

A Note on the Estimation of Asset Pricing
Models Using Simple Regression Betas

Raymond Kan and Cesare Robotti

Working Paper 2009-12
March 2009

A Note on the Estimation of Asset Pricing Models Using Simple Regression Betas

Raymond Kan and Cesare Robotti

Working Paper 2009-12

March 2009

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 tool for estimating and testing beta asset pricing models. In this paper, we focus on the case in which simple regression betas are used as regressors in the second-pass CSR. Under general distributional assumptions, we derive asymptotic standard errors of the risk premia estimates that are robust to model misspecification. When testing whether the beta risk of a given factor is priced, our misspecification robust standard error and the Jagannathan and Wang (1998) standard error (which is derived under the correctly specified model) can lead to different conclusions.

JEL classification: G12

Key words: two-pass cross-sectional regressions, risk premia, model misspecification, simple regression betas, multivariate betas

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. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, or Cesare Robotti (contact author), Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, 404-498-8543, 404-498-8810 (fax), cesare.robotti@atl.frb.org.

Federal Reserve Bank of Atlanta working papers, including revised versions, are available on the Atlanta Fed's Web site at www.frbatlanta.org. Click "Publications" and then "Working Papers." Use the WebScriber Service (at www.frbatlanta.org) to receive e-mail notifications about new papers.

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, its 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 estimated betas obtained 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 the risk premia.

Usually, asset betas are defined as the OLS slope coefficients in the multiple regression of asset returns on factors and are referred to as *multiple regression* or *multivariate* betas. However, there is a potential issue with the use of multiple regression betas: unless the factors are uncorrelated, the beta of an asset with respect to a particular factor in general depends on what other factors are included in the first-pass time series OLS regression. As a result, a factor can possess additional explanatory power for the cross-sectional differences in expected returns but yet have a zero risk premium in a model with multiple factors. This makes it problematic to use the risk premium of a factor for the purpose of model selection. To overcome this problem, Chen, Roll, and Ross (1986) and Jagannathan and Wang (1996, 1998) define the beta of an asset with respect to a given factor as the OLS slope coefficient in a simple regression of its return on the factor. These betas are normally referred to as *simple regression* or *univariate* betas. In models with simple regression betas, adding or deleting a factor in a model will not change the values of the betas corresponding to the other factors and selecting models based on risk premia becomes more meaningful.¹

Jagannathan and Wang (1998) present an asymptotic theory for models with simple regression betas.² However, their asymptotic results rest on the assumption that expected returns are exactly

¹Alternatively, as emphasized by Cochrane (2005) and Kan, Robotti, and Shanken (2009), one can use covariance risks instead of beta risks in the CSR.

²An asymptotic theory for models with simple regression betas is indeed necessary. Although each of the two beta types (as well as the risk premia) is a linear transformation of the other, Jagannathan and Wang (1998) show that the standard errors obtained by applying the transformation to the asymptotic variance for models with multiple

linear in the betas, i.e., the beta pricing model is correctly specified. It is difficult to justify this assumption when estimating the zero-beta rate and risk premia parameters for many different models because some (if not all) of the models are bound to be misspecified. Since asset pricing models are, at best, approximations of reality, it is inevitable that we will often, knowingly or unknowingly (since asset pricing tests have limited power), estimate an expected return relation that departs from exact linearity in the betas. The main contribution of this paper is to propose, under general distributional assumptions, misspecification robust asymptotic standard errors of the estimated zero-beta rate and risk premia for models with simple regression betas. A nice feature of our robust standard errors is that they are applicable to both correctly specified and misspecified models. Our analysis generalizes and simplifies the results of Jagannathan and Wang (1998) that are derived under the assumption that the beta pricing model is correctly specified. 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 asymptotic 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 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.

After describing the econometric methodology, we provide an empirical example to demonstrate the relevance of our results. We focus on the conditional capital asset pricing model (CAPM) of Lettau and Ludvigson (2001), where the scaling variable is the lagged consumption-wealth ratio (CAY). We examine whether model misspecification substantially affects the standard errors of the risk premia estimates that are based on simple regression betas. When using the Jagannathan and Wang (1998) standard error, we find that the scaled market factor (i.e., the market return multiplied by CAY) is significantly priced in the OLS and weighted least squares (WLS) cases. However, using our misspecification robust standard errors, the t -ratios becomes substantially smaller and lead us to conclude that the scaled market factor is not priced.

The rest of the paper is organized as follows. Section 1 presents an asymptotic analysis of the zero-beta rate and risk premia estimates for models with simple regression betas under potential

regression betas would be incorrect for models with simple regression betas.

model misspecification. Section 2 presents a brief empirical example. The final section summarizes our findings and the appendix contains proofs of all propositions.

1. Asymptotic Analysis under Potentially Misspecified Models

1.1 Population Pricing Errors and Risk Premia

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

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

$$V = \text{Var}[Y] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \quad (2)$$

where V is assumed to be positive definite.

To prepare for our presentation of the CSR with simple regression betas, we first describe the CSR that makes use of multiple regression betas. The multiple regression betas of the N test assets with respect to the K factors are defined as $\beta = V_{21}V_{11}^{-1}$. In addition, we denote the covariance matrix of the residuals of the N test assets by $\Sigma = V_{22} - V_{21}V_{11}^{-1}V_{12}$.

When the proposed K -factor beta pricing model is correctly specified, the expected returns of the test assets are exactly linear in β . As a result, the pricing errors, e , of the N test assets are

$$e \equiv \mu_2 - X\gamma = 0_N, \quad (3)$$

where $X = [1_N, \beta]$ is assumed to be of full column rank, 0_N is an N -vector of zeros, 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 However, when the proposed model is misspecified, the pricing error vector of the model will be nonzero regardless of the choice of γ . In that case, it makes sense to choose γ to minimize some form of aggregate pricing errors. Denoting by W an $N \times N$ symmetric positive definite matrix, we define the (pseudo) zero-beta rate and risk premia as the choice of γ that minimizes the quadratic form of pricing errors:

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

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

Under this choice of γ , the pricing errors of the N assets are given by

$$e_W = \mu_2 - X\gamma_W. \quad (5)$$

As suggested by Chen, Roll, and Ross (1986) and Jagannathan and Wang (1996, 1998), there are cases in which it is preferable to use simple regression betas instead of multiple regression betas as regressors in the second-pass CSR. The simple regression betas of the N portfolios with respect to the K factors are defined as $\beta^* = V_{21}D^{-1}$, where $D = \text{Diag}(V_{11})$ is a diagonal matrix of the diagonal elements of V_{11} . Letting $X^* = [1_N, \beta^*]$, the zero-beta rate and the risk premia associated with the simple regression betas are defined as

$$\gamma_W^* \equiv \begin{bmatrix} \gamma_{W,0}^* \\ \gamma_{W,1}^* \end{bmatrix} = \text{argmin}_{\gamma^*} (\mu_2 - X^*\gamma^*)'W(\mu_2 - X^*\gamma^*) = (X^{*'}WX^*)^{-1}X^{*'}W\mu_2. \quad (6)$$

Compared with the usual γ_W in (4), we can easily see that

$$\gamma_{W,0}^* = \gamma_{W,0}, \quad \gamma_{W,1}^* = DV_{11}^{-1}\gamma_{W,1}. \quad (7)$$

In addition, it is easy to see that

$$e_W^* = \mu_2 - X^*\gamma_W^* = \mu_2 - X\gamma_W = e_W, \quad (8)$$

so that the pricing errors from this alternative second-pass CSR are the same as those in (5).

