositions 6 and 8), $\hat{\rho}_A^2 - \hat{\rho}_B^2$ can be either $O_P(T^{-\frac{1}{2}})$ or $O_P(T^{-1})$, respectively, under $H_0 : \rho_A^2 = \rho_B^2$.

C. Discussion

Given the three distinct cases encountered in testing $H_0 : \rho_A^2 = \rho_B^2$ for non-nested models, the approach we have described above entails a sequential test, as suggested by Vuong (1989). In our context, this involves first testing $H_0 : y_A = y_B$ using (62) or (64). If we reject $H_0 : y_A = y_B$, then we use (68) or (69) to test $H_0 : \rho_A^2 = \rho_B^2 = 1$. Finally, if this hypothesis is also rejected, we use the normal test in Proposition 9 to test $H_0 : 0 < \rho_A^2 = \rho_B^2 < 1$. 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 $\max[\alpha_1, \alpha_2, \alpha_3]$.[21] 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%.

Another approach is to simply perform the normal test in Proposition 9. This amounts to assuming that $y_A \neq y_B$ and that both models are misspecified. The first assumption seems reasonable since most of our models only have the constant term in common. Consequently, by Lemma 4, $y_A = y_B$ would imply that the models do not account for any cross-sectional variation in expected returns, an unlikely scenario. The second assumption is sensible because asset pricing models are approximations of reality and we do not expect them to be perfectly specified. In the following empirical application, we conduct both the sequential test and the normal test when comparing non-nested models.

IV. Empirical Analysis

We apply our methodology to several asset pricing models of interest in the asset pricing literature. First, we describe the data used in the empirical analysis and outline the different specifications of the beta pricing models considered. Then we present our results.

A. Data and Beta Pricing Models

The return data are from Kenneth French's website and consist of the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios. For most of our time series, the data are from May 1953 to December 2006 (644 monthly observations). The beginning date of our sample

[21] For the sequential test to reject $\rho_A^2 = \rho_B^2$ all three tests must reject. Consider the first scenario, $y_A = y_B$. P(reject $\rho_A^2 = \rho_B^2 \mid y_A = y_B) \leq$ P(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.

period is dictated by the bond yield data availability from the Board of Governors of the Federal Reserve System.[22]

We analyze six asset pricing models. The first model is the simple static CAPM. The cross-sectional specification is

$$\mu_2 = \gamma_0 + \beta_{vw}\gamma_{vw},$$

where vw is the excess return (in excess of the one-month T-bill rate from Ibbotson Associates) on the value-weighted stock market index (NYSE-AMEX-NASDAQ) from the Center for Research in Security Prices (CRSP).

The second model (CCAPM) is the unconditional consumption CAPM, which implies

$$\mu_2 = \gamma_0 + \beta_{cg}\gamma_{cg},$$

where cg is the growth rate in real nondurables consumption (from the Bureau of Economic Analysis). For consumption growth, the monthly data start in February 1959 (575 monthly observations).

The third model (FF3) is the Fama-French (1993) empirical three-factor model with

$$\mu_2 = \gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{smb}\gamma_{smb} + \beta_{hml}\gamma_{hml},$$

where vw is the stock market factor, smb is the return difference between portfolios of small and large stocks and hml is the return difference between portfolios of high and low book-to-market ratios (from Kenneth French's website).

The fourth model (C-LAB) is the conditional CAPM of Jagannathan and Wang (1996). The cross-sectional specification is

$$\mu_2 = \gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{lab}\gamma_{lab} + \beta_{prem}\gamma_{prem},$$

where vw is the stock market factor, lab is the growth rate in per capita labor income and $prem$ is the lagged yield spread between BAA and AAA rated corporate bonds. Per capita labor income, L, is defined as the difference between total personal income and dividend payments, divided by the total population (from the Bureau of Economic Analysis). Following Jagannathan and Wang (1996), we use a two-month moving average to construct the growth rate in per capita labor income,

[22] All bond yield data are from this source unless noted otherwise.

$lab_t = (L_{t-1} + L_{t-2})/(L_{t-2} + L_{t-3}) - 1$, for the purpose of minimizing the influence of measurement error.

The fifth model (C-CCAPM) is a *conditional* CCAPM, with a cross-sectional specification of the form

$$\mu_2 = \gamma_0 + \beta_{dy}\gamma_{dy} + \beta_{cg}\gamma_{cg} + \beta_{cg \cdot dy}\gamma_{cg \cdot dy},$$

where dy, the conditioning variable, is the lagged dividend yield of the NYSE-AMEX-NASDAQ value-weighted portfolio (from CRSP). This specification is obtained by scaling the constant term and the cg factor of a linearized CCAPM by a constant and dy. Scaling factors by instruments is one popular way of allowing factor risk premia and betas to vary over time. Examples of this type of practice are found in Shanken (1990), Ferson and Schadt (1996), Cochrane (1996), and Lettau and Ludvigson (2001), among others. We choose the lagged dividend yield as an instrument because of its frequent use in the literature.

The last model (ICAPM) is the five-factor intertemporal CAPM proposed by Petkova (2006), which implies

$$\mu_2 = \gamma_0 + \beta_{vw}\gamma_{vw} + \beta_{term}\gamma_{term} + \beta_{def}\gamma_{def} + \beta_{div}\gamma_{div} + \beta_{rf}\gamma_{rf},$$

where vw is the stock market factor, $term$ is the difference between the yields of ten-year and one-year government bonds, def is the difference between the yields of long-term corporate Baa bonds and long-term government bonds (from Ibbotson Associates), div is the dividend yield on the CRSP value-weighted stock market portfolio, and rf is the one-month T-bill yield (from CRSP, Fama Risk Free Rates). Following Petkova (2006), the actual $term$, def, div and rf factors are zero mean innovations from a VAR(1) system of seven state variables that include vw, smb, and hml.[23]

B. Results

We start by estimating the sample cross-sectional R^2s of the different pricing models considered. Then we analyze the impact of potential model misspecification on the statistical properties of the estimated γ and λ parameters. Finally, we present the results of our pairwise tests of equality of

[23]In contrast to Petkova (2006), we do not orthogonalize the innovations since the R^2 of the model is the same whether we orthogonalize or not. The results for the parameter estimates using the orthogonalized innovations are available upon request.

the cross-sectional R^2s for different models.

1. Sample Cross-Sectional R^2s of the Models

One of the main contributions of this paper is to provide the asymptotic distribution of the sample cross-sectional R^2. In Table 1, we report $\hat{\rho}^2$ for each model and use the results in Proposition 4 to investigate whether the model can do a good job of explaining the cross-section of expected returns. We denote the p-value of the specification test of $H_0 : \rho^2 = 1$ by $p(\rho^2 = 1)$, and the p-value of the test of $H_0 : \rho^2 = 0$ by $p(\rho^2 = 0)$. The asymptotic standard error of the sample cross-sectional R^2 computed under the assumption that $0 < \rho^2 < 1$ is $\mathrm{se}(\hat{\rho}^2)$.

In addition, we report a generalized version of the CSRT of Shanken (1985), $\hat{Q}_c = \hat{e}\hat{V}(\hat{e})^{+}\hat{e}$, where $\hat{V}(\hat{e})^{+}$ stands for the pseudo-inverse of $\hat{V}(\hat{e})$, which is a consistent estimator of the asymptotic variance of the sample pricing errors. When the model is correctly specified (i.e., $e = 0_N$ or $\rho^2 = 1$), we have $T\hat{Q}_c \overset{A}{\sim} \chi^2_{N-K-1}$.[24] Following Shanken (1985), we also consider an approximate F-test which is given by $\hat{Q}_c \overset{\text{app.}}{\sim} \left(\frac{N-K-1}{T-N+1}\right) F_{N-K-1,T-N+1}$.[25] The two p-values associated with testing $H_0 : Q_c = 0$ are $p_1(Q_c = 0)$, the asymptotic p-value, and $p_2(Q_c = 0)$, the p-value for the approximate F-test.[26] Finally, the number of parameters in each asset pricing model is No. of par.

Table 1 about here

In Panels A and B of Table 1, we provide results for the OLS and GLS CSRs, respectively. First, we consider the specification tests based on R^2s. It turns out that several models are rejected at the 5% level: three out of six in the OLS case and four out of six using GLS. Consistent with the empirical findings of Petkova (2006), the ICAPM delivers the highest OLS and GLS R^2s and passes the corresponding specification tests.

Interestingly, not all models with high cross-sectional R^2s pass the specification test. For example, the FF3 model has the second highest OLS R^2 (0.769) but is rejected with p-value 0.000 (it is

[24] Our $\hat{Q}_c$ is more general than the CSRT of Shanken (1985) because we can use sample pricing errors from any CSR, not just the ones from the GLS CSR. In addition, we allow for conditional heteroskedasticity and autocorrelated errors. Proofs of the results related to $\hat{Q}_c$ are available upon request.

[25] Simulation evidence suggests that this test has better size properties than the asymptotic test, especially when N is large relative to T.

[26] The p-values and standard errors in Table 1 are computed assuming no serial correlation. In a separate set of results (available upon request), we implement the automatic lag selection procedure without prewhitening of Newey and West (1994). Overall, accounting for serial correlation in the data has a fairly minor impact on the results.

also rejected using GLS). This rejection with a modestly smaller R^2 may be due in part to greater precision, as suggested by the smaller standard error estimates (especially GLS) for FF3 compared to ICAPM. Thus, while the specification tests provide information about the validity of a given model, they provide little information about model comparison. Formal tests will be needed to determine whether the ICAPM outperforms the other models.

We note that the C-CCAPM also passes the R^2 specification test in the OLS and GLS cases, while the unconditional CCAPM is strongly rejected. The true R^2s of the conditional model may indeed be higher, and the pricing errors smaller, because the scaling variable dy allows the price of risk and betas to vary with the business cycle. However, we must keep in mind that the use of conditioning variables increases the number of factors and parameters, making the conditional models better able to fit the average returns in any given sample. Again, we will need a formal test, in this case to establish whether going from unconditional to conditional models truly improves the cross-sectional R^2.

In rows five through seven of Table 1, we report the generalized CSRT and corresponding p-values. The asymptotic and approximate finite sample p-values of the CSRT are close to each other and fully support the asymptotic findings based on the sample R^2s. Out of 12 cases in Panels A and B, all specification tests reject the same seven models at the 1% and 5% significance levels.

Assuming that $0 < \rho^2 < 1$, $\mathrm{se}(\hat{\rho}^2)$ captures the sampling variability of $\hat{\rho}^2$. In Table 1, we observe that the $\hat{\rho}^2$s of several models are quite volatile, with the C-CCAPM having a $\mathrm{se}(\hat{\rho}^2)$ of almost 0.45 in the OLS case. This suggests that some of the models pass the specification test simply because of low power. In fact, for the C-CCAPM and using OLS, not only do we fail to reject $H_0 : \rho^2 = 1$, but also $H_0 : \rho^2 = 0$, the hypothesis that the model cannot explain *any* of the cross-sectional differences in expected returns on the 25 size and book-to-market ranked portfolios. We fail to reject $H_0 : \rho^2 = 0$ at the 5% level for the CAPM and CCAPM as well in the OLS case and for the CCAPM and C-LAB using GLS.[27] We will see below that high volatility of the ρ^2 estimates also makes it hard to distinguish between models.

To summarize, several observations emerge from the results in Table 1. First, there is strong evidence of the need to incorporate model misspecification into our statistical analysis. Second,

[27]In computing the p-value of the test of $H_0 : \rho^2 = 0$, we impose the constraint of $\gamma_1 = 0_K$ in the computation of $\hat{V}(\hat{\gamma}_1)$. If we do not impose this constraint, then we fail to reject $H_0 : \rho^2 = 0$ for more models.

there is considerable sampling variability in $\hat{\rho}^2$ and so it is not entirely clear whether one model consistently outperforms the others. Finally, specification test results are sometimes sensitive to the weighting matrix used, and it is not always the case that models with very high $\hat{\rho}^2$s pass the specification test.

2. Properties of the γ and λ Estimates under Correctly Specified and Potentially Misspecified Models

Before turning to model comparison, we investigate whether model misspecification substantially affects the properties of the γ and λ estimators. As far as we know, all previous empirical asset pricing studies except the recent paper by Shanken and Zhou (2007) have used standard errors that assume the model is correctly specified. As we argued in the introduction, it is difficult to justify this practice because some (if not all) of the models are bound to be misspecified. In this subsection, we see whether using an asymptotic standard error that is robust to model misspecification can lead to different inferences.

