NBER WORKING PAPER SERIES

ESTIMATING AND TESTING BETA PRICING MODELS:
ALTERNATIVE METHODS AND THEIR PERFORMANCE IN SIMULATIONS

Jay Shanken
Guofu Zhou

Estimating and Testing Beta Pricing Models: Alternative Methods and their Performance in
Simulations
Jay Shanken and Guofu Zhou
NBER Working Paper No. 12055
February 2006
JEL No. G12

ABSTRACT

In this paper, we conduct a simulation analysis of the Fama and MacBeth (1973) two-pass procedure,

as well as maximum likelihood (ML) and generalized method of moments estimators of cross-

sectional expected return models. We also provide some new analytical results on computational

issues, the relations between estimators, and asymptotic distributions under model misspecification.

The GLS estimator is often much more precise than the usual OLS estimator, but it displays more

bias as well. A "truncated" form of ML performs quite well overall in terms of bias and precision,

but produces less reliable inferences than the OLS estimator.

Jay Shanken
Goizueta Business School
Emory University
1300 Clifton Road
Atlanta, GA 30322
and NBER
jay_shanken@bus.emory.edu

Guofu Zhou
Olin School of Business
Washington University
St. Louis, MO 63130
zhou@wustl.edu

1 Introduction

Explaining cross-sectional differences in asset expected returns is one of the great challenges of modern finance. Theoretical models, such as the CAPM of Sharpe (1964), Lintner (1965), and Black (1972), and the intertemporal/consumption models of Merton (1973), Breeden (1979), and Rubinstein (1976), imply that expected returns should be linear in asset betas with respect to fundamental economic aggregates. Equilibrium extensions of the Ross (1976) APT and the "multivariate proxy" perspective on factor pricing models, e.g. Shanken (1987), yield similar relations. There is an enormous body of empirical research that examines these linear asset pricing relationships. One of the most widely used methodologies is the two-pass regression approach, known as the Fama-MacBeth procedure, developed by Fama and MacBeth (1973) and Black, Jensen and Scholes (1972). The two-pass procedure is used not only in asset pricing, but also in many other areas of finance. For example, Fama and French (1998), Grinstein and Michaely (2002) and Easley, Hvidjkaer and O'Hara (2002) apply it in analyzing corporate finance and market microstructure issues.[1]

The asymptotic statistical properties of the Fama-MacBeth procedure were first established in the study by Shanken (1992) and later extended by Jagannathan and Wang (1998). However, despite its widespread application, surprisingly little is known about the small-sample statistical properties of the methodology. An early simulation study by Amsler and Schmidt (1985) provides some insights, though the main focus of that paper is on multivariate tests of the linear expected return relation. A contemporaneous paper by Chen and Kan (2005) provides analytical results on estimation bias as well as some simulation evidence. Our paper attempts to fill in some of the important gaps in our knowledge, comparing the performance of the usual OLS version of the Fama-MacBeth procedure to the WLS and GLS approaches occasionally seen in the literature.[2]

In addition to the two-pass approach, we explore alternative analytically tractable procedures based on maximum likelihood (ML) estimation and the generalized method of moments (GMM) of Hansen (1982). The ML method is important in that it is asymptotically efficient under the classical independent and identically distributed (iid) multivariate normal returns assumption. The GMM approach is of further interest since serial correlation and conditional heteroscedasticity in the joint distribution of returns and factors is easily accommodated in making asymptotically valid

---

[1] Applications of the procedure in recent years can be found in at least 735 papers that cite Fama and MacBeth (1973), as complied by Google.

[2] Balduzzi and Robotti (2004) consider estimation issues related to the use of mimicking portfolios for non-traded factors.

inferences. These characteristics of the data have typically been ignored in the empirical literature, but will likely receive more attention in the future. The computational simplifications that we introduce facilitate the simulation analysis and should make these methods more accessible to researchers. We also demonstrate an equivalence between ML estimation and one form of the GMM estimator under the classical assumptions.