It should be emphasized that unless the model is correctly specified, γ_W^* and e_W 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),⁴ and $W = \Sigma_d^{-1}$ (WLS CSR), where $\Sigma_d = \text{Diag}(\Sigma)$. In our subsequent analysis, the choice of W is often clear from the context, and we suppress the subscript W from γ_W^* and e_W when there is no source of confusion.

1.2 Sample Estimates of Pricing Errors and Risk Premia

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 net returns on the 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

⁴For the GLS CSR, γ_W^* stays the same whether we use V_{22}^{-1} or Σ^{-1} as the weighting matrix.

and denote the sample moments of Y_t by

$$\hat{\mu} \equiv \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \frac{1}{T} \sum_{t=1}^T Y_t, \quad (9)$$

$$\hat{V} \equiv \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \quad (10)$$

The estimated simple regression betas are given by $\hat{\beta}^* = \hat{V}_{21} \hat{D}^{-1}$. When the weighting matrix W is known (say OLS CSR), we can estimate γ_W^* in (6) by

$$\hat{\gamma}^* = (\hat{X}^{*'} W \hat{X}^*)^{-1} \hat{X}^{*'} W \hat{\mu}_2, \quad (11)$$

where $\hat{X}^* = [1_N, \hat{\beta}^*]$.

In cases like the GLS and WLS CSRs, the weighting matrix W involves unknown parameters. Therefore, we need to use a consistent estimate of W , say \hat{W} , as the weighting matrix. For example, we use $\hat{W} = \hat{V}_{22}^{-1}$ for the GLS CSR and $\hat{W} = \hat{\Sigma}_d^{-1}$ for the WLS CSR, where $\hat{\Sigma}_d$ is a diagonal matrix of the diagonal elements of $\hat{\Sigma} = \hat{V}_{22} - \hat{V}_{21} \hat{V}_{11}^{-1} \hat{V}_{12}$. When \hat{W} is used as the weighting matrix in the second-pass CSR, the estimator of γ_W^* is given by

$$\hat{\gamma}^* = (\hat{X}^{*'} \hat{W} \hat{X}^*)^{-1} \hat{X}^{*'} \hat{W} \hat{\mu}_2. \quad (12)$$

Accordingly, the sample pricing errors of the N test assets are given by

$$\hat{e} = \hat{\mu}_2 - \hat{X}^* \hat{\gamma}^*. \quad (13)$$

1.3 Asymptotic Distribution of $\hat{\gamma}^*$ under Potentially Misspecified Models

When deriving the asymptotic distribution of $\hat{\gamma}^*$, Jagannathan and Wang (1998) in their Theorem 7 assume that the model is correctly specified. However, it is rather difficult to justify the use of asymptotic results under correctly specified models for all cases, especially when many models are estimated and some of them are rejected by the data. Recently, three papers have started to investigate the asymptotic distribution of $\hat{\gamma}$ under potentially misspecified models. Under the i.i.d. normality assumption, Hou and Kimmel (2006) derive the asymptotic distribution of $\hat{\gamma}$ for the case of GLS CSR with a known value of γ_0 , and Shanken and Zhou (2007) present the asymptotic results for the OLS, GLS, and WLS cases with γ_0 unknown. Kan, Robotti, and Shanken (2009)

relax the normality assumption of Shanken and Zhou (2007) and present the asymptotic distribution of $\hat{\gamma}$ under general distributional assumptions. However, all three papers only deal with CSRs with multiple regression betas and their results cannot be used for CSRs with simple regression betas. In order to fill this gap in the literature, the following proposition presents the asymptotic distribution of $\hat{\gamma}^*$ under potentially misspecified models.

Proposition 1. *Under a potentially misspecified model, the asymptotic distribution of $\hat{\gamma}^*$ is given by*

$$\sqrt{T}(\hat{\gamma}^* - \gamma^*) \overset{A}{\underset{\sim}{\sim}} N(0_{K+1}, V(\hat{\gamma}^*)), \quad (14)$$

where

$$V(\hat{\gamma}^*) = \sum_{j=-\infty}^{\infty} E[h_t h'_{t+j}]. \quad (15)$$

To simplify the h_t expressions for the different CSRs, we define $H^* = (X^{*'}WX^*)^{-1}$, $A^* = H^*X^{*'}W$, $\gamma_t^* = A^*R_t$, $z_t^* = [0, (f_t - \mu_1)'D^{-1}]'$, $D_t = \text{Diag}((f_t - \mu_1)(f_t - \mu_1)')$, $G_t^* = [\beta^*D_t - (R_t - \mu_2)(f_t - \mu_1)']$, and $u_t = e'W(R_t - \mu_2)$, where W equals V_{22}^{-1} for the GLS case and Σ_d^{-1} for the WLS case.

(1) For the known weighting matrix W case,

$$h_t = (\gamma_t^* - \gamma^*) + A^*G_t^*D^{-1}\gamma_1^* + H^*z_t^*u_t. \quad (16)$$

(2) For the GLS case,

$$h_t = (\gamma_t^* - \gamma^*) + A^*G_t^*D^{-1}\gamma_1^* + H^*z_t^*u_t - (\gamma_t^* - \gamma^*)u_t. \quad (17)$$

(3) For the WLS case,

$$h_t = (\gamma_t^* - \gamma^*) + A^*G_t^*D^{-1}\gamma_1^* + H^*z_t^*u_t - A^*\Psi_t\Sigma_d^{-1}e, \quad (18)$$

where $\Psi_t = \text{Diag}(\epsilon_t\epsilon_t')$ and $\epsilon_t = R_t - \mu_2 - \beta(f_t - \mu_1)$.

When the model is correctly specified, we have

$$h_t = (\gamma_t^* - \gamma^*) + A^*G_t^*D^{-1}\gamma_1^*. \quad (19)$$

It is easy to show that our expressions in Equations (15) and (19) coincide with the expression given by Jagannathan and Wang (1998) in their Theorem 7. However, our expression for $V(\hat{\gamma}^*)$ under the correctly specified model represents a substantial simplification over the one provided by Jagannathan and Wang (1998). In addition, from our Proposition 1 it is immediately clear how to construct a consistent estimator of $V(\hat{\gamma}^*)$. For example, in the known W case, this can be accomplished by replacing h_t with

$$\hat{h}_t = (\hat{\gamma}_t^* - \hat{\gamma}^*) + \hat{A}^* \hat{G}_t^* \hat{D}^{-1} \hat{\gamma}_1^* + \hat{H}^* \hat{z}_t^* \hat{u}_t, \quad (20)$$

