ON THE ESTIMATION OF BETA-PRICING MODELS

by

Jay Shanken
University of Rochester

MR 91-06

March, 1991

Send reprints to Jay Shanken, Simon School of Business,
University of Rochester, Rochester NY 14627.

This paper is an extension of the second chapter of my doctoral dissertation at Carnegie-Mellon University. Recent versions were presented in seminars at Washington University, the University of Iowa, the University of Pennsylvania, Concordia University, Cornell University, the 1989 American Finance Association meetings in Atlanta, and the First International Conference of the Center for Research in Finance-IMI Group in Rome, Italy. I am grateful to seminar participants and to J. Bodurtha, N.F. Chen, P. Dybvig, W. Ferson, M. Gibbons, C. Harvey, P. Kearns, L. Kryzanowski, S. Y. Hu, G. W. Schwert, A. Sweeney, M. Weinstein, two anonymous referees, and especially my dissertation advisor, R. Thompson, for helpful comments and discussions.

Abstract

This paper presents an integrated econometric view of maximum likelihood methods and more traditional two-pass approaches to estimating beta-pricing models. Several aspects of the well-known "errors-in-variables problem" are considered, and an earlier conjecture concerning the merits of simultaneous estimation of beta and price of risk parameters is evaluated. The traditional inference procedure is found, under standard assumptions, to overstate the precision of price of risk estimates and an asymptotically valid correction is derived. Modifications to accommodate serial correlation in marketwide factors are also discussed.

On the Estimation of Beta-Pricing Models

Sharpe (1964) and Lintner (1965) demonstrate that, in equilibrium, a financial asset's expected return must be positively linearly related to its "beta", a measure of systematic risk or comovement with the market portfolio return:[1]

$$E_i = \gamma_0 + \gamma_1\beta_i \qquad \text{for all assets i,} \tag{1}$$

where

E_i = the expected return on asset i,

β_i = asset i's market beta,

γ_0 = the expected return on a "zero-beta" portfolio,

and

γ_1 = the market risk premium.

Since β_i is the slope coefficient from the regression of asset i's returns on those of the market portfolio, it seems natural to estimate the risk-return relation of this capital asset pricing model (CAPM) in two stages. In the first pass, beta estimates are obtained from separate time-series regressions for each asset. Mean returns are computed for each asset as well, and γ_0 and γ_1 are then estimated from a second-pass cross-sectional regression (CSR):

$$\bar{R}_i = \gamma_0 + \gamma_1\hat{\beta}_i + \eta_i. \tag{2}$$

This approach is adopted by Lintner, who includes a measure of residual risk in the CSR's in order to test the hypothesis that "only systematic risk matters."[2] Inference is complicated, for security returns are cross-sectionally correlated and heteroskedastic. Therefore, conclusions based on the usual ordinary least squares (OLS) standard errors for the regression (2) can be misleading.[3]

Fama and MacBeth (1973) introduce an alternative technique for estimating the precision of CSR estimates.[4] They run a CSR of returns on beta estimates in each month of the sample period:

$$R_{it} = \gamma_{0t} + \gamma_{1t}\hat{\beta}_{it} + \eta_{it}. \tag{3}$$

The "rolling" betas are estimated from several years of data prior to each CSR. In this way, a time series of estimates is generated for γ_j, $j=0,1$. The sample mean of these $\hat{\gamma}_{jt}$'s is taken as the final estimate and standard errors are computed in the usual manner, as if the series of estimates is independent and identically distributed with mean γ_j. The independence assumption is not strictly satisfied, however, due to measurement error in the overlapping beta estimates. Nonetheless, this time series procedure is an improvement over the naive OLS approach since the nonscalar covariance structure of returns is reflected in the variance of the monthly estimates.

Despite the fundamental role played by the two-pass methodology in modern asset-pricing empirical work, not much is known about its statistical properties. Since the independent variable in the CSR is measured with error, the second-pass estimator is subject to an errors-in-variables (EIV) problem, rendering it biased in small samples. If a single beta is estimated for each asset over the entire period, however, measurement error in beta declines as the time-series sample size, T, increases. In this case, we show that the second-pass estimator is T-consistent; i.e., it converges to the true value of gamma as $T \rightarrow \infty$.

The fact that estimation error in the betas disappears in the limit does not mean it can be ignored, even in large-sample inference. What matters is the rate of convergence, as reflected in the asymptotic variance of the second-pass estimator. The asymptotic variance is derived here, permitting an evaluation of the validity of the traditional inference procedure.[5] We find that the Fama-MacBeth time-series procedure for computing standard errors fails to reflect measurement error in the betas and overstates the precision of the gamma estimates. Examples given in the paper indicate that adjustments for this problem can be important in practice.

Given the limiting distribution of the second-pass estimator, as $T \rightarrow \infty$, its precision can be compared to that of the maximum likelihood estimation (MLE) techniques used by Gibbons (1982) and Stambaugh (1982). Not surprisingly, for a given set of assets the commonly used OLS version of the second-pass estimator need not be asymptotically efficient. However, when an estimator of the covariance matrix of returns is incorporated in a GLS version of the second-pass estimator, two-pass and (efficient) MLE methods are asymptotically equivalent. In fact, the GLS estimator is identical to the Gauss-Newton estimator that Gibbons (1982) uses in actual calculations.[6]

Gibbons emphasizes that use of the MLE approach eliminates the EIV problem since the gammas and betas are estimated simultaneously. Apparently, this advantage is lost in going to the linearized system since, as just noted, the resulting Gauss-Newton estimator turns out to be a second-pass regression estimator and thus subject to an EIV bias. Although

Gibbons' claim for true MLE is intuitively reasonable, the precise sense in which the problem is eliminated is not clear, since the standard asymptotic properties of this approach (as $T \rightarrow \infty$) are shared by the second-pass GLS regression estimator. To gain further insight, we look at the behavior of the estimators when the number of securities is large.

Given some restrictions on the covariance matrix of returns, classical EIV analysis implies that the usual (OLS) second-pass estimator is not N-consistent; i.e., it is not consistent when the time-series length is fixed and the number of securities, N, is allowed to increase without bound.[7] Aware of this problem, Black, Jensen, and Scholes (1972), Blume and Friend (1973), and Fama and MacBeth (1973) employ elaborate portfolio grouping procedures designed to minimize measurement error while maintaining considerable cross-sectional variation in the portfolio betas. Assuming independence of the market model residuals, Black, Jensen, and Scholes show that the second-pass estimator based on a fixed number of well-diversified asset portfolios is N-consistent.

If there is an inherent advantage to maximum likelihood estimation, then one might expect it to be manifested analytically in terms of the N-consistency property, even without portfolio grouping. To explore this possibility, we consider a framework in which the residual covariance matrix is restricted, since the unrestricted covariance matrix estimate is not invertible when the number of assets is sufficiently large. This scenario is challenging econometrically, given that standard results for maximum likelihood estimation apply as the time series increases, but not as the number of assets increases.[8] It is also interesting economically in light of the growing literature on asset pricing in infinite asset or sequential economies, the pioneering work being that of Ross (1976) on the arbitrage pricing theory (APT). We find that MLE does possess the desired N-consistency property, thus providing some formal support for Gibbons' intuitive assertion.

Still another approach to estimation of the risk-return relation is employed by Litzenberger and Ramaswamy (1979). They adopt a standard EIV solution that amounts to a simple modification of the second-pass estimator.[9] Hess (1980) questions the value of this procedure, though, emphasizing that it assumes the (true) residual covariance matrix is diagonal and known exactly. Nonetheless, we show that a similar estimator is N-consistent with these conditions relaxed. Furthermore, the asymptotic distribution of the estimator, as $T \to \infty$, is unaltered by the EIV modification. Thus, on the basis of the econometric properties considered here, we cannot distinguish between this modified version of the two-pass methodology and (simultaneous) MLE. Each approach constitutes a promising alternative to the traditional portfolio grouping procedure as a means of dealing with EIV bias.

Having briefly reviewed the literature and summarized our main results, we now outline the remainder of the paper. Section 1 introduces some notation and describes the multifactor asset pricing framework in which our analysis is conducted. A generalized version of the two-pass methodology is developed in Section 2, integrating elements from the classic studies by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973). The asymptotic properties of this methodology are explored in Section 3 and illustrated through examples in Section 4. Section 5 discusses the relation between traditional two-pass techniques and the more recently employed MLE approach, while Section 6 considers solutions to the small-sample EIV problem when "many" assets are available. Some concluding observations are offered in Section 7. Formal proofs and a simple heuristic for dealing with "rolling" beta estimates are given in the Appendix.

1. Assumptions and Notation

1.1 The econometric specification

Although early applications of the two-pass methodology focused on pricing with respect to a single market index, the procedure has been used more recently in the estimation of multifactor pricing relations of the form

$$E_i = \gamma_0 + b_{i1}\gamma_1 + b_{i2}\gamma_2 \qquad i=1,\dots,N \qquad (4)$$

where $E_i = E(R_{it})$ is the expected return on asset i and

$$R_{it} = a_i + b_{i1}F_{1t} + b_{i2}F_{2t} + \epsilon_{it} \qquad t=1,\dots,T \qquad (5)$$

F_{1t} is a k_1-vector of general factors and F_{2t} is a k_2-vector of factors that are portfolio returns. For example, F_{1t} might contain innovations in certain macroeconomic variables whereas F_{2t} could include stock and bond index returns. Many studies in the literature do not incorporate the additional pricing restriction that is implied (see below) when a given factor is a portfolio return. Such a factor would formally be viewed here as an element of the F_1 vector. The b_i's are row vectors of factor model regression coefficients or systematic risk measures, while γ_1 and γ_2 are the associated "price of risk" vectors of dimension k_1 and k_2, respectively. The ϵ_{it}'s are factor model disturbances.

The econometric analysis will be facilitated by the use of matrix notation. Let R_t (E) be the vector of returns (expected returns) on the N assets for period t, and ε_t the N-vector of residual returns. "Vectors" are column vectors unless noted otherwise. B_1 and B_2 are Nxk_1 and Nxk_2 matrices of regression coefficients associated with the factors F_1 and F_2, respectively, and a is the N-vector of factor model intercepts. Equations (4) and (5) can now be written as

$$E = [1_N : B]\Gamma \qquad (6)$$

and

$$R_t = a + BF_t + \varepsilon_t \qquad t=1,\ldots,T \qquad (7)$$

where

$$B = [B_1 : B_2] \qquad Nxk$$

$$F_t' = (F_{1t}' \quad F_{2t}') \qquad 1xk$$

and

$$\Gamma' = (\gamma_0, \gamma_1', \gamma_2') \qquad 1x(k+1)$$

with $k = k_1+k_2$.