In Table 2, we focus on the zero-beta rate and the risk premia estimates, $\hat{\gamma}$, of the beta pricing models. For each model, we report $\hat{\gamma}$ and associated t-ratios under correctly specified and potentially misspecified models.[28] For correctly specified models, we give the t-ratios of Fama and MacBeth (1973), followed by those of Shanken (1992) and Jagannathan and Wang (1998) which account for estimation error in the betas. Last, are the t-ratios under potentially misspecified models, based on our results in Propositions 1 and 2. The various t-ratios are identified by subscripts fm, s, jw, and pm, respectively.

Table 2 about here

Consistent with our theoretical results, we find that the t-ratios under correctly specified and potentially misspecified models are similar for traded factors, while they can differ substantially for factors that have low correlations with asset returns. Included in the latter category are the macroeconomic factors lab and cg, the financial factors $term$, def, $prem$, div, rf, as well as the factors scaled by dy. Consider, for example, the OLS results for the FF3 model in Panel A. The t-

[28]The t-ratios are computed by assuming that the errors have no serial correlation. In a separate set of results (available upon request), we implement the automatic lag selection procedure without prewhitening of Newey and West (1994). Overall, accounting for serial correlation in the data has a minor impact on the standard errors of $\hat{\gamma}$.

ratios on $\hat{\gamma}_{vw}$, $\hat{\gamma}_{smb}$ and $\hat{\gamma}_{hml}$ for correctly specified and potentially misspecified models are generally very close, as the factors are all mimicked well by the returns on the test assets. For the vw factor, the t-ratio$_{fm} = -2.96$ and the t-ratio$_{pm} = -2.58$ are approximately the same, while the t-ratios for smb and hml hardly vary at all across methods. The GLS results in Panel B deliver a similar message.

When we consider models with factors that are weakly correlated with asset returns, the picture changes substantially. For example, for the dy factor in the C-CCAPM, in Panel A we have t-ratio$_{fm} = -5.30$, t-ratio$_{s} = -2.71$, t-ratio$_{jw} = -2.82$, and t-ratio$_{pm} = -1.07$. Even in the GLS case, the standard error of $\hat{\gamma}_{dy}$ increases by more than 40% when we incorporate potential model misspecification. Finally, the ICAPM provides another example of the different conclusions that one can reach by using misspecification-robust standard errors. While the t-ratios under correctly specified models in Panel A suggest that $\hat{\gamma}_{term}$ is statistically significant (t-ratio$_{fm} = 3.97$, t-ratio$_{s} = 2.50$ and t-ratio$_{cs} = 2.55$), the t-ratio of 1.81 under potentially misspecified models provides much weaker evidence.

To summarize, we find that for factors that are weakly correlated with the returns on the test assets, all of the t-ratios under potentially misspecified models are smaller (in absolute value) than the Fama and MacBeth (1973) t-ratios. In addition, most of the misspecification-robust t-ratios are smaller (in absolute value) than the t-ratios of Shanken (1992) and Jagannathan and Wang (1998). Finally, the latter two are close to each other and substantially smaller (in absolute value) than the Fama-MacBeth t-ratios. Thus, both model misspecification and beta estimation error materially affect inference about the expected return relation.

As discussed in Section III.A, there are issues with testing whether an individual factor risk premium is zero or not in a multi-factor model. Unless the factors are uncorrelated or simple regression betas are used, only the price of covariance risk (elements of λ_1) allows us to identify factors that improve the explanatory power of the expected return model (the usual risk premium for a given factor does not). To investigate whether the covariance risks of the factors are priced, in Table 3 we present estimation results for λ. Similar to Table 2, we report $\hat{\lambda}$ and associated t-ratios, with the OLS results in Panel A and the GLS results in Panel B.[29] First we have the t-ratios

[29] The t-ratios are computed by assuming that the errors have no serial correlation. A separate set of results (available upon request) considers the automatic lag selection procedure without prewhitening of Newey and West (1994). Overall, accounting for serial correlation in the data has a modest impact on the standard errors of $\hat{\lambda}$.

of Fama-MacBeth (t-ratio$_{fm}$), then t-ratios that account for estimation error in the covariances with the model correctly specified (t-ratio$_{cs}$), and finally the t-ratios under potentially misspecified models (t-ratio$_{pm}$). All are based on our results in Proposition 3.[30]

Table 3 about here

To illustrate our point that risk premia and prices of covariance risk can deliver different messages, consider the FF3 model. In both Panels A and B, $\hat{\lambda}_{smb}$ is statistically significant at the 1% level, as all t-ratios are close to or greater than three. In contrast, $\hat{\gamma}_{smb}$ in Table 2 is not significant at the 5% level using either OLS or GLS, with all t-ratios smaller than 1.7. Hence, by focusing on the risk premium, one might think that *smb* is not an important factor in the FF3 model. However, results for the price of covariance risk, $\hat{\lambda}_{smb}$, imply that *smb* has explanatory power for the cross-section of expected returns above and beyond the other factors in the FF3 model.[31]

To summarize, accounting for model misspecification can often make a qualitative difference in determining whether estimates of the risk premium or the price of covariance risk are statistically significant, especially when the factor has low correlation with asset returns. This would typically be the case with macroeconomic or scaled factors. Unless one is confident about a model, potential model misspecification should be accounted for when computing standard errors. In addition, focusing on the $\hat{\gamma}$s, rather than $\hat{\lambda}$s, can lead to erroneous conclusions as to whether or not a factor is helpful in explaining the cross-section of expected returns.

3. Tests of Equality of the Cross-Sectional R^2s of Two Competing Models

Recall that a p-value is the probability, *under the null hypothesis*, of obtaining a test statistic at least as extreme as the one observed. As such, the p-value provides no direct information about alternative hypotheses and the extent of deviations from the null. Therefore, p-values from the specification tests do not allow us to formally compare models. In this subsection, we explore relative goodness of fit by empirically testing whether competing beta pricing models exhibit significantly

[30] We also examined t-ratios for correctly specified models under the normality assumption. These t-ratios were usually close to t-ratio$_{cs}$, which is computed under general distributional assumptions.

[31] There are also situations where the opposite happens — a risk premium is statistically significant in Table 2, while the corresponding price of covariance risk is not in Table 3. As expected, for one-factor models, $\hat{\gamma}_1$ and $\hat{\lambda}_1$ result in similar inferences. In this case, the t-ratios of the $\hat{\gamma}_1$ and $\hat{\lambda}_1$ would be identical if we imposed the null hypotheses of $\gamma_1 = 0$ and $\lambda_1 = 0$, so that the EIV adjustment terms drop out of the analysis.

different sample cross-sectional R^2s.

In Section III, we showed that the asymptotic distribution of the difference between the sample cross-sectional R^2s depends on whether the two competing models are correctly specified or not and whether they are nested or non-nested. For nested models, we use Proposition 5 to test for equality of cross-sectional R^2s.[32] For non-nested models, we use the normal test in Proposition 9 as well as the sequential test described in Section III.C. However, for ease of comparison, we only present results for the normal test, which produces just one more rejection than the sequential test.[33]

In Table 4, we report pairwise tests of equality of cross-sectional R^2s for different models, some nested and others non-nested. Panel A is for the OLS CSR and Panel B is for the GLS CSR. Each panel shows the differences between the sample cross-sectional R^2s for various pairs of models and the associated p-values (in parentheses).[34]

Table 4 about here

The main findings can be summarized as follows. First, the results show that only the unconditional CAPM and CCAPM are often outperformed by other models at the 5% level. Specifically, the CAPM is dominated by C-LAB and FF3 in Panel A, and by FF3 in Panel B, with R^2 differences around 50 percentage points in the OLS cases. In addition, the FF3 and ICAPM fare better relative to the CCAPM in Panel A, while only the C-CCAPM outperforms the CCAPM in Panel B.[35]

Second, there is no strong evidence that conditional models outperform unconditional models. For example, there is no statistically significant evidence that the ICAPM of Petkova (2006) outperforms the FF3 model in terms of OLS and GLS cross-sectional R^2s. Surprisingly, we cannot even strongly conclude that the ICAPM dominates the simple static CAPM, although the OLS p-value of 0.064 is suggestive. In addition, out of eight comparisons involving C-LAB and C-CCAPM

[32]When computing the misspecification-robust $\hat{V}(\hat{\lambda}_{A,2})$, we impose the null hypothesis $H_0 : \lambda_{A,2} = 0_{K_2}$. However, the p-values remain virtually unchanged when we do not impose the null hypothesis. Results obtained using the Wald test in (52) (not reported in the paper) are consistent with the ones shown in Table 4.

[33]The sequential test we implement is based on Lemma 4 and Propositions 7 and 9. We also experimented with a sequential test based on Propositions 6, 8, and 9, and found that both tests reject the same models.

[34]Note that in the case of non-nested models, the reported p-values are two-tailed p-values.

[35]All the p-values in Table 4 are computed assuming no serial correlation. A separate set of results (available upon request) considers the automatic lag selection procedure without prewhitening of Newey and West (1994). We find that most of the p-values of the test statistics become slightly larger and differences between models even harder to detect. Two of the four rejections of equality in Panel A of Table 4 are reversed at the 5% level.

and the unconditional models, CAPM and CCAPM, the C-LAB model of Jagannathan and Wang (1996) dominates the CAPM only in the OLS case (p-value 0.020) and the C-CCAPM outperforms the CCAPM only in the GLS case (p-value 0.025). Of course, failure to reject may, in some cases, be due to low power; as we saw earlier, the precision of the C-CCAPM sample R^2 is particularly low.

We also explored the effect of including the three Fama-French factors, along with the 25 portfolios, as test assets in the various model comparisons. For models that include one or more of these traded factors, inclusion requires that the estimated price of risk conform to the corresponding model restriction (i.e., equal the expected market premium over the zero-beta rate or the expected spread return for *hml* and *smb*) either exactly (GLS) or approximately (OLS), as discussed by Lewellen, Nagel, and Shanken (2009). For the most part, our inferences are unchanged.[36] We do find, however, that the ICAPM specification can now be rejected (GLS R^2 p-value 0.011), with the price of covariance risk for the *term* factor no longer significant after allowing for model misspecification (the GLS t-ratio declines from 2.03 to 1.03).

Finally, since the population R^2 depends on the choice of test assets when a model is misspecified, we consider the robustness of our conclusions to an alternative set of asset portfolios — 25 size-beta sorted portfolios.[37] The main differences in model comparison results are observed in the OLS case, with the C-CCAPM and ICAPM both outperforming the CAPM and the CCAPM at the 5% level. Moreover, we find that the ICAPM now outperforms the FF3 model as well, with a spread in OLS R^2 of 0.203 and a p-value of 0.033.

V. Conclusion

We have provided a systematic analysis of the asymptotic statistical properties of the traditional cross-sectional regression methodology and the associated R^2 goodness-of-fit measure when an

[36]In this context, the vector 1_N in the X matrix and in Equation (7) is modified to have entries of zero corresponding to the *hml* and *smb* test assets. Adding five industry portfolios (from Kenneth French's website) as test assets likewise has little effect on our main conclusions. The results of the analyses with portfolio restrictions and industry portfolios are available upon request.

[37]The 25 portfolios are determined by first forming size quintiles based on market capitalization rankings of all NYSE-AMEX-NASDAQ common stocks (from CRSP), and then by forming beta quintiles within each size quintile. This is similar to the approach of Fama and French (1992). We use quintiles, rather than deciles, to mitigate potential finite-sample issues related to the inversion of a large sample covariance matrix. The results of this analysis are available upon request.

underlying beta pricing model fails to hold exactly. Our misspecification-robust standard errors for the zero-beta rate and factor risk premia are derived under very general distributional assumptions, extending the previous results of Shanken and Zhou (2007) derived under normality. A nice feature of these standard errors is that they can be used whether the model is correctly specified or not.

When factors and returns are multivariate elliptically distributed, we show analytically that with GLS cross-sectional regressions, the standard errors under model misspecification are always larger than the standard errors that assume the model is correctly specified. We also show, in the GLS case, that the misspecification adjustment depends, among other things, on the correlation between the factor and the test asset returns. This adjustment can be very large when the underlying factor is poorly mimicked by asset returns.