where $\hat{H}^* = (\hat{X}^{*'} W \hat{X}^*)^{-1}$, $\hat{A}^* = \hat{H}^* \hat{X}^{*'} W$, $\hat{\gamma}_t^* = \hat{A}^* R_t$, $\hat{u}_t = \hat{e}' W (R_t - \hat{\mu}_2)$, $\hat{D} = \text{Diag}(\hat{V}_{11})$, $\hat{D}_t = \text{Diag}[(f_t - \hat{\mu}_1)(f_t - \hat{\mu}_1)']$, $\hat{G}_t^* = [\hat{\beta}^* \hat{D}_t - (R_t - \hat{\mu}_2)(f_t - \hat{\mu}_1)']$, and $\hat{z}_t^* = [0, (f_t - \hat{\mu}_1)' \hat{D}^{-1}]'$. Similarly, one needs to replace the population quantities in Equations (17)–(18) with their sample counterparts in order to obtain a consistent estimator of $V(\hat{\gamma}^*)$ for the GLS and WLS cases. 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 $\hat{V}(\hat{\gamma}^*) = \frac{1}{T} \sum_{t=1}^T \hat{h}_t \hat{h}_t'$. 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 Equation (16) in Proposition 1 reveals that there are three sources that contribute to the asymptotic variance of $\hat{\gamma}^*$. The first term, $\gamma_t^* - \gamma^*$, measures the asymptotic variance of $\hat{\gamma}^*$ when the *true* betas are used in the CSR. For example, when 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 of computing standard errors for the risk premia estimates. However, since $\hat{\beta}^*$ is used in place of β^* in the second-pass CSR, there is an errors-in-variables (EIV) problem. The second term, $A^* G_t^* D^{-1} \gamma_1^*$, is the EIV adjustment term that accounts for the estimation error in $\hat{\beta}^*$. These two terms together give us the $V(\hat{\gamma}^*)$ under the correctly specified model. When the model is misspecified ($e \neq 0_N$), there is a third term $H^* z_t^* u_t$, which we call the misspecification adjustment term, that contributes to the asymptotic variance of $\hat{\gamma}^*$. This term has been ignored by Jagannathan and Wang (1998) and other researchers when computing standard errors for $\hat{\gamma}^*$.⁵

⁵For the estimated GLS and WLS cases, the misspecification adjustment term contains the additional quantities $-(\gamma_t^* - \gamma^*) u_t$ and $-A^* \Psi_t \Sigma_d^{-1} e$, respectively. These additional terms are due to the estimation error in \hat{W} . Under the correctly specified model, it can be readily shown that the use of \hat{W} instead of W does not alter the asymptotic distribution of $\hat{\gamma}^*$ (the proof of this result is available upon request), but this is not case when the model is misspecified.

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. *Suppose that factors and returns are i.i.d. multivariate elliptically distributed with kurtosis parameter κ .⁶ Let $\tilde{D} = \begin{bmatrix} 0 & 0'_K \\ 0_K & D^{-1}V_{11}D^{-1} \end{bmatrix}$ and \odot denote the Hadamard product. Define*

$$\Upsilon_w = [1 + (1 + \kappa)(\gamma_1^{*'}D^{-1}V_{11}D^{-1}\gamma_1^*)]A^*V_{22}A^{*'} + (1 + \kappa) \times \begin{bmatrix} 0 & 0'_K \\ 0_K & 2(D^{-1}\gamma_1^*\gamma_1^{*'}D^{-1}) \odot V_{11} \odot V_{11} - 4\text{Diag}(\gamma_1^*\gamma_1^{*'}D^{-1}V_{11}) + \gamma_1^*\gamma_1^{*'} \end{bmatrix}, \quad (21)$$

$$\Upsilon_{w1} = -(1 + \kappa)A^*V_{22}We[0, \gamma_1^{*'}D^{-1}V_{11}D^{-1}]H^*, \quad (22)$$

where $W = V_{22}^{-1}$ for the GLS case and $W = \Sigma_d^{-1}$ for the WLS case. The asymptotic variance of $\hat{\gamma}^*$ is given by

$$V(\hat{\gamma}^*) = \Upsilon_w + \Upsilon_{w1} + \Upsilon'_{w1} + \Upsilon_{w2}, \quad (23)$$

where Υ_w is the asymptotic variance of $\hat{\gamma}^*$ when the model is correctly specified, and $\Upsilon_{w1} + \Upsilon'_{w1} + \Upsilon_{w2}$ is the adjustment term due to model misspecification.

(1) For the known weighting matrix W case,

$$\Upsilon_{w2} = (1 + \kappa)(e'WV_{22}We)H^*\tilde{D}H^*. \quad (24)$$

(2) For the GLS case, Υ_{w1} vanishes and

$$\Upsilon_{w2} = (1 + \kappa)(e'V_{22}^{-1}e)(\tilde{H}\tilde{D}\tilde{H} + \tilde{H}), \quad (25)$$

where $\tilde{H} = (X^{*'}\Sigma^{-1}X^*)^{-1}$.

(3) For the WLS case,

$$\Upsilon_{w2} = (1 + \kappa) \left[(e'\Sigma_d^{-1}V_{22}\Sigma_d^{-1}e)H^*\tilde{D}H^* + 2A^*\Phi A^{*'} \right], \quad (26)$$

where Φ is an $(N \times N)$ matrix with its (i, j) -th element equal to $\rho_{ij}^2 e_i e_j$ and $\rho_{ij} = \text{Corr}[\epsilon_{it}, \epsilon_{jt}]$.

⁶The kurtosis parameter for an elliptical distribution is defined as $\kappa = \mu_4/(3\sigma^4) - 1$, where σ^2 and μ_4 are the second and fourth central moments of the elliptical distribution, respectively.

For the known weighting matrix W and the WLS cases, the misspecification adjustment term $\Upsilon_{w1} + \Upsilon'_{w1} + \Upsilon_{w2}$ is not necessarily positive semidefinite. However, for the GLS case, the misspecification adjustment term $\Upsilon_{w2} = (1 + \kappa)(e'V_{22}^{-1}e)(\tilde{H}\tilde{D}\tilde{H} + \tilde{H})$ is positive definite. This is because $1 + \kappa > 0$ (see Bentler and Berkane (1986)) and $\tilde{H}\tilde{D}\tilde{H} + \tilde{H}$ is positive definite. Note that for the GLS case, the misspecification adjustment term is positively related to the aggregate pricing errors $e'V_{22}^{-1}e$ and the kurtosis parameter κ . In addition, it is worth noting that the misspecification adjustment term in the GLS case crucially depends on the correlation between factors and returns. To show this, note that the misspecification adjustment term for $V(\hat{\gamma}_1^*)$ is

$$(1 + \kappa)(e'V_{22}^{-1}e)H_{22}^*D^{-1}[V_{11} - V_{12}V_{22}^{-1}V_{21} + V_{12}V_{22}^{-1}1_N(1'_N V_{22}^{-1}1_N)^{-1}1'_N V_{22}^{-1}V_{21}]D^{-1}H_{22}^*, \quad (27)$$

where H_{22}^* is the lower right $K \times K$ submatrix of H^* . Note that the term $V_{11} - V_{12}V_{22}^{-1}V_{21}$ is 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), the impact of this term and hence of the misspecification adjustment on the asymptotic variance of $\hat{\gamma}_1^*$ can be very large.

2. An Empirical Example

We apply our methodology to the same data and asset pricing models considered by Lettau and Ludvigson (2001).⁷ In the interest of brevity, we report results for only two beta pricing models: the Fama-French (1993) three-factor model (FF3) and the scaled CAPM.⁸ Lettau and Ludvigson (2001) show that scaling the fundamental factors of a given beta pricing model by the consumption-wealth ratio (CAY) helps to explain the cross-section of average returns.

The return data consist of the quarterly net returns on the 25 Fama-French size and book-to-market ranked portfolios from 1963 Q4 to 1998 Q3 (140 quarterly observations). The FF3 implies a cross-sectional specification of the form

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

where vw is the net return on the value-weighted stock market index (NYSE-AMEX-NASDAQ) from the Center for Research in Security Prices (CRSP), smb is the return difference between

⁷We thank Martin Lettau for making his data available to us.