Application of the estimation methods considered here implicitly assumes that the expected return relation is well specified. One approach to testing this specification is to include additional cross-sectional regressors in the relation and see whether such variables are significantly related to returns. For example, Fama-MacBeth add beta-squared and residual variance in their study. This approach requires that the researcher have some idea as to the nature of the departure from the asset pricing relation and is vulnerable to the influence of data mining. In contrast, multivariate tests focus on the overall magnitude of estimated pricing deviations and have the potential to reject the null for a broad range of general alternatives. The downside, however, may be a reduction in power against particular alternatives of interest. Moreover, since the tests typically require some sort of sorting of securities into portfolios, data mining can still be a problem, e.g., Lo and MacKinlay (1990). We evaluate the actual size of various tests and power against several alternatives of interest.

One could argue that, strictly speaking, all models are false and are, at best, close approximations to reality. Moreover, even if we entertain the possibility that a given asset pricing model does hold exactly, the limited power of pricing tests implies that we cannot always know whether the model is well-specified. Thus, it is inevitable that we will often, knowingly or unknowingly, estimate an expected return relation that departs from exact linearity in the betas. We derive the asymptotic distribution of the Fama-MacBeth estimator in this context, thereby providing insight into the extent to which misspecification can affect inferences about factor risk premia.

The paper is organized as follows. Section 2 reviews the statistical specification of the asset pricing model and the different versions of the Fama-MacBeth procedures. We then present the asymptotic theory for the second-pass risk premium estimators when the asset pricing restrictions are violated. In Section 3, we study the ML method and show how to obtain the parameter estimates analytically. In Section 4, we develop two analytical GMM methods that are robust to conditional heteroscedasticity and serial correlation. Extensive simulations are conducted in Section 5, to study the finite sample properties of the various approaches to estimation and testing. An empirical application is discussed in Section 6. Conclusions are offered in the final section.

2 OLS, WLS and GLS two-pass procedures

In this section, first we briefly review the standard cross-sectional model and the associated OLS, WLS and GLS two-pass estimators and tests. Then we provide asymptotic distributional results for the estimators under the alternative hypothesis that the cross-sectional pricing restrictions are not valid.

2.1 Model, estimation and tests

We assume that the asset returns are governed by a multi-factor model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{iK} f_{Kt} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where

$\quad R_{it}$ = the return on asset i in period t ($1 \leq i \leq N$),

$\quad f_{jt}$ = the realization of the j-th factor in period t ($1 \leq j \leq K$),

$\quad \epsilon_{it}$ = the disturbances or random errors,

and T is the number of time-series observations. Like most studies, we maintain the assumption that the disturbances are independent over time and jointly distributed each period with mean zero and a nonsingular residual covariance matrix Σ, conditional on the factors. The maximum likelihood analysis in Section 3 will also impose joint normality of the disturbances, conditional on the factors. The factors are assumed to be independent and identically distributed (iid) over time though, as demonstrated in Shanken (1992), this assumption can easily be relaxed in deriving asymptotic results for the second-pass estimators. The GMM approaches presented later in Section 4 relax the iid assumption on the disturbances as well.

The asset pricing hypothesis underlying the standard two-pass procedure is[3]

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1 + \cdots + \gamma_K \beta_K, \tag{2}$$

where $E[R_t]$ is the N-vector of expected returns on the assets and $\beta_1, \ldots, \beta_K$ are N-vectors of the multiple-regression betas. In the first stage of the two-pass procedure, estimates of the betas are obtained by applying OLS to equation (1) for each asset. Let $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_K)$ be the resulting $N \times K$ matrix of OLS slope estimates. For each period t, one then runs a cross-sectional

---

[3]It may be advisable to impose additional constraints if some of the factors are portfolios. See Shanken (1992). However, such constraints have usually been ignored in the two-pass literature. Our analytical results can be extended to accommodate these constraints.

regression of $R_t = (R_{1t}, \ldots, R_{Nt})'$ on $\hat{X} = [1_N, \hat{\beta}]$ in the second stage to get an estimator of $\Gamma = (\gamma_0, \gamma_1, \ldots, \gamma_K)' = (\gamma_0, \gamma_a')'$,

$$\hat{\Gamma}_t = (\hat{X}'\hat{X})^{-1}\hat{X}'R_t. \tag{3}$$

The average,

$$\hat{\Gamma}^{\text{OLS}} = \sum_{t=1}^{T} \hat{\Gamma}_t/T = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R}, \tag{4}$$

is taken as the final estimator of Γ, where $\bar{R}$ is the N-vector of sample means of the asset returns.[4]

