Evaluation of Asset Pricing Models Using Two-Pass Cross-Sectional Regressions

Raymond Kan[1] and Cesare Robotti[2]

[1] Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6 kan@chass.utoronto.ca
[2] Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St. N.E., Atlanta, Georgia, 30309, USA cesare.robotti@atl.frb.org

Abstract: This chapter provides a review of the two-pass cross-sectional regression methodology, which over the years has become the most popular approach for estimating and testing linear asset pricing models. We focus on some of the recent developments of this methodology and highlight the importance of accounting for model misspecification in estimating risk premia and in comparing the performance of competing asset pricing models.

1 Introduction

Since Black, Jensen, and Scholes [1] and Fama and MacBeth [2], the two-pass cross-sectional regression (CSR) methodology has become the most popular approach for estimating and testing linear asset 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 from ordinary least squares (OLS) time series regressions of returns on some common factors. In the second pass, the returns on the test assets are regressed on the betas estimated from the first pass. The intercept and the slope coefficients from the second-pass CSR are then used as estimates of the zero-beta rate and factor risk premia. In addition, the R^2 from the second-pass CSR is a popular measure of goodness-of-fit and is often used to compare the performance of competing asset pricing models.

Although the two-pass CSR approach is easy to implement, conducting robust statistical inference under this method is not trivial. In this article, we survey the existing asymptotic techniques and provide some new results. While we are not the first to review the CSR methodology (see Shanken [3] and Jagannathan, Skoulakis, and Wang [4]), our summary of this approach is more current and emphasizes the role played by model misspecification in estimating risk premia and in comparing the performance of competing asset pricing models.

The remainder of the article is organized as follows. Sect. 2 presents the notation and introduces the two-pass CSR methodology. Sect. 3 discusses statistical inference under correctly specified models. Sect. 4 shows how to conduct statistical inference under potentially misspecified models. Sect. 5 reviews some popular measures of model misspecification and analyzes their statistical properties. Sect. 6 discusses some subtle issues associated with the two-pass CSR methodology that are often overlooked by researchers. The focus of Sect. 7 and 8 is on pairwise and multiple model comparison tests, respectively. Sect. 9 concludes and discusses several avenues for future research.

2 The Two-Pass Cross-Sectional Regression Methodology

Let f_t be a K-vector of factors at time t and R_t a vector of returns on N test assets at time t. We define $Y_t = [f_t', \ R_t']'$ and its unconditional mean and covariance matrix as

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

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

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 $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$. Throughout the article, we assume that the time series Y_t is jointly stationary and ergodic, with finite fourth moment.

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). In general, asset pricing models only require the linear relationship in (3) to hold conditionally. However, most empirical studies estimate an unconditional version of (3). This can be justified on the following grounds. First, the stochastic process of the conditional betas could be specified such that the K-factor beta pricing model holds unconditionally. See for example, Chan and Chen [5] and Jagannathan and Wang [6]. Second, one could let γ be linear in a set of instruments. This will then lead to an expanded unconditional beta pricing model, which includes the instruments and the original factors multiplied by instruments as additional factors.

Suppose that we have T observations on Y_t and denote the sample mean and covariance matrix of Y_t by

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{4}$$

$$\hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{5}$$

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

$$R_t = \alpha + \beta f_t + \epsilon_t, \quad t = 1, \ldots, T. \tag{6}$$

The estimated betas from this first-pass time-series regression are given by the matrix $\hat{\beta} = \hat{V}_{21}\hat{V}_{11}^{-1}$.

In the second pass, we run a single CSR of $\hat{\mu}_2$ on $\hat{X} = [1_N,\ \hat{\beta}]$ to estimate γ. Note that some studies allow $\hat{\beta}$ to change throughout the sample period. For example, in the original Fama and MacBeth [2] 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. The second-pass CSR estimators will depend on the weighting matrix W. Popular choices of W in the literature are $W = I_N$ (OLS), $W = V_{22}^{-1}$ (generalized least squares, GLS), and $W = \Sigma_d^{-1}$ (weighted least squares, WLS), where $\Sigma_d = \text{Diag}(\Sigma)$ is a diagonal matrix containing the diagonal elements of Σ.

When W is known (say OLS CSR), we can estimate γ in (3) by

$$\hat{\gamma} = (\hat{X}'W\hat{X})^{-1}\hat{X}'W\hat{\mu}_2. \tag{7}$$

In the feasible GLS and WLS cases, W contains unknown parameters and one needs to substitute a consistent estimate of W, say $\hat{W}$, in (7). 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. As pointed out by Lewellen, Nagel, and Shanken [7], the estimates of γ are the same regardless of whether we use $W = V_{22}^{-1}$ or $W = \Sigma^{-1}$ as the weighting matrix for the GLS CSR. However, it should be noted that the cross-sectional R^2s 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.

3 Statistical Inference under Correctly Specified Models

In this section, we present the asymptotic distribution of $\hat{\gamma}$ when the model is correctly specified, i.e., (3) holds exactly.

We first consider the special case in which the true betas are used in the second-pass CSR. The estimate of γ is given by

$$\hat{\gamma} = A\hat{\mu}_2, \tag{8}$$

where $A = (X'WX)^{-1}X'W$. Eq. (8) shows that the randomness of $\hat{\gamma}$ is entirely driven by the randomness of $\hat{\mu}_2$. Under the joint stationarity and ergodicity assumptions, we have

$$\sqrt{T}(\hat{\mu}_2 - \mu_2) \overset{A}{\sim} N\left(0_N, \sum_{j=-\infty}^{\infty} E[(R_t - \mu_2)(R_{t+j} - \mu_2)']\right). \tag{9}$$

It follows that

$$\sqrt{T}(\hat{\gamma} - \gamma) \overset{A}{\sim} N(0_{K+1}, V(\hat{\gamma})), \tag{10}$$

where

$$V(\hat{\gamma}) = \sum_{j=-\infty}^{\infty} E[h_t h'_{t+j}], \tag{11}$$

with

$$h_t = \gamma_t - \gamma, \tag{12}$$

and $\gamma_t \equiv [\gamma_{0t},\ \gamma'_{1t}]' = AR_t$ is the period-by-period estimate of γ from regressing R_t on X.

If R_t is serially uncorrelated, then h_t is serially uncorrelated and

$$V(\hat{\gamma}) = AV_{22}A'. \tag{13}$$

For statistical inference, we need a consistent estimator of $V(\hat{\gamma})$. This can be accomplished by replacing h_t with

$$\hat{h}_t = \gamma_t - \hat{\gamma}. \tag{14}$$

When h_t is serially uncorrelated, a consistent estimator of the asymptotic variance of $\hat{\gamma}$ is given by

$$\hat{V}(\hat{\gamma}) = \frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{h}'_t. \tag{15}$$

Eq. (15) yields the popular standard error of $\hat{\gamma}$ due to Fama and MacBeth [2], which is obtained using the standard deviation of the time series $\{\gamma_t\}$. When h_t is autocorrelated, one can use Newey and West's [8] method to obtain a consistent estimator of $V(\hat{\gamma})$.

In the general case, the betas are estimated with error in the first-pass time series regression and 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 [9] shows that 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 [9] addresses this by developing an asymptotically valid EIV adjustment of the standard errors. Jagannathan and Wang [10] extend Shanken's asymptotic analysis by relaxing the assumption that the returns are homoskedastic conditional on the factors.

It turns out that one can easily deal with the EIV problem by replacing h_t in (12) with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t, \tag{16}$$

where $\phi_t = [\gamma_{0t},\ (\gamma_{1t} - f_t)']'$, $\phi = [\gamma_0,\ (\gamma_1 - \mu_1)']'$, and $w_t = \gamma_1' V_{11}^{-1}(f_t - \mu_1)$. The second term, $(\phi_t - \phi)w_t$, is the EIV adjustment term that accounts for the estimation errors in $\hat{\beta}$. To estimate $V(\hat{\gamma})$, we replace h_t with its sample counterpart

$$\hat{h}_t = (\hat{\gamma}_t - \hat{\gamma}) - (\hat{\phi}_t - \hat{\phi})\hat{w}_t, \tag{17}$$

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

When h_t is serially uncorrelated and $\text{Var}[R_t|f_t] = \Sigma$ (conditional homoskedasticity case), we can simplify $V(\hat{\gamma})$ to

$$V(\hat{\gamma}) = AV_{22}A' + \gamma_1' V_{11}^{-1}\gamma_1 A\Sigma A', \tag{18}$$

which is the expression given in Shanken [9]. Using the fact that

$$V_{22} = \Sigma + \beta V_{11}\beta' = \Sigma + X \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} \end{bmatrix} X', \tag{19}$$

we can also write (18) as

$$V(\hat{\gamma}) = (1 + \gamma_1' V_{11}^{-1}\gamma_1)A\Sigma A' + \begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11} \end{bmatrix}. \tag{20}$$

In the above analysis, we have treated W as a known weighting matrix. Under a correctly specified model, the asymptotic distribution of $\hat{\gamma}$ does not depend on whether we use W or its consistent estimator $\hat{W}$ as the weighting matrix. Therefore, the asymptotic results in this section also hold for the GLS CSR and WLS CSR cases.

