THE JOURNAL OF FINANCE • VOL. LIII, NO. 4 • AUGUST 1998

An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression

RAVI JAGANNATHAN and ZHENYU WANG*

ABSTRACT

Without the assumption of conditional homoskedasticity, a general asymptotic distribution theory for the two-stage cross-sectional regression method shows that the standard errors produced by the Fama–MacBeth procedure do not necessarily overstate the precision of the risk premium estimates. When factors are misspecified, estimators for risk premiums can be biased, and the t-value of a premium may converge to infinity in probability even when the true premium is zero. However, when a beta-pricing model is misspecified, the t-values for firm characteristics generally converge to infinity in probability, which supports the use of firm characteristics in cross-sectional regressions for detecting model misspecification.

LINEAR BETA-PRICING MODELS have received wide attention in finance literature. Although sophisticated econometric methods are available for evaluating linear beta-pricing models, it is difficult to interpret statistical rejections obtained from these methods. The two-stage cross-sectional regression method, which is used in the two classic studies of the CAPM—one by Black, Jensen, and Scholes (1972) and the other by Fama and MacBeth (1973)—is still preferred in many empirical studies. With the cross-sectional regression method, it is rather straightforward to interpret the results in economic terms. It is also convenient to examine model misspecification by checking whether firm characteristics such as relative size and book-to-market value explain any residual variation in average returns across firms. The method is intuitive and easy to implement. Because of this, empirical studies of linear beta-pricing models still use the cross-sectional regression method as a diagnostic tool.[1]

*Jagannathan is from the Kellogg Graduate School of Management, Northwestern University, and the Carlson School of Management, University of Minnesota. Wang is from the Graduate School of Business, Columbia University. We appreciate comments from David Backus, Sid Browne, Edwin Elton, Pierre Hillion, Bob Hodrick, and Raymond Kan, as well as seminar participants at the 1998 meeting of the American Finance Association, Columbia University, Northwestern University, University of Minnesota, Federal Reserve Bank of New York, and Stockholm School of Economics. Errors are our own. We thank Laarni Bulan for research assistance. Ravi Jagannathan acknowledges partial financial support from the National Science Foundation, grant SBR-9409824.

[1] Examples are the studies by Chan et al. (1991), Jegadeesh (1992), Fama and French (1992), Davis (1994), Kothari, Shanken, and Sloan (1995), Fama and French (1996), Jagannathan and Wang (1996), and Daniel and Titman (1997).

The statistical properties of parameter estimates obtained by using the cross-sectional regression method are rather complex. Black et al. (1972) derive the sampling distribution of the estimators under the assumption that measurement errors in the betas obtained in the first stage are small enough to be ignored. Fama and MacBeth (1973) suggest conducting cross-sectional regression for each period and then treating the estimates as independent samples of the estimated parameter. Shanken (1992) shows how to take into account the sampling errors in the betas obtained in the first stage. Under the assumption that, given the realization of factors, asset returns have a conditional joint distribution with constant covariance matrix (conditional homoskedasticity), he shows that the standard errors obtained by using the Fama–MacBeth procedure overstate the precision of the estimated parameters.

In this paper, without assuming conditional homoskedasticity, we derive the asymptotic distribution of the estimators in the cross-sectional regression method. When observations on returns and factors are identically and independently distributed, conditional homoskedasticity holds if the distribution is normal. However, the assumption of conditional homoskedasticity does not hold if the returns and factors are independent and have an identical t-distribution. The asymptotic distribution developed in this paper only requires that the fourth moments of the distribution exist. We show that when the assumption of conditional homoskedasticity is violated, the standard errors obtained from the Fama–MacBeth procedure need not necessarily overstate the precision of the estimates.

A beta-pricing model is often misspecified. When expected returns are not exactly linear in the factor betas, as prescribed by theoretical models, a linear beta-pricing model will be an incorrect specification. For example, if expected return is a linear function of the covariance between the return and the growth rate of consumption, as described by the consumption-based CAPM, Sharpe's (1964) version of the CAPM with betas relative to the market return will be misspecified. Even when expected asset returns are linear in their factor betas, the econometric specification could be wrong if some of the factors are unobservable and an econometrician uses arbitrarily chosen observable variables in place of the unobservable factors. A classic example of this type of misspecification is the use of a stock market index as a proxy for the unobservable market return in the CAPM, which is criticized by Roll (1977).

In this paper, we examine statistical properties of the estimators in cross-sectional regressions when linear beta-pricing models are misspecified. We show that when a model is misspecified in this way, the cross-sectional regression method will generally not provide consistent estimates of the premiums on factors. The estimator of the premiums corresponding to the arbitrarily chosen proxy-factors, as well as the correctly specified factors, may be biased.

In empirical studies, researchers often use the t-statistics associated with factor risk premiums to let the data identify priced factors. A beta with a t-value larger than two is chosen to be included in the model, and a beta with a small t-value is excluded from the model. A recent critique of this practice comes from Kan and Zhang (1997). They conduct simulations in which true asset returns are generated from a one-factor model but the factor in the two-stage cross-sectional regression is misspecified as a random variable uncorrelated with asset returns. They refer to such a misspecified factor as a "useless factor," and show that the t-value of this "useless factor" converges to a large value in the cross-sectional regression.

In this paper, we demonstrate that for any multifactor linear beta-pricing models, when factors are misspecified the t-value of a premium may converge to infinity in probability even when the true premium is zero. This can happen even when the misspecified factors are correlated with returns, that is, the misspecified factors are not "useless" in the sense of Kan and Zhang (1997). Hence, in the absence of economic theory, the practice of using t-statistics to identify price factors could generally lead to erroneous conclusions—a beta can be included in a model simply because of misspecification.

Firm characteristics are often used for detecting misspecification errors. If a linear beta-pricing model is correct, variables of firm characteristics added to the model should not explain the cross-sectional variation of expected returns after controlling the betas prescribed by the model. In this case, the t-values of firm characteristics should be insignificant. Therefore, a significant t-value is viewed as a rejection of the linear beta-pricing model. The common firm characteristic used for this purpose is firm size, which is defined as the log of market capitalization of a firm. Banz (1981) is the first to use firm size to examine the CAPM. Berk (1995) theoretically shows why firm size should be correlated with expected returns in the cross-section. Another commonly used firm characteristic is the ratio of book value to market value of a firm. Chan, Hamao, and Lakonishok (1991) and Fama and French (1992) show that this variable explains a large fraction of the cross-sectional variation in expected returns. Based on this evidence, Fama and French (1993) propose a three-factor model for stock returns. Daniel and Titman (1997) add firm size and book-to-market ratio to the three-factor model and find that the t-values associated with these two firm characteristics are still significant in the cross-sectional regression.

In this paper, we take a close look at the behavior of t-values associated with firm characteristics. When a linear beta-pricing model holds, those t-values converge to a normal distribution with a finite variance. However, when factors or models are misspecified, we show that the t-values associated with firm characteristics generally converge to infinity in probability. In particular, we show that, when a firm characteristic such as size is added to the one-factor model with a misspecified "useless factor," the t-value of the firm characteristic converges to infinity in probability. Therefore, large t-values associated with firm characteristics reject the linear beta-pricing model. These results support the use of firm characteristics in cross-sectional regression for detecting model misspecification.

Usually, betas of assets are defined as slope coefficients in the multiple regression of asset returns on factors. However, it is sometimes convenient to define each beta as the slope coefficient in the simple regression of return on each single factor. For convenience, we refer to these betas as *univariate betas*, while referring to the first type as *multivariate betas*. Chen, Roll, and Ross (1986) report that they tested their factors with both multivariate and univariate betas. From a conditional single-factor model (CAPM), Jagannathan and Wang (1996) arrive at an unconditional beta-pricing model with two univariate betas. When using cross-sectional regression as a diagnostic tool, it is helpful to see how an asset pricing model's performance improves when additional factors are applied. For this purpose, it is convenient to use betas obtained from univariate regression of asset returns on individual factors. In a model with multivariate betas, when a new factor is applied, the values of all the other betas will change unless the new factor is uncorrelated with the old factors. However, in models with univariate betas, adding or deleting a factor will not change the values of betas corresponding to other factors. This makes it more convenient to compare performance of alternative specifications for factor models.

In this paper, we derive an asymptotic distribution for models with univariate betas. Since each of the two beta types (as well as premiums) is simply a linear transformation of the other, one might think such an asymptotic distribution for models with univariate betas is unnecessary because a model with univariate betas can be transformed to a model with multivariate betas. Nevertheless, a distribution theory for models with univariate betas is desirable for the following reason. We show that a nonzero risk premium for a factor in a model with univariate betas does not imply a nonzero risk premium on the same factor in the corresponding model with multivariate betas. Thus, if a theoretical model is in the form of univariate betas, we cannot examine the premium for the factor estimated with multivariate betas. Additionally, we demonstrate that the standard errors obtained by applying the transformation to the asymptotic variance for models with multivariate betas are incorrect for models with univariate betas.