If the factor portfolios also satisfy the linear risk-return relation (4), then

$$\gamma_2 = E(F_2) - \gamma_0 1_{k_2}, \qquad (8)$$

i.e., the associated prices of risk equal the expected portfolio returns in excess of the zero-beta rate.[10] Averaging (7) over time, imposing (6) and (8), and noting that $E = a + B_1E(F_1) + B_2E(F_2)$, yields[11]

$$\bar{R} = [1_N : B]\bar{\Gamma} + \bar{\varepsilon}, \tag{9}$$

where

$$\bar{\Gamma}' = (\gamma_0, \bar{\gamma}_1', \bar{\gamma}_2'),$$

$$\bar{\gamma}_1 = \gamma_1 + \bar{F}_1 - E(F_1),$$

and

$$\bar{\gamma}_2 = \gamma_2 + \bar{F}_2 - E(F_2) = \bar{F}_2 - \gamma_0 1_{k_2}.$$

By (9), expected return is still linear in the asset betas conditional on the factor outcomes. The random coefficient vectors in this relation, $\bar{\gamma}_1$ and $\bar{\gamma}_2$, are referred to, accordingly, as "ex post prices of risk." They equal the ex ante prices plus the (unconditionally) unexpected factor outcomes. In the factor portfolio case, this reduces to an ex post mean excess return. Given (9), cross-sectional covariation in returns arises either from the marketwide random coefficients or from covariation between the factor model disturbances. Note, also, that the variance of each ex post price of risk is just the variance of the corresponding factor sample mean.

The following assumptions are used at various points in the analysis below.

Assumption 1. The vector ε_t is independently and identically distributed over time, conditional on (the time series of values for) F, with

$$E(\varepsilon_t \mid F) = 0$$

and

$$\mathrm{Var}(\varepsilon_t \mid F) = \Sigma \quad \text{(rank N)}$$

Assumption 2. F_t is generated by a stationary process such that the first and second sample moments converge in probability, as $T \to \infty$, to the true moments which are finite. Also, $\bar{F}$ is asymptotically normally distributed.

Although often used in the literature, Assumption 1 is a strong condition which will be relaxed in future work.

Some additional notation will be used throughout the paper. In general, a "^" over a parameter indicates an estimator of that parameter. In particular, $\hat{B}$ and $\hat{\Sigma}$ are the usual unbiased multivariate linear regression estimators of B and Σ, the matrix of factor model regression coefficients and residual covariance matrix, respectively. These estimators are obtained from separate OLS time-series regressions for each asset and cross-products of the OLS residuals. Σ_F ($\hat{\Sigma}_F$) is the population (unbiased sample) covariance matrix of the factors and $S_F = (T-1)T^{-1}\hat{\Sigma}_F$.

The expression "plim_T" is an abbreviation for "the probability limit as $T \to \infty$". "$\to_T$" indicates convergence in distribution as $T \to \infty$. The "asymptotic covariance matrix" of an estimator $\hat{\theta}$, for a parameter θ, refers to the covariance matrix of the limiting distribution of $\sqrt{T}(\hat{\theta}-\theta)$. In particular, $\Sigma_{\bar{F}}$ denotes the asymptotic covariance matrix of $\bar{F}$ given by Assumption 2 above. With serially uncorrelated factors, the covariance matrix of $\bar{F}$ is Σ_F/T; hence the asymptotic covariance matrix is Σ_F. Finally,

$X \otimes Y$ denotes the Kronecker product of two matrices, X and Y, whereby each element of X is multiplied by Y.[12]

1.2 The underlying pricing model

Although the APT often serves as the theoretical motivation for multifactor empirical analyses, the formulation of testable hypotheses requires a richer theoretical model and stronger empirical assumptions than those typically made in the APT literature.[13] The general framework developed in Shanken (1987) supposes that the financial markets are in equilibrium, and that the econometrician observes a "multivariate proxy" for the associated equilibrium benchmark, which is unobservable. Think of the factors, F_1 and F_2 in (5), as the proxy, and consider a hypothetical regression of the equilibrium benchmark on these factors. Let m be the benchmark and ε_m its residual component. Then we have

Proposition 1. The beta pricing relation (4) holds if (i) the benchmark residual, ε_m, is uncorrelated with the returns, R_{it}, $i = 1,\ldots,N$, or equivalently if (ii) the factor model residuals, ε_{it}, $i = 1,\ldots,N$, are uncorrelated with the benchmark.[14]

Condition (i) says that the proxy captures the "relevant" variation in the benchmark, whereas (ii) ensures that the factor model disturbances are not a source of "priced" risk. A sufficient condition for (i) or (ii) to hold is that the benchmark be spanned by the factors. This spanning idea goes back to the multibeta interpretations of the traditional CAPM in Rosenberg and Guy (1976), Ross (1976), and Sharpe (1977), where the market portfolio serves as the benchmark. It also plays an important role, but in reverse, in Breeden's (1979) demonstration that Merton's (1973) multibeta intertemporal CAPM "collapses" into a single-beta consumption model. In this case, unanticipated changes in consumption are spanned by innovations in aggregate wealth and the state variables that describe the opportunity set.

Regarding (i) or (ii) as an implicit joint hypothesis in asset pricing tests is a logical extension of Roll's (1977) important critique of the early CAPM tests. Similar conditions are explicitly introduced in the recent equilibrium versions of Ross's APT, although the empirical significance of these assumptions is sometimes overlooked. For example, Connor (1984) assumes that the market return is spanned by the given factors,[15] whereas Dybvig (1983) and Grinblatt and Titman (1983) make assumptions ensuring that the factor model disturbance for each asset is (approximately) uncorrelated with aggregate wealth.[16]

Another issue that arises in connection with pricing relations like (4) concerns the level of conditioning in the analysis. Equilibrium models typically assume that investors condition on all available information when forming expectations and making decisions. However, the iterated expectation arguments of Hansen and Singleton (1982) and Grossman and Shiller (1982) provide a basis for conditioning on a subset of the information set in empirical work. In this paper we focus on the unconditional joint distribution of returns and factors. Conditional pricing relations will be considered in future work.[17]

While we assume that the relevant unconditional moments are constant, this does not preclude variation in the conditional moments. In particular, Assumption 2 allows for serial correlation in the factors. This could arise, for example, if (conditional) expectations about the factors are changing through time in a persistent manner. Persistent time variation in conditional betas, while not ruled out in theory, would appear to be generally inconsistent with our assumption that the disturbance terms in the unconditional factor model are serially independent.[18]

2. A Generalized Two-Pass Procedure

In this section a general version of the two-pass procedure is developed that, apart from some details of implementation, generalizes the approaches of Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), henceforth BJS and FM. BJS impose the factor-portfolio constraint on the market index in their study, while FM do not. Also, FM employ a series of cross-sectional regressions (CSR's) with rolling betas, while the BJS estimator amounts to a single CSR with each asset's beta estimated using all of the given data.[19] Our analysis follows BJS in this respect. Extensions to deal with rolling betas are considered in the Appendix. Both BJS and FM use time-series procedures to obtain standard errors for the CSR estimates. These procedures are considered later in Section 3.2.

A purported advantage of the "predictive" (rolling beta) approach is that it avoids the problem of a spurious cross-sectional relation arising from statistical dependence between returns and estimated betas.[20] Under assumptions typically made in this literature, though, contemporaneously estimated betas and mean returns are uncorrelated, a point not previously noted. Moreover, assuming joint normality of the asset returns and factors, the estimated betas and mean returns are statistically independent. These observations follow from part (i) of Lemma 1 below.

It may also appear that, unlike the contemporaneous approach, the use of rolling beta estimates allows for changes in the true betas over time. However, by assuming that betas are constant for, say, overlapping five-year periods, one is effectively assuming constancy over the entire period. Nonetheless, it is possible that the predictive approach with annual updating will prove more robust with respect to misspecifications of our underlying assumptions.

Combining the multibeta expected return relation (6) with the restriction on γ_2 in (8) yields[21]

$$E - B_2E(F_2) = X\Gamma_{01}, \qquad (10)$$

where

$$X = [1_N - B_2 1_{k_2} : B_1] \qquad Nx(k_1+1)$$

and

$$\Gamma'_{01} = (\gamma_0, \gamma'_1) \qquad 1x(k_1+1).$$

This relation suggests the following procedure for estimating Γ_{01}; replace B_1 and B_2 with time-series estimates $\hat{B}_1$ and $\hat{B}_2$, respectively, and regress $\bar{R} - \hat{B}_2\bar{F}_2$ on the resulting $\hat{X}$. γ_2 can then be estimated as the difference between the elements of $\bar{F}_2$ and the CSR estimate of γ_0.

If $k_2=0$, so that no factor-portfolio constraints are imposed, the procedure above amounts to a CSR of $\bar{R}$ on $\hat{B}_1$ and a constant. This is similar, apart from estimation of betas, to the approach of FM. If $k_1=0$ then, since $\bar{R} - \hat{B}_2\bar{F}_2 = \hat{a}$, we have a CSR of $\hat{a}$ on $1_N - \hat{B}_2 1_{k_2}$. In the single index case considered by BJS, the corresponding restriction is $a_i = \gamma_0(1-b_{i2})$, the familiar "CAPM" constraint on the market model intercepts.[22]

Although ordinary least squares (OLS) has been employed most often in the literature, weighted least squares (WLS) and generalized least squares (GLS) have also been used in second-pass regressions.[23] Each of these estimators for Γ_{01} is of the general form

$$\hat{\Gamma}_{01} = \hat{A}[\bar{R} - \hat{B}_2\bar{F}_2]$$

with

$$\hat{A}\hat{X} = I_{k_1+1}$$

for some random $(k_1+1)xN$ matrix $\hat{A}$. In particular,

$$\hat{A} = (\hat{X}'\hat{X})^{-1}\hat{X}'$$

in the OLS case while

$$\hat{A} = (\hat{X}'\hat{\Sigma}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Sigma}^{-1}$$

for GLS. The off-diagonal elements of $\hat{\Sigma}$ are set equal to zero in WLS.

The asymptotic analysis of the general form of the second-pass estimator is presented in the next section. We conclude this section with an important lemma that provides the statistical foundation for many of our later results.[24]

Lemma 1. Let B_v ($\hat{B}_v$) be an Nk-vector obtained by transposing the rows of

B ($\hat{B}$) and stacking these N k-vectors on top of one another. Let H be the random (N+Nk+k)-vector $[\bar{\varepsilon}', (\hat{B}_v - B_v)', (\bar{F} - E(F))']'$ and let g be a constant k-vector. Then, given Assumptions 1 and 2,

(i) $\hat{B}$, $\bar{\varepsilon}$, and $\bar{F}$ are uncorrelated unconditionally and conditional on F. If ε_t is jointly normally distributed conditional on F then $\hat{B}$, $\bar{\varepsilon}$, and $\hat{\Sigma}$ are independent conditional on F, while $\bar{\varepsilon}$ and $\hat{\Sigma}$ are independent of $\bar{F}$. If F_t is jointly normally distributed independently over time, then $\hat{B}$ and $\bar{F}$ are also independent.