We also provide a general asymptotic theory for the sample OLS and GLS cross-sectional R^2s. In particular, we believe our study is the first to consider (in any manner) the important sampling distribution of the *difference* between the sample R^2s of two competing models. As we show, the asymptotic distribution of this difference depends on whether the models are correctly specified and whether they are nested or non-nested.

Our econometric results are used to analyze a variety of asset pricing models that have been proposed in the literature, focusing mainly on the commonly employed 25 size and book-to-market ranked portfolios as test assets. We find that the significance of risk premia for several non-traded factors is substantially reduced once potential model misspecification is taken into account. For example, the OLS t-ratio on the risk premium for the lagged dividend yield in a conditional version of the consumption CAPM goes from -2.82 to -1.07, a reduction in magnitude of 60% (the traditional Fama-MacBeth t-ratio is -5.30).

Our empirical findings suggest that the sample cross-sectional R^2 measure can be too noisy to permit a conclusion that one model outperforms another, in that very large differences in R^2 are sometimes statistically insignificant. The estimated standard errors for the sample OLS R^2s range from 0.099 for the Fama and French (1993) three-factor model to 0.447 for a conditional version of the consumption CAPM. These findings imply that the common approach of informally relying solely on the sample R^2 and ignoring its sampling variability in comparing models can be dangerous. In this respect, our work reinforces the simulation-based conclusion of Lewellen, Nagel, and Shanken (2009), while providing a more formal framework to evaluate statistical precision and conduct inference.

Finally, the intertemporal CAPM of Petkova (2006) and the Fama and French (1993) three-factor model perform best in our model comparison tests, while the CAPM and the unconditional consumption CAPM are frequently dominated by other models. Furthermore, the intertemporal CAPM of Petkova (2006), the conditional CAPM of Jagannathan and Wang (1996) and a conditional version of the consumption CAPM are never outperformed at the 5% level.[38]

Our analysis could be extended in a number of ways. For instance, since we find that the zero-beta rate estimates of all models are unreasonably large, it would be interesting to perform model comparison under the constraint that the zero-beta rate equals the risk-free rate. Other metrics for comparing models besides the R^2 measure could also be considered. Finally, although we have made substantial progress in deriving asymptotic results, future research should also address the small sample properties of the test statistics proposed in this paper.

[38]This includes comparisons that employ the 25 size and book-to-market ranked portfolios and the 25 size and beta ranked portfolios as test assets.

Appendix

Proof of Propositions 1 and 2: We only provide the proof of Proposition 2 here, as the proof of Proposition 1 is very similar. 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

$$\varphi = \begin{bmatrix} \mu \\ \mathrm{vec}(V) \end{bmatrix}, \qquad \hat{\varphi} = \begin{bmatrix} \hat{\mu} \\ \mathrm{vec}(\hat{V}) \end{bmatrix}. \tag{A1}$$

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

$$r_t(\varphi) = \begin{bmatrix} Y_t - \mu \\ \mathrm{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}. \tag{A2}$$

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:[39]

$$\sqrt{T}(\hat{\varphi} - \varphi) \overset{A}{\sim} N(0_{(N+K)(N+K+1)}, S_0), \tag{A3}$$

where

$$S_0 = \sum_{j=-\infty}^{\infty} E[r_t(\varphi) r_{t+j}(\varphi)']. \tag{A4}$$

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

$$\sqrt{T}(\hat{\gamma} - \gamma) \quad \overset{A}{\sim} \quad N\left(0_{K+1}, \left[\frac{\partial \gamma}{\partial \varphi'}\right] S_0 \left[\frac{\partial \gamma}{\partial \varphi'}\right]'\right). \tag{A5}$$

It is straightforward to obtain:

$$\frac{\partial \gamma}{\partial \mu_1'} = 0_{(K+1)\times K}, \qquad \frac{\partial \gamma}{\partial \mu_2'} = A. \tag{A6}$$

For the derivative of γ with respect to $\mathrm{vec}(V)$, we first need to show that

$$\frac{\partial x}{\partial \mathrm{vec}(V)'} = \left([0_K,\ V_{11}^{-1}]',\ 0_{(K+1)\times N}\right) \otimes [-\beta,\ I_N], \tag{A7}$$

[39] 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 $[\hat{\mu}',\ \mathrm{vech}(\hat{V})']'$, but the results are the same under both specifications.

where $x = \text{vec}(X)$. In order to prove this identity, we write:

$$V_{11} = [I_K,\ 0_{K\times N}]V[I_K,\ 0_{K\times N}]', \qquad V_{21} = [0_{N\times K},\ I_N]V[I_K,\ 0_{K\times N}]' \tag{A8}$$

to obtain

$$\begin{aligned}
\frac{\partial \text{vec}(V_{11})}{\partial \text{vec}(V)'} &= [I_K,\ 0_{K\times N}] \otimes [I_K,\ 0_{K\times N}], &\text{(A9)}\\
\frac{\partial \text{vec}(V_{21})}{\partial \text{vec}(V)'} &= [I_K,\ 0_{K\times N}] \otimes [0_{N\times K},\ I_N]. &\text{(A10)}
\end{aligned}$$

With the following identity

$$\begin{aligned}
\frac{\partial \text{vec}(V_{11}^{-1})}{\partial \text{vec}(V)'} &= \frac{\partial \text{vec}(V_{11}^{-1})}{\partial \text{vec}(V_{11})'}\frac{\partial \text{vec}(V_{11})}{\partial \text{vec}(V)'}\\
&= -(V_{11}^{-1} \otimes V_{11}^{-1})\left([I_K,\ 0_{K\times N}] \otimes [I_K,\ 0_{K\times N}]\right)\\
&= [V_{11}^{-1},\ 0_{K\times N}] \otimes [-V_{11}^{-1},\ 0_{K\times N}], &\text{(A11)}
\end{aligned}$$

we can use the product rule to obtain

$$\begin{aligned}
\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(V)'} &= (V_{11}^{-1} \otimes I_N)\frac{\partial \text{vec}(V_{21})}{\partial \text{vec}(V)'} + (I_K \otimes V_{21})\frac{\partial \text{vec}(V_{11}^{-1})}{\partial \text{vec}(V)'}\\
&= [V_{11}^{-1},\ 0_{K\times N}] \otimes [0_{N\times K},\ I_N] + [V_{11}^{-1},\ 0_{K\times N}] \otimes [-\beta,\ 0_{N\times N}]\\
&= [V_{11}^{-1},\ 0_{K\times N}] \otimes [-\beta,\ I_N]. &\text{(A12)}
\end{aligned}$$

Finally, using the identity

$$\frac{\partial x}{\partial \text{vec}(\beta)'} = [0_K,\ I_K]' \otimes I_N, \tag{A13}$$

we obtain:

$$\frac{\partial x}{\partial \text{vec}(V)'} = \frac{\partial x}{\partial \text{vec}(\beta)'}\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(V)'} = \left([0_K,\ V_{11}^{-1}]',\ 0_{(K+1)\times N}\right) \otimes [-\beta,\ I_N]. \tag{A14}$$

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:

$$\frac{\partial \gamma}{\partial \text{vec}(V)'} = (\mu_2' V_{22}^{-1} X \otimes I_{K+1})\frac{\partial \text{vec}(H)}{\partial \text{vec}(V)'} + (\mu_2' V_{22}^{-1} \otimes H)\frac{\partial \text{vec}(X')}{\partial \text{vec}(V)'} + (\mu_2' \otimes HX')\frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'}. \tag{A15}$$

The last two terms are given by

$$\begin{aligned}
(\mu_2' V_{22}^{-1} \otimes H)\frac{\partial \text{vec}(X')}{\partial \text{vec}(V)'} &= [H\left[0_K,\ V_{11}^{-1}\right]', 0_{(K+1)\times N}] \otimes [-\mu_2' V_{22}^{-1}\beta,\ \mu_2' V_{22}^{-1}], &\text{(A16)}\\
(\mu_2' \otimes HX')\frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'} &= -[0_K',\ \mu_2' V_{22}^{-1}] \otimes [0_{(K+1)\times K},\ A]. &\text{(A17)}
\end{aligned}$$

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

$$\begin{aligned}
& (\mu_2' V_{22}^{-1} X \otimes I_{K+1}) \frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(V)'} \\
= & (\mu_2' V_{22}^{-1} X \otimes I_{K+1}) \frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(H^{-1})'} \frac{\partial \mathrm{vec}(H^{-1})}{\partial \mathrm{vec}(V)'} \\
= & -(\mu_2' V_{22}^{-1} X \otimes I_{K+1})(H \otimes H) \Bigg[(X' V_{22}^{-1} \otimes I_{K+1}) K_{N,K+1} \frac{\partial x}{\partial \mathrm{vec}(V)'} \\
& + (X' \otimes X') \frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'} + (I_{K+1} \otimes X' V_{22}^{-1}) \frac{\partial x}{\partial \mathrm{vec}(V)'} \Bigg] \\
= & -(\gamma' \otimes H) \left\{ \left([-X' V_{22}^{-1} \beta,\ X' V_{22}^{-1}] \otimes \left[[0_K,\ V_{11}^{-1}]',\ 0_{(K+1)\times N} \right] \right) K_{N+K} \right. \\
& - [0_{(K+1)\times K},\ X' V_{22}^{-1}] \otimes [0_{(K+1)\times K},\ X' V_{22}^{-1}] \\
& \left. + \left[[0_K,\ V_{11}^{-1}]',\ 0_{(K+1)\times N} \right] \otimes [-X' V_{22}^{-1} \beta,\ X' V_{22}^{-1}] \right\} \\
= & \left[H [0_K,\ V_{11}^{-1}]',\ 0_{(K+1)\times N} \right] \otimes [\gamma' X' V_{22}^{-1} \beta,\ -\gamma' X' V_{22}^{-1}] \\
& + [0_K',\ \gamma' X' V_{22}^{-1}] \otimes [0_{(K+1)\times K},\ A] - [\gamma_1' V_{11}^{-1},\ 0_N'] \otimes [-A\beta,\ A]. \qquad \text{(A18)}
\end{aligned}$$

Combining the three terms and using the first order condition $\beta' V_{22}^{-1} e = 0_K$, we have:

$$\begin{aligned}
\frac{\partial \gamma}{\partial \mathrm{vec}(V)'} = & \left[H [0_K,\ V_{11}^{-1}]',\ 0_{(K+1)\times N} \right] \otimes \left[0_K',\ e' V_{22}^{-1} \right] \\
& - \left[\gamma_1' V_{11}^{-1},\ 0_N' \right] \otimes [-A\beta,\ A] - \left[0_K',\ e' V_{22}^{-1} \right] \otimes \left[0_{(K+1)\times K},\ A \right]. \qquad \text{(A19)}
\end{aligned}$$

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

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi) h_{t+j}(\varphi)'], \qquad \text{(A20)}$$

where

$$\begin{aligned}
h_t(\varphi) = & \frac{\partial \gamma}{\partial \varphi'} r_t(\varphi) \\
= & A(R_t - \mu_2) + \mathrm{vec}\left([0_K',\ e' V_{22}^{-1}][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} [0_K,\ V_{11}^{-1}] H \\ 0_{N \times (K+1)} \end{bmatrix} \right) \\
& - \mathrm{vec}\left([-A\beta,\ A][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} V_{11}^{-1} \gamma_1 \\ 0_N \end{bmatrix} \right) \\
& - \mathrm{vec}\left([0_{(K+1)\times K},\ A][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} 0_K \\ V_{22}^{-1} e \end{bmatrix} \right) \\
= & (\gamma_t - \gamma) + H[0_K,\ V_{11}^{-1}]'(f_t - \mu_1) u_t - A[(R_t - \mu_2) - \beta(f_t - \mu_1)](f_t - \mu_1)' V_{11}^{-1} \gamma_1 \\
& - A(R_t - \mu_2) u_t - H[0_K,\ V_{11}^{-1}]' V_{12} V_{22}^{-1} e - A\beta\gamma_1 + A\beta\gamma_1 + Ae \\
= & (\gamma_t - \gamma) + H z_t u_t - (\phi_t - \phi) w_t - (\gamma_t - \gamma) u_t. \qquad \text{(A21)}
\end{aligned}$$

The last equality follows from the first order condition $X'V_{22}^{-1}e = 0_{K+1}$ (which implies $\beta' V_{22}^{-1}e = 0_K$ and $Ae = 0_{K+1}$) and the fact that $A\beta = AX[0_K,\ I_K]' = [0_K,\ I_K]'$ gives us

$$A(R_t - \mu_2) - A\beta(f_t - \mu_1) = \gamma_t - \gamma - \begin{bmatrix} 0 \\ f_t - \mu_1 \end{bmatrix} = \phi_t - \phi. \tag{A22}$$

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

$$h_t(\varphi) = (\gamma_t - \gamma) - (\phi_t - \phi)w_t. \tag{A23}$$

This completes the proof.