The rest of the paper is organized as follows. In Section I, we describe the cross-sectional regression method in detail and set up the necessary notations. In Section II, we derive the asymptotic distribution of estimators from the cross-sectional regression method, and we examine the effect of conditional homoskedasticity and heteroskedasticity. In Section III, we discuss the bias when factors are misspecified and show why firm characteristics such as size can be used as specification tests. In Section IV, we develop the asymptotic theory when betas are estimated in the first stage using simple regressions instead of multiple regressions. In Section V, we provide some concluding remarks. The Appendix contains proofs of theorems.

I. Cross-Sectional Regression

Let $R = (R_1,\dots,R_N)'$ be the vector of gross returns on N assets with covariance matrix Φ. As observed by Hansen and Richard (1987), an asset pricing model implies a stochastic discount factor m such that $E[mR_i] = 1$, for all $i = 1,\dots,N$. Let $y = (y_1,\dots,y_K)'$ be a vector of K factors. Suppose the stochastic discount factor m can be linearly spanned by a vector of factors y; that is, $m = \theta_0 + \theta' y$, where θ_0 is a constant number and θ is a K-dimensional constant vector. Then $E[mR_i] = 1$, for all $i = 1,\ldots,N$, can be rewritten in the following form:

$$E[R_i] = a_0 + b'\beta_i \qquad i = 1,\ldots,N, \tag{1}$$

$$\beta_i = \Omega^{-1} E[(y - E[y])(R_i - E[R_i])], \tag{2}$$

where Ω is the variance matrix of y. In fact, it can be verified that, for a given vector y of random variables, asset returns satisfy the beta-pricing equations (1) and (2) for some constants a_0 ($\neq 0$) and b if and only if there are some constants θ_0 and θ such that $E[(\theta_0 + \theta' y)R_i] = 1$ for all i. An important example is the CAPM, where Dybvig and Ingersoll (1982) show that m is a linear function of return on the market portfolio. When returns have a linear factor structure, Ross (1976) shows that expected returns will be approximately linear in factor betas, and in that case equation (1) can be viewed as an approximation.

To examine whether the linear beta pricing model holds using the cross-sectional regression method, the model in equation (2) is expanded to include other asset characteristics such as relative market capitalization, book-to-market ratio, price-earnings ratio, etc. Let z_i be a vector of L observable asset characteristics. Adding these to equation (2), we obtain

$$E[R_i] = a_0 + a' z_i + b'\beta_i. \tag{3}$$

If the assumption that m is a linear function of the factors is correct, then a should be a vector of zeros. This is often the null hypothesis in an empirical investigation. Financial economists also investigate if a particular factor is priced by looking at the t-value of the corresponding element in b.

Equation (3) can be concisely written using matrix notation. Let $B = (\beta_1,\ldots,\beta_N)'$. Then,

$$B = E[(R - E[R])(y - E[y])']\Omega^{-1}. \tag{4}$$

Let $c = (a_0, a', b')'$, $Z = (z_1,\ldots,z_N)'$, and $X = (\iota, Z, B)$, where ι is the N-dimensional vector of ones. The cross-sectional regression equation is

$$E[R] = \iota a_0 + Za + Bb = Xc. \tag{5}$$

We assume the rank of matrix X is K. Then, the unknown parameter c can be expressed as

$$c = (X'X)^{-1} X' E[R]. \tag{6}$$

Suppose T samples of returns and factors are observed:

$$(R_t', y_t') = (R_{1t}, \ldots, R_{Nt}, y_{1t}, \ldots, y_{Kt}), \qquad t = 1, \ldots, T. \tag{7}$$

These samples are usually time-series observations. We assume they are stationary and the sample moments of returns and factors converge to the corresponding unconditional population moments. Then, the covariance matrices Φ and Ω can be consistently estimated by

$$\hat{\Phi} = \frac{1}{T} \sum_{t=1}^{T} (R_t - \bar{R})(R_t - \bar{R})', \tag{8}$$

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y})(y_t - \bar{y})', \tag{9}$$

where $\bar{R}$ and $\bar{y}$ are the time-series average of observations on the vectors of returns and factors. In the rest of the paper, we use ϕ_{ij} ($\hat{\phi}_{ij}$) to denote the element of Φ ($\hat{\Phi}$) at the ith row and jth column, and ω_{kl} ($\hat{\omega}_{kl}$) to denote the element of Ω ($\hat{\Omega}$) at the kth row and lth column. The consistent estimators for $E[R_i]$ and β_i are

$$\bar{R}_i = \frac{1}{T} \sum_{t=1}^{T} R_{it} \tag{10}$$

$$\hat{\beta}_i = \hat{\Omega}^{-1} \sum_{t=1}^{T} (y_t - \bar{y})(R_{it} - \bar{R}_i). \tag{11}$$

We use $\hat{B}$ to represent the matrix of the estimated betas and let $\hat{X} = (\iota, Z, \hat{B})$. The cross-sectional OLS estimate for c is

$$\hat{c} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R}. \tag{12}$$

To estimate standard errors, Fama and MacBeth (1973) suggest doing a cross-sectional regression for each t; that is,

$$\hat{c}_t = (\hat{X}'\hat{X})^{-1}\hat{X}'R_t, \tag{13}$$

and treating c_t as an independent sample: the average is $\hat{c}$ and the variance of $\sqrt{T}(\hat{c} - c)$ is estimated by

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} (\hat{c}_t - \hat{c})(\hat{c}_t - \hat{c})'. \tag{14}$$

The Fama–MacBeth procedure has two advantages. First, it is very intuitive and easy to implement. Second, when we only need the t-values of the cross-sectional variables, it does not require us to estimate covariances among returns on different assets.

In recent years, several researchers have recommended GLS over OLS for the cross-sectional regression. It is well known that GLS can be more efficient than OLS. There are also economic reasons for using GLS. Kandel and Stambaugh (1995) argue that, if the weighting matrix is chosen to be the covariance of returns, the slope coefficient and the R square in the cross-sectional GLS for the CAPM indicate the degree of inefficiency of the index portfolio, while the slope coefficient and the R square in the cross-sectional OLS can potentially be any value even when the index portfolio is almost efficient. Let Q be a symmetric positive definite matrix and $\hat{Q}$ a consistent estimator of Q. The GLS estimate of c can be written as

$$\hat{c} = (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}\bar{R}. \tag{15}$$

As T becomes large, $\hat{c}$ is a consistent estimator for c in view of equation (15). If we use GLS in the Fama–MacBeth method, the time-series of estimated c should have the form of

$$\hat{c}_t = (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}R_t. \tag{16}$$

Substituting equations (15) and (16) into equation (14), we obtain

$$\hat{V} = (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}\left[\frac{1}{T}\sum_{t=1}^{T}(R_t - \bar{R})(R_t - \bar{R})'\right]\hat{Q}\hat{X}(\hat{X}'\hat{Q}\hat{X})^{-1}. \tag{17}$$

In the rest of the paper, this is referred to as the variance from the Fama–MacBeth method. When observations on returns are identically and independently distributed, the matrix $\hat{V}$ converges in probability to the limit

$$(X'QX)^{-1}X'Q\Phi QX(X'QX)^{-1}$$

as the sample size T becomes infinitely large.

II. Asymptotic Distribution of the Estimators

In order to understand the properties of estimated covariance matrix $\hat{V}$ suggested by Fama and MacBeth (1973), it is convenient to use equation (5) to write the average return as

$$\bar{R} = \hat{X}c + (\bar{R} - E[R]) - (\hat{B} - B)b. \tag{18}$$

Substituting equation (18) into equation (15), we obtain

$$\sqrt{T}(\hat{c} - c) = (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}\sqrt{T}(\bar{R} - E[R]) \tag{19}$$

$$- (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}\sqrt{T}(\hat{B} - B)b \tag{20}$$

which decomposes the estimation error in $\hat{c}$ into two parts: the first part comes from replacement of expected returns by average returns, and the second part comes from replacement of true betas by their estimates.

The error in the estimated expected returns is

$$\sqrt{T}(\bar{R} - E[R]) = T^{-1/2}\sum_{t=1}^{T}(R_t - E[R]). \tag{21}$$

Let us assume the Central Limit Theorem applies to the averages of observations on returns; that is, as $T \to \infty$, the random variable $\sqrt{T}(\bar{R} - E[R])$ converges in distribution to a random vector r of multivariate distribution with mean zero and variance Ψ. If the time series of returns is serially uncorrelated, we will have $\Psi = \Phi$.