(ii) $\sqrt{T}H$ converges in distribution, as $T \to \infty$, to a normal distribution with mean zero and a block diagonal covariance matrix with blocks $(\Sigma, \Sigma\otimes\Sigma_F^{-1}, \Sigma_{\bar{F}})$.

(iii) $\sqrt{T}[\hat{B} - B]g$ converges in distribution, as $T \to \infty$, to a normal distribution with mean zero and covariance matrix $(g'\Sigma_F^{-1}g)\Sigma$.

3. Asymptotic Distribution of the Second-Pass Estimator

3.1 The main theorem

In conducting inferences about a pricing theory, we are generally concerned with the variability of the second-pass estimator about the true gamma vector. It is apparent from (9), though, that the CSR can also be viewed as a procedure for estimating the ex post price of risk vector. This might be relevant, for example, in evaluating mutual fund performance. The theorem below considers the asymptotic distribution from both perspectives, yielding some interesting insights.

Theorem 1. Let $\hat{\Gamma}_{01} = \hat{A}[\bar{R} - \hat{B}_2\bar{F}_2]$ with $\hat{A}\hat{X} = I_{k_1+1}$ and $\text{plim}_T\ \hat{A} = A$. Let $\hat{\gamma}_2 = \bar{F}_2 - \hat{\gamma}_0 1_{k_2}$. Then, given Assumptions 1 and 2,

(i) $\hat{\Gamma}$ is a T-consistent estimator for Γ.

(ii) $\sqrt{T}[\hat{\Gamma} - \bar{\Gamma}]$ converges in distribution, as $T \rightarrow \infty$, to a normal distribution with mean zero and covariance matrix $(1+c)\Omega$ (dimension $k+1$), where $c = \Gamma_{12}'\Sigma_{\bar{F}}^{-1}\Gamma_{12}$. The upper left block of Ω equals $A\Sigma A'$ (dimension k_1+1). Let v be the first column of this block and v_0 the first element of v. The lower right block of Ω (dimension k_2) equals $v_0 1_{k_2} 1_{k_2}'$ and the upper right block is a $(k_1+1) \times k_2$ matrix with each column equal to $-v$.[25]

(iii) $\sqrt{T}[\hat{\Gamma} - \Gamma]$ converges in distribution, as $T \rightarrow \infty$, to a normal distribution with mean zero and covariance matrix $(1+c)\Omega + \Sigma_{\bar{F}}^{*}$ (dimension $k+1$), where $\Sigma_{\bar{F}}^{*}$ (the "bordered version" of $\Sigma_{\bar{F}}$) consists of a top row and left column of zeroes, with $\Sigma_{\bar{F}}$ (dimension k) in the lower right block.

Several aspects of Theorem 1 deserve emphasis. In standard regression analysis, the condition $AX = I$ is necessary and sufficient for the corresponding linear regression estimator to be unbiased. In the present EIV context, this condition is equivalent to the T-consistency of the second-pass regression estimator.[26]

The matrix Ω captures the residual component of variance for the second-pass estimator. The sub-matrix $A\Sigma A'$ is the usual expression for the covariance matrix of a linear regression estimator. It measures the variation of $\hat{\Gamma}_{01}$ about the ex post parameter vector $\bar{\Gamma}_{01}$, ignoring estimation error in the betas. Similarly, other elements of Ω reflect the variation of $\hat{\gamma}_2$ around $\bar{\gamma}_2$. Given the factor-portfolio constraint (8), this is just the variation in $\hat{\gamma}_0$.

When estimating the ex ante price of risk, the factor-related variation in the ex post price must also be considered. As indicated in (iii), this factor component of variance is measured by the variance of the factor sample mean. If the factors are serially uncorrelated, the asymptotic covariance matrix of $\bar{F}$ is just Σ_F. More generally, an "autocorrelation-consistent" estimate of $\Sigma_{\bar{F}}$ can be obtained using time-series procedures that account for serial dependence.[27] Note that the variance of $\hat{\gamma}_0$ has no factor-related component, as this estimator equals the return on a portfolio with betas of zero (asymptotically) on each factor.[28]

The scalar c is an asymptotic adjustment for EIV and $c\Omega$ is the corresponding EIV component of variance. The adjustment reflects the fact that the variance of the beta estimates is directly related to residual variance and inversely related to factor variability. In addition, since the contribution of systematic risk to expected return is proportional to the price of risk, so is the impact of measurement error in beta. Therefore, the variance adjustment increases with the squared price of risk. The precise expression for c follows from part (iii) of Lemma 1 and corresponds to a linear combination of the measurement errors, with weights equal to the respective prices of risk. Given a single factor-portfolio, c is just the squared value of the well-known Sharpe measure of performance.

Consider the scenario in which Σ is a scalar matrix, $\sigma^2 I_N$, and $A = (X'X)^{-1}X'$. In this case, $A\Sigma A'$ equals $\sigma^2(X'X)^{-1}$, the "standard" expression for the covariance matrix of an OLS estimator. Given the observations above, "t-statistics" naively computed from this formula must overstate the statistical significance of the second-pass OLS estimates. Estimation error in the betas and factor-related variation in the price of risk estimator are both ignored by this approach. Recognition of the latter (random coefficients) problem by BJS led to the development of the time-series methodology which we now consider.

3.2 The traditional time-series approach

In our general framework, the time-series methodology entails running a CSR of $R_t - \hat{B}_2F_{2t}$ on $\hat{X}$ each period, thereby generating a series of estimates for Γ_{01}. As earlier, the regression for a given period also provides an estimate of that period's ex post price of risk, i.e., the ex ante price plus the (unconditionally) unexpected factor outcome. Taking the mean of the series yields the CSR estimator $\hat{\Gamma}_{01}$, computed with $\bar{R} - \hat{B}_2\bar{F}_2$ as the dependent variable, as in Theorem 1. Inferences about the mean have traditionally been based on the sample variance of the estimates. Our next result shows that this procedure neglects the EIV problem and may not accurately reflect the factor-related variation in the estimator.

Theorem 2. Let $\hat{\Gamma}_{01t} = \hat{A}[R_t - \hat{B}_2F_{2t}]$ $t=1,\ldots,T$ with $\hat{A}\hat{X} = I_{k_1+1}$ and $\text{plim}_T \hat{A} = A$. Let $\hat{\gamma}_{2t} = F_{2t} - \hat{\gamma}_{0t}1_{k_2}$. Then, given Assumptions 1 and 2, the sample covariance matrix, $\hat{W}$, of the $\hat{\Gamma}_t$ series converges in probability, as $T \to \infty$, to $W = \Omega + \Sigma_F^*$, where Σ_F^* is the bordered version of Σ_F. Thus, a T-consistent estimator for the asymptotic covariance matrix in (iii) of Theorem 1 is given by

$$(1+\hat{c})[\hat{W} - \hat{\Sigma}_F^*] + \hat{\Sigma}_{\overline{F}}^* \tag{11}$$

where $\hat{c} = \hat{\Gamma}_{12}'\hat{\Sigma}_F^{-1}\hat{\Gamma}_{12}$ and $\hat{\Sigma}_{\overline{F}}^*$ is an autocorrelation-consistent estimator for $\Sigma_{\overline{F}}^*$.

Consider the usual case in which the factors are assumed to be serially independent, so that $\Sigma_{\bar{F}} = \Sigma_F$. It is straightforward to show that, if the true betas were used in the second-pass CSR, the covariance matrix of the CSR estimator would equal W, divided by T.[29] This is intuitively clear, given that W is obtained by setting the EIV adjustment term, c, equal to zero in (iii) of Theorem 1. Since the asymptotic covariance matrix exceeds W by a positive definite matrix when $c>0$, the traditional time-series approach generally overstates the precision of the second-pass estimator. The problem is easily corrected, however, using the consistent covariance estimator in (11).

There is one important context in which the EIV adjustment is not needed, as c equals zero. This is when testing the joint hypothesis that all prices of risk equal zero, i.e., $\Gamma_{12} = 0$. If any element of Γ_{12} is nonzero, however, then the asymptotic variance of every element of $\hat{\Gamma}$ is affected, since $c>0$ in this case. Thus, in testing whether a particular factor is priced in a multifactor analysis, the EIV adjustment must be included. Also, the adjustment is needed in constructing asymptotically valid confidence intervals for the gamma parameters.

If the factors are serially dependent, there is an additional misstatement of precision when using the traditional procedure, since $\Sigma_{\bar{F}}$ need not equal Σ_F. This is due to the fact that the variance of an average is not simply the population variance divided by the number of observations, but depends on the covariances as well. Positive autocorrelation will cause the traditional procedure to overstate the precision (understate the variance) of the CSR estimator while negative correlation will have the opposite effect. This is considered further in Section 4.2.

3.3 Expanding the risk-return relation

In some applications, the risk-return relation (4) may be expanded to include additional variables like firm size or variance of return. Suppose there are k_3 such variables. Let Z be the corresponding $N x k_3$ matrix of values for the N assets and γ_3 a k_3-vector of associated expected return premia. We assume Z is constant over time and measured without error.[30] In this case, Theorems 1 and 2 remain valid provided that $X = [1_N - B_2 1_{k_2} : B_1 : Z]$, Γ (and $\bar{\Gamma}$) are expanded to include γ_3, Γ_{01} is replaced by Γ_{013}, and Σ_F^* is supplemented by k_3 rows and columns of zeroes for entries associated with γ_3.

The expanded risk-return relation provides a simple framework in which to test the beta-pricing hypotheses (4) and (8). The joint hypothesis is rejected if any of the additional variables is "priced", i.e., if the CSR estimate of γ_3 is significantly different from zero. Note that adjustment of standard errors for EIV is necessary in this context since Γ_{12} need not equal zero under the null hypothesis.

There is a sense in which the EIV adjustment is greatest when the pricing hypotheses (4) and (8) hold and relatively less important when γ_3 is nonzero for some additional variable. Suppose all of the factors are portfolio returns and a riskless asset exists. Thus, we can talk about the minimum-variance frontier generated by the factors. The quadratic, c, is the square of the slope of the familiar tangent to this frontier, drawn from the riskless asset in mean-standard deviation space.[31]

If (4) and (8) hold, then the factor tangency portfolio will also be the tangency portfolio for the frontier constructed from the larger asset universe that includes the N assets as well as the k factor portfolios. In this case, c is the maximum squared Sharpe measure (ratio of risk premium to standard deviation of return) over all portfolios of the N+k investments. On the other hand, if γ_3 is nonzero then some such portfolio will dominate the factor tangency portfolio, and c will be less than this maximum possible value.