Under a correctly specified model, it is interesting to derive the optimal (in the sense that it minimizes $V(\hat{\gamma})$) weighting matrix W in the second-pass CSR. Ahn, Gadarowski, and Perez [11] provide an analysis of this problem. Using the fact that $\gamma_t - \gamma = A(R_t - \mu_2)$ and $\phi_t - \phi = A\epsilon_t$, where $\epsilon_t = (R_t - \mu_2) - \beta(f_t - \mu_1)$, we can write

$$h_t = Al_t, \tag{21}$$

where $l_t \equiv R_t - \mu_2 - \epsilon_t w_t$. It follows that

$$V(\hat{\gamma}) = AV_lA' = (X'WX)^{-1}X'WV_lWX(X'WX)^{-1}, \tag{22}$$

where

$$V_l = \sum_{j=-\infty}^{\infty} E[l_t l'_{t+j}]. \tag{23}$$

From this expression, it is obvious that we can choose $W = V_l^{-1}$ to minimize $V(\hat{\gamma})$ and we have $\min_W V(\hat{\gamma}) = (X'V_l^{-1}X)^{-1}$. However, it is important to note that V_l^{-1} is not the only choice of W that minimizes $V(\hat{\gamma})$. Using a lemma in Kan and Zhou [12], it is easy to show that any W that is of the form $(aV_l + XCX')^{-1}$, where a is a positive scalar and C is an arbitrary symmetric matrix, will also yield the lowest $V(\hat{\gamma})$.

There are a few cases in which the GLS CSR will give us the lower bound of $V(\hat{\gamma})$. The first case arises when h_t is serially uncorrelated and $\mathrm{Var}[R_t|f_t] = \Sigma$ (conditional homoskedasticity case). In this scenario, we have

$$V_l = E[l_t l_t'] = (1 + \gamma_1' V_{11}^{-1}\gamma_1)\Sigma + \beta V_{11}\beta' = V_{22} + \gamma_1' V_{11}^{-1}\gamma_1\Sigma. \tag{24}$$

It can be readily shown that

$$(X'V_l^{-1}X)^{-1} = (X'V_{22}^{-1}X)^{-1} + \gamma_1' V_{11}^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}. \tag{25}$$

The second case arises when $Y_t = [f_t',\ R_t']'$ is i.i.d. multivariate elliptically distributed with multivariate excess kurtosis parameter κ. In this case, we have

$$V_l = E[l_t l_t'] = [1+(1+\kappa)\gamma_1' V_{11}^{-1}\gamma_1]\Sigma + \beta V_{11}\beta' = V_{22} + (1+\kappa)\gamma_1' V_{11}^{-1}\gamma_1\Sigma \tag{26}$$

and

$$(X'V_l^{-1}X)^{-1} = (X'V_{22}^{-1}X)^{-1} + (1+\kappa)\gamma_1' V_{11}^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1}. \tag{27}$$

In general, the GLS CSR is not the optimal CSR to be used in the second pass. The best choice of W is V_l^{-1}. To use the optimal two-pass CSR, one needs to obtain a consistent estimator of V_l. This can be accomplished with a two-step procedure: (1) Obtain a consistent estimate of γ_1 using, for example, the OLS CSR. (2) Estimate V_l using $\hat{l}_t = (R_t - \hat{\mu}_2) - \hat{\epsilon}_t\hat{w}_t$.

4 Statistical Inference under Potentially Misspecified Models

Standard inference using the two-pass CSR methodology typically 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. In this section, we discuss how to conduct statistical inference on $\hat{\gamma}$ when the model is potentially misspecified. The results that we present here are mostly drawn from Kan, Robotti, and Shanken [13], which generalizes the earlier results of Hou and Kimmel [14] and Shanken and Zhou [15] that are obtained under a normality assumption.

When the model is misspecified, the pricing-error vector, $\mu_2 - X\gamma$, will be nonzero for all values of γ. For a given weighting matrix W, we define the (pseudo) zero-beta rate and risk premia as the choice of γ that minimizes the quadratic form of pricing errors:

$$\gamma_W \equiv \begin{bmatrix} \gamma_{W,0} \\ \gamma_{W,1} \end{bmatrix} = \operatorname{argmin}_\gamma (\mu_2 - X\gamma)'W(\mu_2 - X\gamma) = (X'WX)^{-1}X'W\mu_2. \quad (28)$$

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

$$e_W = \mu_2 - X\gamma_W. \quad (29)$$

It should be emphasized that unless the model is correctly specified, γ_W and e_W depend on the choice of W. To simplify the notation, we suppress the subscript W from γ_W and e_W when the choice of W is clear from the context.

Unlike the case of correctly specified models, the asymptotic variance of $\hat{\gamma}$ under a misspecified model depends on whether we use W or $\hat{W}$ as the weighting matrix. As a result, we need to separate these two cases when presenting the asymptotic distribution of $\hat{\gamma}$. For the known weighting matrix case, the asymptotic variance of $\hat{\gamma}$ is obtained by replacing h_t in (16) with

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + (X'WX)^{-1}z_t u_t, \quad (30)$$

where $z_t = [0,\ (f_t - \mu_1)'V_{11}^{-1}]'$ and $u_t = e'W(R_t - \mu_2)$.

For the GLS case that uses $W = V_{22}^{-1}$ as the weighting matrix, h_t has the following expression:

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + (X'V_{22}^{-1}X)^{-1}z_t u_t - (\gamma_t - \gamma)u_t. \quad (31)$$

For the WLS case, h_t is given by

$$h_t = (\gamma_t - \gamma) - (\phi_t - \phi)w_t + (X'\Sigma_d^{-1}X)^{-1}z_t u_t - A\Psi_t\Sigma_d^{-1}e, \quad (32)$$

where $\Psi_t = \mathrm{Diag}(\epsilon_t\epsilon_t')$ and $\epsilon_t = (R_t - \mu_2) - \beta(f_t - \mu_1)$. As before, we can obtain a consistent estimator of $V(\hat{\gamma})$ by replacing h_t with its sample counterpart.

Note that model misspecification adds extra terms to h_t and this could have a serious impact on the standard error of $\hat{\gamma}$. For example, when h_t is serially uncorrelated and the conditional homoskedasticity assumption holds, we can show that for the GLS CSR

$$V(\hat{\gamma}) = (X'V_{22}^{-1}X)^{-1} + \gamma_1'V_{11}^{-1}\gamma_1(X'\Sigma^{-1}X)^{-1} + e'V_{22}^{-1}e\left[(X'\Sigma^{-1}X)^{-1}\begin{bmatrix} 0 & 0_K' \\ 0_K & V_{11}^{-1}\end{bmatrix}(X'\Sigma^{-1}X)^{-1} + (X'\Sigma^{-1}X)^{-1}\right]. \quad (33)$$

We call the last term in (33) the misspecification adjustment term. When $e'V_{22}^{-1}e > 0$, the misspecification adjustment term is positive definite since it is the sum of two matrices, the first positive semidefinite and the second positive definite. It can also be shown 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 correlations with the returns (e.g., macroeconomic factors), the impact of the misspecification adjustment term on the asymptotic variance of $\hat{\gamma}_1$ can be very large.

5 Specification Tests and Measures of Model Misspecification

One of the earliest problems in empirical asset pricing has been to determine whether a proposed model is correctly specified or not. This can be accomplished by using various specification tests, which are typically aggregate measures of sample pricing errors. However, some of these specification tests aggregate the pricing errors using weighting matrices that are model dependent, and these test statistics cannot be used to perform model comparison. Therefore, researchers are often interested in a normalized goodness-of-fit measure that uses the same weighting matrix across models. One such measure is the cross-sectional R^2. Following Kandel and Stambaugh [16], this is defined as

$$\rho^2 = 1 - \frac{Q}{Q_0}, \quad (34)$$

where

$$Q_0 = \min_{\gamma_0}(\mu_2 - 1_N\gamma_0)'W(\mu_2 - 1_N\gamma_0) = \mu_2'W\mu_2 - \mu_2'W1_N(1_N'W1_N)^{-1}1_N'W\mu_2, \quad (35)$$

$$Q = e'We = \mu_2'W\mu_2 - \mu_2'WX(X'WX)^{-1}X'W\mu_2. \quad (36)$$

In order for ρ^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 $0 \leq \rho^2 \leq 1$ and it is a decreasing function of the aggregate pricing-error measure $Q = e'We$. Thus, ρ^2 is a natural measure of goodness of fit. However, it should be emphasized that unless the model is correctly specified, ρ^2 depends on the choice of W. Therefore, it is possible that a model with a good fit under the OLS CSR provides a very poor fit under the GLS CSR.

The sample measure of ρ^2 is similarly defined as

$$\hat{\rho}^2 = 1 - \frac{\hat{Q}}{\hat{Q}_0}, \tag{37}$$

where $\hat{Q}_0$ and $\hat{Q}$ are consistent estimators of Q_0 and Q in (35) and (36), respectively. When W is known, we estimate Q_0 and Q using

$$\hat{Q}_0 = \hat{\mu}_2' W \hat{\mu}_2 - \hat{\mu}_2' W 1_N (1_N' W 1_N)^{-1} 1_N' W \hat{\mu}_2, \tag{38}$$

$$\hat{Q} = \hat{\mu}_2' W \hat{\mu}_2 - \hat{\mu}_2' W \hat{X} (\hat{X}' W \hat{X})^{-1} \hat{X}' W \hat{\mu}_2. \tag{39}$$

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

To test the null hypothesis of correct model specification, i.e., $e = 0_N$ (or, equivalently, $Q = 0$ and $\rho^2 = 1$), we typically rely on the sample pricing errors $\hat{e}$. Therefore, it is important to obtain the asymptotic distribution of $\hat{e}$ under the null hypothesis. For a given weighting matrix W (or $\hat{W}$ with a limit of W), let P be an $N \times (N - K - 1)$ orthonormal matrix with its columns orthogonal to $W^{\frac{1}{2}} X$. Kan, Robotti, and Shanken [13] derive the asymptotic distribution of $\hat{e}$ under the null hypothesis:

$$\sqrt{T}\hat{e} \overset{A}{\sim} N(0_N, V(\hat{e})), \tag{40}$$

where

$$V(\hat{e}) = \sum_{j=-\infty}^{\infty} E[q_t q_{t+j}'], \tag{41}$$

with

$$q_t = W^{-\frac{1}{2}} P P' W^{\frac{1}{2}} \epsilon_t y_t, \tag{42}$$

and $y_t = 1 - \gamma_1' V_{11}^{-1} (f_t - \mu_1)$.