To assess the error in the estimated betas, we should consider the time-series regression model

$$R_t = A + By_t + u_t, \tag{22}$$

where $A = E[R_t] - BE[y_t]$ and $u_t = R_t - A - By_t$. It follows that

$$E[u_t] = 0 \quad \text{and} \quad E[u_t y_t'] = 0. \tag{23}$$

Let $Y_t = y_t - \bar{y}$, $Y = (Y_1, \ldots, Y_T)'$, and $U = (u_1, \ldots, u_T)'$. Notice that $\hat{B} - B = U'Y(Y'Y)^{-1}$. Under very general conditions, the average of the product of the factor deviations and the residual u converges to a normal distribution, that is, as $T \to \infty$, the random variable $\mathrm{vec}(Y'U)/\sqrt{T}$ converges in distribution to a normal random vector with zero mean. (Here, the operator "vec" stacks all the columns in a matrix into a long vector.) Let $h_i = \lim T^{-1/2}\sum Y_t u_{it}$ and $H = (h_1, \ldots, h_N)$. Denote the covariance between h_i and h_j by $\Pi_{ij} = E[h_i h_j']$. The covariance $E[\mathrm{vec}(H)\mathrm{vec}(H)']$ is then a matrix Π with $N \times N$ blocks such that the block on the ith row and jth column is Π_{ij}. We express this kind of construction by $\Pi = (\Pi_{ij})_{i,j=1,\ldots,N}$.

To consider both the sampling errors in estimated expected returns and estimated betas, we must consider the covariance between r and H. Let r_i be the ith element of vector r and define $\gamma_{ij} \equiv E[r_i h_j]$. If we construct a matrix $\Gamma = (\gamma_{ij}')_{i,j=1,\ldots,N}$ in a similar way as Π, the matrix Γ is the covariance between r and $\mathrm{vec}(H)$.

We summarize these limiting distributions in the following assumption:

ASSUMPTION 1: *As $T \to \infty$, the random variable*

$$(T^{1/2}(\bar{R} - E[R])', T^{-1/2}(\mathrm{vec}(Y'U))')'$$

converges in distribution to a normal random vector with zero mean and covariance matrix

$$\begin{pmatrix} \Psi & \Gamma \\ \Gamma' & \Pi \end{pmatrix}.$$

Assumption 1 can be obtained from a Central Limit Theorem. For example, if the observations of returns and factors are identically and independently distributed over time, this assumption can be obtained from the Lindeberg–Lèvy Central Limit Theorem by assuming the existence of the fourth moments of returns and factors. If the time series $(R_t',(\mathrm{vec}(y_t u_t'))')'$ is also covariance stationary and ergodic, a Central Limit Theorem such as Proposition 7.11 in Hamilton (1994) can be used to obtain Assumption 1, in which the asymptotic variance is the sum of autocovariances of the time series $(R_t',(\mathrm{vec}(y_t u_t'))')'$.

The asymptotic distribution of the error in the estimated $\hat{c}$ is summarized as the following theorem:

THEOREM 1: *Suppose X has full rank. Under Assumption 1 $\sqrt{T}(\hat{c} - c)$ converges in distribution, as $T \to \infty$, to a normal distribution with mean zero and variance $S = V + W - G$, where V, W, and G are defined as*

$$V = (X'QX)^{-1}X'Q\Psi QX(X'QX)^{-1}, \tag{24}$$

$$W = (X'QX)^{-1}X'Q(I_N \otimes (\Omega^{-1}b))'\Pi(I_N \otimes (\Omega^{-1}b))QX(X'QX)^{-1}, \tag{25}$$

$$G = (X'QX)^{-1}X'Q[\Gamma(I_N \otimes \Omega^{-1}b) + (I_N \otimes \Omega^{-1}b)'\Gamma']QX(X'QX)^{-1}. \tag{26}$$

The matrix V is positive definite, and the matrix W is positive semi-definite, but the matrix G is generally not definite, depending on the value of b as well as Ω and Γ. When observations on returns are i.i.d., we have $\Psi = \Phi$. In this case, the matrix V is the limit of the estimated covariance matrix generated by the Fama–MacBeth procedure. Since W and G are generally nonzero, the Fama–MacBeth procedure produces biased asymptotic covariances. However, since G is not positive definite, the direction of the bias in the Fama–MacBeth method is not clear. Nonetheless, if the null hypothesis includes $b = 0$, both W and G are zero under the null hypothesis and thus the original estimate of the standard error in the Fama–MacBeth method will be correct.[2]

In his asymptotic theory for the Fama–MacBeth method, Shanken (1992) assumes asset returns are homoskedastic. To be specific, Shanken's assumption is indicated in Assumption 2. (Also see Assumption 1 in his paper.)

ASSUMPTION 2: *Suppose X has full rank. Conditional on factor values, the residual in the time-series regression model has zero mean and constant covariance; that is,*

$$E[\text{vec}(U)|Y] = 0 \quad \text{and} \quad E[\text{vec}(U)(\text{vec}(U))'|Y] = \Delta \otimes I_T, \tag{27}$$

where Δ is an N-dimensional constant symmetric positive definite matrix and I_T is the T-dimensional identity matrix.

This assumption holds if returns and factors are contemporaneously i.i.d. and have a joint normal distribution.

Conditional homoskedasticity has direct consequences in the assessment of estimation errors. These are summarized in the following theorem:

THEOREM 2: *Under Assumption 2, the random variable $\sqrt{T}(\hat{c} - c)$ converges, as $T \to \infty$, to a normal distribution with mean zero and variance $S = V + \widetilde{W}$, where V is defined in equation (24), and $\widetilde{W}$ is defined as*

$$\widetilde{W} = (b'\Omega^{-1}b)(X'QX)^{-1}X'Q\Delta QX(X'QX)^{-1}. \tag{28}$$

Since $\widetilde{W}$ is positive definite, the standard error obtained from the Fama–MacBeth method always overstates the precision of the estimates, under the assumption of conditional homoskedasticity. However, this is generally not true when a time series is conditionally heteroskedastic.

Numerous studies have presented evidence for nonnormality and heteroskedasticity. Early works by Fama (1965) and Blattberg and Gonedes (1974) document nonnormality, and works by Barone-Adesi and Talwar (1983), Schwert and Sequin (1990), and Bollerslev, Engle, and Wooldridge (1988) document conditional heteroskedasticity. Using contemporaneously i.i.d. time series of returns from a t-distribution of more than four degrees of freedom, MacKinlay and Richardson (1991) demonstrate that returns are conditionally heteroskedastic and the test of mean-variance efficiency will be biased under the assumption of conditional homoskedasticity. Their bootstrapping experiment further demonstrates that stock returns are not homoskedastic. For these reasons, MacKinlay and Richardson (1991) advocate the GMM method developed by Hansen (1982), which does not require conditional homoskedasticity.

[2] This is first pointed out by Shanken (1992, p. 14), in view of the asymptotic variance derived in his paper. Although the asymptotic variance derived here is different from Shanken's, his observation still holds true.

Nevertheless, Assumption 1 may be satisfied by many stationary time series that are not conditionally homoskedastic. As we have pointed out, it follows from the Lindeberg–Lèvy Central Limit Theorem that any contemporaneously i.i.d. time series of returns and factors with finite fourth moments satisfies Assumption 1, while it might not satisfy the assumption of conditional homoskedasticity unless the time series is also normally distributed. Clearly, the i.i.d. time series of returns with t-distribution in MacKinlay and Richardson (1991) is an example of this kind. Notice the GMM also requires the existence of fourth moments (see Hansen (1982)).