4. Empirical Examples

4.1 EIV-adjusted standard errors

In this section we take a break from the econometric analysis and briefly look at some data. This will serve to illustrate the previous theorems and also provide a sense of the relative magnitudes of the different components of variance discussed earlier. Both studies cited below employ OLS in the second-pass regressions. The betas for a given monthly CSR are estimated using five (or more) years of prior data and are updated annually. Neither study incorporates factor-portfolio constraints. For simplicity, we ignore the slight modification, suggested in the Appendix, for dealing with rolling betas. Alternatively, the calculations below can be viewed as hypothetical illustrations of the results in Section 3 with representative parameter values.

In their classic paper, FM consider a single-factor model, with the return on the equally-weighted CRSP stock index as a proxy for the market portfolio. Several specifications are explored, both with and without the additional variables beta-squared and residual variance. For the overall period from 1935 to June 1968, the estimated market prices of risk range from .72 (percent per month) to 1.14. The standard deviation of return on the index for this period is 6.11. Therefore, the largest EIV adjustment (see Theorem 1) is $(1.14/6.11)^2 = .035$. Clearly, this will not have much effect on the inferences drawn.

Chen, Roll and Ross (1986), henceforth CRR, present empirical results for a number of multifactor models. We shall focus on the estimates for a five-factor model in Panel B of Table 4 of their paper. Two of the five factors are return related, two others involve inflation, and the remaining factor is a growth rate of industrial production.[32] Combining the CRR price of risk estimates for the period 1958-1984 with the sample covariance matrix of the factors, yields an EIV adjustment term, $\hat{c}$, equal to .36.[33] Recall that c is directly related to the magnitude of each price of risk and inversely related to factor variance. On the other hand, the t-statistic for a given price coefficient is a ratio of the estimated price to its standard error, one determinant of which is the corresponding factor variation. Thus, the fact that $\hat{c}$ is much greater for the multifactor model than the single index model is not too surprising, given the highly significant (unadjusted) t-statistics that CRR report.

The t-statistic for the zero-beta rate, computed in the traditional manner, is 1.361 with (two-sided) p-value .17. According to Theorem 1, with $\hat{c} = .36$, the standard error of the estimate should be multiplied, and the t-statistic divided by $\sqrt{1.36}$, to account for EIV. The adjusted t-statistic is 1.167 with p-value .24. Coefficients on additional variables like firm size would, if included, be treated in the same way.

The precision of the price of risk estimates is affected proportionately less than the zero-beta rate, since only the residual component of variance is "grossed up" by the EIV adjustment. As an example, consider the unanticipated inflation (UI) factor which, in principle, should be serially uncorrelated. CRR report a point estimate of -.0629 (percent per month) for the price of UI risk, with a t-statistic of -1.979 (p-value = .048). Thus, the unadjusted standard error is .0318. To compute the adjusted variance, we divide the appropriate diagonal element of (11) by T, the number of monthly observations.[34] The result is

$$1.36[(.0318)^2 - .0561/324] + .0561/324 = .001313$$

where .0561 is the variance of UI and T = 324. Note that .0318 already reflects division by $\sqrt{T}$. The adjusted standard error is .0362, an increase of about 14 percent. The adjusted t-statistic is 1.738 with p-value .082.

Based on the numbers above, factor-related variation accounts for 13% of the total adjusted variance of the UI price of risk estimator. The residual component (unadjusted variance minus factor variance) explains 64% and the EIV component the remaining 23%. In contrast, most of the variance of the FM market price of risk estimator is factor-related. Percentages for the basic specification are: factor-85%, residual-14%, and EIV-less than 1%. As a result, the ex post market price of risk is estimated quite precisely, the point estimate being .85 (percent per month) with standard error .13, as compared to the standard error of .33 for the ex ante price. These differences between FM and CRR reflect, among other things, the greater time-series explanatory power of the CRSP index and multicollinearity in the multibeta CSR's.

4.2 Correlation-corrected standard errors

To illustrate the potential impact of factor serial correlation, consider a one-factor model with the (real) return, R_{mt}, on the equally-weighted CRSP index as the factor. As discussed in Section 3.2, the variance of $\bar{R}_m$ need not equal the value implied under independence if the returns are serially correlated. An estimate of this correlation effect can be derived from the variance ratio statistics, VR(k), reported in Table 2 of Poterba and Summers (1988). It is straightforward to show that the ratio of $\mathrm{var}(\bar{R}_m)$ to $\mathrm{var}(R_m)/T$ is just VR(T)/VR(1). For example, if T = 96 months, the longest horizon considered by Poterba and Summers, then VR(96)/VR(1) = .353/.809 = .436, reflecting the negative autocorrelations found at relatively long lags.[35] As $1/\sqrt{.436} = 1.51$, the t-statistic on the sample mean must be increased by about fifty percent to reflect autocorrelation in the index.

The t-statistic for a two-pass CSR estimate of the market risk premium is affected less, since the residual component of variance is not affected at all. If v is the EIV-adjusted asymptotic variance (ignoring serial correlation) of monthly risk-premia estimates, and the market variance over the same (eight year) period is, say, 85 percent of v, then the correlation-corrected variance is

$$.436(.85v) + .15v = .52v.$$

Thus, we "discount" the factor component of variance and then add back the rest of v, which reflects the residual and EIV-related variation in the estimator. The required t-statistic correction is now about forty percent $(1.39 = 1/\sqrt{.52})$. When data from the Depression years is excluded in estimating the variance ratios, a smaller correction is indicated - about fifteen percent. Note that asymptotic standard errors for the zero-beta rate and "additional variable" risk premia are not affected by factor autocorrelation.

The numbers used in the computations above are not estimated with much precision and may be subject to small-sample biases.[36] The correlation corrections suggested are nontrivial, however, and warrant further attention in future research.

5. The Relation Between Two-Pass and Maximum Likelihood Methods

Section 3 analyzed the asymptotic properties of a class of T-consistent second-pass regression estimators. Small-sample considerations aside, one would prefer to use the most efficient estimator available. From the form of the asymptotic covariance matrix in Theorem 1, we see that factor-related variance and the EIV adjustment term are the same for each estimator. Thus, the relative precision of second-pass estimators depends only on their residual components of variance. An appeal to the Gauss-Markov Theorem of classical regression analysis implies that the residual component is minimized by the GLS estimator described in Section 2.[37] While the GLS estimator of Γ is thus efficient in a limited sense, it remains to be shown that it is still efficient when a broader class of estimators is considered.

Under the "usual assumptions" of joint normality and serial independence for the factors and asset returns, standard results imply that MLE is asymptotically efficient with respect to the larger class of all T-consistent asymptotically normal estimators. The MLE approach entails maximizing the constrained multivariate normal density function over the parameters B, Γ, Σ, E(F), and Σ_F. A computationally simpler approach linearizes the restriction (4) about initial consistent estimates of the parameters and then applies multivariate linear regression methods. This Gauss-Newton technique is known to be asymptotically equivalent (as $T \rightarrow \infty$) to MLE.

Recent work by Kandel (1984), extended in Shanken (1986), makes MLE computations straightforward when all factors are portfolio returns. In this case, the zero-beta estimate is the solution to a quadratic equation. Our next theorem addresses the general case in which some factors are not portfolio returns and relaxes the assumption, made in previous work, that the factors are serially uncorrelated and normally distributed. Although a closed form formula is not obtained, the computational burden is reduced considerably since maximization over the entire parameter space is reduced to the problem of minimizing a nonlinear function of just k_1+1 variables. The numerator of this function is the familiar quadratic which is minimized by the GLS regression estimator, while the denominator corresponds to the EIV adjustment encountered earlier in Theorem 1.

Theorem 3. Under Assumptions 1 and 2, and given joint normality of R_t conditional on F,

(i) the MLE for γ_0 and $\gamma_1^* = \gamma_1 - E(F_1)$ minimizes the function

$$f(\gamma_0, \gamma_1^*) = e'\hat{\Sigma}^{-1}e/(1 + \Gamma_{12}'S_F^{-1}\Gamma_{12}) \tag{12}$$

where, in this context, $\gamma_2 = \bar{F}_2 - \gamma_0 1_{k_2}$ and $e = \hat{a} - \gamma_0[1_N - \hat{B}_2 1_{k_2}] - \hat{B}_1\gamma_1^*$.

(ii) If, in addition, the MLE for E(F) is $\bar{F}$, then the MLE for γ_1 equals the MLE for γ_1^* plus $\bar{F}_1$, and the MLE for γ_2 equals $\bar{F}_2$ minus the MLE for γ_0 times 1_{k_2}.[38]

Our final result in this section reveals a surprising relation between the two-pass procedure and more conventional statistical methods, thereby resolving an issue raised earlier.

Theorem 4. The second-pass GLS estimator for Γ equals the Gauss-Newton estimator, provided that $\hat{B}$ and $\hat{\Sigma}$ are taken as the initial T-consistent estimates for B and Σ. In this case, the initial consistent estimate for Γ is irrelevant.[39]

It follows from Theorem 4 that GLS is asymptotically equivalent to MLE and hence is efficient under the "usual" assumptions. This is somewhat unexpected, since the second-pass estimator employs the unconstrained beta estimates which are asymptotically inefficient. Additional results on efficient estimation of beta and expected return can be found in Shanken (1982b).

Although we omit the details, the equivalence between two-pass GLS and MLE continues to hold in the general framework of Theorems 1 and 2. Thus, MLE standard errors "automatically" incorporate the EIV adjustment discussed earlier. However, if independence is assumed when, in fact, the factors are serially correlated, an explicit correlation correction will be needed. The procedure is identical to that described in Section 4.2 for the two-pass CSR estimator.

Two practical limitations of Theorem 4 should be noted. For one thing, concern that factor model parameters may be changing over time generally restricts the time-series length, T, employed in empirical applications. As a result, the number of securities available for use in the CSR's typically exceeds T. In such cases, unless additional structure is imposed on Σ, the usual MLE and GLS estimators are undefined, as $\hat{\Sigma}$ is singular. While a reduced set of securities or portfolios could be used, Theorem 4 says nothing about the relative precision of the resulting estimators and, say, OLS based on the original set of securities.[40]

There is also reason to doubt the validity of the asymptotic approximations for GLS and MLE when N, the number of assets, is "relatively large" compared to T. Intuitively, this is due to the large number of covariance parameters that must be estimated. More formally, it can be shown that the variance of each component of $\hat{\Sigma}^{-1}$ is inversely related to T-N. I have demonstrated elsewhere that this can lead to striking discrepancies between asymptotic and small-sample inferences in multivariate tests of the linear pricing relation.[41] The precise effect on the variability of the estimators is less clear and requires further investigation, perhaps through simulations. Additional concerns about small-sample issues motivate the analysis in the next section.