In the second pass, Shanken (1985), among others, proposes the use of the following GLS estimator,

$$\hat{\Gamma}^{\text{GLS}} = (\hat{X}'\widehat{\Sigma}^{-1}\hat{X})^{-1}\hat{X}'\widehat{\Sigma}^{-1}\bar{R}, \tag{5}$$

where $\widehat{\Sigma}$ is the estimator of the residual covariance matrix computed as the cross product of the fitted factor model residuals divided by T.[5] Shanken (1992) proves that the GLS estimator is asymptotically efficient.

The OLS estimator is not asymptotically efficient, in general. However, since the GLS estimator requires estimation of the inverse of the covariance matrix, it might not be expected to perform well when the sample size is small. An alternative that would seem to have potential is a weighted least-squares (WLS) estimator,

$$\hat{\Gamma}^{\text{WLS}} = (\hat{X}'\widehat{\Sigma}_d^{-1}\hat{X})^{-1}\hat{X}'\widehat{\Sigma}_d^{-1}\bar{R}, \tag{6}$$

where the weighting matrix $\widehat{\Sigma}_d$ consists of the diagonal elements of $\widehat{\Sigma}$. Litzenberger and Ramaswamy (1979) were perhaps the first to use this type of estimator in testing the CAPM.

The issue of whether a specific factor is priced has been of primary interest in the literature. The OLS t-ratios traditionally used for assessing the significance of factor pricing are computed as

$$\hat{t}_j = \frac{\hat{\gamma}_j^{\text{OLS}}}{\hat{s}_j/\sqrt{T}}, \qquad j = 1, \ldots, K, \tag{7}$$

where $\hat{\gamma}_j^{\text{OLS}}$ and $\hat{s}_j$ are the sample mean and standard deviation of the j-th component of the time series $\hat{\Gamma}_t, t = 1, \ldots, T$. Similar ratios can be computed for the WLS and GLS estimators. The p-values are usually computed from a t distribution with degrees of freedom $T-1$ or from a standard normal distribution. The advantage of evaluating significance in this manner is that cross-sectional

---

[4]In some studies, betas are estimated from a rolling window of past data, further complicating the econometric analysis.

[5]We assume $T > (N + K)$ so that the usual sample covariance estimator is invertible. Shanken (1985) shows that the same estimator is obtained using the sample covariance matrix of returns.

heteroskedasticity and correlation are implicitly taken into account, as these characteristics of the distribution influence the precision of each estimator and, therefore, are reflected in the time-series variability of the estimators. However, this approach ignores estimation error in the betas.

Note that all of the two-pass estimators can be written in the following form,

$$\hat{\Gamma}_w = \hat{A}_w \bar{R}, \quad \hat{A}_w = (\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W}, \tag{8}$$

where $\hat{W}$ is a symmetric weighting matrix. For example, the OLS estimator is obtained with $\hat{W}$ equal to the identity matrix. Under the standard iid assumption, Shanken (1992) provides the asymptotic covariance matrix for this type of estimator,

$$\Upsilon_w = \text{ACov}(\hat{\Gamma}_w) = (1+c)\Omega_w + \Sigma_f^*, \tag{9}$$

where $c = \gamma_a'\Sigma_f^{-1}\gamma_a$, γ_a, as defined earlier, is Γ excluding the first element, $\Omega_w = A_w\Sigma A_w'$, A_w is the probability limit of $\hat{A}_w$, and

$$\Sigma_f^* = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_f \end{bmatrix}, \tag{10}$$

with Σ_f the population covariance matrix of the K factors.[6] Asymptotically standard normal "t-ratios" are then obtained by dividing the estimates by their asymptotic standard errors. It follows that the Fama-MacBeth standard errors, computed as the times series standard errors of the estimated gammas, understate the true asymptotic standard errors by the amount $c\Omega_w$. Although this difference is fairly small in our simulations, it may be important elsewhere; see, e.g., Section 4.1 of Shanken (1992).