Remark 1. Under the correctly specified model, the asymptotic distribution of $\hat{e}$ does not depend on whether we use W or $\hat{W}$ as the weighting matrix.

Remark 2. Under the correctly specified model, q_t in (42) can also be written as

$$q_t = W^{-\frac{1}{2}} P P' W^{\frac{1}{2}} R_t y_t. \tag{43}$$

Remark 3. $V(\hat{e})$ is a singular matrix and some linear combinations of $\sqrt{T}\hat{e}$ are not asymptotically normally distributed. As a result, one has to be careful when relying on individual sample pricing errors to test the validity of a model because some of them may not be asymptotically normally distributed. Gospodinov, Kan, and Robotti [17] provide a detailed analysis of this problem. For our subsequent analysis, it is easier to work with

$$\tilde{e} = P' W^{\frac{1}{2}} \hat{e}. \tag{44}$$

The reason is that the asymptotic variance of $\tilde{e}$ is given by

$$V(\tilde{e}) = \sum_{j=-\infty}^{\infty} E[\tilde{q}_t \tilde{q}_{t+j}'], \tag{45}$$

where

$$\tilde{q}_t = P'W^{\frac{1}{2}} \epsilon_t y_t, \tag{46}$$

and $V(\tilde{e})$ is nonsingular.

Given (40), we can obtain the asymptotic distribution of any quadratic form of sample pricing errors. For example, let Ω be an $N \times N$ positive definite matrix, and let $\hat{\Omega}$ be a consistent estimator of Ω. When the model is correctly specified, we have

$$T\hat{e}'\hat{\Omega}\hat{e} \overset{A}{\sim} \sum_{i=1}^{N-K-1} \xi_i x_i, \tag{47}$$

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

$$(P'W^{-\frac{1}{2}}\Omega W^{-\frac{1}{2}}P)V(\tilde{e}). \tag{48}$$

Using an algorithm due to Imhof [18] and later improved by Davies [20] and Lu and King [19], one can easily compute the cumulative distribution function of a linear combination of independent χ^2 random variables. As a result, one can use (47) as a specification test of the model.

There are several interesting choices of $\hat{\Omega}$. The first one is $\hat{\Omega} = \hat{W}$, and the test statistic is simply given by $T\hat{e}'\hat{W}\hat{e} = T\hat{Q}$. In this case, the ξ_i's are the eigenvalues of $V(\tilde{e})$. The second one is $\hat{\Omega} = \hat{V}(\hat{e})^+$, where $\hat{V}(\hat{e})$ is a consistent estimator of $V(\hat{e})$ and $\hat{V}(\hat{e})^+$ stands for its pseudo-inverse. This choice of $\hat{\Omega}$ yields the following Wald test statistic:

$$J_W = T\hat{e}'\hat{V}(\hat{e})^+\hat{e} = T\tilde{e}'\hat{V}(\tilde{e})^{-1}\tilde{e} \overset{A}{\sim} \chi^2_{N-K-1}, \tag{49}$$

where $\hat{V}(\tilde{e})$ is a consistent estimator of $V(\tilde{e})$. The advantage of using J_W is that its asymptotic distribution is simply χ^2_{N-K-1} and does not involve the computation of the distribution of a linear combination of independent χ^2 random variables. The disadvantage of using J_W is that the weighting matrix is model dependent, making it problematic to compare the J_W's of different models.

When q_t is serially uncorrelated and $\mathrm{Var}[R_t|f_t] = \Sigma$ (conditional homoskedasticity case), we can show that

$$V(\tilde{e}) = (1 + \gamma_1' V_{11}^{-1}\gamma_1)P'W^{\frac{1}{2}}\Sigma W^{\frac{1}{2}}P. \tag{50}$$

For the special case of $W = V_{22}^{-1}$ or $W = \Sigma^{-1}$, we have

$$V(\tilde{e}) = (1 + \gamma_1' V_{11}^{-1}\gamma_1)I_{N-K-1}. \tag{51}$$

If we estimate $V(\tilde{e})$ using $\hat{V}(\tilde{e}) = (1 + \hat{\gamma}_1'\hat{V}_{11}^{-1}\hat{\gamma}_1)I_{N-K-1}$, the Wald test in (49) becomes

$$J_W = T\tilde{e}'\hat{V}(\tilde{e})^{-1}\tilde{e} = \frac{T\hat{e}'\hat{V}_{22}^{-1}\hat{e}}{1 + \hat{\gamma}_1'\hat{V}_{11}^{-1}\hat{\gamma}_1} = \frac{T\hat{e}'\hat{\Sigma}^{-1}\hat{e}}{1 + \hat{\gamma}_1'\hat{V}_{11}^{-1}\hat{\gamma}_1} \overset{A}{\sim} \chi^2_{N-K-1}, \qquad (52)$$

and J_W coincides with the cross-sectional regression test (CSRT) proposed by Shanken [21]. Better finite sample properties of the Wald test can be obtained, as suggested by Shanken [21], by using the following approximate F-test:

$$J_W \overset{\text{app.}}{\sim} \frac{T(N-K-1)}{T-N+1} F_{N-K-1,T-N+1}. \qquad (53)$$

Using the general result in (47), one can show that when the model is correctly specified,

$$T(\hat{\rho}^2 - 1) \overset{A}{\sim} \sum_{i=1}^{N-K-1} -\frac{\xi_i}{Q_0} x_i, \qquad (54)$$

and the sample cross-sectional R^2 can be used as a specification test.

When the model is misspecified, i.e., $\rho^2 < 1$, there are two possible asymptotic distributions for $\hat{\rho}^2$. When $\rho^2 = 0$, we have

$$T\hat{\rho}^2 \overset{A}{\sim} \sum_{i=1}^{K} \tilde{\xi}_i x_i, \qquad (55)$$

where the x_i's are independent χ^2_1 random variables and the $\tilde{\xi}_i$'s are the eigenvalues of

$$[\beta'W\beta - \beta'W1_N(1_N'W1_N)^{-1}1_N'W\beta]V(\hat{\gamma}_1), \qquad (56)$$

where $V(\hat{\gamma}_1)$ is the asymptotic covariance matrix of $\hat{\gamma}_1$ under potentially misspecified models (i.e., based on the expressions of h_t in (30)–(32)). This asymptotic distribution permits a test of whether the model has any explanatory power for expected returns. It can be shown that $\rho^2 = 0$ if and only if $\gamma_1 = 0_K$. Therefore, one can also test $H_0 : \rho^2 = 0$ using a Wald test of $H_0 : \gamma_1 = 0_K$.

When $0 < \rho^2 < 1$, the asymptotic distribution of $\hat{\rho}^2$ is given by

$$\sqrt{T}(\hat{\rho}^2 - \rho^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[n_t n_{t+j}]\right), \qquad (57)$$

where

$$n_t = 2\left[-u_t y_t + (1-\rho^2)v_t\right]/Q_0 \quad \text{for known } W, \qquad (58)$$

$$n_t = \left[u_t^2 - 2u_t y_t + (1-\rho^2)(2v_t - v_t^2)\right]/Q_0 \quad \text{for } \hat{W} = \hat{V}_{22}^{-1}, \qquad (59)$$

$$n_t = \left[e'\Gamma_t e - 2u_t y_t + (1-\rho^2)(2v_t - e_0\Gamma_t e_0)\right]/Q_0 \quad \text{for } \hat{W} = \hat{\Sigma}_d^{-1}, \qquad (60)$$

with $v_t = e_0' W(R_t - \mu_2)$ and $\Gamma_t = \Sigma_d^{-1}\text{Diag}(\epsilon_t\epsilon_t')\Sigma_d^{-1}$.

In the $0 < \rho^2 < 1$ case, $\hat{\rho}^2$ is asymptotically normally distributed around its true value. It is readily verified that the expressions for n_t approach zero when $\rho^2 \to 0$ or $\rho^2 \to 1$. Consequently, the standard error of $\hat{\rho}^2$ tends to be lowest when ρ^2 is close to zero or one, and thus it 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). 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 $\rho^2 = 1$).

6 Some Subtle Issues

In this section, we discuss two issues related to the two-pass CSR methodology that are worth clarifying. The first point is about testing whether the risk premium associated with an individual factor is equal to zero. The second point is about the assumption of full column rank on the matrix $X = [1_N,\ \beta]$.

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 in the context of model selection can be problematic.

For example, suppose that a model has two factors f_1 and f_2. We are often interested in determining whether f_2 is needed in the model. Some researchers have tried to answer this question by performing a test of $H_0 : \gamma_2 = 0$, where γ_2 is the risk premium associated with factor 2. When the null hypothesis is rejected by the data, they typically conclude that factor 2 is important, and when the null hypothesis is not rejected, they conclude that factor 2 is unimportant. In the following, we provide two numerical examples that illustrate that the test of $H_0 : \gamma_2 = 0$ does not answer the question of whether factor 2 helps to explain the cross-sectional differences in expected returns on the test assets.