6. Consistent Estimation when the Number of Assets is Large

The analysis in Sections 3 and 5 indicates that two-pass estimation leaves little to be desired with regard to its large-sample properties, as $T \to \infty$. All forms of the estimator are T-consistent and the GLS estimator is asymptotically efficient. Its asymptotic variance is larger than it would be if the true betas were known; however, this loss of precision is not inherent in the two-pass approach since MLE has the same large-sample distribution. Thus, the asymptotic analysis, as $T \to \infty$, does not reveal an EIV "problem".

In this section we examine the estimators from a different analytical perspective. We assume T is fixed and consider limiting behavior as N, the number of assets, approaches infinity.[42] This perspective is particularly relevant given the large number of securities for which returns data are available. We cannot reasonably hope to consistently estimate the ex ante price of risk with T fixed, since increasing the number of assets affects residual variation but does not resolve uncertainty about the unanticipated factor realizations. Therefore, we define an estimator to be N-consistent if it converges in probability to the ex post parameter vector, $\overline{\Gamma}$, as $N \to \infty$. Such an estimator is consistent for γ_0 and asymptotically unbiased with respect to the ex ante price of risk parameters. Classical EIV analysis implies, in this context, that the second-pass estimator is biased and inconsistent, provided the residual covariance matrix is assumed to be scalar.

Black, Jensen, and Scholes (1972) suggest two-pass estimation with portfolio grouping as a means of dealing with the EIV problem. Assuming independence of the market (factor) model disturbances, grouping permits the construction of a finite set of portfolios that are well-diversified in the limit as $N \to \infty$. Since residual variation is eliminated, the portfolio betas are free of estimation error. It follows that the limiting value of the portfolio-based estimator equals the ex post price vector, i.e., the estimator is N-consistent. Actually, strict independence is not essential for this result since one can appeal to versions of the law of large numbers that allow for "weak" dependence between the factor model disturbances. For example, a block-diagonal residual covariance structure is permitted.[43]

Litzenberger and Ramaswamy (1979) employ a modified version of the WLS second-pass estimator that is N-consistent without portfolio grouping. Although they assume the residual covariance matrix is diagonal, again this strict condition is not necessary. An assumption of greater concern is that the security residual variances are known exactly. This condition is relaxed in the development below. We also show how to incorporate factor-portfolio pricing restrictions and multiple factors in the modified two-pass procedure.

Our modified OLS estimator is based on the following lemma. Since the focus of this section is on convergence to $\bar{\Gamma}$, the statistical analysis is conditioned on the factor values. Therefore, Assumption 2 is not needed.

Lemma 2. Given Assumption 1, we have the following expectations (conditional on F)

$$E(\hat{X}'\hat{X}) = X'X + qM'\hat{\Sigma}_F^{-1}M \tag{13}$$

and

$$E[\hat{X}'(\bar{R} - \hat{B}_2\bar{F}_2)] = (X'X)\bar{\Gamma}_{01} - qM'\hat{\Sigma}_F^{-1}\bar{F}_2^* \tag{14}$$

where

$$q = tr(\Sigma)/(T-1)$$

Here, $\bar{F}_2^*$ is a k-vector consisting of k_1 zeroes followed by $\bar{F}_2$. M is a $kx(k_1+1)$ matrix with -1_{k_2} in the lower left corner, I_{k_1} in the upper right corner, and zeroes elsewhere.

Under certain conditions, (13) and (14) allow us to eliminate the systematic effects of measurement error in the betas, thereby yielding "consistent" expressions for $(X'X)^{-1}$ and $(X'X)\bar{\Gamma}_{01}$. These expressions are combined to form the modified estimator. Subtracting the estimator of γ_0 from $\bar{F}_2$ then provides an N-consistent estimator for $\bar{\gamma}_2$. The situation is simpler when $k_2=0$, as M is an identity matrix and $\bar{F}_2^*$ equals zero in this case. In general, we have the following result.

Theorem 5. Under Assumption 1 and given sufficiently weak cross-sectional dependence between the residual disturbances,

$$[\hat{X}'\hat{X} - \hat{q}M'\hat{\Sigma}_F^{-1}M]^{-1}[\hat{X}'(\bar{R} - \hat{B}_2\bar{F}_2) + \hat{q}M'\hat{\Sigma}_F^{-1}\bar{F}_2^*] \tag{15}$$

is an N-consistent estimator for $\bar{\Gamma}_{01}$, where $\hat{q} = tr(\hat{\Sigma})/(T-1)$.[44]

An N-consistent modification of the WLS estimator can also be obtained, provided that the joint distribution of returns, conditional on the factors, is normal. Recall that WLS amounts to dividing each element of $\hat{X}$ and $\bar{R} - \hat{B}_2\bar{F}_2$ by the corresponding estimated residual standard deviation and then running an OLS CSR. The modified WLS estimator, in terms of these standardized variables, is of the same form as (15), except that $tr(\hat{\Sigma})$ is replaced by $N(T-k-1)/(T-k-3)$. If the residual variances were known, $tr(\Sigma)$ would just be replaced by N, since standardized residual variance equals one in this case.[45] The factor $(T-k-1)/(T-k-3)$ is thus an adjustment for sampling error in the variance estimates. To generate a modified GLS estimator under the joint normality assumption, $\hat{X}'$ should be replaced by $\hat{X}'\hat{\Sigma}^{-1}$ and $tr(\hat{\Sigma})$ by $N(T-k-1)/(T-N-k-2)$ in (15). Large-sample properties, as $N \rightarrow \infty$, cannot be analyzed, however, since $\hat{\Sigma}$ is singular for large N.[46]

We explore the large-sample benefits of simultaneous estimation of betas and gammas by examining an "OLS version" of MLE. Thus, the estimator is computed as if the residual covariance matrix is scalar. This amounts to replacing the numerator of the function in equation (12) by the simple sum of squares, $e'e$. Our final theorem considers the behavior of this estimator.

Theorem 6. Under Assumption 1 and given sufficiently weak cross-sectional dependence between residual disturbances, the "OLS version" of MLE is N-consistent for $\bar{\Gamma}_{01}$ when $k_1=0$.

The proof of Theorem 6, given in the Appendix, makes use of the fact that the estimator for γ_0 is the root of a quadratic equation. Limiting values of the coefficients in this equation are derived and it is verified that γ_0 is the appropriate root of the limiting equation. A similar proof is possible when $k_2=0$ and $k_1=1$. Interestingly, normality of the residuals is not needed for these results, even though it is assumed in generating the estimator. With normality, one can show that a WLS version of MLE is also N-consistent. The extension to a general factor model is left to future research. While Theorem 6 does not say anything directly about the usual GLS version of MLE, it does suggest that joint estimation of betas and gammas may be beneficial.

7. Conclusions

This paper has presented an integrated econometric view of maximum likelihood methods and two-pass approaches to estimating beta-pricing models. The usual framework has been expanded in a simple manner to allow for serial correlation in the underlying market factors. In addition, two aspects of the "errors-in-variables (EIV) problem" have been considered. The first concerns a downward bias in commonly used standard errors for price of risk and zero-beta estimates. The second involves the "classical" EIV problem, i.e., small-sample bias in the second-pass cross-sectional regression estimates due to measurement error in the betas. With regard to the latter bias, our results provide some support for the use of modified two-pass methods or maximum likelihood estimation (MLE) as alternatives to traditional two-pass estimation with portfolio grouping. A few words of caution are in order, however.

The modified version of the second-pass estimator incorporates an unbiased estimator of a certain positive definite matrix. The matrix estimate is not constrained to be positive definite, however, so the diagonal elements could be negative in a given sample. One would not have confidence in the usefulness of the modification in such a case. Shanken and Weinstein (1988) encounter this problem in a multifactor context,[47] while Banz (1981) alludes to "serious problems in applying the Litzenberger-Ramaswamy estimator" in his analysis of the size effect. The MLE approach may also be subject to problems. In particular, Amsler and Schmidt (1985) report that MLE was "occasionally off by spectacularly large amounts" in simulations where the true values were known.

Given these observations, as well as the limitations inherent in any asymptotic statistical analysis, the results of this paper cannot be taken as an absolute prescription for econometric practice. The procedures developed here should, nonetheless, provide valuable information to complement that obtained by the usual methods. More extensive simulations will probably be needed to fill the gaps in our knowledge. Hopefully, this paper has clarified some of the questions that should be addressed and offers useful benchmarks against which simulation results can be interpreted.

References

Amsler, D., and P. Schmidt, 1985, "A Monte Carlo Investigation of the Accuracy of Multivariate CAPM Tests," *Journal of Financial Economics*, 14, 359-375.

Banz, R., 1981, "The Relations Between Returns and Market Values of Common Stocks," *Journal of Financial Economics*, 9, 3-18.

Black, F., 1972, "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, 45, 444-454.

Black, F., and M. Scholes, 1974, "The Effects of Dividend Yield and Dividend Policy on Common Stock Prices and Returns," *Journal of Financial Economics* 1, 1-22.

Black, F., M. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Findings," in M. Jensen, (ed.), *Studies in the Theory of Capital Markets* (Praeger, New York) 79-121.

Blume, M., and I. Friend, 1973, "A New Look at the Capital Asset Pricing Model," *Journal of Finance*, 28, 19-34.

Breeden, D., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities, *Journal of Financial Economics* 7, 265-296.

Breeden, D., M. Gibbons, and R. Litzenberger, 1989, "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance*, 44, 231-262.

Brock, W., 1982, "Asset Prices in a Production Economy," *The Economics of Information and Uncertainty*, John J. McCall, ed., University of Chicago Press.

Chamberlain, G., and M. Rothschild, 1983, "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281-1304.

Chan, K.C., N. Chen, and D. Hsieh, 1985, "An Exploratory Investigation of the Firm Size Effect," *Journal of Financial Economics*, 14, 451-471.

Chen, N., and J. Ingersoll, Jr., 1983, "Exact Pricing in Linear Factor Models with Finitely Many Assets: A Note," *Journal of Finance*, 38, 985-88.

Chen, N., R. Roll, and S. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business*, 59, 383-403.

Connor, G., 1984, "A Unified Beta Pricing Theory," *Journal of Economic Theory*, 34, 13-31.

Connor, G., and R. Korajczyk, 1988, "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21, 255-289.

Douglas, G., 1969, "Risk in Equity Markets: An Empirical Appraisal of Market Efficiency," *Yale Economic Essays*, 9, 3-45.
Dybvig, P., 1983, "An Explicit Bound on Deviations from APT Pricing in a Finite Economy," *Journal of Financial Economics*, 12, 483-496.

Dybvig, P., and S. Ross, 1985, "Yes, the APT is Testable," *Journal of Finance*, 40, 1173-1188.

Fama, E.F., 1976, *Foundations of Finance* (Basic Books, New York).

Fama, E.F., and K. French, 1988, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246-273.

Fama, E., and J. MacBeth, 1973, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607-636.

Garman, M., and J. Ohlson, 1980, "Information and the Sequential Valuation of Assets in Arbitrage-Free Economies," *Journal of Accounting Research*, 18, 420-440.