⁸The full set of estimation results is available upon request.

portfolios of small and large stocks and hml is the return difference between portfolios of high and low book-to-market ratios.

The scaled CAPM is obtained by scaling the constant term and the vw factor described above by a constant and lagged CAY. 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.

When various asset pricing models are estimated, we argued in the introduction that it is not very sensible to compute the standard error of $\hat{\gamma}^*$ by assuming that all models are correctly specified. Therefore, in our empirical example, we mainly examine whether model misspecification substantially affects the standard error of $\hat{\gamma}^*$.

In Table 1, we report $\hat{\gamma}^*$ and associated t -ratios of the FF3 and the scaled CAPM under correctly specified and potentially misspecified models.⁹ For the t -ratios under the assumption of correctly specified models, we first report the Fama and MacBeth (1973) ones followed by the Jagannathan and Wang (1998) ones which account for estimation error in the betas. Last, are the t -ratios under potentially misspecified models, based on our results in Proposition 1. The various ratios are identified by subscripts fm , vw , and hml , respectively.¹⁰

Table 1 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 non-traded factors such as the scaled market return and the lagged state variable CAY. Consider, for example, the OLS results for the FF3 model in Panel A. The three t -ratios of $\hat{\gamma}_{vw}^*$, $\hat{\gamma}_{smb}^*$, and $\hat{\gamma}_{hml}^*$ are similar as the factors are all mimicked well by the returns on the test assets. The GLS and WLS results in Panels B and C deliver a similar message.

However, when we consider the scaled CAPM, the picture changes substantially. For example, consider the scaled market factor in Panel A. Under the correctly specified model, t -ratio $_{fm}$ is 3.63

⁹The 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 very minor impact on the results.

¹⁰A set of Matlab programs to implement our formulae is available upon request.

and $t\text{-ratio}_{jw}$ is 2.70. But, once we account for potential model misspecification, the t -ratio goes down to 1.27. Even for the GLS and WLS cases, the standard errors of $\hat{\gamma}_{vw.cay}$ increase substantially when we incorporate potential model misspecification.

In summary, we find that for non-traded factors all of the t -ratios under potentially misspecified models are smaller (in absolute value) than the Fama-MacBeth (1973) and Jagannathan and Wang (1998) t -ratios. Using our robust standard errors, the estimates of the risk premia on the scaled factors are never statistically significant at the 5% level.

3. Conclusion

In this paper, we present an asymptotic analysis of the two-pass cross-sectional regression methodology that makes use of simple regression betas instead of multiple regression betas. We contribute to the existing literature by proposing a simple method for computing standard errors for the estimated zero-beta rate and risk premia that are robust to model misspecification. A nice feature of our misspecification robust standard errors is that they can be used whether the model is correctly specified or not. When returns and factors are multivariate elliptically distributed, we are able to show analytically that with GLS cross-sectional regressions the standard errors under misspecified models are 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. Our empirical example suggests that ignoring potential model misspecification can lead to the incorrect conclusion that a given risk factor is priced.

Appendix

Proof of Proposition 1: In the following, we provide the proof of Proposition 1 for the estimated GLS and WLS cases as the proof for the known weighting matrix W case 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

$$\varphi = \begin{bmatrix} \mu \\ \text{vec}(V) \end{bmatrix}, \quad \hat{\varphi} = \begin{bmatrix} \hat{\mu} \\ \text{vec}(\hat{V}) \end{bmatrix}. \quad (\text{A1})$$

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

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

Since this is an exactly identified system of moment conditions, it is straightforward to verify that under the assumptions that Y_t is stationary and ergodic with finite fourth moments, we have¹¹

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

where

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

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

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

For both the GLS and the WLS cases, it is straightforward to obtain

$$\frac{\partial \gamma^*}{\partial \mu'_1} = 0_{(K+1) \times K}, \quad \frac{\partial \gamma^*}{\partial \mu'_2} = A^*. \quad (\text{A6})$$

In order to obtain the derivative of $\gamma^* = H^* X^{*'} V_{22}^{-1} \mu_2$ with respect to $\text{vec}(V)$ for the GLS case, we need to first obtain $\partial x^* / \partial \text{vec}(V)'$, where $x^* = \text{vec}(X^*)$. We begin by writing write V_{11} and V_{21} as

$$V_{11} = [I_K, 0_{K \times N}] V [I_K, 0_{K \times N}]', \quad V_{21} = [0_{N \times K}, I_N] V [I_K, 0_{K \times N}]' \quad (\text{A7})$$

¹¹Note that S_0 is a singular matrix as \hat{V} is symmetric, so there are redundant elements in $\hat{\varphi}$. We could have written $\hat{\varphi}$ as $[\hat{\mu}', \text{vech}(\hat{V})']'$, but the results are the same under both specifications.

to obtain

$$\frac{\partial \text{vec}(V_{11})}{\partial \text{vec}(V)'} = [I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}], \quad (\text{A8})$$

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

Let Θ_1 be a $K^2 \times K^2$ matrix such that $\text{vec}(D) = \Theta_1 \text{vec}(V_{11})$,¹² we can use the chain rule to derive the derivative of $\text{vec}(D^{-1})$ with respect to $\text{vec}(V)$ as

$$\begin{aligned} \frac{\partial \text{vec}(D^{-1})}{\partial \text{vec}(V)'} &= \frac{\partial \text{vec}(D^{-1})}{\partial \text{vec}(D)'} \frac{\partial \text{vec}(D)}{\partial \text{vec}(V_{11})'} \frac{\partial \text{vec}(V_{11})}{\partial \text{vec}(V)'} \\ &= -(D^{-1} \otimes D^{-1}) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]). \end{aligned} \quad (\text{A10})$$

Then using the product rule, we obtain

$$\begin{aligned} \frac{\partial \text{vec}(\beta^*)}{\partial \text{vec}(V)'} &= (D^{-1} \otimes I_N) \frac{\partial \text{vec}(V_{21})}{\partial \text{vec}(V)'} + (I_K \otimes V_{21}) \frac{\partial \text{vec}(D^{-1})}{\partial \text{vec}(V)'} \\ &= [D^{-1}, 0_{K \times N}] \otimes [0_{N \times K}, I_N] - (D^{-1} \otimes \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]) \end{aligned} \quad (\text{A11})$$

Finally, using the identity

$$\frac{\partial x^*}{\partial \text{vec}(\beta^*)'} = [0_K, I_K]' \otimes I_N, \quad (\text{A12})$$

we obtain

$$\begin{aligned} \frac{\partial x^*}{\partial \text{vec}(V)'} &= \frac{\partial x^*}{\partial \text{vec}(\beta^*)'} \frac{\partial \text{vec}(\beta^*)}{\partial \text{vec}(V)'} \\ &= [0_K, I_K]' [D^{-1}, 0_{K \times N}] \otimes [0_{N \times K}, I_N] \\ &\quad - ([0_K, I_K]' D^{-1} \otimes \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]). \end{aligned} \quad (\text{A13})$$

With this identity, we now define $K_{m,n}$ as 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. Then using the product rule, we obtain

$$\frac{\partial \gamma^*}{\partial \text{vec}(V)'} = (\mu_2' V_{22}^{-1} X^* \otimes I_{K+1}) \frac{\partial \text{vec}(H^*)}{\partial \text{vec}(V)'} + (\mu_2' V_{22}^{-1} \otimes H^*) \frac{\partial \text{vec}(X^{*'})}{\partial \text{vec}(V)'} + (\mu_2' \otimes H^* X^{*'}) \frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'} \quad (\text{A14})$$