III. Misspecification Bias

In the previous analysis, we assume the null hypothesis model is correctly specified. When the model is correctly specified, the estimator in cross-sectional regression is consistent if the time series of returns and factors is stationary and ergodic. However, if the null hypothesis model is misspecified, the estimator in cross-sectional regression will be asymptotically biased. Assume that $\tilde{y}$ is a different vector of factors than y; that is, at least some of the factors in $\tilde{y}$ and y are different. Suppose the true model is

$$E[R] = \iota a_0 + Za + \tilde{B}b, \quad \text{where } \tilde{B} = E[(R - E[R])(\tilde{y} - E[\tilde{y}])']\tilde{\Omega}^{-1}. \tag{29}$$

A researcher who incorrectly specifies the model as

$$E[R] = \iota a_0 + Za + Bb, \quad \text{where } B = E[(R - E[R])(y - E[y])']\Omega^{-1} \tag{30}$$

will estimate betas by regressing returns on the vector of misspecified factors y and then will estimate the risk premiums in cross-sectional regression. The bias is given in the following theorem:

THEOREM 3: *Assume that $X = (\iota, Z, B)$ has full rank. If equation (29) holds with time series $(R_t', \tilde{y}_t')'$ but betas are estimated from the time series (R_t', y_t'), then the vector of estimated coefficients $\hat{c}$ converges to $c + (X'QX)^{-1}X'Q(\tilde{B} - B)b$ in probability.*

In view of this theorem, the estimator in cross-sectional regression is asymptotically biased if and only if $X'Q(\tilde{B} - B)b \neq 0$. Notice that not only the estimates for the premium on the misspecified betas can be biased, but also the estimates for the premium on those correctly specified betas can be biased when some other factor is misspecified.

When an estimator in cross-sectional regression is biased, the practice of using t-values to judge if a factor should be included in the model is problematic. For example, suppose we wish to check if factor y_k should be included in the model by looking at the t-value of $\hat{b}_k$. Let $l = 1 + L + k$ and $\hat{s}_{ll}$ be a consistent estimate for the lth diagonal element of the variance matrix

S in Theorem 1, and ζ_l be a $1 + L + K$ dimensional vector with the lth element as 1 and other elements as zero. The t-value for the kth element of $\hat{b}$ is then

$$\begin{aligned} T^{1/2}\hat{s}_{ll}^{-1/2}\hat{b}_k = {} & T^{1/2}\hat{s}_{ll}^{-1/2}[b_k + \zeta_l'(\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}\hat{Q}'(\tilde{B} - B)b] \\ & + \hat{s}_{ll}^{-1/2}\zeta_l'(\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q} \\ & \times T^{1/2}[(\bar{R} - E[R]) - (\hat{B} - B)b]. \end{aligned} \tag{31}$$

Notice that the second term on the right-hand side of the equation will converge to a finite distribution when T becomes infinitely large. It follows that $T^{1/2}\hat{s}_{ll}^{-1/2}\hat{b}_k$ converges to infinity in probability if $\zeta_l(X'QX)^{-1}X'Q(\tilde{B} - B)b \neq 0$, even when $b_k = 0$. In this case, the probability of concluding that factor y_k is priced converges to 1, although the true premium on factor $\tilde{y}_k$ is zero. Therefore, the cross-sectional regression method is not suitable for identifying factors in the model.[3]

Kan and Zhang (1997) study a related case for a one-factor model in which $\tilde{y}_1$ is misspecified as a factor y_1 that is uncorrelated with all the returns, while the true beta $\tilde{\beta}_1$ is nonzero and has nonzero premium. In their case, the misspecified beta is a zero vector ($\beta_1 = 0$). Then, we have $X = (\iota,0)$ and it does not have full rank. Under the assumption that returns and factors are i.i.d. and normal, Kan and Zhang conduct simulations showing that the t-value for the slope of the misspecified beta obtained in the Fama–MacBeth procedure tends to become large as observations increase.

Here we offer an alternative characterization of Kan and Zhang's (1997) observation. To simplify the algebra, let us assume $\hat{Q} = Q = I_N$ for the rest of this section. The estimated variance obtained from the Fama–MacBeth procedure is a two-dimensional symmetric matrix

$$\hat{V} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\Phi}\hat{X}(\hat{X}'\hat{X})^{-1}, \tag{32}$$

where $\hat{\Phi}$ is the sample estimate of Φ and $\hat{X} = (\iota, \hat{\beta}_1)$. Like Kan and Zhang, we also assume the estimated $\hat{X}$ has full rank, although the true $X = (\iota,0)$ is degenerate. If returns and factors are i.i.d., the matrix $\hat{X}$ should have full rank with probability 1. The t-value calculated from the Fama–MacBeth procedure is

$$t = (T/\hat{v}_{22})^{1/2}\hat{b}_1, \tag{33}$$

where $\hat{v}_{22}$ is the lower-right element of the matrix $\hat{V}$ in equation (32). The probability distribution of this t-value is given in the following theorem.

[3] The generalized method of moments developed by Hansen (1982) is more suitable for identifying model misspecification. Systematic approaches to evaluation of misspecified models are proposed by Hansen and Jagannathan (1991, 1997).

THEOREM 4: *Suppose $E[R] = a_0\iota + b_1\tilde{\beta}_1$ and $\tilde{\beta}_1 \neq 0$. Assume that observations of returns and the misspecified factor are i.i.d. with finite fourth moments. In addition, assume the misspecified factor is uncorrelated with all the returns. Then, the t-value in equation (33), which is calculated from the misspecified factor using the Fama–MacBeth procedure, converges to infinity in probability (i.e., $\lim P[|t| > \delta] = 1$ for any $\delta > 0$) if and only if $b_1 \neq 0$.*

This theorem shows that a misspecified factor, even when it is not correlated with returns, is likely to give a t-value larger than a given value such as two. For this reason, Kan and Zhang (1997) argue for caution in using cross-sectional regression to test beta pricing models. However, we do not believe this is a major concern, for the following reasons: (1) economic intuition tells us that Chen–Roll–Ross factors and the market factor used by Fama and French should be correlated with asset returns, and thus the true betas of those factors cannot be zero; (2) empirical experience also shows us high correlation between returns and those factors; (3) if a factor is suspected to be uncorrelated with returns, a statistical test can be easily conducted, and a zero-beta factor should not be used in the cross-sectional test in the first place; (4) even if a factor is misspecified, Theorem 4 tells us a large t-value on the factor still indicates that $b_1 \neq 0$, and thus a model with $b_1 = 0$ is correctly rejected; and (5) as we will show in Theorem 6, a factor uncorrelated with asset returns cannot stand up against a test with a cross-sectional variable z such as firm size.

In empirical tests of asset pricing models, we often check risk-based variables against nonrisk-based variables such as firm size or book-to-market. This practice works well even when factors are misspecified. Let us explain this with one beta and one firm characteristic, such as firm size. Suppose the vector of N average returns has the asymptotic properties

$$\operatorname{plim} \bar{R} = E[R] \quad \text{and} \quad \lim \sqrt{T}(\bar{R} - E[R]) \sim N(0, \Phi). \tag{34}$$

Suppose the expected return is a linear span of the vector $\tilde{\beta}$ of true betas on the N assets,

$$E[R] = b\tilde{\beta}, \tag{35}$$

where b is a nonzero real number. Let z be the vector of N firm sizes. The alternative model for the cross-sectional expected returns can be specified as

$$E[R] = az + b\tilde{\beta}, \tag{36}$$

where a is a real number. The null hypothesis in the specification test is $a = 0$.

Let $\hat{\beta}$ be a vector of estimated betas, which has the asymptotic properties

$$\operatorname{plim} \hat{\beta} = \beta \quad \text{and} \quad \lim \sqrt{T}(\hat{\beta} - \beta) \sim N(0, \Sigma), \tag{37}$$

where β is an N dimensional vector and Σ is an N dimensional symmetric positive definite matrix. Let $\hat{X} = (z, \hat{\beta})$. Assume $\hat{X}$ has full rank with probability one. The cross-sectional regression equation is

$$\bar{R} = az + b\hat{\beta} + \epsilon. \tag{38}$$

The cross-sectional estimator for (a, b) is

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R}. \tag{39}$$

The variance based on the Fama–MacBeth procedure is asymptotically equivalent to

$$\hat{V} = (\hat{X}'\hat{X})^{-1}\hat{X}'\Phi\hat{X}(\hat{X}'\hat{X})^{-1}. \tag{40}$$

The t-statistic for a is $t_a = \sqrt{T}\hat{v}_{11}^{-1/2}\hat{a}$.

The asymptotic distribution of t_a depends on the relationship between the limit of $\hat{\beta}$ and the true beta $\tilde{\beta}$. The following theorem summarizes this relationship.