Proof of Lemma 1: 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:

$$\begin{aligned} E[X_iX_jX_k] &= 0, && \text{(A24)} \\ E[X_iX_jX_kX_l] &= (1+\kappa)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), && \text{(A25)} \end{aligned}$$

where $\sigma_{ij} = \mathrm{Cov}[X_i, X_j]$. We first note that since γ_t, ϕ_t, z_t, w_t, and u_t are all linear functions of R_t and f_t, they are also jointly elliptically distributed. In addition, using (A22), we have $\phi_t - \phi = A\epsilon_t$, where $\epsilon_t = R_t - \mu_2 - \beta(f_t - \mu_1)$, which is uncorrelated with f_t. Using this result, we can easily show that

$$\begin{aligned} \mathrm{Var}[\gamma_t] &= AV_{22}A', && \text{(A26)} \\ \mathrm{Var}[\phi_t] &= A\Sigma A', && \text{(A27)} \\ \mathrm{Var}[z_t] &= \tilde{V}_{11}^{-1}, && \text{(A28)} \\ \mathrm{Var}[w_t] &= \gamma_1' V_{11}^{-1}\gamma_1, && \text{(A29)} \\ \mathrm{Var}[u_t] &= e'WV_{22}We, && \text{(A30)} \\ \mathrm{Cov}[\phi_t, z_t'] &= 0_{(K+1)\times(K+1)}, && \text{(A31)} \\ \mathrm{Cov}[\phi_t, w_t] &= 0_{K+1}, && \text{(A32)} \\ \mathrm{Cov}[\phi_t, u_t] &= A\Sigma We = AV_{22}We, && \text{(A33)} \\ \mathrm{Cov}[w_t, z_t'] &= [0,\ \gamma_1' V_{11}^{-1}], && \text{(A34)} \\ \mathrm{Cov}[z_t, u_t] &= 0_{K+1}, && \text{(A35)} \\ \mathrm{Cov}[w_t, u_t] &= 0. && \text{(A36)} \end{aligned}$$

Using these second moments, we can then apply (A24) and (A25) to obtain

$$
\begin{aligned}
E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] &= 0_{(K+1)\times(K+1)}, && \text{(A37)}\\
E[(\gamma_t-\gamma)z_t'u_t] &= 0_{(K+1)\times(K+1)}, && \text{(A38)}\\
E[z_tz_t'u_t^2] &= (1+\kappa)e'WV_{22}We\tilde{V}_{11}^{-1}, && \text{(A39)}\\
E[(\phi_t-\phi)z_t'w_tu_t] &= (1+\kappa)AV_{22}We[0,\ \gamma_1'V_{11}^{-1}], && \text{(A40)}\\
E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] &= (1+\kappa)\gamma_1'V_{11}^{-1}\gamma_1A\Sigma A'. && \text{(A41)}
\end{aligned}
$$

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

$$
\begin{aligned}
V(\hat{\gamma}) &= E[h_th_t']\\
&= \mathrm{Var}[\gamma_t] - E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] + E[(\gamma_t-\gamma)z_t'u_t]H\\
&\quad + E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] - E[(\phi_t-\phi)(\gamma_t-\gamma)'w_t] - E[(\phi_t-\phi)z_t'w_tu_t]H\\
&\quad + HE[z_tz_t'u_t^2]H + HE[z_t(\gamma_t-\gamma)'u_t] - HE[z_t(\phi_t-\phi)'u_tw_t]\\
&= AV_{22}A' + (1+\kappa)(\gamma_1'V_{11}^{-1}\gamma_1)A\Sigma A' + (1+\kappa)e'WV_{22}WeH\tilde{V}_{11}^{-1}H\\
&\quad - (1+\kappa)AV_{22}We[0,\ \gamma_1'V_{11}^{-1}]H - (1+\kappa)H[0,\ \gamma_1'V_{11}^{-1}]'e'WV_{22}A'. \qquad \text{(A42)}
\end{aligned}
$$

This completes the proof.

Proof of Lemma 2: Under the i.i.d. assumption, the expression for $V(\hat{\gamma})$ is given by

$$
\begin{aligned}
E[h_th_t'] &= \mathrm{Var}[\gamma_t] - E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] + E[(\gamma_t-\gamma)z_t'u_t]H - E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t]\\
&\quad + E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] - E[(\phi_t-\phi)(\gamma_t-\gamma)'w_t] - E[(\phi_t-\phi)z_t'w_tu_t]H\\
&\quad + E[(\phi_t-\phi)(\gamma_t-\gamma)'w_tu_t] + HE[z_tz_t'u_t^2]H + HE[z_t(\gamma_t-\gamma)'u_t]\\
&\quad - HE[z_t(\phi_t-\phi)'u_tw_t] - HE[z_t(\gamma_t-\gamma)'u_t^2] + E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t^2]\\
&\quad - E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t] + E[(\gamma_t-\gamma)(\phi_t-\phi)'w_tu_t] - E[(\gamma_t-\gamma)z_t'u_t^2]H. \qquad \text{(A43)}
\end{aligned}
$$

Following the proof of Lemma 1, we have:

$$
\begin{aligned}
\mathrm{Var}[\gamma_t] &= H, && \text{(A44)}\\
E[(\gamma_t-\gamma)(\phi_t-\phi)'w_t] &= 0_{(K+1)\times(K+1)}, && \text{(A45)}\\
E[(\gamma_t-\gamma)z_t'u_t] &= 0_{(K+1)\times(K+1)}, && \text{(A46)}\\
E[z_tz_t'u_t^2] &= (1+\kappa)Q\tilde{V}_{11}^{-1}, && \text{(A47)}
\end{aligned}
$$

$$
\begin{aligned}
E[(\phi_t-\phi)z_t'w_tu_t] &= 0_{(K+1)\times(K+1)}, && \text{(A48)}\\
E[(\phi_t-\phi)(\phi_t-\phi)'w_t^2] &= (1+\kappa)\gamma_1'V_{11}^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}, && \text{(A49)}\\
E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t] &= 0_{(K+1)\times(K+1)}, && \text{(A50)}\\
E[(\phi_t-\phi)(\gamma_t-\gamma)'w_tu_t] &= 0_{(K+1)\times(K+1)}, && \text{(A51)}\\
E[(\gamma_t-\gamma)(\gamma_t-\gamma)'u_t^2] &= (1+\kappa)QH, && \text{(A52)}\\
E[z_t(\gamma_t-\gamma)'u_t^2] &= (1+\kappa)Q\begin{bmatrix} 0 & 0_K' \\ 0_K & I_K \end{bmatrix}. && \text{(A53)}
\end{aligned}
$$

By partitioning H as

$$
H=\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \tag{A54}
$$

where H_{11} is the $(1,1)$ element of H, and using (A44)–(A53), we can write:

$$
\begin{aligned}
E[h_th_t'] &= H+(1+\kappa)\gamma_1'V_{11}^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}+(1+\kappa)QH\tilde{V}_{11}^{-1}H\\
&\quad -(1+\kappa)QH\begin{bmatrix} 0 & 0_K' \\ 0_K & I_K \end{bmatrix}+(1+\kappa)QH-(1+\kappa)Q\begin{bmatrix} 0 & 0_K' \\ 0_K & I_K \end{bmatrix}H\\
&= \Upsilon_w+(1+\kappa)Q\left(H\tilde{V}_{11}^{-1}H+\begin{bmatrix} H_{11} & 0_K' \\ 0_K & -H_{22} \end{bmatrix}\right)\\
&= \Upsilon_w+(1+\kappa)Q\begin{bmatrix} H_{12}V_{11}^{-1}H_{21}+H_{11} & H_{12}V_{11}^{-1}H_{22} \\ H_{22}V_{11}^{-1}H_{21} & H_{22}V_{11}^{-1}H_{22}-H_{22} \end{bmatrix}. \qquad \text{(A55)}
\end{aligned}
$$

By applying the identity $(X'\Sigma^{-1}X)^{-1}=H-\tilde{V}_{11}$, where $\tilde{V}_{11}=\begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} \end{bmatrix}$, we can verify that the expression of Υ_{w2} in Lemma 2 is the same as the second term in (A55) as follows:[40]

$$
\begin{aligned}
(X'\Sigma^{-1}X)^{-1}\tilde{V}_{11}^{-1}(X'\Sigma^{-1}X)^{-1}+(X'\Sigma^{-1}X)^{-1} &= (H-\tilde{V}_{11})\tilde{V}_{11}^{-1}(H-\tilde{V}_{11})+H-\tilde{V}_{11}\\
&= H\tilde{V}_{11}^{-1}H+\begin{bmatrix} H_{11} & 0_K' \\ 0_K & -H_{22} \end{bmatrix}. \qquad \text{(A56)}
\end{aligned}
$$

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

$$
\begin{aligned}
&(1+\kappa)Q(H_{22}V_{11}^{-1}H_{22}-H_{22})\\
=\ &(1+\kappa)QH_{22}V_{11}^{-1}(V_{11}-V_{11}H_{22}^{-1}V_{11})V_{11}^{-1}H_{22}\\
=\ &(1+\kappa)QH_{22}V_{11}^{-1}[V_{11}-V_{12}V_{22}^{-1}V_{21}+V_{12}V_{22}^{-1}1_N(1_N'V_{22}^{-1}1_N)^{-1}1_N'V_{22}^{-1}V_{21}]V_{11}^{-1}H_{22}, \quad \text{(A57)}
\end{aligned}
$$

[40]By comparing $V(\hat{\gamma})$ for the estimated GLS case with the $V(\hat{\gamma})$ for the true GLS case in (29), it is easy to see that the use of $\hat{V}_{22}^{-1}$ instead of V_{22}^{-1} as weighting matrix increases the asymptotic variance of $\hat{\gamma}_0$ but reduces the asymptotic variance of $\hat{\gamma}_1$.

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

$$H_{22}^{-1} = \beta' V_{22}^{-1}\beta - \beta' V_{22}^{-1} 1_N (1_N' V_{22}^{-1} 1_N)^{-1} 1_N' V_{22}^{-1}\beta. \tag{A58}$$

This completes the proof.

Proof of Proposition 4: (1) $\rho^2 = 1$: We first derive the asymptotic distribution of

$$T\hat{Q} = T(\hat{\mu}_2'\hat{W}\hat{\mu}_2 - \hat{\mu}_2'\hat{W}\hat{X}(\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W}\hat{\mu}_2) \tag{A59}$$

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 $\theta_1 = (\alpha',\ \mathrm{vec}(\beta)')'$ and $\theta_2 = \gamma$. Define

$$g_t(\theta) \equiv \begin{bmatrix} g_{1t}(\theta_1) \\ g_{2t}(\theta) \end{bmatrix} = \begin{bmatrix} l_t \otimes \epsilon_t \\ R_t - X\gamma \end{bmatrix}, \tag{A60}$$

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

$$\bar{g}_T(\theta) = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T g_{1t}(\theta_1) \\ \frac{1}{T}\sum_{t=1}^T g_{2t}(\theta) \end{bmatrix}. \tag{A61}$$

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

$$\hat{\theta}_2 = \hat{\gamma} = (\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W}\hat{\mu}_2 \tag{A62}$$

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

$$B_T\bar{g}_T(\theta) = 0_{p+K+1}, \tag{A63}$$

where

$$B_T = \begin{bmatrix} I_p & 0_{p\times N} \\ 0_{(K+1)\times p} & \hat{X}'\hat{W} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} I_p & 0_{p\times N} \\ 0_{(K+1)\times p} & X'W \end{bmatrix} \equiv B. \tag{A64}$$