In the first example, we consider two factors with

$$V_{11} = \begin{bmatrix} 15 & -10 \\ -10 & 15 \end{bmatrix}. \tag{61}$$

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

$$\mu_2 = [2,\ 3,\ 4,\ 5]', \qquad V_{12} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 5 & 2 & 1 \end{bmatrix}. \tag{62}$$

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:

$$\beta = V_{21}V_{11}^{-1} = \begin{bmatrix} 0.36\ 0.64\ 0.52\ 0.56 \\ 0.44\ 0.76\ 0.48\ 0.44 \end{bmatrix}'. \tag{63}$$

Simple calculations give $\gamma = [1,\ 15,\ -10]'$ and γ_2 is nonzero even though f_2 is irrelevant. This suggests that when the capital asset pricing model is true, it does not imply that the betas with respect to the other two Fama and French [44] factors should not be priced. See Grauer and Janmaat [25] for a discussion of this point.

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 R^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 R^2 of the full model is 100%. Simple calculations give us $\gamma = [1,\ 25,\ 0]'$ and $\gamma_2 = 0$, even though f_2 is needed in the factor model, along with f_1, to explain μ_2.

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:

$$\lambda_W \equiv \begin{bmatrix} \lambda_{W,0} \\ \lambda_{W,1} \end{bmatrix} = \mathrm{argmin}_\lambda (\mu_2 - C\lambda)'W(\mu_2 - C\lambda) = (C'WC)^{-1}C'W\mu_2. \tag{64}$$

Given (28) and (64), there is a one-to-one correspondence between γ_W and λ_W:

$$\lambda_{W,0} = \gamma_{W,0}, \qquad \lambda_{W,1} = V_{11}^{-1}\gamma_{W,1}. \tag{65}$$

To simplify the notation, we suppress the subscript W from λ_W when the choice of W is clear from the context. It is easy to see that the pricing errors from this alternative second-pass CSR are the same as the pricing errors from the CSR that uses the betas as regressors. It follows that the ρ^2 for these two CSRs are also identical. However, it is important to note that unless V_{11} is a diagonal matrix, $\lambda_{1,i} = 0$ does not imply $\gamma_{1,i} = 0$, and vice versa. If interest lies in determining whether a particular factor i contributes to the explanatory power of the model, the correct hypothesis to test is $H_0 : \lambda_{1,i} = 0$ and not $H_0 : \gamma_{1,i} = 0$. This issue is also discussed in Jagannathan and Wang [10] and Cochrane ([22], Chapter 13.4). 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 [23] and Jagannathan and Wang [6], [10]. Kan and Robotti [24] provide asymptotic results for the CSR with simple regression betas under potentially misspecified models.

Let $\hat{C} = [1_N,\ \hat{V}_{21}]$. The estimate of λ from the second-pass CSR is given by

$$\hat{\lambda} = (\hat{C}'W\hat{C})^{-1}\hat{C}'W\hat{\mu}_2. \quad (66)$$

For the GLS and WLS cases, one needs to replace W with $\hat{W}$ in the expression for $\hat{\lambda}$.

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

$$\sqrt{T}(\hat{\lambda} - \lambda) \stackrel{A}{\sim} N(0_{K+1}, V(\hat{\lambda})), \quad (67)$$

where

$$V(\hat{\lambda}) = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t\tilde{h}'_{t+j}]. \quad (68)$$

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

$$\tilde{h}_t = (\lambda_t - \lambda) + \tilde{A}\tilde{G}_t\lambda_1 + (C'WC)^{-1}\tilde{z}_t u_t \quad \text{for known } W, \quad (69)$$

$$\tilde{h}_t = (\lambda_t - \lambda) + \tilde{A}\tilde{G}_t\lambda_1 + (C'V_{22}^{-1}C)^{-1}\tilde{z}_t u_t - (\lambda_t - \lambda)u_t \quad \text{for } \hat{W} = \hat{V}_{22}^{-1}, \quad (70)$$

$$\tilde{h}_t = (\lambda_t - \lambda) + \tilde{A}\tilde{G}_t\lambda_1 + (C'\Sigma_d^{-1}C)^{-1}\tilde{z}_t u_t - \tilde{A}\Psi_t\Sigma_d^{-1}e \quad \text{for } \hat{W} = \hat{\Sigma}_d^{-1}, \quad (71)$$

where $\Psi_t = \text{Diag}(\epsilon_t\epsilon'_t)$. Besides allowing us to test whether a given factor is important in a model, the asymptotic distribution of $\hat{\lambda}$ is necessary for the implementation of model comparison, a topic that will be discussed in Sect. 7 and 8. To test $H_0 : \lambda_{1,i} = 0$, one needs to obtain a consistent estimator of $V(\hat{\lambda})$. This can be easily accomplished by replacing $\tilde{h}_t$ with its sample counterpart.

The second issue that is often overlooked by researchers is related to the full column rank assumption on $X = [1_N,\ \beta]$. In the two-pass CSR methodology, we need to assume that X has full rank to ensure that γ_W (or λ_W) is uniquely defined. This assumption is often taken for granted and most researchers do not examine its validity before performing the two-pass CSR. When X does not have full column rank, the asymptotic results will break down, leading to misleading statistical inference on $\hat{\gamma}$ and $\hat{\rho}^2$, especially when the model is misspecified. For example, Kan and Zhang [26] show that there is a high probability that the estimated risk premium on a useless factor is significantly different from zero. This happens because, when the factor is useless, $\beta = 0_N$ and $X = [1_N,\ \beta]$ does not have full column rank. As a result, γ_W is not uniquely defined and the usual asymptotic standard error of $\hat{\gamma}$ is no longer valid. Note that the useless factors scenario is not completely unreasonable since many macroeconomic factors exhibit very low correlations with asset returns. Even when the full column rank condition is satisfied in population, Kleibergen [27] shows that there can still be serious finite sample problems with the asymptotic results if the factors have low correlations with the returns and the beta estimates are noisy.

When the factors have very low correlations with the returns, it is sensible to test whether X has full column rank before running the two-pass CSR. Note that testing $H_0 : \text{rank}(X) = K$ is the same as testing $H_0 : \text{rank}(\Pi) = K - 1$, where $\Pi = P'\beta$ and P is an $N \times (N-1)$ orthonormal matrix with its columns orthogonal to 1_N. When $K = 1$, it is easy to perform this test because the null hypothesis is simply $H_0 : P'\beta = 0_{N-1}$.

A simple Wald test of $H_0 : P'\beta = 0_{N-1}$ can be performed using the following test statistic:

$$J_1 = T\hat{\beta}' P (P' \hat{V}(\hat{\beta}) P)^{-1} P' \hat{\beta} \overset{A}{\sim} \chi^2_{N-1}, \tag{72}$$

where $\hat{V}(\hat{\beta})$ is a consistent estimator of $V(\hat{\beta})$, the asymptotic covariance of $\hat{\beta}$. Under the stationarity and ergodicity assumptions on Y_t,

$$V(\hat{\beta}) = \sum_{j=-\infty}^{\infty} E[x_t x'_{t+j}], \tag{73}$$

where

$$x_t = V_{11}^{-1}(f_t - \mu_1)\epsilon_t. \tag{74}$$

If we further assume that x_t is serially uncorrelated and $\text{Var}[R_t|f_t] = \Sigma$ (conditional homoskedasticity case),

$$V(\hat{\beta}) = V_{11}^{-1}\Sigma, \tag{75}$$

and we can use the following Wald test:

$$J_2 = T\hat{V}_{11}\hat{\beta}' P (P' \hat{\Sigma} P)^{-1} P' \hat{\beta} \overset{A}{\sim} \chi^2_{N-1}. \tag{76}$$

When ϵ_t is i.i.d. multivariate normal, we have the following exact test:

$$F_2 = \frac{(T-N)J_2}{(N-1)T} \sim F_{N-1,T-N}. \tag{77}$$

When ϵ_t is not normally distributed, this F-test is only an approximate test but it generally works better than the asymptotic one.

When $K > 1$, the test of $H_0 : \text{rank}(\Pi) = K - 1$ is more complicated. Several tests of the rank of a matrix have been proposed in the literature. In the following, we describe the test of Cragg and Donald [28]. Let $\hat{\Pi} = P'\hat{\beta}$, we have

$$\sqrt{T}\text{vec}(\hat{\Pi} - \Pi) \overset{A}{\sim} N(0_{(N-1)K}, S_{\hat{\Pi}}), \tag{78}$$

where

$$S_{\hat{\Pi}} = \sum_{j=-\infty}^{\infty} E[\tilde{x}_t \tilde{x}'_{t+j}], \tag{79}$$

and

$$\tilde{x}_t = V_{11}^{-1}(f_t - \mu_1) \otimes P'\epsilon_t. \tag{80}$$

Denoting by $\hat{S}_{\hat{\Pi}}$ a consistent estimator of $S_{\hat{\Pi}}$, Cragg and Donald [28] suggest that we can test $H_0 : \text{rank}(\Pi) = K - 1$ using

$$J_3 = T \min_{\Pi \in \Gamma(K-1)} \text{vec}(\hat{\Pi} - \Pi)' \hat{S}_{\hat{\Pi}}^{-1} \text{vec}(\hat{\Pi} - \Pi) \stackrel{A}{\sim} \chi^2_{N-K}, \tag{81}$$

where $\Gamma(K-1)$ is the space of an $(N-1) \times K$ matrix with rank $K-1$. This test is not computationally attractive in general since we need to optimize over $N(K-1)$ parameters. Gospodinov, Kan, and Robotti [29] propose an alternative expression for J_3 that greatly reduces the computational burden. Their test is given by

$$J_3 = T \min_c [-1,\ c'] \hat{\Pi}' (C \hat{S}_{\hat{\Pi}} C')^{-1} \hat{\Pi} [-1,\ c']' \stackrel{A}{\sim} \chi^2_{N-K}, \tag{82}$$

where c is a $(K-1)$-vector and $C = [-1,\ c'] \otimes I_{N-1}$. With this new expression, one can easily test whether X has full column rank even when K is large.