The second term is given by

$$(\mu_2' V_{22}^{-1} \otimes H^*) \frac{\partial \text{vec}(X^{*'})}{\partial \text{vec}(V)'} \quad (\text{A14})$$

¹²Specifically, Θ_1 is a matrix with its (i, i) -th element equals to one, where $i = 1, 1+1(K+1), 1+2(K+1), \dots, 1+(K-1)(K+1)$, and zero elsewhere.

$$\begin{aligned}
&= (\mu'_2 V_{22}^{-1} \otimes H^*) K_{N, K+1} \frac{\partial x^*}{\partial \text{vec}(V)'} \\
&= [H^*[0_K, D^{-1}]', 0_{(K+1) \times N}] \otimes [0'_K, \mu'_2 V_{22}^{-1}] \\
&\quad - (H^*[0_K, D^{-1}]' \otimes \mu'_2 V_{22}^{-1} \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]). \tag{A15}
\end{aligned}$$

The third term is given by

$$(\mu'_2 \otimes H^* X^{*'}) \frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'} = -[0'_K, \mu'_2 V_{22}^{-1}] \otimes [0_{(K+1) \times K}, A^*]. \tag{A16}$$

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

$$\begin{aligned}
&(\mu'_2 V_{22}^{-1} X^* \otimes I_{K+1}) \frac{\partial \text{vec}(H^*)}{\partial \text{vec}(V)'} \\
&= (\mu'_2 V_{22}^{-1} X^* \otimes I_{K+1}) \frac{\partial \text{vec}(H^*)}{\partial \text{vec}(H^{*-1})'} \frac{\partial \text{vec}(H^{*-1})}{\partial \text{vec}(V)'} \\
&= -(\mu'_2 V_{22}^{-1} X^* \otimes I_{K+1}) (H^* \otimes H^*) \left[(X^{*'} V_{22}^{-1} \otimes I_{K+1}) \frac{\partial \text{vec}(X^{*'})}{\partial \text{vec}(V)'} + (I_{K+1} \otimes X^{*'} V_{22}^{-1}) \frac{\partial \text{vec}(X^*)}{\partial \text{vec}(V)'} \right. \\
&\quad \left. + (X^{*'} \otimes X^{*'}) \frac{\partial \text{vec}(V_{22}^{-1})}{\partial \text{vec}(V)'} \right] \\
&= -(H^*[0_K, I_K]' [D^{-1}, 0_{K \times N}] \otimes [0'_K, \gamma^{*'} X^{*'} V_{22}^{-1}]) \\
&\quad + (H^*[0_K, D^{-1}]' \otimes \gamma^{*'} X^{*'} V_{22}^{-1} \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]) \\
&\quad - (\gamma^{*'} [0_K, I_K]' [D^{-1}, 0_{K \times N}] \otimes [0_{(K+1) \times K}, A^*]) \\
&\quad + (\gamma^{*'} [0_K, D^{-1}]' \otimes A^* \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]) \\
&\quad + [0'_K, \gamma^{*'} X^{*'} V_{22}^{-1}] \otimes [0_{(K+1) \times K}, A^*]. \tag{A17}
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial \gamma^*}{\partial \text{vec}(V)'} &= [H^*[0_K, D^{-1}]', 0_{(K+1) \times N}] \otimes [0'_K, e' V_{22}^{-1}] - [\gamma_1^{*'} D^{-1}, e' V_{22}^{-1}] \otimes [0_{(K+1) \times K}, A^*] \\
&\quad + (\gamma_1^{*'} D^{-1} \otimes A^* \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]). \tag{A18}
\end{aligned}$$

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

$$V(\hat{\gamma}^*) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi) h'_{t+j}(\varphi)], \tag{A19}$$

where

$$h_t(\varphi) = \frac{\partial \gamma^*}{\partial \varphi'} r_t(\varphi)$$

$$\begin{aligned}
&= A^*(R_t - \mu_2) + \text{vec} \left([0'_K, e'V_{22}^{-1}][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} [0_K, D^{-1}]H^* \\ 0_{N \times (K+1)} \end{bmatrix} \right) \\
&\quad - \text{vec} \left([0_{(K+1) \times K}, A^*][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} D^{-1}\gamma_1^* \\ V_{22}^{-1}e \end{bmatrix} \right) \\
&\quad + (\gamma_1^{*\prime} D^{-1} \otimes A^* \beta^*) \Theta_1 \text{vec} \left([I_K, 0_{K \times N}][(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} I_K \\ 0_{N \times K} \end{bmatrix} \right) \\
&= A^*(R_t - \mu_2) + H^* z_t^* (R_t - \mu_2)' V_{22}^{-1} e - A^*(R_t - \mu_2) (f_t - \mu_1)' D^{-1} \gamma_1^* \\
&\quad - A^*(R_t - \mu_2) (R_t - \mu_2)' V_{22}^{-1} e + A^* \beta^* \gamma_1^* + A^* \beta^* D_t D^{-1} \gamma_1^* - A^* \beta^* \gamma_1^* \\
&= A^*(R_t - \mu_2) - A^* [(R_t - \mu_2) (f_t - \mu_1)' - \beta^* D_t] D^{-1} \gamma_1^* + H^* z_t^* u_t - A^*(R_t - \mu_2) u_t \\
&= (\gamma_t^* - \gamma^*) + A^* G_t^* D^{-1} \gamma_1^* + H^* z_t^* u_t - (\gamma_t^* - \gamma^*) u_t. \tag{A20}
\end{aligned}$$

We now turn to the WLS case where $W = \Sigma_d^{-1}$ and $\gamma^* = H^* X^{*\prime} \Sigma_d^{-1} \mu_2$. Using the product rule, we obtain

$$\frac{\partial \gamma^*}{\partial \text{vec}(V)'} = (\mu_2' \Sigma_d^{-1} X^* \otimes I_{K+1}) \frac{\partial \text{vec}(H^*)}{\partial \text{vec}(V)'} + (\mu_2' \Sigma_d^{-1} \otimes H^*) \frac{\partial \text{vec}(X^{*\prime})}{\partial \text{vec}(V)'} + (\mu_2' \otimes H^* X^{*\prime}) \frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(V)'}. \tag{A21}$$

The second term is given by

$$\begin{aligned}
&(\mu_2' \Sigma_d^{-1} \otimes H^*) \frac{\partial \text{vec}(X^{*\prime})}{\partial \text{vec}(V)'} \\
&= (\mu_2' \Sigma_d^{-1} \otimes H^*) K_{N, K+1} \frac{\partial x^*}{\partial \text{vec}(V)'} \\
&= [H^* [0_K, D^{-1}]', 0_{(K+1) \times N}] \otimes [0'_K, \mu_2' \Sigma_d^{-1}] \\
&\quad - (H^* [0_K, D^{-1}]' \otimes \mu_2' \Sigma_d^{-1} \beta^*) \Theta_1 ([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]). \tag{A22}
\end{aligned}$$