Writing

$$l_t \otimes \epsilon_t = \mathrm{vec}(\epsilon_t l_t') = (l_t \otimes I_N)\mathrm{vec}(\epsilon_t), \tag{A65}$$

$$\epsilon_t = R_t - \alpha - \beta f_t = R_t - (l_t' \otimes I_N)\theta_1, \tag{A66}$$

$$\beta\gamma_1 = (\gamma_1' \otimes I_N)\mathrm{vec}(\beta), \tag{A67}$$

we have:

$$
\begin{aligned}
\frac{\partial g_{1t}(\theta_1)}{\partial \theta_1'} &= -l_t l_t' \otimes I_N, & \text{(A68)} \\
\frac{\partial g_{1t}(\theta_1)}{\partial \theta_2'} &= 0_{p\times(K+1)}, & \text{(A69)} \\
\frac{\partial g_{2t}(\theta)}{\partial \theta_1'} &= [0,\ -\gamma_1'] \otimes I_N, & \text{(A70)} \\
\frac{\partial g_{2t}(\theta)}{\partial \theta_2'} &= -X. & \text{(A71)}
\end{aligned}
$$

Let

$$
\begin{aligned}
D_T &= \frac{\partial \bar{g}_T(\theta)}{\partial \theta'} \\
&= \begin{bmatrix} -\left(\frac{1}{T}\sum_{t=1}^T l_t l_t'\right) \otimes I_N & 0_{p\times(K+1)} \\ [0,\ -\gamma_1'] \otimes I_N & -X \end{bmatrix} \\
&\xrightarrow{\text{a.s.}} \begin{bmatrix} -E[l_t l_t'] \otimes I_N & 0_{p\times(K+1)} \\ [0,\ -\gamma_1'] \otimes I_N & -X \end{bmatrix} \equiv D. \qquad \text{(A72)}
\end{aligned}
$$

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

$$
\sqrt{T}\bar{g}_T(\hat{\theta}) \overset{A}{\sim} N(0_{p+N}, [I_{p+N} - D(BD)^{-1}B]S[I_{p+N} - D(BD)^{-1}B]'), \qquad \text{(A73)}
$$

where

$$
S = \sum_{j=-\infty}^{\infty} E[g_t(\theta) g_{t+j}(\theta)']. \qquad \text{(A74)}
$$

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

$$
E[l_t l_t']^{-1} = \begin{bmatrix} 1 + \mu_1' V_{11}^{-1}\mu_1 & -\mu_1' V_{11}^{-1} \\ -V_{11}^{-1}\mu_1 & V_{11}^{-1} \end{bmatrix}. \qquad \text{(A75)}
$$

It follows that

$$
\begin{aligned}
BD &= \begin{bmatrix} -E[l_t l_t'] \otimes I_N & 0_{p\times(K+1)} \\ [0,\ -\gamma_1'] \otimes X'W & -H^{-1} \end{bmatrix}, & \text{(A76)} \\
(BD)^{-1} &= \begin{bmatrix} -E[l_t l_t']^{-1} \otimes I_N & 0_{p\times(K+1)} \\ [-\gamma_1' V_{11}^{-1}\mu_1,\ \gamma_1' V_{11}^{-1}] \otimes A & -H \end{bmatrix}, & \text{(A77)} \\
D(BD)^{-1}B &= \begin{bmatrix} I_p & 0_{p\times N} \\ [-\gamma_1' V_{11}^{-1}\mu_1,\ \gamma_1' V_{11}^{-1}] \otimes (I_N - XA) & -XA \end{bmatrix}, & \text{(A78)} \\
I_N - D(BD)^{-1}B &= \begin{bmatrix} 0_{p\times p} & 0_{p\times N} \\ [\gamma_1' V_{11}^{-1}\mu_1,\ -\gamma_1' V_{11}^{-1}] \otimes (I_N - XA) & I_N - XA \end{bmatrix}. & \text{(A79)}
\end{aligned}
$$

We now provide a simplification of the asymptotic distribution of $\bar{g}_{2T}(\hat{\theta})$. From (A73), we have:

$$\sqrt{T}\bar{g}_{2T}(\hat{\theta}) \stackrel{A}{\sim} N(0_N, V_q), \tag{A80}$$

where

$$V_q = \sum_{j=-\infty}^{\infty} E[q_t(\theta)q_{t+j}(\theta)'], \tag{A81}$$

and

$$\begin{aligned}
q_t(\theta) &= [0_{N\times p},\ I_N](I_N - D(BD)^{-1}B)g_t(\theta) \\
&= -(I_N - XA)\epsilon_t\gamma_1'V_{11}^{-1}(f_t-\mu_1) + (I_N - XA)(R_t - X\gamma) \\
&= (I_N - XA)[R_t - \epsilon_t\gamma_1'V_{11}^{-1}(f_t-\mu_1)] \\
&= (I_N - X(X'WX)^{-1}X'W)[\alpha + \beta f_t + \epsilon_t - \epsilon_t\gamma_1'V_{11}^{-1}(f_t-\mu_1)] \\
&= W^{-\frac{1}{2}}[I_N - W^{\frac{1}{2}}X(X'WX)^{-1}X'W^{\frac{1}{2}}]W^{\frac{1}{2}}\epsilon_t y_t \\
&= W^{-\frac{1}{2}}[I_N - W^{\frac{1}{2}}C(C'WC)^{-1}C'W^{\frac{1}{2}}]W^{\frac{1}{2}}\epsilon_t y_t \\
&= W^{-\frac{1}{2}}PP'W^{\frac{1}{2}}\epsilon_t y_t,
\end{aligned} \tag{A82}$$

where $y_t = 1 - \gamma_1'V_{11}^{-1}(f_t - \mu_1) = 1 - \lambda_1'(f_t - \mu_1)$ follows from (10). The fifth equality holds because $\alpha = \mu_2 - \beta\mu_1 = 1_N\gamma_0 + \beta(\gamma_1 - \mu_1) = X\phi$ when the model is correctly specified and $\beta = X[0_K,\ I_K]$. Therefore, both α and βf_t vanish when premultiplied by $I_N - XA$. With this expression of q_t, we can write V_q as

$$V_q = W^{-\frac{1}{2}}PP'W^{\frac{1}{2}}SW^{\frac{1}{2}}PP'W^{-\frac{1}{2}}, \tag{A83}$$

where S is the asymptotic covariance matrix of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t y_t$. Having derived the asymptotic distribution of $\bar{g}_{2T}(\hat{\theta})$, the asymptotic distribution of $\hat{Q}$ is given by

$$T\hat{Q} = T\bar{g}_{2T}(\hat{\theta})'\hat{W}\bar{g}_{2T}(\theta) \stackrel{A}{\sim} \sum_{j=1}^{N-K-1}\xi_j x_j, \tag{A84}$$

where the x_j's are independent χ^2_1 random variables, and the ξ_j's are the $N-K-1$ nonzero eigenvalues of

$$W^{\frac{1}{2}}V_qW^{\frac{1}{2}} = PP'W^{\frac{1}{2}}SW^{\frac{1}{2}}PP'. \tag{A85}$$

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

$$T(\hat{\rho}^2 - 1) = -\frac{T\hat{Q}}{\hat{Q}_0} \stackrel{A}{\sim} -\sum_{j=1}^{N-K-1}\frac{\xi_j}{Q_0}x_j. \tag{A86}$$

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

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \tag{A87}$$

where

$$n_t = \frac{\partial \rho^2}{\partial \varphi'} r_t(\varphi). \tag{A88}$$

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:

$$\frac{\partial \rho^2}{\partial \mu_1} = 0_K, \tag{A89}$$

$$\frac{\partial \rho^2}{\partial \mu_2} = 2Q_0^{-1}W[(1-\rho^2)e_0 - e]. \tag{A90}$$

Equation (A89) follows because ρ^2 does not depend on μ_1. For (A90), using the first order conditions $1_N'We_0 = 0$ and $X'We = 0_{K+1}$ and letting $Q_0 = e_0'We_0$, we have:

$$\frac{\partial Q_0}{\partial \mu_2} = 2We_0, \qquad \frac{\partial Q}{\partial \mu_2} = 2We. \tag{A91}$$

It follows that

$$\frac{\partial \rho^2}{\partial \mu_2} = -Q_0^{-1}\frac{\partial Q}{\partial \mu_2} + Q_0^{-2}Q\frac{\partial Q_0}{\partial \mu_2} = -2Q_0^{-1}We + 2QQ_0^{-2}We_0 = 2Q_0^{-1}W[(1-\rho^2)e_0 - e]. \tag{A92}$$

The expression for $\partial\rho^2/\partial\mathrm{vec}(V)'$, 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 $\mathrm{vec}(V)$, we obtain:

$$\frac{\partial Q}{\partial \mathrm{vec}(V)'} = 2e'W\frac{\partial(\mu_2 - X\gamma)}{\partial \mathrm{vec}(V)'} = -2e'W\left[(\gamma' \otimes I_N)\frac{\partial x}{\partial \mathrm{vec}(V)'} + X\frac{\partial \gamma}{\partial \mathrm{vec}(V)'}\right]. \tag{A93}$$

Note that the second term vanishes because of the first order condition $X'We = 0_{K+1}$. Using (A7) for the first term and the fact that $\beta'We = 0_K$ gives

$$\frac{\partial Q}{\partial \mathrm{vec}(V)'} = -2e'W\left([\gamma_1'V_{11}^{-1},\ 0_N'] \otimes [-\beta,\ I_N]\right) = -2\left([\gamma_1'V_{11}^{-1},\ 0_N'] \otimes [0_K',\ e'W]\right). \tag{A94}$$

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

$$\frac{\partial \rho^2}{\partial \mathrm{vec}(V)'} = -Q_0^{-1}\frac{\partial Q}{\partial \mathrm{vec}(V)'} = 2Q_0^{-1}\left[\gamma_1'V_{11}^{-1},\ 0_N'\right] \otimes \left[0_K',\ e'W\right]. \tag{A95}$$

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

$$\begin{aligned}
n_t &= \frac{\partial \rho^2}{\partial \varphi'} r_t(\varphi) \\
&= 2Q_0^{-1}[(1-\rho^2)e_0' - e']W(R_t - \mu_2) + 2Q_0^{-1}e'W(R_t-\mu_2)(f_t-\mu_1)'V_{11}^{-1}\gamma_1 \\
&= 2Q_0^{-1}[-u_t y_t + (1-\rho^2)v_t]. \qquad \text{(A96)}
\end{aligned}$$

We now turn to the $\hat{W} = \hat{V}_{22}^{-1}$ case. Differentiating $Q = e'V_{22}^{-1}e$ with respect to $\mathrm{vec}(V)$, we obtain:

$$\begin{aligned}
\frac{\partial Q}{\partial \mathrm{vec}(V)'} &= 2e'V_{22}^{-1}\frac{\partial(\mu_2 - X\gamma)}{\partial \mathrm{vec}(V)'} + (e'\otimes e')\frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'} \\
&= -2\left([\gamma_1'V_{11}^{-1},\ 0_N'] \otimes [0_K',\ e'V_{22}^{-1}]\right) - (e'\otimes e')\left([0_{N\times K},\ V_{22}^{-1}] \otimes [0_{N\times K},\ V_{22}^{-1}]\right) \\
&= -[2\gamma_1'V_{11}^{-1},\ e'V_{22}^{-1}] \otimes [0_K',\ e'V_{22}^{-1}]. \qquad \text{(A97)}
\end{aligned}$$

Similarly, we have:

$$\frac{\partial Q_0}{\partial \mathrm{vec}(V)'} = -[0_K',\ e_0'V_{22}^{-1}] \otimes [0_K',\ e_0'V_{22}^{-1}]. \qquad \text{(A98)}$$

It follows that

$$\begin{aligned}
\frac{\partial \rho^2}{\partial \mathrm{vec}(V)'} &= -Q_0^{-1}\frac{\partial Q}{\partial \mathrm{vec}(V)'} + Q_0^{-2}Q\frac{\partial Q_0}{\partial \mathrm{vec}(V)'} \\
&= Q_0^{-1}\left[2\gamma_1'V_{11}^{-1},\ e'V_{22}^{-1}\right] \otimes \left[0_K',\ e'V_{22}^{-1}\right] \\
&\quad - Q_0^{-1}(1-\rho^2)\left[0_K',\ e_0'V_{22}^{-1}\right] \otimes \left[0_K',\ e_0'V_{22}^{-1}\right]. \qquad \text{(A99)}
\end{aligned}$$

Therefore, we have:

$$\begin{aligned}
n_t &= \frac{\partial \rho^2}{\partial \varphi'} r_t(\varphi) \\
&= 2Q_0^{-1}[(1-\rho^2)e_0' - e']V_{22}^{-1}(R_t-\mu_2) + Q_0^{-1}e'V_{22}^{-1}(R_t-\mu_2)[2\gamma_1'V_{11}^{-1}(f_t-\mu_1) \\
&\quad + e'V_{22}^{-1}(R_t-\mu_2)] - Q_0^{-1}(1-\rho^2)[e_0'V_{22}^{-1}(R_t-\mu_2)]^2 - Q_0^{-1}Q + Q_0^{-1}(1-\rho^2)Q_0 \\
&= Q_0^{-1}[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)]. \qquad \text{(A100)}
\end{aligned}$$