If we further assume that $\tilde{x}_t$ is serially uncorrelated and $\text{Var}[R_t | f_t] = \Sigma$ (conditional homoskedasticity case),

$$S_{\hat{\Pi}} = V_{11}^{-1} \otimes P' \Sigma P, \tag{83}$$

and we have the following simple test of $H_0 : \text{rank}(\Pi) = K - 1$:

$$J_4 = T \xi_K \stackrel{A}{\sim} \chi^2_{N-K}, \tag{84}$$

where ξ_K is the smallest eigenvalue of $\hat{V}_{11} \hat{\beta}' P (P' \hat{\Sigma} P)^{-1} P' \hat{\beta}$.

7 Pairwise Model Comparison Tests

One way to think about pairwise model comparison is to ask whether two competing beta pricing models have the same population cross-sectional R^2. Kan, Robotti, and Shanken [13] show that the asymptotic distribution of the difference between the sample cross-sectional R^2s of two models depends on whether the models are nested or non-nested and whether the models are correctly specified or not. In this section, 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. Since the weighting matrix of the WLS CSR is model dependent, it is not meaningful to compare the WLS cross-sectional R^2s of two or more models. Therefore, we do not consider the WLS cross-sectional R^2 in the remainder of the article. Our analysis in this section is based on the earlier work of Vuong [30], Rivers and Vuong [31], and Golden [32].

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

$$\mu_2 = 1_N\lambda_{1,0} + \mathrm{Cov}[R_t, f_{1t}']\lambda_{1,1} + \mathrm{Cov}[R_t, f_{2t}']\lambda_{1,2} = C_1\lambda_1, \tag{85}$$

where $C_1 = [1_N,\ \mathrm{Cov}[R_t, f_{1t}'],\ \mathrm{Cov}[R_t, f_{2t}']]$ and $\lambda_1 = [\lambda_{1,0},\ \lambda_{1,1}',\ \lambda_{1,2}']'$. Model 2 requires that expected returns are linear in the betas or covariances with respect to f_{1t} and f_{3t}, i.e.,

$$\mu_2 = 1_N\lambda_{2,0} + \mathrm{Cov}[R_t, f_{1t}']\lambda_{2,1} + \mathrm{Cov}[R_t, f_{3t}']\lambda_{2,3} = C_2\lambda_2, \tag{86}$$

where $C_2 = [1_N,\ \mathrm{Cov}[R_t, f_{1t}'],\ \mathrm{Cov}[R_t, f_{3t}']]$ and $\lambda_2 = [\lambda_{2,0},\ \lambda_{2,1}',\ \lambda_{2,3}']'$.

In general, both models can be misspecified. The λ_i that maximizes the ρ^2 of model i is given by

$$\lambda_i = (C_i'WC_i)^{-1}C_i'W\mu_2, \tag{87}$$

where C_i is assumed to have full column rank, $i = 1, 2$. For each model, the pricing-error vector e_i, the aggregate pricing-error measure Q_i, and the corresponding goodness-of-fit measure ρ_i^2 are all defined as in Sect. 4 and 5.

When $K_2 = 0$, model 2 nests model 1 as a special case. Similarly, when $K_3 = 0$, model 1 nests model 2. When both $K_2 > 0$ and $K_3 > 0$, the two models are non-nested.

We study the nested models case next and deal with non-nested models later in the section. Without loss of generality, we assume $K_3 = 0$, so that model 1 nests model 2. Since $\rho_1^2 = \rho_2^2$ if and only if $\lambda_{1,2} = 0_{K_2}$ (this result is applicable even when the models are misspecified), testing whether the models have the same ρ^2 is equivalent to testing $H_0 : \lambda_{1,2} = 0_{K_2}$. Under the null hypothesis,

$$T\hat{\lambda}_{1,2}'\hat{V}(\hat{\lambda}_{1,2})^{-1}\hat{\lambda}_{1,2} \overset{A}{\sim} \chi^2_{K_2}, \tag{88}$$

where $\hat{V}(\hat{\lambda}_{1,2})$ is a consistent estimator of the asymptotic covariance of $\sqrt{T}(\hat{\lambda}_{1,2} - \lambda_{1,2})$ given in Sect. 6. This statistic can be used to test $H_0 : \rho_1^2 = \rho_2^2$. 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 1 is correctly specified. Instead, we need to rely on the misspecification-robust standard error of $\hat{\lambda}$ given in Sect. 6.

Alternatively, one can derive the asymptotic distribution of $\hat{\rho}_1^2 - \hat{\rho}_2^2$ and use this statistic to test $H_0 : \rho_1^2 = \rho_2^2$. Partition $\tilde{H}_1 = (C_1'WC_1)^{-1}$ as

$$\tilde{H}_1 = \begin{bmatrix} \tilde{H}_{1,11} & \tilde{H}_{1,12} \\ \tilde{H}_{1,21} & \tilde{H}_{1,22} \end{bmatrix}, \tag{89}$$

where $\tilde{H}_{1,22}$ is $K_2 \times K_2$. Under the null hypothesis $H_0 : \rho_1^2 = \rho_2^2$,

$$T(\hat{\rho}_1^2 - \hat{\rho}_2^2) \overset{A}{\sim} \sum_{i=1}^{K_2} \frac{\xi_i}{Q_0} x_i, \tag{90}$$

where the x_i's are independent χ^2_1 random variables and the ξ_i's are the eigenvalues of $\tilde{H}^{-1}_{1,22}V(\hat{\lambda}_{1,2})$. Once again, it is worth emphasizing that the misspecification-robust version of $V(\hat{\lambda}_{1,2})$ should be used to test $H_0 : \rho^2_1 = \rho^2_2$. Model misspecification tends to create additional sampling variation in $\hat{\rho}^2_1 - \hat{\rho}^2_2$. Without taking this into account, one might mistakenly reject the null hypothesis when it is true. In actual testing, we replace ξ_i with its sample counterpart $\hat{\xi}_i$, where the $\hat{\xi}_i$'s are the eigenvalues of $\hat{\tilde{H}}^{-1}_{1,22}\hat{V}(\hat{\lambda}_{1,2})$, and $\hat{\tilde{H}}_{1,22}$ and $\hat{V}(\hat{\lambda}_{1,2})$ are consistent estimators of $\tilde{H}_{1,22}$ and $V(\hat{\lambda}_{1,2})$, respectively.

The test of $H_0 : \rho^2_1 = \rho^2_2$ is more complicated for non-nested models. The reason is that under H_0, there are three possible asymptotic distributions for $\hat{\rho}^2_1 - \hat{\rho}^2_2$, depending on why the two models have the same cross-sectional R^2. To see this, first let us define the normalized stochastic discount factors at time t for models 1 and 2 as

$$y_{1t} = 1 - (f_{1t} - E[f_{1t}])'\lambda_{1,1} - (f_{2t} - E[f_{2t}])'\lambda_{1,2}, \tag{91}$$

$$y_{2t} = 1 - (f_{1t} - E[f_{1t}])'\lambda_{2,1} - (f_{3t} - E[f_{3t}])'\lambda_{2,3}. \tag{92}$$

Kan, Robotti, and Shanken [13] show that $y_{1t} = y_{2t}$ implies that the two models have the same pricing errors and hence $\rho^2_1 = \rho^2_2$. If $y_{1t} \neq y_{2t}$, there are additional cases in which $\rho^2_1 = \rho^2_2$. A second possibility is that both models are correctly specified (i.e., $\rho^2_1 = \rho^2_2 = 1$). This occurs, for example, if model 1 is correctly specified and the factors f_{3t} in model 2 are given by $f_{3t} = f_{2t} + \epsilon_t$, where ϵ_t is pure "noise" — a vector of measurement errors with mean zero, independent of returns. In this case, we have $C_1 = C_2$ 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 aggregation of pricing errors is the same in each case ($\rho^2_1 = \rho^2_2 < 1$). As it turns out, each of these three scenarios results in a different asymptotic distribution for $\hat{\rho}^2_1 - \hat{\rho}^2_2$.

For non-nested models, Kan, Robotti, and Shanken [13] show that $y_{1t} = y_{2t}$ if and only if $\lambda_{1,2} = 0_{K_2}$ and $\lambda_{2,3} = 0_{K_3}$. This result, which is applicable even when the models are misspecified, implies that we can test $H_0 : y_{1t} = y_{2t}$ by testing the joint hypothesis $H_0 : \lambda_{1,2} = 0_{K_2},\ \lambda_{2,3} = 0_{K_3}$. Let $\psi = [\lambda'_{1,2},\ \lambda'_{2,3}]'$ and $\hat{\psi} = [\hat{\lambda}'_{1,2},\ \hat{\lambda}'_{2,3}]'$. Under $H_0 : y_{1t} = y_{2t}$, the asymptotic distribution of $\hat{\psi}$ is given by

$$\sqrt{T}(\hat{\psi} - \psi) \overset{A}{\sim} N(0_{K_2+K_3}, V(\hat{\psi})), \tag{93}$$

where

$$V(\hat{\psi}) = \sum_{j=-\infty}^{\infty} E[\tilde{q}_t\tilde{q}'_{t+j}], \tag{94}$$

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 1 and 2, respectively, where $\tilde{h}_t$ is given in Sect. 6.

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

$$T\hat{\psi}'\hat{V}(\hat{\psi})^{-1}\hat{\psi} \overset{A}{\sim} \chi^2_{K_2+K_3}, \tag{95}$$

and this statistic can be used to test $H_0 : y_{1t} = y_{2t}$. As in the nested models case, it is important to conduct this test using the misspecification-robust standard error of $\hat{\psi}$.