The usual interpretation of standard tests for factor pricing presumes that expected returns can indeed be expressed as a linear function of the betas. However, a risk premium parameter can be significantly different from zero based on the t-test, even if there are large pricing errors. Therefore, it is important to separately test the validity of the model. As mentioned earlier, a common approach to testing linearity is to include additional variables in the cross-sectional regression and evaluate the significance of these variables via the Fama-MacBeth method.

Shanken (1985) proposes a cross-sectional specification test of (2) against a general alternative,

$$Q_c = T(\bar{R} - \hat{\gamma}_0^{\text{GLS}}1_N - \hat{\beta}\hat{\gamma}_a^{\text{GLS}})'\hat{\Sigma}^{-1}(\bar{R} - \hat{\gamma}_0^{\text{GLS}}1_N - \hat{\beta}\hat{\gamma}_a^{\text{GLS}})/(1+\hat{c}), \tag{11}$$

---

[6]This is the limiting covariance matrix for $\sqrt{T}$ times the difference between the estimator and the true parameter value.

5

where $\hat{\Gamma}^{\text{GLS}} = (\hat{\gamma}_0^{\text{GLS}}, \hat{\gamma}_a^{\text{GLS}\prime})'$ and $\hat{c}$ is a consistent estimator of c with the GLS weighting matrix. Q_c summarizes the estimated pricing errors across assets weighted by $\widehat{\Sigma}^{-1}$, with an adjustment in the denominator for errors in the betas. The larger the pricing errors, the larger the observed value for Q_c. Intuitively, if (2) is true, Q_c should not be "too far" from zero, as the errors will be random. On the other hand, if (2) does not hold then there will be systematic deviations as well, resulting in larger values of the observed test statistic. Roll (1985) provides an interesting geometric interpretation of Q_c when the factor is the return on a benchmark portfolio.[7]

2.2 Estimation under the alternative

Standard inference using the two-pass procedures implicitly assumes that the asset pricing restriction (2) is true. What happens when this restriction is violated, as is likely to be the case in practice? The purpose of this subsection is to provide both the asymptotic distributions of the two-pass estimators under the alternative and a model specification test.

Whether the restriction is true or not, we can always project the expected return vector $E[R_t]$ onto $X = [1_N, \beta]$,

$$E[R_t] = X\Gamma_w + \eta_w, \tag{12}$$

where

$$\Gamma_w = (X'WX)^{-1}X'WE[R_t] \tag{13}$$

is the coefficient vector of the weighted projection and η_w is the projection residual or (true) pricing error vector. As the sample size, T, gets large, $\hat{\Gamma}_w$ will converge to Γ_w. This was noted previously by Kandel and Stambaugh (1995), who study the cross-section of returns when the benchmark portfolio is inefficient. We go beyond this consistency result and derive the asymptotic distribution of the second-pass estimator when the expected return relation (2) is misspecified, extending Theorem 1 of Shanken (1992).[8]

Proposition 1   Given the assumptions of Section 2.1 and the additional joint normality assump-

---

[7]An identical test statistic is obtained if the sample covariance matrix of asset returns is substituted for the residual covariance matrix in Q_c.

[8]An independent paper by Kimmel (2003) considers the GLS case with the (excess) zero-beta rate constrained to equal zero. Chen, Kan and Zhang (1999) analyze the relation between statistical significance and explanatory power for expected returns allowing for model misspecification of the sort considered here. They do not address issues related to estimation error in the betas, however.