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

$$\begin{aligned}
Q_0 - Q &= \mu_2'WX(X'WX)^{-1}X'W\mu_2 - \mu_2'W1_N(1_N'W1_N)^{-1}1_N'W\mu_2 \\
&= \mu_2'WX(X'WX)^{-1}X'W\mu_2 - \mu_2'WX\begin{bmatrix} (1_N'W1_N)^{-1} & 0_K' \\ 0_K & 0_{K\times K} \end{bmatrix} X'W\mu_2
\end{aligned}$$

$$
\begin{aligned}
&= \gamma'(X'WX)\gamma - \gamma'(X'WX)\begin{bmatrix} (1_N'W1_N)^{-1} & 0_K' \\ 0_K & 0_{K\times K}\end{bmatrix}(X'WX)\gamma \\
&= \gamma'(X'WX)\gamma - \gamma'\begin{bmatrix} 1_N'W1_N & 1_N'W\beta \\ \beta'W1_N & \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta\end{bmatrix}\gamma \\
&= \gamma_1'[\beta'W\beta - \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta]\gamma_1. \qquad \text{(A101)}
\end{aligned}
$$

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 (A101) also holds for its sample counterpart, so we can write $\hat{\rho}^2$ as

$$
\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0} = \frac{\hat{Q}_0 - \hat{Q}}{\hat{Q}_0} = \frac{\hat{\gamma}_1'[\hat{\beta}'\hat{W}\hat{\beta} - \hat{\beta}'\hat{W}1_N(1_N'\hat{W}1_N)^{-1}1_N'\hat{W}\hat{\beta}]\hat{\gamma}_1}{\hat{Q}_0}. \qquad \text{(A102)}
$$

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

$$
\sqrt{T}\hat{\gamma}_1 \overset{A}{\sim} N(0_K, V(\hat{\gamma}_1)), \qquad \text{(A103)}
$$

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

$$
\hat{\beta}'\hat{W}\hat{\beta} - \hat{\beta}'\hat{W}1_N(1_N'\hat{W}1_N)^{-1}1_N'\hat{W}\hat{\beta} \xrightarrow{\text{a.s.}} \beta'W\beta - \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta, \qquad \text{(A104)}
$$

it follows that

$$
T\hat{\rho}^2 \overset{A}{\sim} \sum_{j=1}^{K} \frac{\xi_j}{Q_0} x_j, \qquad \text{(A105)}
$$

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

$$
[\beta'W\beta - \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta]V(\hat{\gamma}_1). \qquad \text{(A106)}
$$

This completes the proof.

Proof of Lemma 3: 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

$$
\begin{aligned}
Q_B - Q_A &= \mu_2'WC_A(C_A'WC_A)^{-1}C_A'W\mu_2 - \mu_2'WC_B(C_B'WC_B)^{-1}C_B'W\mu_2 \\
&= \mu_2'WC_A(C_A'WC_A)^{-1}C_A'W\mu_2 - \mu_2'WC_A\begin{bmatrix} (C_{Aa}'WC_{Aa})^{-1} & 0_{(K_1+1)\times K_2} \\ 0_{K_2\times(K_1+1)} & 0_{K_2\times K_2}\end{bmatrix}C_A'W\mu_2
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda_A'(C_A'WC_A)\lambda_A - \lambda_A'(C_A'WC_A)\left[\begin{array}{cc} (C_{Aa}'WC_{Aa})^{-1} & 0_{(K_1+1)\times K_2} \\ 0_{K_2\times(K_1+1)} & 0_{K_2\times K_2} \end{array}\right](C_A'WC_A)\lambda_A \\
&= \lambda_{A,2}'[C_{Ab}'WC_{Ab} - C_{Ab}'WC_{Aa}(C_{Aa}'WC_{Aa})^{-1}(C_{Aa}'WC_{Ab})]\lambda_{A,2} \\
&= \lambda_{A,2}'\tilde{H}_{A,22}^{-1}\lambda_{A,2}, \qquad \text{(A107)}
\end{aligned}
$$

where $\tilde{H}_{A,22}$ is the lower right $K_2 \times K_2$ submatrix of $\tilde{H}_A$. 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 $\lambda_{A,2} = 0_{K_2}$. This completes the proof.

Proof that $y_A = y_B$ implies $e_A = e_B$: Using the first order conditions for model A, we obtain:

$$
0 = 1_N'We_A = 1_N'W\mu_2 - 1_N'W1_N\lambda_{A,0} - 1_N'W\text{Cov}[R, f_1']\lambda_{A,1} - 1_N'W\text{Cov}[R, f_2']\lambda_{A,2}. \qquad \text{(A108)}
$$

This implies that the (pseudo) zero-beta rate of model A is

$$
\lambda_{A,0} = (1_N'W1_N)^{-1}1_N'W(\mu_2 - \text{Cov}[R, f_1']\lambda_{A,1} - \text{Cov}[R, f_2']\lambda_{A,2}) = (1_N'W1_N)^{-1}1_N'WE[Ry_A], \qquad \text{(A109)}
$$

and the pricing errors of model A can be written as

$$
e_A = [I_N - 1_N(1_N'W1_N)^{-1}1_N'W]E[Ry_A]. \qquad \text{(A110)}
$$

Similarly, the pricing errors of model B can be written as

$$
e_B = [I_N - 1_N(1_N'W1_N)^{-1}1_N'W]E[Ry_B]. \qquad \text{(A111)}
$$

Therefore, when $y_A = y_B$, we have $e_A = e_B$. This completes the proof.

Proof of Lemma 4: 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 $\lambda_{A,1} = \lambda_{B,1}$. For model A, premultiplying both sides of (51) by $C_A'WC_A$, we obtain:

$$
\left[\begin{array}{cc} C_{Aa}'WC_{Aa} & C_{Aa}'WC_{Ab} \\ C_{Ab}'WC_{Aa} & C_{Ab}'WC_{Ab} \end{array}\right]\left[\begin{array}{c} \lambda_{A,0} \\ \lambda_{A,1} \\ \lambda_{A,2} \end{array}\right] = \left[\begin{array}{c} C_{Aa}'W\mu_2 \\ C_{Ab}'W\mu_2 \end{array}\right]. \qquad \text{(A112)}
$$

When $\lambda_{A,2} = 0_{K_2}$, the first block of this equation gives us

$$
\left[\begin{array}{c} \lambda_{A,0} \\ \lambda_{A,1} \end{array}\right] = (C_{Aa}'WC_{Aa})^{-1}C_{Aa}'W\mu_2. \qquad \text{(A113)}
$$

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

$$\begin{bmatrix} \lambda_{B,0} \\ \lambda_{B,1} \end{bmatrix} = (C_{Ba}'WC_{Ba})^{-1}C_{Ba}'W\mu_2, \tag{A114}$$

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.

Proof of Propositions 5 and 6: Since Proposition 5 is a special case of Proposition 6 when $K_3 = 0$, we only provide the proof of Proposition 6 here. We first derive a simplified expression for $Q_B - Q_A$. The aggregate pricing errors of model A are given by

$$Q_A = e_A'We_A = \mu_2'W\mu_2 - \mu_2'WC_A(C_A'WC_A)^{-1}C_A'W\mu_2. \tag{A115}$$

We now introduce a model M that uses only f_1 as factors. The aggregate pricing errors of model M are given by

$$Q_M = e_M'We_M = \mu_2'W\mu_2 - \mu_2'WC_M(C_M'WC_M)^{-1}C_M'W\mu_2, \tag{A116}$$

where $C_M = [1_N,\ \mathrm{Cov}[R, f_1']]$. Using the fact that the $C_{Aa} = C_{Ba} = C_M$ and (A107), we can write the difference between Q_M and Q_A as

$$Q_M - Q_A = \lambda_{A,2}'\tilde{H}_{A,22}^{-1}\lambda_{A,2}. \tag{A117}$$

Similarly, we have:

$$Q_M - Q_B = \lambda_{B,3}'\tilde{H}_{B,33}^{-1}\lambda_{B,3}. \tag{A118}$$

Subtracting (A118) from (A117), we obtain:

$$Q_B - Q_A = \lambda_{A,2}'\tilde{H}_{A,22}^{-1}\lambda_{A,2} - \lambda_{B,3}'\tilde{H}_{B,33}^{-1}\lambda_{B,3} = \psi' \begin{bmatrix} \tilde{H}_{A,22}^{-1} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}_{B,33}^{-1} \end{bmatrix} \psi, \tag{A119}$$

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 $\sqrt{T}V(\hat{\psi})^{-\frac{1}{2}}\hat{\psi} \overset{A}{\sim} N(0_{K_2+K_3}, I_{K_2+K_3})$. It follows that

$$T(\hat{Q}_B - \hat{Q}_A) \overset{A}{\sim} \sum_{j=1}^{K_2+K_3} \xi_j x_j, \tag{A120}$$

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

$$\begin{bmatrix} \tilde{H}_{A,22}^{-1} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}_{B,33}^{-1} \end{bmatrix} V(\hat{\psi}). \tag{A121}$$

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

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{K_2+K_3} \frac{\xi_j}{Q_0} x_j. \tag{A122}$$

This completes the proof.

Proof of Propositions 7 and 8: In the proof of Proposition 4, we show that when the model is correctly specified,

$$\sqrt{T}\hat{e}_A \overset{A}{\sim} N(0_N, V_{q_A}), \tag{A123}$$

where

$$V_{q_A} = \sum_{j=-\infty}^{\infty} E[q_{At} q'_{A,t+j}], \tag{A124}$$

with

$$q_{At} = W^{-\frac{1}{2}} P_A P'_A W^{\frac{1}{2}} \epsilon_{At} y_{At} = W^{-\frac{1}{2}} P_A P'_A W^{\frac{1}{2}} g_{At}. \tag{A125}$$

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

$$\sqrt{T} \begin{bmatrix} \hat{e}_A \\ \hat{e}_B \end{bmatrix} \overset{A}{\sim} N(0_{2N}, V_q), \tag{A126}$$

where

$$V_q = \sum_{j=-\infty}^{\infty} E[q_t q'_{t+j}], \tag{A127}$$

and

$$q_t = \begin{bmatrix} q_{At} \\ q_{Bt} \end{bmatrix} = \begin{bmatrix} W^{-\frac{1}{2}} P_A P'_A W^{\frac{1}{2}} g_{At} \\ W^{-\frac{1}{2}} P_B P'_B W^{\frac{1}{2}} g_{Bt} \end{bmatrix}. \tag{A128}$$

We can simplify V_q as

$$V_q = \begin{bmatrix} W^{-\frac{1}{2}} P_A P'_A W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A P'_A W^{-\frac{1}{2}} & W^{-\frac{1}{2}} P_A P'_A W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B P'_B W^{-\frac{1}{2}} \\ W^{-\frac{1}{2}} P_B P'_B W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A P'_A W^{-\frac{1}{2}} & W^{-\frac{1}{2}} P_B P'_B W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B P'_B W^{-\frac{1}{2}} \end{bmatrix}. \tag{A129}$$

It follows that

$$z = \sqrt{T} \begin{bmatrix} \hat{P}'_A \hat{W}^{\frac{1}{2}} \hat{e}_A \\ \hat{P}'_B \hat{W}^{\frac{1}{2}} \hat{e}_B \end{bmatrix} \overset{A}{\sim} N(0_{n_A+n_B}, V_z), \tag{A130}$$

where

$$V_z = \begin{bmatrix} P'_A W^{\frac{1}{2}} S_{AA} W^{\frac{1}{2}} P_A & P'_A W^{\frac{1}{2}} S_{AB} W^{\frac{1}{2}} P_B \\ P'_B W^{\frac{1}{2}} S_{BA} W^{\frac{1}{2}} P_A & P'_B W^{\frac{1}{2}} S_{BB} W^{\frac{1}{2}} P_B \end{bmatrix}. \tag{A131}$$

Then, we have:

$$z'\hat{V}_z^{-1}z \overset{A}{\sim} \chi^2_{n_A+n_B}. \tag{A132}$$

This completes the proof of Proposition 7.