Alternatively, one can derive the asymptotic distribution of $\hat{\rho}_1^2 - \hat{\rho}_2^2$ given $H_0 : y_{1t} = y_{2t}$. Let $\tilde{H}_1 = (C_1'WC_1)^{-1}$ and $\tilde{H}_2 = (C_2'WC_2)^{-1}$, and partition them as

$$\tilde{H}_1 = \begin{bmatrix} \tilde{H}_{1,11} & \tilde{H}_{1,12} \\ \tilde{H}_{1,21} & \tilde{H}_{1,22} \end{bmatrix}, \qquad \tilde{H}_2 = \begin{bmatrix} \tilde{H}_{2,11} & \tilde{H}_{2,13} \\ \tilde{H}_{2,31} & \tilde{H}_{2,33} \end{bmatrix}, \tag{96}$$

where $\tilde{H}_{1,11}$ and $\tilde{H}_{2,11}$ are $(K_1 + 1) \times (K_1 + 1)$. Under the null hypothesis $H_0 : y_{1t} = y_{2t}$,

$$T(\hat{\rho}_1^2 - \hat{\rho}_2^2) \overset{A}{\sim} \sum_{i=1}^{K_2+K_3} \frac{\xi_i}{Q_0} x_i, \tag{97}$$

where the x_i's are independent χ^2_1 random variables and the ξ_i's are the eigenvalues of

$$\begin{bmatrix} \tilde{H}_{1,22}^{-1} & 0_{K_2\times K_3} \\ 0_{K_3\times K_2} & -\tilde{H}_{2,33}^{-1} \end{bmatrix} V(\hat{\psi}). \tag{98}$$

Note that we can think of the earlier nested models scenario as a special case of testing $H_0 : y_{1t} = y_{2t}$ with $K_3 = 0$. The only difference is that the ξ_i's in (90) are all positive whereas some of the ξ_i's in (97) are negative. As a result, we need to perform a two-sided test based on $\hat{\rho}_1^2 - \hat{\rho}_2^2$ in the non-nested models case.

If we fail to reject $H_0 : y_1 = y_2$, we are finished since equality of ρ_1^2 and ρ_2^2 is implied by this hypothesis. Otherwise, we need to consider the case $y_{1t} \neq y_{2t}$. As noted earlier, when $y_{1t} \neq y_{2t}$, the asymptotic distribution of $\hat{\rho}_1^2 - \hat{\rho}_2^2$ given $H_0 : \rho_1^2 = \rho_2^2$ depends on whether the models are correctly specified or not. A simple chi-squared statistic can be used for testing whether models 1 and 2 are both correctly specified. As this joint specification test focuses on the pricing errors, it can be viewed as a generalization of the CSRT of Shanken [21], which tests the validity of the expected return relation for a single pricing model.

Let $n_1 = N - K_1 - K_2 - 1$ and $n_2 = N - K_1 - K_3 - 1$. Also let P_1 be an $N \times n_1$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}}C_1$ and P_2 be an $N \times n_2$ orthonormal matrix with columns orthogonal to $W^{\frac{1}{2}}C_2$. Define

$$g_t(\theta) = \begin{bmatrix} g_{1t}(\lambda_1) \\ g_{2t}(\lambda_2) \end{bmatrix} = \begin{bmatrix} \epsilon_{1t}y_{1t} \\ \epsilon_{2t}y_{2t} \end{bmatrix}, \tag{99}$$

where ϵ_{1t} and ϵ_{2t} are the residuals of models 1 and 2, respectively, $\theta = (\lambda_1', \lambda_2')'$, and

$$S \equiv \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[g_t(\theta)g_{t+j}(\theta)']. \tag{100}$$

If $y_{1t} \neq y_{2t}$ and the null hypothesis $H_0 : \rho_1^2 = \rho_2^2 = 1$ holds, then

$$T \begin{bmatrix} \hat{P}_1'\hat{W}^{\frac{1}{2}}\hat{e}_1 \\ \hat{P}_2'\hat{W}^{\frac{1}{2}}\hat{e}_2 \end{bmatrix}' \begin{bmatrix} \hat{P}_1'\hat{W}^{\frac{1}{2}}\hat{S}_{11}\hat{W}^{\frac{1}{2}}\hat{P}_1 & \hat{P}_1'\hat{W}^{\frac{1}{2}}\hat{S}_{12}\hat{W}^{\frac{1}{2}}\hat{P}_2 \\ \hat{P}_2'\hat{W}^{\frac{1}{2}}\hat{S}_{21}\hat{W}^{\frac{1}{2}}\hat{P}_1 & \hat{P}_2'\hat{W}^{\frac{1}{2}}\hat{S}_{22}\hat{W}^{\frac{1}{2}}\hat{P}_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{P}_1'\hat{W}^{\frac{1}{2}}\hat{e}_1 \\ \hat{P}_2'\hat{W}^{\frac{1}{2}}\hat{e}_2 \end{bmatrix} \overset{A}{\sim} \chi^2_{n_1+n_2}, \tag{101}$$

where $\hat{e}_1$ and $\hat{e}_2$ are the sample pricing errors of models 1 and 2, and $\hat{P}_1$, $\hat{P}_2$, and $\hat{S}$ are consistent estimators of P_1, P_2, and S, respectively.

An alternative specification test makes use of the cross-sectional R^2s. If $y_{1t} \neq y_{2t}$ and the null hypothesis $H_0 : \rho_1^2 = \rho_2^2 = 1$ holds, then

$$T(\hat{\rho}_1^2 - \hat{\rho}_2^2) \overset{A}{\sim} \sum_{i=1}^{n_1+n_2} \frac{\xi_i}{Q_0} x_i, \tag{102}$$

where the x_i's are independent χ_1^2 random variables and the ξ_i's are the eigenvalues of

$$\begin{bmatrix} -P_1'W^{\frac{1}{2}}S_{11}W^{\frac{1}{2}}P_1 & -P_1'W^{\frac{1}{2}}S_{12}W^{\frac{1}{2}}P_2 \\ P_2'W^{\frac{1}{2}}S_{21}W^{\frac{1}{2}}P_1 & P_2'W^{\frac{1}{2}}S_{22}W^{\frac{1}{2}}P_2 \end{bmatrix}. \tag{103}$$

Note that the ξ_i's are not all positive because $\hat{\rho}_1^2 - \hat{\rho}_2^2$ can be negative. Thus, again, we need to perform a two-sided test of $H_0 : \rho_1^2 = \rho_2^2$.

If the hypothesis that both models are correctly specified is not rejected, we are finished, as the data are consistent with $H_0 : \rho_1^2 = \rho_2^2 = 1$. Otherwise, we need to determine whether $\rho_1^2 = \rho_2^2$ for some value less than one. As in the earlier analysis for $\hat{\rho}^2$, the asymptotic distribution of $\hat{\rho}_1^2 - \hat{\rho}_2^2$ changes when the models are misspecified. Suppose $y_{1t} \neq y_{2t}$ and $0 < \rho_1^2 = \rho_2^2 < 1$. Then,

$$\sqrt{T}(\hat{\rho}_1^2 - \hat{\rho}_2^2) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}]\right). \tag{104}$$

When the weighting matrix W is known,

$$d_t = 2Q_0^{-1}(u_{2t}y_{2t} - u_{1t}y_{1t}), \tag{105}$$

where $u_{1t} = e_1'W(R_t - \mu_2)$ and $u_{2t} = e_2'W(R_t - \mu_2)$. With the GLS weighting matrix $\hat{W} = \hat{V}_{22}^{-1}$,

$$d_t = Q_0^{-1}(u_{1t}^2 - 2u_{1t}y_{1t} - u_{2t}^2 + 2u_{2t}y_{2t}). \tag{106}$$

Note that if $y_{1t} = y_{2t}$, then $\rho_1^2 = \rho_2^2$, $u_{1t} = u_{2t}$, and hence $d_t = 0$. Or, if $y_{1t} \neq y_{2t}$, but both models are correctly specified (i.e., $u_{1t} = u_{2t} = 0$ and

$\rho_1^2 = \rho_2^2 = 1$), then again $d_t = 0$. Thus, the normal test cannot be used in these cases.

Given the three distinct cases encountered in testing $H_0 : \rho_1^2 = \rho_2^2$ for non-nested models, the approach we have described above entails a sequential test, as suggested by Vuong [30]. In our context, this involves first testing $H_0 : y_{1t} = y_{2t}$ using (95) or (97). If we reject $H_0 : y_{1t} = y_{2t}$, then we use (101) or (102) to test $H_0 : \rho_1^2 = \rho_2^2 = 1$. Finally, if this hypothesis is also rejected, we use the normal test in (104) to test $H_0 : 0 < \rho_1^2 = \rho_2^2 < 1$. 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 $\max[\alpha_1, \alpha_2, \alpha_3]$.

Another approach is to simply perform the normal test in (104). This amounts to assuming that $y_{1t} \neq y_{2t}$ and that both models are misspecified. The first assumption rules out the unlikely scenario that the additional factors are completely irrelevant for explaining cross-sectional variation in expected returns. The second assumption is sensible because asset pricing models are approximations of reality and we do not expect them to be perfectly specified.

8 Multiple Model Comparison Tests

Thus far, we have considered comparison of two competing models. However, given a set of models of interest, one may want to test whether one model, the "benchmark," has the highest ρ^2 of all models in the set. This gives rise to a common problem in applied work — if we focus on the statistic that provides the strongest evidence of rejection, without taking into account the process of searching across alternative specifications, there will be a tendency to reject the benchmark more often than the nominal size of the tests suggests. In other words, the true p-value will be larger than the one associated with the most extreme statistic.