Gibbons, M., 1980, "Estimating the Parameters of the Capital Asset Pricing Model: A Minimum Expected Loss Approach," unpublished manuscript, Graduate School of Business, Stanford University.

Gibbons, M., 1982, "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, 10, 3-27.

Gibbons, M., S. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121-1152.

Grinblatt, M., and S. Titman, 1983, "Factor Pricing in a Finite Economy," *Journal of Financial Economics*, 12, 497-507.

Grossman, S., and R. Shiller, 1982, "Consumption Correlatedness and Risk Management in Economies with Non-Traded Assets and Heterogeneous Information," *Journal of Financial Economics*, 10, 195-210.

Hansen, L.P., and K. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-1286.

Hess, P., 1980, "Dividend Yields and Stock Returns: A Test for Tax Effects," Ph.D. dissertation, Graduate School of Business, University of Chicago.

Huang, C., and R. Litzenberger, 1988, *Foundations for Financial Economics* (North-Holland, New York).

Jobson, J., and B. Korkie, 1982, "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics*, 10, 433-466.

Kandel, S., 1984, "The Likelihood Ratio Test Statistic of Mean-Variance Efficiency without a Riskless Asset," *Journal of Financial Economics*, 13, 575-592.

Lintner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistsics* 47, 13-37.

Litzenberger, R.H., and K. Ramaswamy, 1979, "The Effect of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence," *Journal of Financial Economics*, 7, 163-196.

Lo, A., and A.C. MacKinlay, 1988, "Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 41-66.

MacKinlay, C., 1987, "On Multivariate Tests of the CAPM," *Journal of Financial Economics*, 18, 341-371.

McElroy, M., and E. Burmeister, 1988, Arbitrage Pricing Theory as a Restricted Nonlinear Multivariate Regression Model," *Journal of Business and Economic Statistics*, 6, 29-42.

Merton, R., 1973, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 44, 867-887.

Miller, M.H., and M. Scholes, 1972, "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings," in M.C. Jensen, (ed.), *Studies in the Theory of Capital Markets* (Praeger, New York) 47-78.

Poterba, J., and L. Summers, 1988, "Mean Reversion in Stock Prices," *Journal of Financial Economics*, 22, 27-59.

Richardson, M., and J. Stock, 1989, "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," *Journal of Financial Economics*, 25, 323-348.

Roll, R., 1977, "A Critique of the Asset Pricing Theory's Tests, Part I", *Journal of Financial Economics*, 4, 129-176.

Rosenberg, B., and J. Guy, 1976, "Prediction of Beta from Investment Fundamentals," *Financial Analysts Journal* 32, 60-72.

Ross, S.A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341-360.

Rubinstein, M., 1973, "The Fundamental Theorem of Parameter-Preference Security Valuation," *Journal of Financial and Quantitative Analysis*, January 1973, 61-69.

Shanken, J., 1982a, "The Arbitrage Pricing Theory: Is It Testable?," *Journal of Finance*, 37, 1129-1140.

Shanken, J., 1982b, "An Asymptotic Analysis of the Traditional Risk-Return Model," Chapter 2 of doctoral dissertation, Carnegie-Mellon University.

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

Shanken, J., 1985b, "Multi-beta CAPM or Equilibrium APT?: A Reply," *Journal of Finance*, 40, 1189-1196.

Shanken, J., 1986, "Testing Portfolio Efficiency when the Zero-Beta Rate is Unknown: A Note," *Journal of Finance*, 41, 269-276.

Shanken, J., 1987, "Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique," *Journal of Financial Economics*, 18, 91-110.

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

Shanken, J., and M. Weinstein, 1990, Macroeconomic Variables and Asset Pricing: Further Results," working paper, University of Rochester.

Sharpe, W., 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance* 19, 425-442.

Sharpe, W., 1977, "The Capital Asset Pricing Model: A 'Multibeta' Interpretation," in: H. Levy and M. Sarnat, eds., "Financial Decision Making Under Uncertainty," (Academic Press, New York).

Stambaugh, R., 1982, "On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis," *Journal of Financial Economics*, 10, 237-268.

Theil, H., 1971, *Principles of Econometrics* (Wiley, New York).

Appendix

Proof of Lemma 1

Let F_d be a Txk matrix with columns equal to the factor time-series realizations expressed as deviations from the time-series means. Let ε_i be a T-vector of factor model disturbances for asset i and let ε be an NT-vector obtained by stacking the ε_i on top of one another, for $i=1,\ldots,N$. Also, let $U \equiv \hat{B} - B$ and U_v equal the Nk-vector, $\hat{B}_v - B_v$. Then

$$U_v = [I_N \otimes (F_d'F_d)^{-1}F_d']\varepsilon \qquad \text{Nkx1} \qquad \text{(A1)}$$

and

$$\bar{\varepsilon} = T^{-1}[I_N \otimes 1_T']\varepsilon \qquad \text{Nx1} \qquad \text{(A2)}$$

Therefore,

$$E(U_v\bar{\varepsilon}'|F) = [I_N \otimes (F_d'F_d)^{-1}F_d'](\Sigma \otimes I_T)T^{-1}[I_N \otimes 1_T] \qquad \text{(A3)}$$

$$= T^{-1}\Sigma \otimes (F_d'F_d)^{-1}F_d'1_T = 0$$

since $F_d'1_T = 0$. Also, $E(U|F) = 0$ and $E(\bar{\varepsilon}|F) = 0$. It follows that $\hat{B}$ and $\bar{\varepsilon}$ are conditionally (and unconditionally) uncorrelated and, since mean independence implies uncorrelatedness, both are uncorrelated with $\bar{F}$.

Let $\hat{\varepsilon}$ be the NT-vector of OLS residuals corresponding to the NT-vector of disturbances, ε. Then there is a TxT symmetric idempotent matrix M (not to be confused with M in Section 6), that depends only on the factor realizations [see Theil (1971), P. 113] such that $M1_T = 0$, $MF_d = 0$, and

$$\hat{\varepsilon} = (I_N \otimes M)\varepsilon \qquad (A4)$$

Proceeding as in (A3) above, it follows that $\hat{\varepsilon}$ is conditionally (and unconditionally) uncorrelated with $\bar{\varepsilon}$ and $\hat{B}$. Assuming joint normality of ε_t conditional on F, and noting that $\hat{\Sigma}$ is a function of the vector, $\hat{\varepsilon}$, the above results imply that $\hat{B}$, $\bar{\varepsilon}$, and $\hat{\Sigma}$ are conditionally independent. The remaining assertions in (i) are standard regression results.

To prove (ii) for the case in which F_t is independent and identically distributed (i.i.d.) consider the (N+Nk+k)-vector

$$Q_t = [\varepsilon_t', \ \varepsilon_t' \otimes (F_t - E(F))'\Sigma_F^{-1}, (F_t - E(F))']'. \qquad (A5)$$

Since $E(\varepsilon_t|F) = 0$ and $Var(\varepsilon_t|F) = \Sigma$ is constant, the components of Q_t are uncorrelated and have zero means. Thus Var(Q) is block diagonal with blocks Σ, $\Sigma \otimes \Sigma_F^{-1}$, and Σ_F, respectively. Assumptions 1, 2, and the central limit theorem for i.i.d. random vectors then imply that $\sqrt{T}\bar{Q}$ converges in distribution, as $T \to \infty$, to a multivariate normal distribution with mean zero and covariance matrix Var(Q). The same result is obtained if $(F_t - E(F))'$ is replaced by F_{dt} (row t of F_d) and Σ_F by S_F in the middle component of Q_t. This is true since $\sqrt{T}$ times the average difference between Q_t and the modified Q_t converges in probability to zero. The desired conclusion now follows from the observation that U_v' equals the (modified) middle component of $\bar{Q}$. The general case is proved similarly using the last part of Assumption 2 and a more general central limit theorem. Part (iii) follows easily from part (ii) since

$$Ug = (I_N \otimes g')U_v.$$

Proof of Theorem 1

Equations (9) and (10) imply that

$$\bar{R} - B_2\bar{F}_2 = X\bar{\Gamma}_{01} + \bar{\varepsilon} \qquad \text{(A6)}$$

Letting $U = \hat{B} - B$,

$$\bar{R} - \hat{B}_2\bar{F}_2 = \hat{X}\bar{\Gamma}_{01} + [\bar{\varepsilon} - U\bar{\Gamma}_{12}] \qquad \text{(A7)}$$

Since $\hat{A}\hat{X} = I_{k_1+1}$, (A7) implies that

$$\hat{\Gamma}_{01} - \bar{\Gamma}_{01} = \hat{A}[\bar{\varepsilon} - U\bar{\Gamma}_{12}] \qquad \text{(A8)}$$

Assumption 2 implies that $\text{plim}_T\, \bar{\Gamma} = \Gamma$. Hence, (ii) and (iii) of Lemma 1 imply that

$$\sqrt{T}[\bar{\varepsilon} - U\bar{\Gamma}_{12}] \rightarrow_T N[0, \Sigma(1+c)] \qquad \text{(A9)}$$

Therefore, using (A8), (A9), and the fact that $\text{plim}_T\, \hat{A} = A$,

$$\sqrt{T}[\hat{\Gamma}_{01} - \bar{\Gamma}_{01}] \rightarrow_T N[0, A\Sigma A'(1+c)] \qquad \text{(A10)}$$

The rest of part (ii) of Theorem 1 follows from the observation that, by definition,

$$\hat{\gamma}_2 - \bar{\gamma}_2 = -(\hat{\gamma}_0 - \gamma_0)1_{k_2} \qquad \text{(A11)}$$

Part (ii) implies that $\text{plim}_T\, \hat{\Gamma} - \bar{\Gamma} = 0$. This and the fact that $\text{plim}_T\, \bar{\Gamma} = \Gamma$ imply (i).

To prove (iii), write $\hat{\Gamma} - \Gamma$ as

$$\hat{\Gamma} - \Gamma = (\hat{\Gamma} - \bar{\Gamma}) + (\bar{\Gamma} - \Gamma) \tag{A12}$$

In general, if one sequence of random variables converges in probability to a constant and another converges in distribution to some random variable, then the sequence of products converges in distribution to the product of the constant and the limiting random variable. Using the multivariate generalization of this result and the fact that $\bar{F}$ is uncorrelated with both $\bar{\varepsilon}$ and U (Lemma 1), (A8) and (A11) imply that $\hat{\Gamma} - \bar{\Gamma}$ is asymptotically uncorrelated with $\bar{F}$. Since the variation in $\bar{\Gamma}$ is simply that of $\bar{F}$, the expressions in parentheses in (A12) are asymptotically uncorrelated and (iii) follows.