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

$$\begin{aligned} T\hat{Q}_A &= T\hat{e}_A'\hat{W}^{\frac{1}{2}}[\hat{P}_A\hat{P}_A' + \hat{W}^{\frac{1}{2}}\hat{C}_A(\hat{C}_A'\hat{W}\hat{C}_A)^{-1}\hat{C}_A'\hat{W}^{\frac{1}{2}}]\hat{W}^{\frac{1}{2}}\hat{e}_A \\ &= T\hat{e}_A'\hat{W}^{\frac{1}{2}}\hat{P}_A\hat{P}_A'\hat{W}^{\frac{1}{2}}\hat{e}_A \\ &= z_A'z_A, \end{aligned} \tag{A133}$$

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

$$V_z^{\frac{1}{2}}\begin{bmatrix} -I_{n_A} & 0_{n_A\times n_B} \\ 0_{n_B\times n_A} & I_{n_B}\end{bmatrix}V_z^{\frac{1}{2}}, \tag{A134}$$

where $\Xi = \mathrm{Diag}(\xi_1, \cdots, \xi_{n_A+n_B})$ is a diagonal matrix of the eigenvalues of (A134) or, equivalently, of the eigenvalues of (70). 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:

$$T(\hat{Q}_B - \hat{Q}_A) = z'\begin{bmatrix} -I_{n_A} & 0_{n_A\times n_B} \\ 0_{n_B\times n_A} & I_{n_B}\end{bmatrix}z = z'V_z^{\frac{1}{2}}Q\Xi Q'V_z^{\frac{1}{2}}z = \sum_{j=1}^{n_A+n_B}\xi_j x_j, \tag{A135}$$

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

$$T(\hat{\rho}_A^2 - \hat{\rho}_B^2) \overset{A}{\sim} \sum_{j=1}^{n_A+n_B}\frac{\xi_j}{Q_0}x_j. \tag{A136}$$

This completes the proof of Proposition 8.

Proof of Proposition 9: We start from the known weighting matrix case. Using the results of Proposition 4, we obtain the following expressions for models A and B:

$$\begin{aligned} n_{At}(\varphi) &= \left[\frac{\partial\rho_A^2}{\partial\varphi}\right]' r_t(\varphi) = 2Q_0^{-1}[-u_{At}y_{At} + (1-\rho_A^2)v_t], &\text{(A137)} \\ n_{Bt}(\varphi) &= \left[\frac{\partial\rho_B^2}{\partial\varphi}\right]' r_t(\varphi) = 2Q_0^{-1}[-u_{Bt}y_{Bt} + (1-\rho_B^2)v_t]. &\text{(A138)} \end{aligned}$$

Now, using the delta method and equations (A1)–(A4), the asymptotic distribution of $\hat{\rho}_A^2 - \hat{\rho}_B^2$ when both models are misspecified is given by

$$\sqrt{T}(\hat{\rho}_A^2 - \hat{\rho}_B^2 - (\rho_A^2 - \rho_B^2)) \stackrel{A}{\sim} N\left(0, \left[\frac{\partial(\rho_A^2 - \rho_B^2)}{\partial\varphi}\right]' S_0 \left[\frac{\partial(\rho_A^2 - \rho_B^2)}{\partial\varphi}\right]\right). \tag{A139}$$

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

$$\sum_{j=-\infty}^{\infty} E[d_t(\varphi)d_{t+j}(\varphi)], \tag{A140}$$

where

$$d_t(\varphi) = \left(\frac{\partial\rho_A^2}{\partial\varphi} - \frac{\partial\rho_B^2}{\partial\varphi}\right)' r_t(\varphi) = n_{At}(\varphi) - n_{Bt}(\varphi). \tag{A141}$$

Under $H_0 : \rho_A^2 = \rho_B^2$, we have:

$$d_t(\varphi) = 2Q_0^{-1}(u_{Bt}y_{Bt} - u_{At}y_{At}). \tag{A142}$$

Using the same type of proof for the GLS case with $\hat{W} = \hat{V}_{22}^{-1}$, we obtain:

$$d_t(\varphi) = Q_0^{-1}(u_{At}^2 - 2u_{At}y_{At} - u_{Bt}^2 + 2u_{Bt}y_{Bt}). \tag{A143}$$

This completes the proof.

References

Ahn, Seung C., and Christopher Gadarowski. 2003. "Two-Pass Cross-Sectional Regression of Factor Pricing Models: Minimum Distance Approach." Working Paper, Arizona State University.

Bentler, Peter M., and Maia Berkane. 1986. "Greatest Lower Bound to the Elliptical Theory Kurtosis Parameter." *Biometrika* 73: 240–241.

Black, Fischer, Michael C. Jensen, and Myron Scholes. 1972. "The Capital Asset Pricing Model: Some Empirical Findings." In Jensen, M.C. (Ed.), Studies in the Theory of Capital Markets. Praeger, New York.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets.* Princeton University Press, Princeton.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross. 1986. "Economic Forces and the Stock Market." *Journal of Business* 59: 383–404.

Chen, Robert, and Raymond Kan. 2003. "Finite Sample Analysis of Two-Pass Cross-Sectional Regressions." Working Paper, University of Toronto.

Cochrane, John H. 1996. "A Cross-Sectional Test of an Investment-Based Asset Pricing Model." *Journal of Political Economy* 104: 572-621.

Cochrane, John H. 2005. *Asset Pricing.* Princeton University Press, Princeton.

Fama, Eugene F., and James D. MacBeth. 1973. "Risk, Return and Equilibrium: Empirical Tests." *Journal of Political Economy* 71: 607-636.

Fama, Eugene. F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47: 427–466.

Fama, Eugene. F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33: 3–56.

Ferson, Wayne E., and Rudi W. Schadt. 1996. "Measuring Fund Strategy and Performance in Changing Economic Conditions." *Journal of Finance* 51: 425–461.

Golden, Richard M. 2003. "Discrepancy Risk Model Selection Test Theory for Comparing Possibly Misspecified or Nonnested Models." *Psychometrika* 68: 229–249.

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*, forthcoming.

Hall, Alastair R., and Atsushi Inoue. 2003. "The Large Sample Behaviour of the Generalized Method of Moments Estimator in Misspecified Models." *Journal of Econometrics* 114: 361–394.

Hansen, Lars P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50: 1029–1054.

Hansen, Lars P., and Ravi Jagannathan. 1997. "Assessing Specification Errors in Stochastic Discount Factor Models." *Journal of Finance* 52: 557–590.

Hou, Kewei, and Robert Kimmel. 2006. "On the Estimation of Risk Premia in Linear Factor Models." Working Paper, Ohio State University.

Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara. 1998. "Relationship Between Labor-Income Risk and Average Return: Empirical Evidence From the Japanese Stock Market." *Journal of Business* 71: 319–348.

Jagannathan, Ravi, Georgios Skoulakis, and Zhenyu Wang. 2008. "The Analysis of the Cross Section of Security Returns." *Handbook of Financial Econometrics*, forthcoming.

Jagannathan, Ravi, and Zhenyu Wang. 1996. "The Conditional CAPM and the Cross-section of Expected Returns." *Journal of Finance* 51: 3–53.

Jagannathan, Ravi, and Zhenyu Wang. 1998. "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression." *Journal of Finance* 53: 1285–1309.

Kan, Raymond, and Cesare Robotti. 2008. "Model Comparison Using the Hansen-Jagannathan Distance." *Review of Financial Studies*, forthcoming.

Kan, Raymond, and Cesare Robotti. 2009. "A Note on the Estimation of Asset Pricing Models Using Simple Regression Betas." Working Paper, University of Toronto.

Kan, Raymond, and Chu Zhang. 1999. "Two-Pass Tests of Asset Pricing Models with Useless Factors." *Journal of Finance* 54: 203–235.

Kandel, Shmuel, and Robert F. Stambaugh. 1995. "Portfolio Inefficiency and the Cross-Section of Expected Returns." *Journal of Finance* 50: 157–184.

Lettau, Martin, and Sydney C. Ludvigson. 2001. "Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying." *Journal of Political Economy* 109: 1238–1287.

Lewellen, Jonathan W., Stefan Nagel, and Jay Shanken. 2009. "A Skeptical Appraisal of Asset-Pricing Tests." *Journal of Financial Economics*, forthcoming.

Li, Haitao, Yuewu Xu, and Xiaoyan Zhang. 2009. "Evaluating Asset Pricing Models Using the Second Hansen-Jagannathan Distance." *Journal of Financial Economics*, forthcoming.

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.

Newey, Whitney K. 1984. "A Method of Moments Interpretation of Sequential Estimators." *Economics Letters* 14: 201–206.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55: 703–708.

Newey, Whitney K., and Kenneth D. West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies* 61: 631–653.

Pagan, Adrian. 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors." *International Economic Review* 25: 221–247.

Petkova, Ralitsa. 2006. "Do the Fama-French Factors Proxy for Innovations in Predictive Variables?" *Journal of Finance* 61: 581–612.

Rivers, Douglas, and Quang H. Vuong. 2002. "Model Selection Tests for Nonlinear Dynamic Models." *Econometrics Journal* 5: 1–39.

Roll, Richard, and Stephen A. Ross. 1994. "On the Cross-sectional Relation between Expected Returns and Betas." *Journal of Finance* 49: 101–121.

Shanken, Jay. 1985. "Multivariate Tests of the Zero-Beta CAPM." *Journal of Financial Economics* 14: 327–348.

Shanken, Jay. 1990. "Intertemporal Asset Pricing: An Empirical Investigation." *Journal of Econometrics* 45: 99–120.

Shanken, Jay. 1992. "On the Estimation of Beta-Pricing Models." *Review of Financial Studies* 5: 1–33.

Shanken, Jay, and Guofu Zhou. 2007. "Estimating and Testing Beta Pricing Models: Alternative Methods and their Performance in Simulations." *Journal of Financial Economics* 84: 40–86.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57: 307–333.

White, Halbert L. 1994. *Estimation, Inference and Specification Analysis.* Cambridge, New York.

TABLE 1

SAMPLE CROSS-SECTIONAL R^2S AND SPECIFICATION TESTS OF THE MODELS

PANEL A: OLS

	CAPM	CCAPM	FF3	C-LAB	C-CCAPM	ICAPM
$\hat{\rho}^2$	0.213	0.036	0.769	0.691	0.526	0.793
$p(\rho^2 = 1)$	0.000	0.000	0.000	0.099	0.207	0.207
$\text{se}(\hat{\rho}^2)$	0.236	0.118	0.099	0.156	0.447	0.115
$p(\rho^2 = 0)$	0.099	0.545	0.004	0.007	0.367	0.004
$\hat{Q}_c$	0.101	0.128	0.075	0.040	0.022	0.030
$p_1(Q_c = 0)$	0.000	0.000	0.001	0.207	0.911	0.427
$p_2(Q_c = 0)$	0.000	0.000	0.001	0.251	0.927	0.475
No. of par.	2	2	4	4	4	6

PANEL B: GLS

	CAPM	CCAPM	FF3	C-LAB	C-CCAPM	ICAPM
$\hat{\rho}^2$	0.127	0.045	0.336	0.158	0.388	0.389
$p(\rho^2 = 1)$	0.000	0.000	0.001	0.000	0.418	0.223
$\text{se}(\hat{\rho}^2)$	0.085	0.076	0.114	0.106	0.229	0.189
$p(\rho^2 = 0)$	0.004	0.228	0.000	0.244	0.004	0.030
$\hat{Q}_c$	0.098	0.123	0.077	0.093	0.040	0.042
$p_1(Q_c = 0)$	0.000	0.000	0.000	0.000	0.340	0.105
$p_2(Q_c = 0)$	0.000	0.000	0.001	0.000	0.397	0.136
No. of par.	2	2	4	4	4	6

Note.–The table presents the sample cross-sectional R^2 ($\hat{\rho}^2$) and the generalized CSRT ($\hat{Q}_c$) of six asset pricing models. The models include the unconditional CAPM (CAPM), the consumption CAPM (CCAPM), the Fama and French (1993) three-factor model (FF3), the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the conditional CCAPM (C-CCAPM), and the intertemporal CAPM (ICAPM) of Petkova (2006). The models are estimated using monthly returns on the 25 Fama-French size and book-to-market ranked portfolios. Most of the data are from May 1953 to December 2006 (644 observations), but the data for the CCAPM and C-CCAPM start in February 1959 (575 observations). $p(\rho^2 = 1)$ is the p-value for the test of $H_0 : \rho^2 = 1$. $\text{se}(\hat{\rho}^2)$ is the standard error of $\hat{\rho}^2$ under the assumption that $0 < \rho^2 < 1$. $p(\rho^2 = 0)$ is the p-value for the test of $H_0 : \rho^2 = 0$. $p_1(Q_c = 0)$ is the p-value for the asymptotic test of $H_0 : Q_c = 0$. $p_2(Q_c = 0)$ is the p-value for the approximate F-test of $H_0 : Q_c = 0$. No. of par. is the number of parameters in the model.