Therefore, in this section we discuss how to perform model comparison when multiple models are involved. Suppose we have p models. Let ρ_i^2 denotes the cross-sectional R^2 of model i. We are interested in testing if model 1 performs as well as models 2 to p. Let $\delta = (\delta_2, \ldots, \delta_p)$, where $\delta_i = \rho_1^2 - \rho_i^2$. We are interested in testing $H_0 : \delta \geq 0_r$, where $r = p - 1$.

We consider two different tests of this null hypothesis. The first one is the multivariate inequality test developed by Kan, Robotti, and Shanken [13]. Numerous studies in statistics focus on tests of inequality constraints on parameters. The relevant work dates back to Bartholomew [33], Kudo [34], Perlman [35], Gourieroux, Holly, and Monfort [36] and Wolak [37], [38]. Following Wolak [38], we state the null and alternative hypotheses as

$$H_0 : \delta \geq 0_r, \qquad H_1 : \delta \in \Re^r. \tag{107}$$

We also consider another test based on the reality check of White [39] that has been used by Chen and Ludvigson [40]. Let $\delta_{\min} = \min_{2 \leq i \leq p} \delta_i$. Define the null and alternative hypotheses as

$$H_0 : \delta_{\min} \geq 0, \qquad H_1 : \delta_{\min} < 0. \tag{108}$$

The null hypotheses presented above suggest that no other model outperforms model 1, whereas the alternative hypotheses suggest that at least one of the other models outperforms model 1.

Let $\hat{\delta} = (\hat{\delta}_2, \ldots, \hat{\delta}_p)$, where $\hat{\delta}_i = \hat{\rho}_1^2 - \hat{\rho}_i^2$. For both tests, we assume

$$\sqrt{T}(\hat{\delta} - \delta) \stackrel{A}{\sim} N(0_r, \Sigma_{\hat{\delta}}). \tag{109}$$

Starting with the multivariate inequality test, its test statistic is constructed by first solving the following quadratic programming problem

$$\min_{\delta} (\hat{\delta} - \delta)' \hat{\Sigma}_{\hat{\delta}}^{-1} (\hat{\delta} - \delta) \quad \text{s.t.} \quad \delta \geq 0_r, \tag{110}$$

where $\hat{\Sigma}_{\hat{\delta}}$ is a consistent estimator of $\Sigma_{\hat{\delta}}$. Let $\tilde{\delta}$ be the optimal solution of the problem in (110). The likelihood ratio test of the null hypothesis is given by

$$LR = T(\hat{\delta} - \tilde{\delta})' \hat{\Sigma}_{\hat{\delta}}^{-1} (\hat{\delta} - \tilde{\delta}). \tag{111}$$

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

$$\min_{\lambda} \lambda' \hat{\delta} + \frac{1}{2} \lambda' \hat{\Sigma}_{\hat{\delta}} \lambda \quad \text{s.t.} \quad \lambda \geq 0_r. \tag{112}$$

Let $\tilde{\lambda}$ be the optimal solution of the problem in (112). The Kuhn-Tucker test of the null hypothesis is given by

$$KT = T \tilde{\lambda}' \hat{\Sigma}_{\hat{\delta}} \tilde{\lambda}. \tag{113}$$

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

To conduct statistical inference, it is necessary to derive the asymptotic distribution of LR. Wolak [38] shows that under $H_0 : \delta = 0_r$ (i.e., the least favorable value of δ under the null hypothesis), LR has a weighted chi-squared distribution:

$$LR \stackrel{A}{\sim} \sum_{i=0}^{r} w_i(\Sigma_{\hat{\delta}}^{-1}) X_i = \sum_{i=0}^{r} w_{r-i}(\Sigma_{\hat{\delta}}) X_i, \tag{114}$$

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

The biggest hurdle in determining the p-value of this multivariate inequality test is the computation of the weights. For a given $r \times r$ covariance matrix $\Sigma = (\sigma_{ij})$, the expressions for the weights $w_i(\Sigma)$, $i = 0, \ldots, r$, are given in Kudo [34]. The weights depend on Σ through the correlation coefficients $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$. When $r = 1$, $w_0 = w_1 = 1/2$. When $r = 2$,

$$w_0 = \frac{1}{2} - w_2, \tag{115}$$

$$w_1 = \frac{1}{2}, \tag{116}$$

$$w_2 = \frac{1}{4} + \frac{\arcsin(\rho_{12})}{2\pi}. \tag{117}$$

When $r = 3$,

$$w_0 = \frac{1}{2} - w_2, \tag{118}$$

$$w_1 = \frac{1}{2} - w_3, \tag{119}$$

$$w_2 = \frac{3}{8} + \frac{\arcsin(\rho_{12\cdot 3}) + \arcsin(\rho_{13\cdot 2}) + \arcsin(\rho_{23\cdot 1})}{4\pi}, \tag{120}$$

$$w_3 = \frac{1}{8} + \frac{\arcsin(\rho_{12}) + \arcsin(\rho_{13}) + \arcsin(\rho_{23})}{4\pi}, \tag{121}$$

where

$$\rho_{ij\cdot k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{[(1-\rho_{ik}^2)(1-\rho_{jk}^2)]^{\frac{1}{2}}}. \tag{122}$$

For $r > 3$, the computation of the weights is more complicated. Following Kudo [34], let $P = \{1, \dots, r\}$. There are 2^r subsets of P, which are indexed by M. Let $n(M)$ be the number of elements in M and M' be the complement of M relative to P. Define Σ_M as the submatrix that consists of the rows and columns in the set M, $\Sigma_{M'}$ as the submatrix that consists of the rows and columns in the set M', $\Sigma_{M,M'}$ the submatrix with rows corresponding to the elements in M and columns corresponding to the elements in M' ($\Sigma_{M',M}$ is similarly defined), and $\Sigma_{M\cdot M'} = \Sigma_M - \Sigma_{M,M'}\Sigma_{M'}^{-1}\Sigma_{M',M}$. Kudo [34] shows that

$$w_i(\Sigma) = \sum_{M:\, n(M)=i} P(\Sigma_{M'}^{-1})P(\Sigma_{M\cdot M'}), \tag{123}$$

where $P(A)$ is the probability for a multivariate normal distribution with zero mean and covariance matrix A to have all positive elements. In the above equation, we use the convention that $P[\Sigma_{\emptyset\cdot P}] = 1$ and $P[\Sigma_{\emptyset}^{-1}] = 1$. Using (123), we have $w_0(\Sigma) = P(\Sigma^{-1})$ and $w_r(\Sigma) = P(\Sigma)$.

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

To overcome this problem, Kan, Robotti, and Shanken [13] rely on a formula for the positive orthant probability due to Childs [41] and Sun [42]. Let $R = (r_{ij})$ be the correlation matrix corresponding to A. Childs [41] and Sun [42] show that

$$P_{2k}(A) = \frac{1}{2^{2k}} + \frac{1}{2^{2k-1}\pi} \sum_{1\le i<j\le 2k} \arcsin(r_{ij}) + \sum_{j=2}^{k} \frac{1}{2^{2k-j}\pi^j} \sum_{1\le i_1<\cdots<i_{2j}\le 2k} I_{2j}\left(R_{(i_1,\ldots,i_{2j})}\right), \quad (124)$$

$$P_{2k+1}(A) = \frac{1}{2^{2k+1}} + \frac{1}{2^{2k}\pi} \sum_{1\le i<j\le 2k+1} \arcsin(r_{ij}) + \sum_{j=2}^{k} \frac{1}{2^{2k+1-j}\pi^j} \sum_{1\le i_1<\cdots<i_{2j}\le 2k+1} I_{2j}\left(R_{(i_1,\ldots,i_{2j})}\right), \quad (125)$$

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

$$I_{2j}(\Lambda) = \frac{(-1)^j}{(2\pi)^j} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{2j} \frac{1}{\omega_i}\right) \exp\left(-\frac{\omega'\Lambda\omega}{2}\right) \mathrm{d}\omega_1 \cdots \mathrm{d}\omega_{2j}, \quad (126)$$

where Λ is a $2j \times 2j$ covariance matrix and $\omega = (\omega_1,\ldots,\omega_{2j})'$. Sun [42] provides a recursive relation for $I_{2j}(\Lambda)$ that allows us to obtain I_{2j} starting from I_2. Sun's formula enables us to compute the $2j$-th order multivariate integral I_{2j} using a $(j-1)$-th order multivariate integral, which can be obtained numerically using the Gauss-Legendre quadrature method. Sun [43] provides a Fortran subroutine to compute $P(A)$ for $r \le 9$. Kan, Robotti, and Shanken [13] improve on Sun's program and are able to accurately compute $P(A)$ and hence $w_i(\Sigma)$ for $r \le 11$.

Turning to the $\delta_{\min}$ test based on White [39], one can use the sample counterpart of $\delta_{\min}$:

$$\hat{\delta}_{\min} = \min_{2\le i\le p} \hat{\delta}_i \quad (127)$$

to test (108). To determine the p-value of $\hat{\delta}_{\min}$, we need to identify the least favorable value of δ under the null hypothesis. It can be easily shown that the least favorable value of δ under the null hypothesis occurs at $\delta = 0_r$. It follows that asymptotically,

$$\begin{aligned} P[\sqrt{T}\delta_{\min} < c] &\to P[\min_{1\le i\le r} z_i < c] \\ &= 1 - P[\min_{1\le i\le r} z_i > c] \\ &= 1 - P[z_1 > c,\ldots,z_r > c] \\ &= 1 - P[z_1 < -c,\ldots,z_r < -c], \end{aligned} \quad (128)$$

where $z = (z_1,\ldots,z_r)' \sim N(0_r, \Psi)$, and the last equality follows from symmetry since $E[z] = 0_r$. Therefore, to compute the asymptotic p-value one needs to evaluate the cumulative distribution function of a multivariate normal distribution.