Proof of Theorem 2

It follows from the definitions of $\hat{\Gamma}_{01t}$ and $\hat{\Gamma}_{01}$ that

$$\hat{\Gamma}_{01t} - \hat{\Gamma}_{01} = \hat{A}[(R_t - \bar{R}) - \hat{B}_2(F_{2t} - \bar{F}_2)] \tag{A13}$$

Letting e_t be the time-series residual N-vector $R_t - \hat{a} - \hat{B}F_t$,

$$\hat{\Gamma}_{01t} - \hat{\Gamma}_{01} = \hat{A}[e_t + \hat{B}_1(F_{1t} - \bar{F}_1)] \tag{A14}$$

Since the time series of residuals is orthogonal to the time series of factor values by construction, the sample covariance matrix of the expression in brackets in (A14) is

$$S_e + \hat{B}_1 S_{F_1} \hat{B}_1' \tag{A15}$$

where S with a subscript denotes the sample covariance matrix (with T in the denominator) of the subscripted variable. Using the fact that $\hat{A}\hat{B}_1$ equals the last k_1 columns of $\hat{A}\hat{X} = I_{k_1+1}$, (A14) and (A15) imply that the sample covariance matrix of the $\hat{\Gamma}_{01t}$'s is

$$\hat{A}S_e\hat{A}' + S^*_{F_1} \tag{A16}$$

where $S^*_{F_1}$ is the appropriate submatrix of S^*_F. Taking the probability limit of (A16) establishes the conclusion of Theorem 2 with respect to the upper left block of $\hat{W}$. The result for the lower right block follows easily since

$$\hat{\gamma}_{2t} - \hat{\gamma}_2 = (F_{2t} - \bar{F}_2) - (\hat{\gamma}_{0t} - \hat{\gamma}_0)1_{k_2}$$

and the series of zero-beta estimates is orthogonal, in sample, to the factors [use (A14) and the fact that $\hat{A}\hat{X} = I$]. The off-diagonal block of $\hat{W}$ is treated similarly.

Proof of Lemma 2

The following analysis is conditioned on F. Let U_i be the ith row of $U = \hat{B} - B$. Thus, U_i is the k-vector of estimation errors in the betas for asset i. Appealing to standard regression results, the covariance matrix of U_i is

$$\sigma_i^2\hat{\Sigma}_F^{-1}/(T-1)$$

where σ_i^2 is the residual variance for asset i. Since E(U'U) equals the sum, over i, of $E(U_i'U_i)$ we have

$$E(U'U) = \hat{\Sigma}_F^{-1}tr(\Sigma)/(T-1) = q\hat{\Sigma}_F^{-1} \tag{A17}$$

Noting that $\hat{X} - X = UM$, equation (13) follows.

To prove (14), first observe that

$$\hat{X}'(\bar{R} - \hat{B}_2\bar{F}_2) = X'(\bar{R} - \hat{B}_2\bar{F}_2) + M'U'(\bar{R} - \hat{B}_2\bar{F}_2) \tag{A18}$$

and recall from (A7) that

$$\bar{R} - \hat{B}_2\bar{F}_2 = \hat{X}\bar{\Gamma}_{01} + [\bar{\varepsilon} - U\bar{\Gamma}_{12}]$$

Since $\bar{\varepsilon}$ and U have zero means, this equation implies that

$$E[X'(\bar{R} - \hat{B}_2\bar{F}_2)] = X'X\bar{\Gamma}_{01} \tag{A19}$$

By (i) of Lemma 1, U is uncorrelated with $\bar{R}$. Therefore,

$$E[U'(\bar{R} - \hat{B}_2\bar{F}_2)] = -E(U'U\bar{F}_2^*) = -q\hat{\Sigma}_F^{-1}\bar{F}_2^* \tag{A20}$$

where the last equality follows from (A17). Combining (A18), (A19), and (A20) yields equation (14).

Proof of Theorem 6

The proof below is for the case $k_1=0$. The statistical analysis is conditioned on F and all limits are taken as N, the number of assets, approaches infinity. By Theorem 3, the OLS version of the MLE for γ_0 minimizes the function

$$f_N(\gamma) = e'e/[1 + (\bar{F} - \gamma 1_k)'S_F^{-1}(\bar{F} - \gamma 1_k)] \tag{A21}$$

Given results in the Appendix of Shanken (1986), it follows that $f_N'(\gamma)$ has the same sign, for all γ, as the quadratic

$$H_N(\gamma) = A_N\gamma^2 + B_N\gamma + C_N \tag{A22}$$

where

$$A_N = [c_1\hat{X}'\hat{a} - c_2\hat{X}'\hat{X}]/N$$

$$B_N = [(1+c_3)\hat{X}'\hat{X} - c_1\hat{a}'\hat{a}]/N$$

and

$$C_N = [-(1+c_3)\hat{X}'\hat{a} + c_2\hat{a}'\hat{a}]/N$$

with

$$c_1 = 1_k' S_F^{-1} 1_k \quad c_2 = 1_k' S_F^{-1}\bar{F} \quad c_3 = \bar{F}' S_F^{-1}\bar{F}$$

and $\hat{a}$ the vector of OLS intercept estimates. Thus, the MLE for a given sample of N assets is a root of H_N. We shall show that the coefficients in (A22) converge to limits, as $N \to \infty$, and that γ_0 is a root of this limiting quadratic.

Let σ^2 denote the limit of the average, over the N assets, of residual variance divided by T and let σ_x^2 be the limit of X'X/N. We assume that both limits exist. By (A17) and the weak law of large numbers

$$\text{plim } U'U/N = \sigma^2 S_F^{-1} \tag{A23}$$

Since $\hat{X} = X - U1_k$, it follows that

$$\text{plim } \hat{X}'\hat{X}/N = \sigma_x^2 + c_1\sigma^2 \tag{A24}$$

By (A7),

$$\hat{a} = \hat{X}\gamma_0 + [\bar{\epsilon} - U(\bar{F} - \gamma_0 1_k)] \tag{A25}$$

Using (A23)-(A25) and the fact that $\bar{\epsilon}$ and U are uncorrelated (Lemma 1),

$$\text{plim } \hat{X}'\hat{a}/N = \gamma_0(\sigma_x^2 + c_1\sigma^2) + (c_2\sigma^2 - \gamma_0 c_1\sigma^2) = \gamma_0\sigma_x^2 + c_2\sigma^2 \tag{A26}$$

By (A25),

$$\hat{a} = X\gamma_0 + [\bar{\varepsilon} - U\bar{F}] \tag{A27}$$

Using (A23), (A27), and the fact the $\bar{\varepsilon}$ and U are uncorrelated,

$$\text{plim } \hat{a}'\hat{a}/N = \gamma_0^2\sigma_x^2 + \sigma^2(1+c_3) \tag{A28}$$

It follows from (A24), (A26), and (A28) that

$$\text{plim } A_N = A = \sigma_x^2(\gamma_0 c_1 - c_2) \tag{A29}$$

$$\text{plim } B_N = B = \sigma_x^2(1+c_3 - c_1\gamma_0^2)$$

and

$$\text{plim } C_N = C = -\sigma_x^2\gamma_0(1+c_3 - c_2\gamma_0)$$

Consider the limiting quadratic

$$H(\gamma) = A\gamma^2 + B\gamma + C$$

and note that, with probability one, A is not equal to zero. Since the roots of a quadratic are continuous functions of the coefficients, (A29) implies that each root of H_N converges in probability to the corresponding root of H. Furthermore, for sufficiently large N and with probability arbitrarily close to one, the MLE for γ_0 will equal the smaller or the larger root of H_N depending on whether A is greater or less than zero [see Shanken (1986), p. 275]. In either case, the MLE is uniquely determined by the condition that f_N', and hence H_N, change sign from negative to positive at the given root. Thus, to establish N-consistency it suffices to show that $H(\gamma_0) = 0$ and that $H'(\gamma_0)$ is positive.

Evaluating H at γ_0 yields a third order polynomial in γ_0 which, one can easily verify, is identically equal to zero! A bit of algebraic manipulation establishes that

$$H'(\gamma_0) = 2A\gamma_0 + B = (c_1\gamma_0^2 - 2c_2\gamma_0 + 1 + c_3)\sigma_x^2 \tag{A30}$$

The discriminant of this quadratic in γ_0 equals σ_x^4 times

$$4c_2^2 - 4c_1(1+c_3) = 4(c_2^2 - c_1c_3) - 4c_1$$

Since $c_1c_3 - c_2^2$ is the determinant of the positive semidefinite matrix, $[1_k,\bar{F}]'S_F^{-1}[1_k,\bar{F}]$, and $c_1>0$, the discriminant is negative. Thus, the quadratic in (A30) has no real roots and must be positive for all values of γ_0. This completes the proof.

The proof of N-consistency for the WLS version of the MLE is similar. In this case, however, a normality assumption is used to ensure (see Lemma 1) that the diagonal elements of $\hat{\Sigma}$ are independent of $\bar{\epsilon}$ and $\dot{U}$.

Accommodating "Rolling Betas"

The proof of Theorem 1 is modified here to derive an approximate covariance matrix for the CSR estimator when the following variation on the traditional methodology is employed: The estimate $(\hat{B}_t)$ of B, used in the CSR for month t of a given year, is obtained from time series regressions over the previous five calendar years. Thus, all betas are based on five years of data and are updated annually. In this context, the CSR estimator need not be T-consistent, as expanding the time series does not eliminate the systematic EIV bias in each CSR. The discussion below is not, therefore, offered as a formal asymptotic analysis but, rather, as a simple heuristic for EIV-adjustment of standard errors when rolling betas are employed.[48]

Letting $\hat{\Gamma}_{01t} = \hat{A}_t[R_t - \hat{B}_{2t}F_{2t}]$ (t=1,...,T), with $\hat{A}_t$ defined in terms of $\hat{B}_t$, we have

$$\hat{\Gamma}_{01t} - \Gamma_{01t} = \hat{A}_t[\varepsilon_t - U_t\Gamma_{12t}] \approx A[\varepsilon_t - U_t\Gamma_{12}]$$

where $\Gamma_{12t} = \Gamma_{12} + F_t - E(F)$ and $\gamma_{0t} = \gamma_0$, with second order error terms ignored in the approximation. Taking time series sample means yields (compare to (A8))

$$\hat{\Gamma}_{01} - \bar{\Gamma}_{01} \approx A[\bar{\varepsilon} - \bar{U}\Gamma_{12}] \tag{A31}$$

By Lemma 1 and Assumption 1, U_t is uncorrelated with $\bar{\varepsilon}$ (conditional on F) provided that the five year period used to estimate $\hat{B}_t$ either precedes or is completely contained within the period over which CSR's are run. In cases of partial overlap, the covariance may be nonzero but is not systematically positive or negative. Thus, we shall treat the covariance between $\bar{\varepsilon}$ and $\bar{U}$ as equal to zero. Therefore, it remains to consider the covariance matrix of $\bar{U}\Gamma_{12}$.