tion for the disturbances conditional on the factors, we have

$$\sqrt{T}(\hat{\Gamma}_w - \Gamma_w) \overset{\text{asy}}{\sim} N(0, \Upsilon_w + \Upsilon_{w1} + \Upsilon'_{w1} + \Upsilon_{w2}) \tag{14}$$

where Υ_w is given in (9),

$$\Upsilon_{w1} = -(X'WX)^{-1} \begin{pmatrix} 0 \\ \Sigma_f^{-1} \gamma_{wa} \eta'_w W\Sigma \end{pmatrix} WX(X'WX)^{-1}, \tag{15}$$

which vanishes for the GLS estimator, and

$$\Upsilon_{w2} = (X'WX)^{-1} \left[\begin{pmatrix} 0 & 0 \\ 0 & (\eta'_w W\Sigma W\eta_w)\Sigma_f^{-1} \end{pmatrix} + (\eta'_w \otimes X')V_w(\eta_w \otimes X) \right] (X'WX)^{-1}, \tag{16}$$

where V_w is the asymptotic covariance matrix of $\text{vec}(\hat{W} - W)$. In the GLS case, the elements of V_w are given by

$$\text{ACov}(\hat{w}_{ij}, \hat{w}_{kl}) = \sigma^*_{ik}\sigma^*_{jl} + \sigma^*_{il}\sigma^*_{jk}, \tag{17}$$

where σ^*_{mn} is the (m,n) element of Σ^{-1}. In the WLS case, the nonzero elements are of the form,

$$\text{ACov}\left(\frac{1}{\hat{\sigma}_{ii}}, \frac{1}{\hat{\sigma}_{jj}}\right) = \frac{2\sigma_{ij}^2}{\sigma_{ii}^2\sigma_{jj}^2}, \tag{18}$$

where σ_{mn} is the (m,n) element of Σ. V_w is zero and joint normality is, therefore, not required in the OLS case.

Proof.   See Appendix A.

Proposition 1 shows that the asymptotic covariance matrix of the two-pass estimator is altered when the null hypothesis is violated. There are two new terms in addition to the usual Υ_w. These relate to the product of the $\hat{A}_w$ matrix and the pricing error vector, η_w. Υ_{w2} is the covariance matrix of this product while Υ_{w1} is its covariance with the original disturbance terms that are present in the absence of misspecification. Although $\Upsilon_w + \Upsilon_{w1} + \Upsilon'_{w1} + \Upsilon_{w2}$ is obviously positive definite, it is not clear whether it is greater than Υ_w or not. Intuitively, the interaction between the pricing errors and the errors in $\hat{\beta}$ introduces additional "noise" that may reduce the precision of the risk premium estimates. While we have not shown this theoretically, our simulation results support this intuition.

It is easy to verify that, when the null (2) is true and $E[R_t] = X\Gamma$, then Γ_w is independent of W, i.e.,

$$\Gamma_w = (X'WX)^{-1}X'WX\Gamma = \Gamma. \tag{19}$$

7

Thus, the various two-pass estimators all converge to the same limit when the null is true. In this case, differences between the estimators are entirely attributable to random estimation errors. On the other hand, when (2) is false, the estimators may differ more systematically, as we will see in the simulation examples presented later. Thus, a test of this *necessary* condition can serve as another means of evaluating expected return linearity. The asymptotic distribution of the difference of two-pass estimators follows easily from the results in Shanken (1992), yielding a simple chi-squared test as shown in the next proposition.

Proposition 2 Under the assumptions of Shanken (1992, Theorem 1), and if the $(K+1) \times N$ matrix, $\Pi_a = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} - (X'X)^{-1}X'$, is of rank $K+1$, then

$$J_\gamma = T(\hat{\Gamma}^{OLS} - \hat{\Gamma}^{\mathrm{GLS}})'[(1+\hat{c})\hat{\Pi}_a\hat{\Sigma}\hat{\Pi}_a']^{-1}(\hat{\Gamma}^{OLS} - \hat{\Gamma}^{\mathrm{GLS}}) \overset{\mathrm{asy}}{\sim} \chi^2_{K+1}, \tag{20}$$

as $T \to \infty$, where $\hat{\Pi}_a = (\hat{X}'\hat{\Sigma}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Sigma}^{-1} - (\hat{X}'\hat{X})^{-1}\hat{X}'$.

Proof.   See Appendix A.