Let Θ be an $N^2 \times N^2$ matrix such that $\text{vec}(\Sigma_d) = \Theta \text{vec}(\Sigma)$. It follows that

$$\frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(V)'} = \frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(\Sigma_d)'} \frac{\partial \text{vec}(\Sigma_d)}{\partial \text{vec}(\Sigma)'} \frac{\partial \text{vec}(\Sigma)}{\partial \text{vec}(V)'} = -(\Sigma_d^{-1} \otimes \Sigma_d^{-1}) \Theta ([-\beta, I_N] \otimes [-\beta, I_N]) \tag{A23}$$

and the third term is given by

$$(\mu_2' \otimes H^* X^{*\prime}) \frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(V)'} = -(\mu_2' \Sigma_d^{-1} \otimes A^*) \Theta ([-\beta, I_N] \otimes [-\beta, I_N]). \tag{A24}$$

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

$$(\mu_2' \Sigma_d^{-1} X^* \otimes I_{K+1}) \frac{\partial \text{vec}(H^*)}{\partial \text{vec}(V)'}$$

$$\begin{aligned}
&= (\mu'_2 \Sigma_d^{-1} X^* \otimes I_{K+1}) \frac{\partial \text{vec}(H^*)}{\partial \text{vec}(H^{*-1})'} \frac{\partial \text{vec}(H^{*-1})}{\partial \text{vec}(V)'} \\
&= -(\mu'_2 \Sigma_d^{-1} X^* \otimes I_{K+1})(H^* \otimes H^*) \left[(X^{*'} \Sigma_d^{-1} \otimes I_{K+1}) \frac{\partial \text{vec}(X^{*'})}{\partial \text{vec}(V)'} + (I_{K+1} \otimes X^{*'} \Sigma_d^{-1}) \frac{\partial \text{vec}(X^*)}{\partial \text{vec}(V)'} \right. \\
&\quad \left. + (X^{*'} \otimes X^{*'}) \frac{\partial \text{vec}(\Sigma_d^{-1})}{\partial \text{vec}(V)'} \right] \\
&= -(H^*[0_K, I_K]'[D^{-1}, 0_{K \times N}] \otimes [0'_K, \gamma^{*'} X^{*'} \Sigma_d^{-1}]) \\
&\quad + (H^*[0_K, D^{-1}]' \otimes \gamma^{*'} X^{*'} \Sigma_d^{-1} \beta^*) \Theta_1([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]) \\
&\quad - (\gamma^{*'}[0_K, I_K]'[D^{-1}, 0_{K \times N}] \otimes [0_{(K+1) \times K}, A^*]) \\
&\quad + (\gamma^{*'}[0_K, D^{-1}]' \otimes A^* \beta^*) \Theta_1([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]) \\
&\quad + (\gamma^{*'} X^{*'} \Sigma_d^{-1} \otimes A^*) \Theta([- \beta, I_N] \otimes [- \beta, I_N]). \tag{A25}
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial \gamma^*}{\partial \text{vec}(V)'} &= [H^*[0_K, D^{-1}]', 0_{(K+1) \times N}] \otimes [0'_K, e' \Sigma_d^{-1}] - [\gamma_1^{*'} D^{-1}, 0'_N] \otimes [0_{(K+1) \times K}, A^*] \\
&\quad + (\gamma_1^{*'} D^{-1} \otimes A^* \beta^*) \Theta_1([I_K, 0_{K \times N}] \otimes [I_K, 0_{K \times N}]) \\
&\quad - (e' \Sigma_d^{-1} \otimes A^*) \Theta([- \beta, I_N] \otimes [- \beta, I_N]). \tag{A26}
\end{aligned}$$

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

$$V(\hat{\gamma}^*) = \sum_{j=-\infty}^{\infty} E[h_t(\varphi) h'_{t+j}(\varphi)], \tag{A27}$$

where

$$\begin{aligned}
h_t(\varphi) &= \frac{\partial \gamma^*}{\partial \varphi'} r_t(\varphi) \\
&= A^*(R_t - \mu_2) + \text{vec} \left([0'_K, e' \Sigma_d^{-1}] [(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} [0_K, D^{-1}] H^* \\ 0_{N \times (K+1)} \end{bmatrix} \right) \\
&\quad - \text{vec} \left([0_{(K+1) \times K}, A^*] [(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} D^{-1} \gamma_1^* \\ 0_N \end{bmatrix} \right) \\
&\quad + (\gamma_1^{*'} D^{-1} \otimes A^* \beta^*) \Theta_1 \text{vec} \left([I_K, 0_{K \times N}] [(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} I_K \\ 0_{N \times K} \end{bmatrix} \right) \\
&\quad - (e' \Sigma_d^{-1} \otimes A^*) \Theta \text{vec} \left([- \beta, I_N] [(Y_t - \mu)(Y_t - \mu)' - V] \begin{bmatrix} -\beta' \\ I_N \end{bmatrix} \right) \\
&= A^*(R_t - \mu_2) + H^* z_t^* (R_t - \mu_2)' V_{22}^{-1} e - A^*(R_t - \mu_2) (f_t - \mu_1)' D^{-1} \gamma_1^* + A^* \beta^{*'} \gamma_1^*
\end{aligned}$$

$$\begin{aligned}
& + A^* \beta^* D_t D^{-1} \gamma_1^* - A^* \beta^* \gamma_1^* - A^* \Psi_t \Sigma_d^{-1} e \\
= & A^* (R_t - \mu_2) - A^* [(R_t - \mu_2)(f_t - \mu_1)' - \beta^* D_t] D^{-1} \gamma_1^* + H^* z_t^* u_t - A^* \Psi_t \Sigma_d^{-1} e \\
= & (\gamma_t^* - \gamma^*) + A^* G_t^* D^{-1} \gamma_1^* + H^* z_t^* u_t - A^* \Psi_t \Sigma_d^{-1} e.
\end{aligned} \tag{A28}$$

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

$$h_t(\varphi) = (\gamma_t^* - \gamma^*) + A^* G_t^* D^{-1} \gamma_1^*. \tag{A29}$$

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

$$E[X_i X_j X_k] = 0, \tag{A30}$$

$$E[X_i X_j X_k X_l] = (1 + \kappa)(\sigma_{ij} \sigma_{kl} + \sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{jk}), \tag{A31}$$

where $\sigma_{ij} = \text{Cov}[X_i, X_j]$.

Starting from the known weighting matrix case, the asymptotic variance of $\hat{\gamma}^*$ is given by

$$V(\hat{\gamma}^*) = E[h_t h_t'], \tag{A32}$$

where

$$h_t = h_{1t} + h_{2t} + h_{3t}, \tag{A33}$$

with

$$h_{1t} = A^* (R_t - \mu_2), \quad h_{2t} = A^* [\beta^* D_t - (R_t - \mu_2)(f_t - \mu_1)'] D^{-1} \gamma_1^*, \quad h_{3t} = H^* z_t^* u_t. \tag{A34}$$

It is straightforward to show that the means of h_{1t} to h_{3t} are all equal to zero and

$$E[h_{1t} h_{1t}'] = A^* V_{22} A^{*'} \tag{A35}$$

In addition, h_{1t} is uncorrelated with h_{2t} and h_{3t} . For h_{2t} , we first write

$$R_t - \mu_2 = \beta(f_t - \mu_1) + \epsilon_t = \beta^* D V_{11}^{-1} (f_t - \mu_1) + \epsilon_t. \tag{A36}$$

Using the fact that $A^*\beta^* = [0_K, I_K]'$, we can write

$$h_{2t} = \begin{bmatrix} 0 \\ q_t \end{bmatrix} - A^*\epsilon_t(f_t - \mu_1)'D^{-1}\gamma_1^*, \quad (\text{A37})$$

where

$$q_t = [D_t - DV_{11}^{-1}(f_t - \mu_1)(f_t - \mu_1)']D^{-1}\gamma_1^*. \quad (\text{A38})$$