THEOREM 5: *Suppose $(z'z)(\beta'\beta) \neq (z'\beta)^2$. If $(z'\tilde{\beta})(\beta'\beta) = (z'\beta)(\beta'\tilde{\beta})$, as $T \to \infty$, the t-statistic t_a converges in distribution to a normal random variable with zero mean and finite variance. If $(z'\tilde{\beta})(\beta'\beta) \neq (z'\beta)(\beta'\tilde{\beta})$, as $T \to \infty$, the probability $P[|t_a| > \delta]$ converges to 1 for every positive number δ.*

As a special case, if $\beta = \tilde{\beta}$; that is, $\hat{\beta}$ is a consistent estimator of the true beta, then the equation $(z'\tilde{\beta})(\beta'\beta) = (z'\beta)(\beta'\tilde{\beta})$ holds; thus the t-statistics t_a have an asymptotic normal distribution with zero mean and finite variance. In this case, t_a will converge to a finite distribution. However, if $(z'\tilde{\beta})(\beta'\beta) \neq (z'\beta)(\beta'\tilde{\beta})$, the theorem says t_a explodes to infinity in probability, although the true value of a is zero. Another way to state the theorem is that lim $P[|t_a| > \delta] = 1$ for any positive δ if and only if $(z'\tilde{\beta})(\beta'\beta) \neq (z'\beta)(\beta'\tilde{\beta})$. Since $(z'\tilde{\beta})(\beta'\beta) \neq (z'\beta)(\beta'\tilde{\beta})$ implies $\beta \neq \tilde{\beta}$, a large t-value on firm size z implies the beta pricing model is likely to be misspecified.

If the limit of the estimated beta is zero, that is, $\beta = 0$, the misspecified factor is uncorrelated with returns and referred to as "a useless factor" by Kan and Zhang (1997). In this case, we never have $(z'z)(\beta'\beta) \neq (z'\beta)^2$, and the limit of the cross-sectional regressor X becomes degenerate. Then we cannot apply Theorem 5. However, firm characteristics such as firm size can still be used to check out this kind of model misspecification. This follows from the next theorem.

THEOREM 6: *If $\beta = 0$, as $T \to \infty$, the probability $P[|t_a| > \delta]$ converges to 1 for any positive number δ.*

This theorem shows that such a useless factor cannot make firm size insignificant in the cross-sectional regression.

IV. Asymptotic Theory for Models with Univariate Betas

Sometimes a beta in asset pricing models takes the form of the slope coefficient in the simple regression of a return on a single factor. For convenience, we refer to such betas as *univariate betas*, while referring to the betas in Section I as *multivariate betas*. To be explicit, a univariate beta of asset i with respect to factor y_k is defined as $\beta^*_{ik} = \mathrm{cov}(R_i, y_k)/\mathrm{var}(y_k)$. Let $\beta^*_k = (\beta^*_{1k}, \ldots, \beta^*_{Nk})'$. We have

$$B^*_k = \omega^{-1}_{kk} E[(y_k - E[y_k])R]. \tag{41}$$

If we define $B^* = (\beta^*_1, \ldots, \beta^*_K)$ and let D be a $K \times K$ diagonal matrix with the same diagonal elements as the matrix Ω, then

$$B^* = E[R(y - E[y])']D^{-1}. \tag{42}$$

A model with univariate betas will be in the form of

$$E[R] = \iota a_0 + Za + B^* b^* = X^* c^*, \tag{43}$$

where $X^* = (\iota, Z, B^*)$ and $c^* = (a_0, a', b^{*\prime})'$.

Univariate betas are simply a linear transformation of multivariate betas. In fact, it follows from equations (4) and (42) that

$$B^* = B\Omega D^{-1}. \tag{44}$$

It then follows from equations (5), (43), and (44) that

$$b^* = D\Omega^{-1}b, \qquad c^* = Fc, \quad \text{and} \quad X^* = XF^{-1}, \tag{45}$$

where

$$F = \begin{pmatrix} I & 0 \\ 0 & D\Omega^{-1} \end{pmatrix}, \tag{46}$$

in which the dimension of the identity matrix I is $L + 1$.

Although univariate and multivariate betas are related by a simple linear transformation, testing if some coefficients in b are zero does not tell us whether some coefficients in b^* are zero. For example, if

$$b^* = (1,2)' \quad \text{and} \quad \Omega = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}, \tag{47}$$

then D is the identity matrix, and

$$b = \Omega D^{-1} b^* = (0, 1.5)'. \tag{48}$$

Notice each element of b^* is positive but the first element of b is zero. Therefore, even if all the coefficients in the univariate-beta model are significant, some of the coefficients in the corresponding multivariate-beta model can still be insignificant. For this reason, it is misleading to look at the t-values in the multivariate-beta model while primary interests are on the coefficients in the univariate-beta model. Therefore, we sometimes have to directly examine estimators for the univariate-beta model.

Since the coefficients of univariate betas are simply a linear transformation of the coefficients of multivariate betas, one is tempted to apply the transformation to the asymptotic variance of estimation errors in Theorem 1 to obtain the asymptotic variance for the case with univariate betas. Nevertheless, this approach is incorrect. If we apply the transformations in (45) to the variance obtained in Theorem 1, we have $F(V + W - G)F'$. However, equation (45) and its estimated version imply

$$\sqrt{T}(\hat{c}^* - c^*) = \hat{F}\sqrt{T}(\hat{c} - c) + \sqrt{T}(\hat{F} - F)c. \tag{49}$$

Under Assumption 1, it follows from Theorem 1 that the first term on the right-hand side converges to a normal distribution with the variance matrix $F(V + W - G)F'$. Therefore, the matrix $F(V + W - G)F'$ is not the variance for the asymptotic distribution of $\sqrt{T}(\hat{c}^* - c^*)$ because it omits the second term in equation (49). Generally, the second term does not converge to zero.

Now, let us present an asymptotic theory for the estimators using univariate betas. For each $k = 1,\ldots,K$ and $t = 1,\ldots,T$, if we define

$$A_k^* = E[R_t] - E[y_{kt}]\beta_k^* \quad \text{and} \quad u_{kt}^* = R_t - A_k^* - y_{kt}\beta_k^*, \tag{50}$$

then

$$R_t = A_k^* + y_{kt}\beta_k^* + u_{kt}^*, \quad E[u_{kt}^*] = 0, \quad \text{and} \quad E[u_{kt}^* y_{kt}] = 0. \tag{51}$$

Let $Y_k^* = (y_{k1} - \bar{y}_k,\ldots,y_{kT} - \bar{y}_k)'$ and $U_k^* = (u_{k1},\ldots,u_{kT})'$. We need an assumption similar to Assumption 1:

ASSUMPTION 3: *As $T \to \infty$, the random variable*

$$\frac{1}{\sqrt{T}}\,(T(\bar{R} - \mu)', U_1^{*\prime} Y_1^*, \ldots, U_K^{*\prime} Y_K^*)'$$

converges to a normal distribution with zero mean.

We need some notations for the covariance matrix of this limiting distribution. Let $r = \lim T^{1/2}(\bar{R}_i - E[R_i])$ and $h_k^* = \lim T^{-1/2} Y_k^{*\prime} U_k^*$. The covariance matrix contains $\Psi = E[rr']$, $\Gamma_k^* = E[h_k^* r']$ and $\Pi_{kl}^* = E[h_k^* h_l^{*\prime}]$. Following the derivation in Section II, we obtain Theorem 7.

THEOREM 7: *Under Assumption 3, the random variable* $\sqrt{T}(\hat{c}^* - c^*)$ *converges, as* $T \to \infty$*, to a normal distribution with mean zero and variance* $S^* = V^* + W^* - G^*$*, where*

$$V^* = (X^{*\prime}QX^*)^{-1}X^{*\prime}Q\Psi QX^*(X^{*\prime}QX^*)^{-1}, \tag{52}$$

$$W^* = (X^{*\prime}QX^*)^{-1}X^{*\prime}Q\left(\sum_{k,l=1}^{K} b_k^*\,\omega_{kk}^{-1}\,\Pi_{kl}^*\,\omega_{ll}^{-1}\,b_l^*\right)QX^*(X^{*\prime}QX^*)^{-1}, \tag{53}$$

$$G^* = (X^{*\prime}QX^*)^{-1}X^{*\prime}Q\left(\sum_{k=1}^{K} b_k^*\,\omega_{kk}^{-1}(\Gamma_k^* + \Gamma_k^{*\prime})\right)QX^*(X^{*\prime}QX^*)^{-1}. \tag{54}$$

Under the assumption of $E[u_{kt}^*|Y_k] = 0$ and $E[u_{kt}^* u_{kt}^{*\prime}|Y_k] = \Delta_k$, where Δ_k is a positive definite matrix independent of the value of Y_k, a derivation similar to those of Theorem 2 shows $G^* = 0$. As we have argued in Section II, conditional homoskedasticity calls for samples from an i.i.d. normal distribution, which is usually violated by financial data. Therefore, for a cross-sectional regression with univariate betas, the matrix G^* is generally nonzero. Jagannathan and Wang (1996) show $\sqrt{T}(\hat{c}^* - c^*)$ converges to a normal distribution with mean zero and variance $V^* + W^*$ (see Theorem 2 in their paper). However, they assume $E[u_{kt}^*|Y] = 0$ and $E[u_{kt}^* u_{kt}^{*\prime}|Y] = \Delta_k$, which cannot be satisfied even by identically, independently, and normally distributed returns and factors. This wrong assumption is corrected by Jagannathan and Wang (1998).