It may be instructive to contrast the notion of model misspecification used here with that analyzed by Jagannathan and Wang (1998). They posit a "true" factor pricing model and consider the situation in which some other set of factors is employed in the estimation. Naturally, the second-pass estimator will not typically converge to the vector of coefficients in the "true" pricing model in this context and, in this sense, the estimator is biased. Consequently, the t-statistic for the risk premium associated with an observed factor can diverge to infinity (a non-zero value divided by a shrinking estimated standard error), even if the corresponding true factor is not priced. We focus, instead, on whether the *observed* factor is priced, i.e., whether betas on that factor are related to expected returns. Our asymptotic results allow the researcher to test hypotheses about that relation, even if the model is misspecified, in the sense that the factor betas do not fully account for the cross-sectional variation in expected returns on the given assets.[9]

3   The maximum likelihood approach

In this section, we provide a detailed discussion of the maximum likelihood (ML) approach as applied to the two-pass regression model. We solve the estimation problem analytically and review the associated asset pricing tests.[10]

---

[9]Factor portfolio constraints can easily be accommodated in this framework when some or all of the factors are portfolio spread returns.

[10]Our analytical expression for the ML estimator has recently been simplified by Chen and Kan (2005).

To apply the ML approach, it is convenient to rewrite the asset pricing restrictions (2) in terms of the alphas,

$$H_0: \qquad \alpha = \lambda_0 1_N + \lambda_1 \beta_1 + \cdots + \lambda_K \beta_K, \qquad (21)$$

where $\lambda_0 = \gamma_0$ and $\lambda_k = \gamma_k - E[f_k]$ for $k = 1, 2, \ldots K$. Then, the ML estimator of the risk premia vector, γ_a, is obtained by adding the factor-mean vector to the estimator of $\lambda_a = (\lambda_1, \ldots, \lambda_K)'$.[11] To find the ML estimator of λ_a, we need to maximize the likelihood function over all the parameters – the alphas, betas and $\lambda = (\lambda_0, \lambda_a')'$. Since the parameters enter the restriction multiplicatively, the constraints are non-linear. As a result, standard procedures are not applicable and special techniques have to be developed to solve the maximization problem.

Gibbons (1982) is the first to suggest using the ML method to estimate a special case of (21) for $K = 1$ with an additional constraint, $\lambda_1 = -\lambda_0$, which is an implication of the zero-beta capital asset pricing model (CAPM). This version of the CAPM can arise if one assumes the borrowing rate differs from the lending rate. Extending a result of Kandel (1984), Shanken (1986) gives an explicit solution for the zero-beta ML estimator. Here we solve for the estimator with only (21) imposed, the case most often considered in the two-pass estimation literature. Stambaugh (1982) uses ML estimation for this specification with $K = 1$.

Following Shanken (1985), one can show that the ML estimator of λ minimizes the function

$$Q(\lambda) = \frac{\tilde{a}' \widehat{\Sigma}^{-1} \tilde{a}}{1 + (\bar{F} + \lambda_a)' \widehat{\Delta}^{-1} (\bar{F} + \lambda_a)}, \qquad (22)$$

where $\tilde{a} = \hat{\alpha} - \lambda_0 1_N - \lambda_1 \hat{\beta}_1 - \cdots - \lambda_K \hat{\beta}_K$ and $\bar{F}$ and $\widehat{\Delta}$ are the sample mean and covariance matrix of the factors. Note that the numerator is a quadratic form in the pricing errors. Minimizing this quadratic form alone yields the GLS cross-sectional regression estimator.

Given (22), the ML estimator can be computed as follows. Let

$$\hat{\alpha}^* = \hat{\alpha} - (1_N' \widehat{\Sigma}^{-1} 1_N)^{-1} (1_N' \widehat{\Sigma}^{-1} \hat{\alpha}) 1_N, \qquad (23)$$

$$\hat{\beta}_j^* = \hat{\beta}_j - (1_N' \widehat{\Sigma}^{-1} 1_N)^{-1} (1_N' \widehat{\Sigma}^{-1} \hat{\beta}_j) 1_N, \qquad (24)$$

for $j = 1, \ldots, K$ and $\hat{\beta}^* = (\hat{\beta}_1^*, \ldots, \hat{\beta}_K^*)$. Then compute $(K + 1) \times (K + 1)$ matrices A and B as