Then using the fact that ϵ_t and f_t are uncorrelated, we have

$$E[h_{2t}h_{2t}'] = \begin{bmatrix} 0 & 0'_K \\ 0_K & E[q_tq_t'] \end{bmatrix} + (1 + \kappa)(\gamma_1^{*'}D^{-1}V_{11}D^{-1}\gamma_1^*)A^*(V_{22} - V_{21}V_{11}^{-1}V_{12})A^{*'}. \quad (\text{A39})$$

Let a be a K -vector. It is easy to show that

$$E[D_taa'D_t] = (1 + \kappa)[Daa'D + 2aa' \odot V_{11} \odot V_{11}], \quad (\text{A40})$$

$$E[D_taa'(f_t - \mu_1)(f_t - \mu_1)'V_{11}^{-1}D] = (1 + \kappa)[Daa'D + 2\text{Diag}(aa'V_{11})D]. \quad (\text{A41})$$

It follows that by choosing $a = D^{-1}\gamma_1^*$, we have

$$\begin{aligned} E[q_tq_t'] &= (1 + \kappa)[2(D^{-1}\gamma_1^*\gamma_1^{*'}D^{-1}) \odot V_{11} \odot V_{11} + \gamma_1^*\gamma_1^{*'}] \\ &\quad + (1 + \kappa)[(\gamma_1^{*'}D^{-1}V_{11}D^{-1}\gamma_1^*)DV_{11}^{-1}D + 2\gamma_1^*\gamma_1^{*'}] \\ &\quad - (1 + \kappa)[4\text{Diag}(D^{-1}\gamma_1^*\gamma_1^{*'}D^{-1}V_{11})D + 2\gamma_1^*\gamma_1^{*'}] \\ &= (1 + \kappa)[2(D^{-1}\gamma_1^*\gamma_1^{*'}D^{-1}) \odot V_{11} \odot V_{11} + (\gamma_1^{*'}D^{-1}V_{11}D^{-1}\gamma_1^*)DV_{11}^{-1}D \\ &\quad - 4\text{Diag}(\gamma_1^*\gamma_1^{*'}D^{-1}V_{11}) + \gamma_1^*\gamma_1^{*'}]. \end{aligned} \quad (\text{A42})$$

The last identity follows because $\text{Diag}(D^{-1}A)D = \text{Diag}(A)$.

As a result, we have

$$E[h_{2t}h_{2t}'] = (1 + \kappa)(\gamma_1^{*'}D^{-1}V_{11}D^{-1}\gamma_1^*)A^*V_{22}A^{*'} + (1 + \kappa) \times \begin{bmatrix} 0 & 0'_K \\ 0_K & 2(D^{-1}\gamma_1^*\gamma_1^{*'}D^{-1}) \odot V_{11} \odot V_{11} - 4\text{Diag}(\gamma_1^*\gamma_1^{*'}D^{-1}V_{11}) + \gamma_1^*\gamma_1^{*'} \end{bmatrix}. \quad (\text{A43})$$

For h_{3t} , it is easy to obtain

$$E[h_{3t}h_{3t}'] = H^* \begin{bmatrix} 0 & 0'_K \\ 0_K & (1 + \kappa)(e'WV_{22}We)D^{-1}V_{11}D^{-1} \end{bmatrix} H^*. \quad (\text{A44})$$

Finally, using the fact that u_t is uncorrelated with f_t , the cross-moment between h_{2t} and h_{3t} is given by

$$E[h_{2t}h_{3t}'] = -(1 + \kappa)A^*V_{22}We[0, \gamma_1^{*'}D^{-1}V_{11}D^{-1}]H^*. \quad (\text{A45})$$

Collecting terms, we obtain

$$\begin{aligned}\Upsilon_w &= E[h_{1t}h'_{1t}] + E[h_{2t}h'_{2t}] \\ &= [1 + (1 + \kappa)(\gamma_1^{*'}D^{-1}V_{11}D^{-1}\gamma_1^*)]A^*V_{22}A^{*'} + (1 + \kappa) \times \\ &\quad \begin{bmatrix} 0 & 0'_K \\ 0_K & 2(D^{-1}\gamma_1^*\gamma_1^{*'}D^{-1}) \odot V_{11} \odot V_{11} - 4\text{Diag}(\gamma_1^*\gamma_1^{*'}D^{-1}V_{11}) + \gamma_1^*\gamma_1^{*'} \end{bmatrix},\end{aligned}\quad (\text{A46})$$

$$\Upsilon_{w1} = E[h_{2t}h'_{3t}] = -(1 + \kappa)A^*V_{22}We[0, \gamma_1^{*'}D^{-1}V_{11}D^{-1}]H^*, \quad (\text{A47})$$

$$\Upsilon_{w2} = E[h_{3t}h'_{3t}] = (1 + \kappa)(e'WV_{22}We)H^*\tilde{D}H^*. \quad (\text{A48})$$

Turning to the GLS case, we have

$$h_t = h_{1t} + h_{2t} + h_{3t} + h_{4t}, \quad (\text{A49})$$

with

$$h_{1t} = A^*(R_t - \mu_2), \quad h_{2t} = A^*[\beta^*D_t - (R_t - \mu_2)(f_t - \mu_1)']D^{-1}\gamma_1^*, \quad (\text{A50})$$

$$h_{3t} = H^*z_t^*u_t, \quad h_{4t} = -(\gamma_t^* - \gamma^*)u_t. \quad (\text{A51})$$

Note that h_{1t} to h_{3t} are the same as those in the known weighting matrix case after setting $W = V_{22}^{-1}$. It follows that the Υ_w and Υ_{w1} expressions are the same as the ones in the known weighting matrix case. For the GLS case, it is also obvious that Υ_{w1} is a zero matrix because $A^*V_{22}V_{22}^{-1}e = A^*e = 0_{K+1}$.

It is easy to see that h_{1t} is uncorrelated with h_{4t} . We now show that h_{2t} is also uncorrelated with h_{4t} . Let a be a K -vector and b be an N -vector. Under the multivariate elliptical distribution assumption, it can be shown that

$$E[D_t a b'(R_t - \mu_2)(R_t - \mu_2)'] = (1 + \kappa)[D a b'V_{22} + 2\text{Diag}(V_{12} b a')V_{12}]. \quad (\text{A52})$$

Using (A52) with $a = D^{-1}\gamma_1^*$ and $b = V_{22}^{-1}e$ and the fact that $A^*e = 0_{K+1}$ and $\beta^{*'}V_{22}^{-1}e = 0_K$, after some algebra we obtain

$$E[h_{2t}h'_{4t}] = 0_{(K+1) \times (K+1)}. \quad (\text{A53})$$

It is straightforward to show that

$$E[h_{3t}h'_{3t}] = (1 + \kappa)(e'V_{22}^{-1}e)H^*\tilde{D}H^*, \quad (\text{A54})$$

$$E[h_{3t}h'_{4t}] = -(1 + \kappa)(e'V_{22}^{-1}e)H^* \begin{bmatrix} 0 & 0'_K \\ 0_K & I_K \end{bmatrix}, \quad (\text{A55})$$

$$E[h_{4t}h'_{4t}] = (1 + \kappa)(e'V_{22}^{-1}e)H^*. \quad (\text{A56})$$

Collecting terms and using the identity

$$H^* = \tilde{H} + \begin{bmatrix} 0 & 0'_K \\ 0_K & DV_{11}^{-1}D \end{bmatrix}, \quad (\text{A57})$$

we obtain

$$\Upsilon_{w2} = E[h_{3t}h'_{3t}] + E[h_{3t}h'_{4t}] + E[h_{4t}h'_{3t}] + E[h_{4t}h'_{4t}] = (1 + \kappa)(e'V_{22}^{-1}e)(\tilde{H}\tilde{D}\tilde{H} + \tilde{H}). \quad (\text{A58})$$