V. Conclusion

Although the cross-sectional regression method is elegant and intuitively appealing, assessing the sampling errors associated with the estimated parameters is not straightforward. Black et al. (1972) and Fama and MacBeth (1973) show how to assess the sampling errors under the assumption that the betas of assets are measured so precisely in the first stage that the estimation error in the betas can be ignored. Shanken (1992) shows how to compute standard errors of the estimated parameters when asset returns, conditional on the factors, have a joint normal distribution with constant variance. Under this assumption, he shows that the Fama–MacBeth procedure overstates the precision of the estimator in the cross-sectional regression.

Based on findings reported in the empirical literature in this area, there are reasons to suspect that asset returns may not be jointly normally distributed and they exhibit conditional heteroskedasticity (see, for example, MacKinlay and Richardson (1991)). In this paper, we derive the asymptotic distribution of the estimators in the cross-sectional regression under the assumption that asset returns follow a stationary and ergodic process. When asset returns exhibit conditional heteroskedasticity, the Fama–MacBeth procedure does not necessarily overstate the precision of the estimates, unlike the case considered by Shanken (1992).

At this point, we have three standard errors for the estimator in the two-stage cross-sectional regression; one from the Fama–MacBeth procedure, which ignores the estimation errors in betas; one provided by Shanken, which assumes conditional homoskedasticity; and one developed in this paper. We need to know the magnitude of numerical differences among the three standard errors. We also need to understand more about the performance of these three standard errors in the samples of finite size we often encounter. These can be examined by Monte Carlo simulation and will be studied in a separate paper.

One advantage of the standard error from the Fama–MacBeth procedure is that it does not need to estimate the covariances among asset returns, which makes it suitable for studying a large cross section of individual stocks. However, the other two types of standard errors, as well as the GMM, require a researcher to estimate the covariance matrix of asset returns. In order to apply these two asymptotic standard errors or the GMM to only several hundred observations, we have to limit our study to a small number of portfolios.

Using a misspecified model, an econometrician may erroneously conclude that a particular factor is priced—that is, it captures some of the pervasive economy-wide risk. This paper demonstrates that when a model is misspecified, the t-statistic for the factor risk premium can converge to infinity under certain conditions even when the true factor risk premium is zero (i.e., the particular factor-risk is not priced). This challenges the practice of using t-values to identify priced factors, since a factor may have a significant t-value simply because the model is misspecified.

However, we confirm that model misspecification can be detected by using firm characteristics, because the t-values associated with those firm characteristics in a misspecified model converge to infinity in probability. Our results also imply that firm characteristics might be significant in the cross-sectional regression because the linear beta-pricing model is misspecified, not because the firm characteristics are priced.

Appendix

Proof of Theorem 1: Since $\hat{B} - B = U'Y(Y'Y)^{-1}$, the error in the estimator can be written as

$$\begin{aligned}\sqrt{T}(\hat{c} - c) = {} & (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}\sqrt{T}(\bar{R} - E[R]) \\ & -(\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}\left[\frac{U'Y}{\sqrt{T}}\right]\left[\frac{Y'Y}{T}\right]^{-1} b. \qquad \text{(A1)}\end{aligned}$$

Since the time series of returns and factors is stationary and ergodic, we have $\hat{X}$ and $Y'Y/T$ converge to X and Ω respectively in probability. It then follows from Assumption 1 and Slutsky's theorem that $\sqrt{T}(\hat{c} - c)$ converges to a normal distribution with zero mean. We therefore only need to prove that the asymptotic covariance matrix is $V + W - G$, as given in Theorem 1.

Assumption 1 implies that $\sqrt{T}(\bar{R} - E[R])$ converges to a normal distribution with mean zero and variance Ψ. Then, the first term on the right-hand side of equation (A1) converges to a normal distribution with mean zero and variance V, where V is given in equation (24).

It follows from Assumption 1 that the limit of the sampling error in betas is

$$\lim \sqrt{T}(\hat{B} - B) = \lim \left[\frac{U'Y}{\sqrt{T}}\right]\left[\frac{Y'Y}{T}\right]^{-1} = H\Omega^{-1}. \tag{A2}$$

The asymptotic covariance matrix between the sampling errors in betas for two assets is therefore

$$b'\Omega^{-1}E[h_i h_j']\Omega^{-1}b = b'\Omega^{-1}\Pi_{ij}\Omega^{-1}b. \tag{A3}$$

Then, the asymptotic covariance matrix of $(\hat{B} - B)b$ can be written as

$$b'\Omega^{-1}E[H'H]\Omega^{-1}b = (I \otimes (\Omega^{-1}b))'\Pi(I \otimes (\Omega^{-1}b)), \tag{A4}$$

which implies that the covariance of the second term on the right-hand side of equation (A1) is the matrix W given in equation (25).

The asymptotic covariance between $\sqrt{T}(\bar{R}_i - E[R_i])$ and $\sqrt{T}(\hat{\beta}_j - \beta_j)'b$ can be calculated as follows:

$$\begin{aligned} &E[\lim \sqrt{T}(\bar{R}_i - E[R_i]) \lim \sqrt{T}(\hat{\beta}_j - \beta_j)'b] \\ &\quad = E\left[\lim \sqrt{T}(\bar{R}_i - E[R_i]) \lim\left(\frac{1}{\sqrt{T}}\sum_t u_{jt}Y_t'\right)\lim(Y'Y/T)^{-1}b\right] \\ &\quad = E[r_i h_j']\Omega^{-1}b \equiv \gamma_{ij}'\Omega^{-1}b. \end{aligned} \tag{A5}$$

We therefore have

$$E[\lim \sqrt{T}(\bar{R} - E[R]) \lim \sqrt{T}(\hat{B} - B)b] = \Gamma(I_N \otimes \Omega^{-1}b), \tag{A6}$$

which implies that the asymptotic covariance between the two terms on the right-hand side of equation (A1) is

$$(\hat{X}'Q\hat{X})^{-1}\hat{X}'Q\Gamma(I_N \otimes \Omega^{-1}b)Q\hat{X}(\hat{X}'Q\hat{X})^{-1}.$$

The sum of this matrix and its transpose is the matrix G given in equation (26). Therefore, the asymptotic covariance of $\hat{c}$ is the sum of V, W, and $-G$. Q.E.D.

Proof of Theorem 2: Let $e_i = (u_{i1}, \ldots, u_{iT})'$ and δ_{ij} be the element of Δ at the ith row and the jth column. It follows from Assumption 2 that

$$\begin{aligned}
E[\lim \sqrt{T} b'(\hat{\beta}_i - \beta_i) \lim \sqrt{T}(\hat{\beta}_j - \beta_j)'b] \\
&= \lim TE[b'(Y'Y)^{-1}Y'e_i e_j' Y(Y'Y)^{-1}b] \\
&= \lim TE[b'(Y'Y)^{-1}Y'E[e_i e_j'|Y]Y(Y'Y)^{-1}b] \\
&= \lim \delta_{ij} b'E\left[\left(\frac{Y'Y}{T}\right)^{-1}\right]b = \delta_{ij} b'\Omega^{-1}b, \qquad \text{(A7)}
\end{aligned}$$

which means the matrix W in Theorem 1 becomes

$$W = (b'\Omega^{-1}b)(X'QX)^{-1}X'Q\Delta QX(X'QX)^{-1}. \qquad \text{(A8)}$$

If we denote by ι_T the T dimensional vector of ones, the error in the estimated expected return can be expressed as

$$\sqrt{T}(\bar{R}_i - E[R_i]) = \sqrt{T}(\bar{y} - E[y])'\beta_i + \iota_T' e_i/\sqrt{T}. \qquad \text{(A9)}$$

It follows from equation (A9) and Assumption 2 that

$$\begin{aligned}
\gamma_{ij}' &= E[\lim \sqrt{T}(\bar{R}_i - E[R_i]) \lim(e_j' Y/\sqrt{T})] \\
&= \lim \beta_i' E[(\bar{y} - E[y])e_j' Y] + \lim \frac{1}{T} E[\iota_T' e_i e_j' Y] \\
&= \lim \beta_i' E[(\bar{y} - E[y])E[e_j'|Y]Y] + \lim \frac{1}{T} E[\iota_T'[e_i e_j'|Y]Y] \\
&= \lim \frac{1}{T} \delta_{ij} E[\iota_T' Y] = 0, \qquad \text{(A10)}
\end{aligned}$$

which implies the matrices Γ and G in Theorem 1 are zero. Q.E.D.