Note that both tests crucially depend on the asymptotic normality assumption in (109). Sufficient conditions for this assumption to hold are i) $0 < \rho_i^2 < 1$, and ii) the implied stochastic discount factors of the different models are distinct. Even though the multivariate normality assumption may not always hold at the boundary point of the null hypothesis (i.e., $\delta = 0_r$), it is still possible to compute the p-value as long as we assume that the true δ is not at the boundary point of the null hypothesis. There are, however, cases in which this assumption does not hold. For example, if model 2 nests model 1, then we cannot have $\delta_2 > 0$. As a result, the null hypothesis $H_0 : \delta_2 \geq 0$ becomes $H_0 : \delta_2 = 0$. Under this null hypothesis, $\sqrt{T}\hat{\delta}_2$ no longer has a multivariate normal distribution and both the multivariate inequality test and the $\delta_{\min}$ test will break down.

Therefore, when nested models are involved, the two tests need to be modified. If model 1 nests some of the competing models, then those models that are nested by model 1 will not be included in the model comparison tests. The reason is that these models are clearly dominated by model 1 and we no longer need to perform tests in presence of these models. If some of the competing models are nested by other competing models, then the smaller models will not be included in the model comparison tests. This is reasonable since if model 1 outperforms a larger model, it will also outperform the smaller models that are nested by the larger model. With these two types of models being eliminated from the model comparison tests, the remaining models will not nest each other and the multivariate asymptotic normality assumption on $\sqrt{T}(\hat{\delta} - \delta)$ can be justified.

Finally, if model 1 is nested by some competing models, one should separate the set of competing models into two subsets. The first subset will include competing models that nest model 1. To test whether model 1 performs as well as the models in this subset, one can construct a model M that contains all the distinct factors in this subset. It can be easily verified that model 1 performs as well as the models in this subset if and only if $\rho_1^2 = \rho_M^2$. In this case, a test of $H_0 : \rho_1^2 = \rho_M^2$ can be simply performed using the model comparison tests for nested models described earlier. The second subset includes competing models that do not nest model 1. For this second subset, we can use the non-nested multiple model comparison tests as before. If we perform each test at a significance level of $\alpha/2$ and accept the null hypothesis if we fail to reject in both tests, then by the Bonferroni inequality, the size of the joint test is less than or equal to α.

9 Conclusion

In this review article, we provide an up-to-date summary of the asymptotic results related to the two-pass CSR methodology, with special emphasis on the role played by model misspecification in estimating risk premia and in comparing the performance of competing models. We also point out some pitfalls with certain popular usages of this methodology that could lead to erroneous conclusions.

There are some issues related to the two-pass CSR methodology that require further investigation. At the top of the list are the finite sample properties of the risk premium and cross-sectional R^2 estimates. At the current stage, we have little understanding of the finite sample biases of these estimates and, as a result, no good way to correct them. This is a serious concern especially when the number of assets is large relative to the length of the time series. An important related issue is how to implement the two-pass CSR methodology when the number of assets is large. In this respect, the standard practice is to simply run an OLS CSR since the GLS CSR becomes infeasible. However, in this scenario, relying on asymptotic results may not be entirely appropriate. Alternatively, one could form portfolios and use the potentially more efficient GLS CSR. How many portfolios should we consider and how should we form them are certainly open questions that we hope future research will address.

While most of the econometric results in this review article are relatively easy to program, some of them require specialized subroutines that may be time consuming to develop. To facilitate this task, a set of Matlab programs is available from the authors upon request.

References

1. Black, F, Jensen MC, Scholes M (1972) The Capital Asset Pricing Model: Some Empirical Findings. In Jensen MC (ed) Studies in the Theory of Capital Markets. Praeger New York
2. Fama, EF, MacBeth JD (1973) Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 71:607–636
3. Shanken J (1996) Statistical Methods in Tests of Portfolio Efficiency: A Synthesis. In: Maddala GS, Rao CR (eds) Handbook of Statistics, Volume 14, 693–711. Elsevier Science BV, Amsterdam
4. Jagannathan R, Skoulakis G, Wang R (2009) The Analysis of the Cross Section of Security Returns. In: Aït-Sahalia Y, Hansen L (eds) Handbook of Financial Econometrics, Volume 2, 73–134. Elsevier Science BV, Amsterdam
5. Chan KC, Chen NF (1988) An Unconditional Asset-Pricing Test and the Role of Firm Size as an Instrumental Variable for Risk. Journal of Finance 43:309–325
6. Jagannathan R, Wang Z (1996) The Conditional CAPM and the Cross-Section of Expected Returns. Journal of Finance 51:3–53
7. Lewellen JW, Nagel S, Shanken J (2010) A Skeptical Appraisal of Asset-Pricing Tests. Journal of Financial Economics 96:175–194
8. Newey WK, West KD (1987) A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55:703–708
9. Shanken J (1992) On the Estimation of Beta-Pricing Models. Review of Financial Studies 5:1–33
10. Jagannathan R, Wang Z (1998) An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression. Journal of Finance 53:1285–1309

11. Ahn S, Gadarowski C, Perez MF (2009) Two-Pass Cross-Sectional Regression of Factor Pricing Models: Minimum Distance Approach. Working paper, Wilfrid Laurier University
12. Kan R, Zhou G (2004) Hansen-Jagannathan Distance: Geometry and Exact Distribution. Working paper, University of Toronto
13. Kan R, Robotti C, Shanken J (2010) Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. Working paper, University of Toronto
14. Hou K, Kimmel R (2006) On the Estimation of Risk Premia in Linear Factor Models. Working Paper, Ohio State University
15. Shanken J, Zhou G (2007) Estimating and Testing Beta Pricing Models: Alternative Methods and their Performance in Simulations. Journal of Financial Economics 84:40–86
16. Kandel S, Stambaugh RF (1995) Portfolio Inefficiency and the Cross-Section of Expected Returns. Journal of Finance 50:157–184
17. Gospodinov N, Kan R, Robotti C (2010) Further Results on the Limiting Distribution of GMM Sample Moment Conditions. Working Paper, Federal Reserve Bank of Atlanta
18. Imhof JP (1961) Computing the Distribution of Quadratic Forms in Normal Variables. Biometrika 48:419–426
19. Lu ZH, King ML (2002) Improving the Numerical Technique for Computing the Accumulated Distribution of a Quadratic Form in Normal Variables. Econometric Reviews 21:149–165
20. Davies RB (1980) Algorithm AS 155: The Distribution of a Linear Combination of χ^2 Random Variables. Applied Statistics 29:323–333
21. Shanken J (1985) Multivariate Tests of the Zero-Beta CAPM. Journal of Financial Economics 14:327–348
22. Cochrane, JH (2005) Asset Pricing. Princeton University Press, Princeton
23. Chen NF, Roll R, Ross S (1986) Economic Forces and the Stock Market. Journal of Business 59:383–404
24. Kan R, Robotti C (2010) On the Estimation of Asset Pricing Models Using Univariate Betas. Economics Letters, forthcoming
25. Grauer RR, Janmaat JJ (2009) On the Power of Cross-Sectional and Multivariate Tests of the CAPM. Journal of Banking and Finance 33:775–787
26. Kan R, Zhang C (1999) Two-Pass Tests of Asset Pricing Models with Useless Factors. Journal of Finance 54:203–235
27. Kleibergen F (2009) Tests of Risk Premia in Linear Factor Models. Journal of Econometrics 149:149–173
28. Cragg JG, Donald SG (1997) Inferring the Rank of a Matrix. Journal of Econometrics 76:223–250
29. Gospodinov N, Kan R, Robotti C (2010) A Simplified Rank Restriction Test. Working Paper, Federal Reserve Bank of Atlanta
30. Vuong QH (1989) Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica 57:307–333
31. Rivers D, Vuong QH (2002) Model Selection Tests for Nonlinear Dynamic Models. Econometrics Journal 5:1–39
32. Golden RM (2003) Discrepancy Risk Model Selection Test Theory for Comparing Possibly Misspecified or Nonnested Models. Psychometrika 68:229–249
33. Bartholomew DJ (1961) A Test of Homogeneity of Means under Restricted Alternatives. Journal of the Royal Statistical Society 23: 239–281

34. Kudo A (1963) A Multivariate Analogue of the One-Sided Test. Biometrika 50:403–418
35. Perlman MD (1969) One-Sided Testing Problems in Multivariate Analyses. Annals of Mathematical Statistics 40:549–567
36. Gourieroux C, Holly A, Monfort A (1982) Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. Econometrica 50:63–80
37. Wolak FA (1987) An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model. Journal of the American Statistical Association 82:782–793
38. Wolak FA (1989) Testing Inequality Constraints in Linear Econometric Models. Journal of Econometrics 31:205–235
39. White HL (2000) A Reality Check For Data Snooping. Econometrica 68:1097–1127
40. Chen X, Ludvigson SC (2009) Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models. Journal of Applied Econometrics 24:1057–1093
41. Childs DR (1967) Reduction of the Multivariate Normal Integral to Characteristic Form. Biometrika 54:293–300
42. Sun HJ (1988a) A General Reduction Method for n-Variate Normal Orthant Probability. Communications in Statistics – Theory and Methods 11:3913–3921
43. Sun HJ (1988b) A Fortran Subroutine for Computing Normal Orthant Probabilities of Dimensions up to Nine. Communications in Statistics – Simulation and Computation 17:1097–1111
44. Fama, EF, French KR (1993) Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33:3–56