As in the proof of Lemma 1, given a kx1 vector g, we have

$$Ug = (I_N \otimes g')U_v = [I_N \otimes g'(F_d'F_d)^{-1}F_d']\varepsilon \tag{A32}$$

for each U (really U_t) in the average $\bar{U}$. Let z_1 and z_2 be kx60 matrices of the form $(F_d'F_d)^{-1}F_d'$, for two (possibly) overlapping five year beta estimation periods, and let ε^1 and ε^2 be the corresponding 60Nx1 disturbance vectors. Assume that entries have been reordered so that values of t from 1 through α correspond to overlapping months, where α is between 0 and 60. It can then be shown that the covariance between $[I_N \otimes g'z_1]\varepsilon^1$ and $[I_N \otimes g'z_2]\varepsilon^2$ is $g'z_1 I^* z_2' g$ times Σ, where I^* is a 60x60 matrix with ones in diagonal entries 1 through α and zeroes elsewhere. If the overlap is complete, $z_1 = z_2$, $I^* = I_{60}$, and the scalar is $g'(F_d'F_d)^{-1}g \approx 60^{-1} g'\Sigma_F^{-1} g$. With partial overlap, the scalar is approximately $\alpha/60$ times this expression, reflecting the proportion of overlapping (and hence correlated) observations.

Let $T = 12n$ and note that, with annual updating, $\bar{U}g$ equals the average of n distinct terms of the form Ug. Thus, the implied covariance matrix for $\bar{U}g$ equals Σ times $(n^{-2})(12)(60^{-2})g'\Sigma_F^{-1}g$ times the sum of overlap (in years) between all n^2 pairs of the n beta estimation periods. This sum is

$$5n + 2[4(n-1) + 3(n-2) + 2(n-3) + 1(n-4)] = 25n - 40.$$

Substituting Γ_{12} for g, the resulting covariance matrix for $\sqrt{T}\bar{U}\Gamma_{12}$ is $c^*\Sigma$, where

$$c^* = T(5n-8)(60n^2)^{-1}c = (1-1.6/n)c$$

and c is defined as in Theorem 1.[49]

The suggested heuristic, then, is to use c^* in place of c in the EIV adjustments described in the text of the paper. For example, with $T=324$ and $n=27$ as in the Chen, Roll, and Ross example of Section 4.1, $c^*=.94c$. Intuitively, the smaller adjustment reflects the fact that, overall, n+5 years of data are now used in estimating betas while n years are used in the CSR's. Clearly, c^* approaches c as $n \to \infty$.

Endnotes

1. Black (1972) and Rubinstein (1973) extend the result to the case in which there is no riskless asset.

2. These unpublished results are reproduced in a related paper by Douglas (1969).

3. This is essentially the "random coefficients problem" pointed out by Black, Jensen and Scholes (1972). See Miller and Scholes (1972) for a discussion of other problems associated with the Douglas-Lintner regressions. Jensen (1972) provides a good review of the early CAPM work.

4. Black, Jensen, and Scholes (1972) introduce a related procedure that will be discussed later in the paper.

5. Gibbons (1980) independently derives the asymptotic distribution for a related estimator developed by Black, Jensen, and Scholes (1972). This result will later be seen to be a special case of our more general analysis.

6. Gauss-Newton estimation involves linearizing the nonlinear constraint (1) about initial consistent estimates. The constraint is nonlinear since the unknown parameters γ_1 and β_i enter multiplicatively.

7. This is pointed out by Miller and Scholes (1972).

8. In particular, the parameter space is no longer fixed in the latter case since each new asset introduces an additional beta parameter.

9. Litzenberger and Ramaswamy refer to their technique as MLE. To avoid confusion, we reserve this designation for the estimator obtained by simultaneously maximizing the likelihood function for betas and gammas as well as the parameters in the residual covariance matrix.

10. This follows from the fact that each factor portfolio has a beta of one on itself and zero on all other factors. A "1" with a subscript refers to a column vector of ones with dimension equal to the subscript.

11. A bar over a variable previously subscripted by t indicates the time series average of that variable.

12. See, for example, Theil (1971), section 7.2.

13. See Shanken (1982a, 1985b) and related discussion in Dybvig and Ross (1985).

14. Condition (i) is discussed in Section 6 of Shanken (1987) while the sufficiency of condition (ii) follows from the proof of Theorem 1 in that paper. The equivalence between (i) and (ii) amounts to the observation that $\mathrm{cov}(m,\varepsilon_i) = \mathrm{cov}(\varepsilon_m,\varepsilon_i) = \mathrm{cov}(\varepsilon_m,R_i)$ for all i, as ε_i

and ε_m are orthogonal to the factors by construction. Conditions (i) and (ii) are not necessary for (4) to hold. For example, adding pure measurement error to an otherwise perfect proxy can cause these conditions to be violated, even though expected return is still linear in the "new" betas. See Breeden, Gibbons, and Litzenberger (1989) for a discussion of measurement error in a single factor case. Huang and Litzenberger (1988, page 308) consider condition (ii) in a single-factor context, together with an assumption that the market proxy has a beta of one on the true market portfolio.

15. Garman and Ohlson (1980), Brock (1982), and Chen and Ingersoll (1983) also use spanning conditions.

16. Specifically, they assume that the disturbance term for a given asset is independent of the returns on all other assets in the economy. They also assume that the asset constitutes a small proportion of aggregate wealth. While the latter assumption can be verified using a subset of the asset universe, the former cannot.

17. See Shanken (1990) for a simple approach to testing conditional pricing relations when the factors are portfolios and the zero-beta rate is known.

18. Constant conditional betas and zero-beta rate, and conditionally serially independent disturbances are *sufficient* to ensure unconditionally serially independent disturbances in the factor portfolio case. I am grateful to a referee for focusing my attention on this issue.

19. See the discussion in Blume and Friend (1973).

20. The concern here is with stochastic dependence between the random components of these variables. Naturally, we expect to observe cross-sectional correlation between the variables if the underlying theory is true.

21. In general, Γ with multiple subscripts refers to the corresponding sub-vector of Γ.

22. Gibbons (1982) imposes this constraint in his MLE approach. Stambaugh (1982) also employs MLE, but does not impose a factor pricing constraint because of the specific nature of his index.

23. See, for example, Litzenberger and Ramaswamy (1979) for WLS and, anticipating Theorem 4, Gibbons (1982) for GLS. This is not true GLS or WLS since the true covariance matrix is unknown. Estimation error in the covariance matrix does not affect the asymptotic results of Section 3, however.

24. Proofs are in the appendix unless noted otherwise.

25. If we let $\gamma_1^* = \gamma_1 - E(F_1)$ and $\hat{\gamma}_1^* = \hat{\gamma}_1 - \bar{F}_1$, then $\hat{\gamma}_1^* - \gamma_1^* = \hat{\gamma}_1 - \gamma_1$ and part (ii) can be reinterpreted in terms of this alternative parameterization. Also see Theorem 3 in Section 5.

26. The proof is similar to that of the first part of the Gauss-Markov Theorem. See, for example, Theil (1971).

27. See, for example, Newey and West (1987).

28. This follows from the condition $AX=I$. See Fama (1976), Chapter 9, for a detailed discussion of the portfolio interpretation of CSR estimators.

29. See the proof of Theorem 1 and let $U=0$.

30. Since $E = a + BE(F)$, time variation in Z would imply that a is not constant if γ_3 is nonzero. This is contrary to our assumptions in Section 1.1. A similar point is made in Hess (1980). Note that neither assumption on Z is required if we are only interested in the null hypothesis that γ_3 equals zero. Also, imposing the constraint (8) would not generally be appropriate unless $\gamma_3 = 0$, although it may be possible to transform the Z values so that (8) is still valid. See, for example, Black and Scholes (1974).

31. See Jobson and Korkie (1982).

32. See Shanken and Weinstein (1990) for a detailed analysis of pricing with respect to the CRR factors. Also see related work by Chan, Chen, and Hsieh (1985) and McElroy and Burmeister (1988).

33. The covariance matrix is not reported by CRR and is obtained separately.

34. Recall that the asymptotic covariance matrix in Theorem 1 is given for the estimator multiplied by $\sqrt{T}$.

35. Also see related work by Fama and French (1988) and Lo and MacKinlay (1988).

36. See Richardson and Stock (1989).

37. This does not require a normality assumption.

38. The proof is available on request.

39. The proof is available on request.

40. See Shanken and Weinstein (1990) for an empirical examination of this issue. Black and Scholes (1974) reject the use of GLS estimation in their dividend study, due to concerns about the precision of covariance estimates.

41. See the discussion in Shanken (1985a) and related analysis in Gibbons, Ross and Shanken (1989). MacKinlay (1987) cites simulation evidence indicating that the true standard deviation of the GLS estimator exceeds the asymptotic standard deviation.

42. As in the theoretical APT literature, this can be formalized in terms of a hypothetical infinite sequence of finite economies [see Ross (1976)] or a single infinite-asset Hilbert space economy [see Chamberlain and Rothschild (1983)].

43. The structure would have infinitely many blocks, as N goes to infinity, and satisfy some additional technical conditions. Connor and Korajczyk (1988) explore the empirical relevance of a block-diagonal structure for several factor models.

44. The proof of Theorem 5 is similar to the analysis in Litzenberger and Ramaswamy (1979). A rough sketch of the proof is as follows. First, divide each of the expressions in brackets in (15) by N, leaving the product unchanged. Assume the disturbances (and hence the measurement errors in the betas) are sufficiently independent that a weak law of large numbers can be applied to cross-sectional averages of (i) the estimates of residual variance and, (ii) products involving measurement errors and/or disturbances. In this case, $\hat{q}/N$ converges in probability, as N goes to infinity, to the limiting average residual variance (assumed to exist). Furthermore, using (13) and (14), the inverse of the first expression converges to the inverse of the limit of X'X/N (assumed to exist) and the second expression converges to the limit of X'X/N times $\bar{\Gamma}_{01}$. The desired conclusion follows.

45. This corresponds to the Litzenberger-Ramaswamy analysis when $k_1=1$ and $k_2=0$. The WLS and GLS versions of the modified estimator are based on properties of the Student t and Hotelling T^2 distributions, respectively, since the residual variances and covariances are unknown. Details are available on request.

46. The results of Section 3 also apply to the modified versions of the two-pass procedure discussed in this section. This follows from the fact that $\sqrt{T}$ times $\hat{q}$ converges in probability to zero as T approaches infinity and that the inverse of a matrix can be expressed in terms of cofactors and determinants. See Shanken (1982b), Appendix E, for a detailed proof in the case $k_1=1$, $k_2=0$.

47. In a personal communication, K. Ramaswamy reports a similar experience with two factors.

48. The innovative econometric methods of Richardson and Stock (1989) might prove useful in a more rigorous analysis.

49. If y years are used to estimate betas than c^* equals 1-(y-1)(y+1)/(3yn) times c.