TABLE 2

Estimates and t-ratios of Zero-Beta Rate and Risk Premia under Correctly Specified and Misspecified Models

Panel A: OLS

	CAPM		CCAPM	
	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_0$	$\hat{\gamma}_{cg}$
Estimate	1.90	−0.66	1.00	0.15
t-ratio$_{fm}$	5.53	−1.72	4.85	0.70
t-ratio$_s$	5.47	−1.70	4.75	0.69
t-ratio$_{jw}$	5.28	−1.67	4.80	0.70
t-ratio$_{pm}$	4.97	−1.60	3.86	0.58

	FF3				C-LAB			
	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{smb}$	$\hat{\gamma}_{hml}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{lab}$	$\hat{\gamma}_{prem}$
Estimate	2.01	−0.99	0.15	0.43	1.95	−1.08	0.11	0.54
t-ratio$_{fm}$	6.94	−2.96	1.22	3.86	5.79	−2.96	0.94	4.31
t-ratio$_s$	6.69	−2.87	1.22	3.86	3.44	−1.89	0.56	2.58
t-ratio$_{jw}$	6.72	−2.85	1.22	3.87	3.46	−1.94	0.58	2.78
t-ratio$_{pm}$	5.91	−2.58	1.23	3.86	3.48	−1.97	0.52	3.21

	C-CCAPM				ICAPM					
	$\hat{\gamma}_0$	$\hat{\gamma}_{dy}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cg\cdot dy}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$
Estimate	1.33	−1.61	0.50	0.01	1.21	−0.16	0.26	−0.11	−0.01	−0.48
t-ratio$_{fm}$	6.82	−5.30	2.94	2.47	3.91	−0.47	3.97	−2.37	−0.54	−3.77
t-ratio$_s$	3.47	−2.71	1.51	1.27	2.44	−0.32	2.50	−1.50	−0.35	−2.38
t-ratio$_{jw}$	3.56	−2.82	1.47	1.26	2.06	−0.28	2.55	−1.29	−0.26	−2.19
t-ratio$_{pm}$	3.13	−1.07	0.53	0.65	1.73	−0.25	1.81	−1.18	−0.20	−1.96

TABLE 2 (Continued)

ESTIMATES AND t-RATIOS OF ZERO-BETA RATE AND RISK PREMIA UNDER CORRECTLY SPECIFIED AND MISSPECIFIED MODELS

PANEL B: GLS

	CAPM		CCAPM	
	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_0$	$\hat{\gamma}_{cg}$
Estimate	1.90	−0.85	1.27	0.21
t-ratio$_{fm}$	8.98	−3.17	8.93	1.92
t-ratio$_s$	8.80	−3.14	8.56	1.85
t-ratio$_{jw}$	8.63	−3.10	8.49	1.85
t-ratio$_{pm}$	7.51	−2.84	7.61	1.19

	FF3				C-LAB			
	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{smb}$	$\hat{\gamma}_{hml}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{lab}$	$\hat{\gamma}_{prem}$
Estimate	1.88	−0.84	0.20	0.40	1.82	−0.78	−0.07	0.06
t-ratio$_{fm}$	7.40	−2.77	1.68	3.61	8.41	−2.84	−0.93	0.83
t-ratio$_s$	7.17	−2.71	1.68	3.61	8.00	−2.75	−0.89	0.80
t-ratio$_{jw}$	7.22	−2.70	1.69	3.62	7.64	−2.67	−0.89	0.76
t-ratio$_{pm}$	6.11	−2.37	1.69	3.61	6.57	−2.32	−0.44	0.40

	C-CCAPM				ICAPM					
	$\hat{\gamma}_0$	$\hat{\gamma}_{dy}$	$\hat{\gamma}_{cg}$	$\hat{\gamma}_{cg\cdot dy}$	$\hat{\gamma}_0$	$\hat{\gamma}_{vw}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$
Estimate	1.18	−0.97	0.50	0.02	1.61	−0.52	0.21	−0.08	0.01	−0.22
t-ratio$_{fm}$	8.18	−5.02	4.03	3.98	5.94	−1.67	3.99	−2.10	0.53	−2.06
t-ratio$_s$	5.29	−3.30	2.66	2.62	4.33	−1.31	2.94	−1.55	0.39	−1.52
t-ratio$_{jw}$	4.89	−3.85	2.61	2.52	3.90	−1.23	2.97	−1.47	0.34	−1.46
t-ratio$_{pm}$	4.69	−2.64	2.08	2.06	3.08	−1.01	2.01	−1.09	0.25	−1.10

Note.–The table presents the estimation results of six beta pricing models. The models include the unconditional CAPM (CAPM), the consumption CAPM (CCAPM), the Fama and French (1993) three-factor model (FF3), the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the conditional CCAPM (C-CCAPM), and the intertemporal CAPM (ICAPM) of Petkova (2006). The models are estimated using monthly returns on the 25 Fama-French size and book-to-market ranked portfolios. Most of the data are from May 1953 to December 2006 (644 observations), but the data for the CCAPM and C-CCAPM start in February 1959 (575 observations). We report parameter estimates $\hat{\gamma}$ (multiplied by 100), the Fama and MacBeth (1973) t-ratios 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$_{pm}$).

TABLE 3

Estimates and t-ratios of Zero-Beta Rate and Prices of Covariance Risk under Correctly Specified and Misspecified Models

Panel A: OLS

	CAPM		CCAPM	
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_0$	$\hat{\lambda}_{cg}$
Estimate	1.90	−3.66	1.00	28.96
t-ratio$_{fm}$	5.53	−1.72	4.85	0.70
t-ratio$_{cs}$	5.28	−1.69	4.80	0.70
t-ratio$_{pm}$	4.97	−1.61	3.86	0.58

	FF3				C-LAB			
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{lab}$	$\hat{\lambda}_{prem}$
Estimate	2.01	−5.43	4.66	3.96	1.95	−7.70	31.48	318.91
t-ratio$_{fm}$	6.94	−2.49	3.09	2.12	5.79	−3.83	0.44	4.50
t-ratio$_{cs}$	6.72	−2.39	2.95	2.02	3.46	−2.41	0.26	2.84
t-ratio$_{pm}$	5.91	−2.13	2.90	1.92	3.48	−2.46	0.23	3.29

	C-CCAPM				ICAPM					
	$\hat{\lambda}_0$	$\hat{\lambda}_{dy}$	$\hat{\lambda}_{cg}$	$\hat{\lambda}_{cg\cdot dy}$	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{term}$	$\hat{\lambda}_{def}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rf}$
Estimate	1.33	−156.29	−58.35	4471.67	1.21	−9.87	286.90	−278.89	−275.39	−91.95
t-ratio$_{fm}$	6.82	−5.38	−0.71	1.65	3.91	−1.13	2.60	−3.05	−0.90	−2.22
t-ratio$_{cs}$	3.56	−2.74	−0.33	0.77	2.06	−0.54	1.80	−1.56	−0.42	−1.22
t-ratio$_{pm}$	3.13	−1.35	−0.11	0.35	1.73	−0.38	1.15	−1.41	−0.28	−1.17

TABLE 3 (Continued)
Estimates and t-ratios of Zero-Beta Rate and Prices of Covariance Risk under Correctly Specified and Misspecified Models

Panel B: GLS

	CAPM		CCAPM	
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_0$	$\hat{\lambda}_{cg}$
Estimate	1.90	−4.73	1.27	40.53
t-ratio$_{fm}$	8.98	−3.17	8.93	1.92
t-ratio$_{cs}$	8.63	−3.15	8.49	1.83
t-ratio$_{pm}$	7.51	−2.81	7.61	1.18

	FF3				C-LAB			
	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{lab}$	$\hat{\lambda}_{prem}$
Estimate	1.88	−4.66	4.95	4.04	1.82	−4.98	−55.12	43.60
t-ratio$_{fm}$	7.40	−2.36	3.41	2.24	8.41	−3.30	−1.12	1.04
t-ratio$_{cs}$	7.22	−2.28	3.26	2.20	7.64	−3.19	−1.07	0.96
t-ratio$_{pm}$	6.11	−1.97	3.21	2.08	6.57	−2.77	−0.54	0.52

	C-CCAPM				ICAPM					
	$\hat{\lambda}_0$	$\hat{\lambda}_{dy}$	$\hat{\lambda}_{cg}$	$\hat{\lambda}_{cg\cdot dy}$	$\hat{\lambda}_0$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{term}$	$\hat{\lambda}_{def}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rf}$
Estimate	1.18	−96.74	−49.83	4369.36	1.61	−4.58	345.64	−155.28	−9.08	4.51
t-ratio$_{fm}$	8.18	−5.24	−0.72	2.03	5.94	−0.59	3.81	−2.16	−0.03	0.14
t-ratio$_{cs}$	4.89	−3.76	−0.49	1.34	3.90	−0.38	2.83	−1.45	−0.02	0.09
t-ratio$_{pm}$	4.69	−2.72	−0.36	1.02	3.08	−0.27	2.03	−1.10	−0.01	0.07

Note.–The table presents the estimation results of six beta pricing models. The models include the unconditional CAPM (CAPM), the consumption CAPM (CCAPM), the Fama and French (1993) three-factor model (FF3), the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the conditional CCAPM (C-CCAPM), and the intertemporal CAPM (ICAPM) of Petkova (2006). The models are estimated using monthly returns on the 25 Fama-French size and book-to-market ranked portfolios. Most of the data are from May 1953 to December 2006 (644 observations), but the data for the CCAPM and C-CCAPM start in February 1959 (575 observations). We report parameter estimates $\hat{\lambda}$ (with $\hat{\lambda}_0$ multiplied by 100), the Fama and MacBeth (1973) t-ratios under correctly specified models (t-ratio$_{fm}$), the t-ratios under correctly specified models that account for the EIV problem (t-ratio$_{cs}$), and model misspecification-robust t-ratios (t-ratio$_{pm}$).

TABLE 4
Tests of Equality of Cross-Sectional R^2s

Panel A: OLS

	CCAPM	FF3	C-LAB	C-CCAPM	ICAPM
CAPM	0.135	−0.555	−0.478	−0.355	−0.580
	(0.686)	(**0.000**)	(**0.020**)	(0.457)	(0.064)
CCAPM		−0.747	−0.585	−0.490	−0.803
		(**0.029**)	(0.129)	(0.321)	(**0.023**)
FF3			0.078	0.256	−0.024
			(0.608)	(0.558)	(0.792)
C-LAB				0.095	−0.102
				(0.849)	(0.543)
C-CCAPM					−0.313
					(0.470)

Panel B: GLS

	CCAPM	FF3	C-LAB	C-CCAPM	ICAPM
CAPM	0.067	−0.209	−0.031	−0.275	−0.261
	(0.588)	(**0.001**)	(0.735)	(0.256)	(0.256)
CCAPM		−0.283	−0.092	−0.342	−0.296
		(0.058)	(0.502)	(**0.025**)	(0.151)
FF3			0.178	−0.059	−0.053
			(0.163)	(0.802)	(0.778)
C-LAB				−0.250	−0.230
				(0.318)	(0.268)
C-CCAPM					0.047
					(0.871)

Note.–The table presents pairwise tests of equality of the OLS and GLS cross-sectional R^2s of six beta pricing models. The models include the unconditional CAPM (CAPM), the consumption CAPM (CCAPM), the Fama and French (1993) three-factor model (FF3), the conditional CAPM (C-LAB) of Jagannathan and Wang (1996), the conditional CCAPM (C-CCAPM), and the intertemporal CAPM (ICAPM) of Petkova (2006). The models are estimated using monthly returns on the 25 Fama-French size and book-to-market ranked portfolios. Most of the data are from May 1953 to December 2006 (644 observations), but the data for the CCAPM and C-CCAPM start in February 1959 (575 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 parentheses) 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.