Finally, for the WLS case, we have

$$h_t = h_{1t} + h_{2t} + h_{3t} + h_{4t}, \quad (\text{A59})$$

with

$$h_{1t} = A^*(R_t - \mu_2), \quad h_{2t} = A^*[\beta^*D_t - (R_t - \mu_2)(f_t - \mu_1)']D^{-1}\gamma_1^*, \quad (\text{A60})$$

$$h_{3t} = H^*z_t^*u_t, \quad h_{4t} = -A^*\Psi_t\Sigma_d^{-1}e. \quad (\text{A61})$$

Note that h_{1t} to h_{3t} are the same as those in the known weighting matrix case after setting $W = \Sigma_d^{-1}$. It follows that the Υ_w and Υ_{w1} expressions are the same as the ones in the known weighting matrix case.

It is easy to verify that h_{4t} is uncorrelated with h_{1t} to h_{3t} . In addition, it can be shown that

$$E[\Psi_t\Sigma_d^{-1}ee'\Sigma_d^{-1}\Psi_t] = (1 + \kappa)(2\Phi + ee'). \quad (\text{A62})$$

Using this identity and the fact that $A^*e = 0_{K+1}$, we obtain

$$E[h_{4t}h'_{4t}] = 2(1 + \kappa)A^*\Phi A^{*'} \quad (\text{A63})$$

and

$$\Upsilon_{w2} = E[h_{3t}h'_{3t}] + E[h_{4t}h'_{4t}] = (1 + \kappa) \left[(e'\Sigma_d^{-1}V_{22}\Sigma_d^{-1}e)H^*\tilde{D}H^* + 2A^*\Phi A^{*'} \right]. \quad (\text{A64})$$

When the model is correctly specified, $e = 0_N$ and as a result both Υ_{w1} and Υ_{w2} vanish and we have $V(\hat{\gamma}^*) = \Upsilon_w$. This completes the proof.

References

- Bentler, P. M., and M. Berkane, 1986, "Greatest Lower Bound to the Elliptical Theory Kurtosis Parameter," *Biometrika*, 73, 240–241.
- Black, F., M. C. Jensen, and M. 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.
- Chen, N., R. Roll, and S. A. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business*, 59, 383–404.
- Cochrane, J. H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104, 572–621.
- Cochrane, J. H., 2005, *Asset Pricing*, Princeton University Press, Princeton.
- Fama, E. F., and J. MacBeth, 1973, "Risk, Returns and Equilibrium: Empirical Tests," *Journal of Political Economy*, 71, 607–636.
- Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.
- Ferson, W. E., and R. W. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance*, 51, 425–461.
- Hou, K., and R. Kimmel, 2006, "On the Estimation of Risk Premia in Linear Factor Models," working paper, Ohio State University.
- Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-section of Expected Returns," *Journal of Finance*, 51, 3–53.
- Jagannathan, R., and Z. Wang, 1998, "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression," *Journal of Finance*, 53, 1285–1309.
- Kan, R., C. Robotti, and J. Shanken, 2009, "Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology," working paper, University of Toronto.

- Lettau, M., and S. Ludvigson, 2001, “Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying,” *Journal of Political Economy*, 109, 1238–1287.
- Newey, W. K., and K. D. West, 1987, “A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- Newey, W. K., and K. D. West, 1994, “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, 61, 631–653.
- Shanken, J., 1990, “Intertemporal Asset Pricing: An Empirical Investigation,” *Journal of Econometrics*, 45, 99–120.
- Shanken, J., and G. Zhou, 2007, “Estimating and Testing Beta Pricing Models: Alternative Methods and their Performance in Simulations,” *Journal of Financial Economics*, 84, 40–86.

Table 1

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

Panel A: OLS

	FF3				Scaled CAPM			
	$\hat{\gamma}_0^*$	$\hat{\gamma}_{vw}^*$	$\hat{\gamma}_{smb}^*$	$\hat{\gamma}_{hml}^*$	$\hat{\gamma}_0^*$	$\hat{\gamma}_{cay}^*$	$\hat{\gamma}_{vw}^*$	$\hat{\gamma}_{vw-cay}^*$
Estimate	1.87	2.92	-0.24	2.19	3.68	-0.33	0.85	0.12
t -ratio _{fm}	1.31	1.19	-0.25	3.22	3.89	-1.22	0.61	3.63
t -ratio _{jw}	1.16	1.09	-0.23	2.97	2.69	-0.64	0.36	2.70
t -ratio _{pm}	0.95	0.89	-0.20	2.50	2.38	-0.19	0.18	1.27

Panel B: GLS

	FF3				Scaled CAPM			
	$\hat{\gamma}_0^*$	$\hat{\gamma}_{vw}^*$	$\hat{\gamma}_{smb}^*$	$\hat{\gamma}_{hml}^*$	$\hat{\gamma}_0^*$	$\hat{\gamma}_{cay}^*$	$\hat{\gamma}_{vw}^*$	$\hat{\gamma}_{vw-cay}^*$
Estimate	3.98	-0.52	0.93	1.31	4.01	0.24	-1.25	0.01
t -ratio _{fm}	4.18	-0.30	1.24	2.25	5.85	1.12	-1.22	0.69
t -ratio _{jw}	4.07	-0.29	1.17	2.33	5.03	1.08	-1.11	0.68
t -ratio _{pm}	2.64	-0.20	0.95	1.87	4.04	0.65	-0.93	0.48

Panel C: WLS

	FF3				Scaled CAPM			
	$\hat{\gamma}_0^*$	$\hat{\gamma}_{vw}^*$	$\hat{\gamma}_{smb}^*$	$\hat{\gamma}_{hml}^*$	$\hat{\gamma}_0^*$	$\hat{\gamma}_{cay}^*$	$\hat{\gamma}_{vw}^*$	$\hat{\gamma}_{vw-cay}^*$
Estimate	2.53	1.85	0.15	1.93	2.82	-0.08	0.91	0.12
t -ratio _{fm}	1.79	0.77	0.16	2.87	3.09	-0.31	0.70	3.91
t -ratio _{jw}	1.66	0.73	0.15	2.78	1.94	-0.15	0.39	2.75
t -ratio _{pm}	1.29	0.57	0.13	2.22	1.82	-0.07	0.24	1.62

The table presents the estimation results of the FF3 and scaled CAPM, where the scaling variables are a constant term and the lagged consumption-wealth ratio (CAY) of Lettau and Ludvigson (2001). The models are estimated using quarterly returns on the 25 Fama-French size and book-to-market ranked portfolios. The data are from 1963 Q4 to 1998 Q3 (140 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 Jagannathan and Wang (1998) t -ratios under correctly specified models (t -ratio _{jw}) that account for the errors-in-variables problem, and our model misspecification robust t -ratios (t -ratio _{pm}).