Proof of Theorem 3: Using equation (29) and $\hat{X} = (\iota, Z, \hat{B})$, we obtain

$$\bar{R} = \hat{X}c + (\bar{R} - E[R]) - (\hat{B} - B)b + (\tilde{B} - B)b. \qquad \text{(A11)}$$

Since $\hat{X}\hat{Q}'\hat{X}$ is assumed to be non-singular, we can multiply the above equation by $\hat{X}'\hat{Q}\hat{X}^{-1}\hat{X}'\hat{Q}$ to obtain the following formula for the estimator of the cross-sectional regression

$$\hat{c} = c + (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}[(\bar{R} - E[R]) - (\hat{B} - B)b] + (\hat{X}'\hat{Q}\hat{X})^{-1}\hat{X}'\hat{Q}(\tilde{B} - B)b. \tag{A12}$$

Since the time series of returns R_t and factors y_t is assumed to be stationary and ergodic, and since X has full rank, it follows from equation (A12) that

$$\lim \hat{c} = c + (X'QX)^{-1}X'Q(\tilde{B} - B)b. \tag{A13}$$

Proof of Theorem 4: Notice that $\hat{X} = (\iota, \hat{\beta}_1)$. Since $\hat{\Phi}$ converges to Φ in probability, the lower-right element v_{22} of matrix $\hat{V}$ in equation (32) is asymptotically equivalent to

$$[N\hat{\beta}_1'\hat{\beta}_1 - (\iota'\hat{\beta}_1)^2]^{-2}[(\iota'\hat{\beta}_1)^2(\iota'\Phi\iota) + N^2(\hat{\beta}_1'\Phi\hat{\beta}_1) - 2N(\iota'\hat{\beta}_1)(\hat{\beta}_1'\Phi\iota)]. \tag{A14}$$

It follows from equation $\beta_1 = 0$ that

$$\hat{b}_1 = b_1 + [N\hat{\beta}_1'\hat{\beta}_1 - (\iota'\hat{\beta}_1)^2]^{-1} \times [N\hat{\beta}_1' - (\iota'\hat{\beta}_1)\iota'][(\bar{R} - E[R]) - \hat{\beta}_1 b_1 + \tilde{\beta}_1 b_1]. \tag{A15}$$

The t-value for the slope of β_1 can be written as $T^{1/2}\hat{v}_{22}^{-1/2}\hat{b}_1$. If we let $\vec{\beta}_1 \equiv T^{1/2}\hat{\beta}_1$ and $\vec{\mu} \equiv T^{1/2}(\bar{R} - E[R])$, it follows from equation (A14) that the t-value for the slope of β_1 is asymptotically equivalent to

$$t = Q_1(\vec{\beta}_1, \vec{\mu}) + Q_2(\vec{\beta}_1, \vec{\mu}) \times b_1 T^{1/2}, \tag{A16}$$

where the functions $Q_1(\cdot,\cdot)$ and $Q_2(\cdot,\cdot)$ from $\mathcal{R}^{2N}$ to $\mathcal{R}^1$ are defined as

$$Q_1(\vec{\beta}_1, \vec{\mu}) = [(\iota'\vec{\beta}_1)^2\iota'\Phi\iota + N^2\vec{\beta}_1'\Phi\vec{\beta}_1 - 2N(\iota'\vec{\beta}_1)\vec{\beta}_1'\Phi\iota]^{-1/2} \times \{[N\vec{\beta}_1'\vec{\beta}_1 - (\iota'\vec{\beta}_1)^2]b_1 + [N\vec{\beta}_1' - (\iota'\vec{\beta}_1)\iota'][\vec{\mu} - \vec{\beta}_1 b_1]\} \tag{A17}$$

$$Q_2(\vec{\beta}_1, \vec{\mu}) = [(\iota'\vec{\beta}_1)^2\iota'\Phi\iota + N^2\vec{\beta}_1'\Phi\vec{\beta}_1 - 2N(\iota'\vec{\beta}_1)\vec{\beta}_1'\Phi\iota]^{-1/2} \times [N\vec{\beta}_1' - (\iota'\vec{\beta}_1)\iota']\tilde{\beta}_1 \tag{A18}$$

when $\vec{\beta} \neq 0$. If $\vec{\beta} = 0$, we define Q_1 and Q_2 to be zero. Under the assumption that returns and factors are i.i.d. with finite fourth moments, the vector $(\vec{\beta}_1', \vec{\mu}')'$ converges in distribution to a vector of normal random variables with zero mean. Let us denote the limiting variable by $(\xi', \eta')'$. The set of discontinuity points of Q_1 and Q_2 is

$$\mathcal{D} \equiv \{(\xi', \eta')' \in \mathcal{R}^{2N} : (\iota'\xi)^2\iota'\Phi\iota + N^2\xi'\Phi\xi - 2N(\iota'\xi)\xi'\Phi\iota = 0\}, \tag{A19}$$

which is obviously a closed subset in $\mathcal{R}^{2N}$. Since ξ has a normal distribution, the probability for the random variable to fall in the set $\mathcal{D}$ is zero. It then follows from Theorem 3.2.5 in Amemiya (1985) that $Q_1(\vec{\beta}_1, \vec{\mu})$ and $Q_2(\vec{\beta}_1, \vec{\mu})$

converge in distribution to $Q_1(\xi,\eta)$ and $Q_2(\xi,\eta)$. Since $\lim T^{1/2} = +\infty$, it follows from equation (A16) that t converges in probability to infinity if and only if $b_1 \neq 0$, which completes the proof.

Proof of Theorem 5: It follows from $E[R] = b\tilde{\beta}$ and $a = 0$ that

$$\bar{R} = (\bar{R} - E[R]) + (z \quad \hat{\beta})\begin{pmatrix} a \\ b \end{pmatrix} + (\tilde{\beta} - \hat{\beta})b. \tag{A20}$$

Multiplying $(\hat{X}'\hat{X})^{-1}\hat{X}'$ to both sides, we obtain

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix} + (\hat{X}'\hat{X})^{-1}\hat{X}'[(\bar{R} - E[R]) + (\tilde{\beta} - \hat{\beta})b]. \tag{A21}$$

Let $\hat{D} = (z'z)(\hat{\beta}'\hat{\beta}) - (z'\hat{\beta})^2$. Since

$$(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\beta} = \frac{1}{\hat{D}}\begin{pmatrix} \hat{\beta}'\hat{\beta} & -z'\hat{\beta} \\ -z'\hat{\beta} & z'z \end{pmatrix}\begin{pmatrix} z' \\ \hat{\beta}' \end{pmatrix}\hat{\beta} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{A22}$$

we have

$$\begin{aligned} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} &= (\hat{X}'\hat{X})^{-1}\hat{X}'[(\bar{R} - E[R]) + \tilde{\beta}b] \\ &= \frac{1}{\hat{D}}\begin{pmatrix} (\hat{\beta}'\hat{\beta})z' - (z'\hat{\beta})\hat{\beta}' \\ (z'z)\hat{\beta}' - (z'\hat{\beta})z' \end{pmatrix}[(\bar{R} - E[R]) + \tilde{\beta}b]. \end{aligned} \tag{A23}$$

Then, the t-statistic t_a is

$$\begin{aligned} t_a &= \sqrt{T}\hat{v}_{11}^{-1/2}\hat{a} \\ &= \hat{v}_{11}^{-1/2}\hat{D}^{-1}[(\hat{\beta}'\hat{\beta})z' - (z'\hat{\beta})\hat{\beta}']\sqrt{T}(\bar{R} - E[R]) \\ &\quad + \sqrt{T}\hat{v}_{11}^{-1/2}\hat{D}^{-1}[(\hat{\beta}'\hat{\beta})(z'\tilde{\beta}) - (z'\hat{\beta})(\hat{\beta}'\tilde{\beta})]b. \end{aligned} \tag{A24}$$

The first term converges in distribution to a normal random variable with zero mean and finite variance. The rest of the proof examines whether the second term converges to a distribution with finite variance.

If $(z'\tilde{\beta})(\beta'\beta) = (z'\beta)(\beta'\tilde{\beta})$, it can be shown that

$$\begin{aligned} &\sqrt{T}[(\hat{\beta}'\hat{\beta})(z'\tilde{\beta}) - (z'\hat{\beta})(\hat{\beta}'\tilde{\beta})] \\ &\quad = [(z'\tilde{\beta})(\hat{\beta}' + \beta') - (z'\beta)\tilde{\beta}' - (\hat{\beta}'\tilde{\beta})z']\sqrt{T}(\hat{\beta} - \beta), \end{aligned} \tag{A25}$$

which converges in distribution to

$$[2(z'\tilde{\beta})\beta' - (z'\beta)\tilde{\beta}' - (\beta'\tilde{\beta})z'] \lim \sqrt{T}(\hat{\beta} - \beta).$$

This expression has a normal distribution. Therefore, t_a converges in distribution to a normal random variable with finite variance.