$$A = \begin{pmatrix} \hat{\alpha}^{*\prime} \widehat{\Sigma}^{-1} \hat{\alpha}^* & -\hat{\alpha}^{*\prime} \widehat{\Sigma}^{-1} \hat{\beta}^* \\ -\hat{\beta}^{*\prime} \widehat{\Sigma}^{-1} \hat{\alpha}^* & \hat{\beta}^{*\prime} \widehat{\Sigma}^{-1} \hat{\beta}^* \end{pmatrix}, \qquad B = \begin{pmatrix} 1 + \bar{F}' \widehat{\Delta}^{-1} \bar{F} & \bar{F}' \widehat{\Delta}^{-1} \\ \widehat{\Delta}^{-1} \bar{F} & \widehat{\Delta}^{-1} \end{pmatrix}. \qquad (25)$$

---

[11]Likewise, the factor covariance matrix is added to the asymptotic covariance matrix of λ_a. This assumes that the factors are iid and that the ML estimator for the factor mean is just the sample mean, as would be the case, for example, under normality.

Now let L be a lower-triangular matrix such that $L'L = B^{-1}$. It is straightforward to compute the eigenvalues and eigenvectors of

$$|LAL' - \zeta I| = 0. \tag{26}$$

Letting $(1, \tilde{\lambda}_a')'$ be the eigenvector (re-scaled to make the first element be one) corresponding to the smallest eigenvalue of (26), we have

Proposition 3 The ML estimator of $\lambda_a = (\lambda_1, \ldots, \lambda_K)'$ equals $\tilde{\lambda}_a$ and the ML estimator of λ_0 is

$$\tilde{\lambda}_0 = (1_N' \widehat{\Sigma}^{-1} 1_N)^{-1} 1_N' \widehat{\Sigma}^{-1} (\hat{\alpha} - \tilde{\lambda}_1 \hat{\beta}_1 - \cdots - \tilde{\lambda}_K \hat{\beta}_K). \tag{27}$$

Proof.   See Appendix A.

Given the ML estimate of λ, the constrained beta estimates are obtained by running time-series regressions of $R_t - \tilde{\lambda}_0 1_N$ on $F_t^* = (f_{1t} + \tilde{\lambda}_1, \ldots, f_{Kt} + \tilde{\lambda}_K)'$. To guard against possible coding errors, it is advisable to verify that the first order derivatives of the likelihood function are identically equal to zero (to the accuracy of rounding error) when evaluated at the constrained estimates, i.e.,

$$\frac{\partial \mathcal{L}_c}{\partial \lambda_j} = \beta_j' \Sigma^{-1} \sum_{t=1}^{T} [R_t - \lambda_0 1_N - \beta_1(f_{1t} + \lambda_1) - \cdots - \beta_K(f_{Kt} + \lambda_K)] = 0, \tag{28}$$

for $j = 1, \ldots, K$. When K=1, we also have the following results for the risk premium estimator.

Proposition 4 The ML estimator of γ_1, $\tilde{\gamma}_1^{\mathrm{ML}}$, is the root of a quadratic equation. Moreover, if the GLS estimator satisfies $\hat{\gamma}_1^{\mathrm{GLS}} > 0$, we have

$$\tilde{\gamma}_1^{\mathrm{ML}} > \hat{\gamma}_1^{\mathrm{GLS}} \tag{29}$$

and the reverse holds if $\hat{\gamma}_1^{\mathrm{GLS}} < 0$.

Proof.   See Appendix A.

Kandel (1984) and Roll (1985) derive a similar result when the single factor is a portfolio and the associated restriction on the risk premium estimator is imposed. Proposition 4 treats the case of a general one-factor model, without any such restriction.

The traditional analysis of linear regression estimation with an independent variable that is measured with error reveals a bias in the estimated slope coefficient toward zero. Chen and Kan (2005) show that this conclusion holds for the second-pass GLS estimator of γ_1 as well. Against this background, Proposition 4 is interesting in that it suggests that (simultaneous) ML estimation of all parameters may eliminate this errors-in-variables bias. In fact, Shanken (1992) demonstrates

that this is true, in a limiting sense, when the number of assets, N, is large. Thus, one might have conjectured that the ML estimator is, indeed, unbiased in finite samples. However, Chen and Kan (2005) have also shown that the ML estimator of γ_1 does not even have a mean, shedding light on Amsler and Schmidt's (1985) observation that the ML estimator sometimes takes on extreme values in their simulations.