If $(z'\tilde{\beta})(\beta'\beta) \neq (z'\beta)(\beta'\tilde{\beta})$, then we have

$$\begin{aligned}&\lim \sqrt{T}\hat{v}_{11}^{-1/2}\hat{D}^{-1}[(\hat{\beta}'\hat{\beta})(z'\tilde{\beta}) - (z'\hat{\beta})(\hat{\beta}'\tilde{\beta})]b \\ &\quad = (\lim \sqrt{T})\{v_{11}^{-1/2}D^{-1}[(\beta'\beta)(z'\tilde{\beta}) - (z'\beta)(\beta'\tilde{\beta})]b\} = \infty,\end{aligned} \tag{A26}$$

which implies that $\lim P[|t_a| > \delta] = 1$ for any $\delta > 0$. This completes the proof. Q.E.D.

Proof of Theorem 6: Let $\hat{X} = (z, \hat{\beta})$. Since $\hat{\Phi}$ converges to Φ in probability, the upper-left element v_{11} of matrix $\hat{V}$ in equation (40) is asymptotically equivalent to

$$[(z'z)(\hat{\beta}'\hat{\beta}) - (z'\hat{\beta})^2]^{-2}\,[(\hat{\beta}'\hat{\beta})^2(z'\Phi z) + (z'\hat{\beta})^2(\hat{\beta}'\Phi\hat{\beta}) - 2(\hat{\beta}'\hat{\beta})(z'\hat{\beta})(\hat{\beta}'\Phi z)]. \tag{A27}$$

Using the same derivation as in the proof of Theorem 5, the t-statistics for the slope of z can be shown to be

$$\begin{aligned}t_a &= \hat{v}_{11}^{-1/2}\hat{D}^{-1}[(\hat{\beta}'\hat{\beta})z' - (z'\hat{\beta})\hat{\beta}']\sqrt{T}(\bar{R} - E[R]) \\ &\quad + \sqrt{T}\hat{v}_{11}^{-1/2}\hat{D}^{-1}[(\hat{\beta}'\hat{\beta})(z'\tilde{\beta}) - (z'\hat{\beta})(\hat{\beta}'\tilde{\beta})]b.\end{aligned} \tag{A28}$$

If we substitute expression (A27) for $\hat{v}_{11}$ and let $\vec{\beta}_1 \equiv T^{1/2}\hat{\beta}_1$ and $\vec{\mu} \equiv T^{1/2}(\bar{R} - E[R])$, the t-value for the slope of z is asymptotically equivalent to

$$t_a = Q_1(\vec{\beta}, \vec{\mu}) + Q_2(\vec{\beta}, \vec{\mu})\cdot b_1 T^{1/2}, \tag{A29}$$

where the functions $Q_1(\cdot,\cdot)$ and $Q_2(\cdot,\cdot)$ from $\mathcal{R}^{2N}$ to $\mathcal{R}^1$ are defined as

$$\begin{aligned}Q_1(\vec{\beta}, \vec{\mu}) &= [(\vec{\beta}'\vec{\beta})^2(z'\Phi z) + (z'\vec{\beta})^2(\vec{\beta}'\Phi\vec{\beta}) - 2(\vec{\beta}'\vec{\beta})(z'\vec{\beta})(\vec{\beta}'\Phi z)]^{-1/2} \\ &\quad \times [(\vec{\beta}'\vec{\beta})z' - (z'\vec{\beta})\vec{\beta}']\vec{\mu}\end{aligned} \tag{A30}$$

$$\begin{aligned}Q_2(\vec{\beta}_1, \vec{\mu}) &= [(\vec{\beta}'\vec{\beta})^2(z'\Phi z) + (z'\vec{\beta})^2(\vec{\beta}'\Phi\vec{\beta}) - 2(\vec{\beta}'\vec{\beta})(z'\vec{\beta})(\vec{\beta}'\Phi z)]^{-1/2} \\ &\quad \times (z'\vec{\beta})\vec{\beta}'[(\vec{\beta}/\sqrt{T}) - \tilde{\beta}]b\end{aligned} \tag{A31}$$

when $\vec{\beta} \neq 0$. If $\vec{\beta} = 0$, we define Q_1 and Q_2 to be zero. The vector $(\vec{\beta}', \vec{\mu}')'$ converges in distribution to a vector of normal random variables with zero mean. Let us denote the limiting variable by $(\xi', \eta')'$. The set of discontinuity points of Q_1 and Q_2 is

$$\mathcal{D} \equiv \{(\xi', \eta')' \in \mathcal{R}^{2N} : (\xi'\xi')^2(z'\Phi z) + (z'\xi)^2(\xi'\Phi\xi) - 2(\xi'\xi)(z'\xi)(\xi'\Phi z) = 0\}, \tag{A32}$$

which is obviously a closed subset in $\mathcal{R}^{2N}$. Since ξ has a normal distribution, the probability for the random variable to fall in the set $\mathcal{D}$ is zero. It then follows from Theorem 3.2.5 in Amemiya (1985) that $Q_1(\vec{\beta}, \vec{\mu})$ and $Q_2(\vec{\beta}, \vec{\mu})$ converge in distribution to $Q_1(\xi, \eta)$ and $Q_2(\xi, \eta)$. Since $\lim T^{1/2} = +\infty$, it follows from equation (A29) that t_a converges in probability to infinity if and only if $b_1 \neq 0$, which completes the proof.

Proof of Theorem 7: Similar to the proof of Theorem 1.

REFERENCES

Amemiya, Takeshi, 1985, *Advanced Econometrics* (Harvard University Press, Cambridge, Mass.).

Barone-Adesi, Giovanni, and Prem Talwar, 1983, Market models and heteroscedasticity of security returns, *Journal of Business and Economic Statistics* 1, 163–168.

Banz, Rolf W., 1981, The relationship between returns and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Berk, Jonathan B., 1995, A critique of size-related anomalies, *Review of Financial Studies* 8, 275–286.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael C. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).

Blattberg, Robert C., and Nicholas J. Gonedes, 1974, A comparison of the stable and student distributions as statistical models of stock prices, *Journal of Business* 47, 244–280.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.

Chan, Louis, Yasushi Hamao, and Josef Lakonishok, 1991, Fundamentals and stock returns in Japan, *Journal of Finance* 46, 1739–1764.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–404.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation of stock returns, *Journal of Finance* 52, 1–33.

Davis, James L., 1994, The cross-section of realized stock returns: The pre-COMPUSTAT evidence, *Journal of Finance* 49, 1579–1593.

Dybvig, Philip H., and Jonathan E. Ingersoll, 1982, Mean-variance theory in complete markets, *Journal of Business* 55, 233–252.

Fama, Eugene F., 1965, The behavior of stock market prices, *Journal of Business* 38, 34–105.

Fama, Eugene F., and Kenneth French, 1992, The cross-section of expected sock returns, *Journal of Finance* 47, 427–466.

Fama, Eugene F., and Kenneth French, 1993, Common risk factors in the returns on bonds and stocks, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press, Princeton, NJ).

Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hansen, Lars P., and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hansen, Lars P., and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Hansen, Lars P., and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jagannathan, Ravi, and Zhenyu Wang, 1998, A note on the asymptotic covariance in Fama–MacBeth regression, *Journal of Finance* 53, 799–801.

Jegadeesh, Narasimhan, 1992, Does market risk really explain the size effects?, *Journal of Financial and Quantitative Analysis* 27, 337–351.

Kan, Raymond, and C. Zhang, 1997, Two-pass tests of asset pricing models with useless factors, *Journal of Finance*, forthcoming.

Kandel, Shmuel, and Robert F. Stambaugh, 1995, Portfolio efficiency and the cross section of expected returns, *Journal of Finance* 50, 157–184.

Kothari, S. P., Jay Shanken, and Richard G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 50, 185–224.

MacKinlay, A. Craig, and Mathew P. Richardson, 1991, Using generalized method of moments to test mean-variance efficiency, *Journal of Finance* 46, 551–527.

Roll, Richard, 1977, A critique of the asset pricing theory's tests; Part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129–176.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Schwert, G. William, and Paul Seguin, 1990, Heteroscedasticity in stock returns, *Journal of Finance* 45, 1129–1155.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.