It is difficult to know what message to take away from all of this in terms of applied work. In particular, it would not make sense to focus on the moments of the ML estimator in simulations. Instead, we explore a simple "truncated" version of the ML estimator; specifically, if the absolute value of the ML estimator is more than twice that of the GLS estimator, we set the truncated estimator equal to the GLS estimator, otherwise it is unchanged.[12] This estimator will have finite moments whenever the GLS estimator does and will have the same asymptotic distribution, as $T \rightarrow \infty$, as the GLS and ML estimators.[13] Its performance in finite samples remains to be evaluated.

One approach to evaluating model specification is to use the standard likelihood ratio test of (2), which can be shown to equal $T \log(|\widetilde{\Sigma}|/|\widehat{\Sigma}|)$, where $\widetilde{\Sigma}$ is the Σ estimator evaluated at the constrained ML estimator of the alphas and betas. This test has an asymptotic chi-squared distribution χ^2_{N-K-1}. Since the chi-squared distribution is only a first-order asymptotic approximation and tends to reject too often in this context, Jobson and Korkie (1982) suggest the use of a Bartlett (1947) correction,

$$\text{LRT} = [T - \frac{1}{2}(N + K + 3)] \log(|\widetilde{\Sigma}|/|\widehat{\Sigma}|) \stackrel{\text{asy}}{\sim} \chi^2_{N-K-1}. \tag{30}$$

Asymptotically, both tests have the same limiting distribution. The Bartlett-corrected test performs much better, however, and hence we will use this throughout. As an alternative, one can substitute the ML estimator of Γ for the GLS estimator in the earlier Q_c statistic and perform an F-test of (2). Shanken (1985) shows that the resulting statistic is actually a monotonic transformation of the likelihood ratio test for this problem.[14]

---

[12]Of course, other truncation rules could be considered. For example, we also examined a 5 times GLS rule with similar results.

[13]We make use of the fact the the mean of a random variable is finite if and only if the mean of its absolute value is finite.

[14]Under additional factor portfolio constraints of the multi-beta CAPM, Velu and Zhou (1999) provide the small sample distribution of the likelihood ratio test, which is shown to depend on nuisance parameters.

4 The GMM Approach

The methods discussed thus far – the different versions of the Fama-MacBeth two-pass methodology and the traditional maximum likelihood approach – assume that returns are independent and identically distributed over time. If returns exhibit heteroskedasticity conditional on the factors or serial correlation, the standard errors of the parameter estimates may not be correct, even asymptotically, and the associated tests may no longer be valid. Shanken (1992) shows how to adjust Fama-MacBeth standard errors for serial correlation in the factor component of returns, while Jagannathan and Wang (1998) derive the asymptotic covariance matrix under conditional heteroskedasticity.

A conceptually simple and more general solution to the problem, advocated by Cochrane (2001), is to use Hansen's (1982) GMM approach, which is robust to both conditional heteroscedasticity and serial correlation in the return residuals as well as the factors. Following MacKinlay and Richardson (1991) and Harvey and Zhou (1993), the factor model moment conditions are

$$E\left[\mathcal{E}_t \otimes \begin{pmatrix} 1 \\ F_t \end{pmatrix}\right] = E\left[\begin{array}{c} R_t - \alpha - \beta F_t \\ (R_t - \alpha - \beta F_t) \otimes F_t \end{array}\right] = 0. \tag{31}$$

These earlier papers explore GMM-based tests of the familiar zero-intercept restriction, which requires that the factors are portfolio excess returns and that the zero-beta rate is known, typically assumed to equal a riskless Treasury bill rate. We consider both estimation and testing under the more general restriction (21).

Let g_T be the sample moments:

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathcal{E}_t(\theta) \otimes Z_t, \quad NL \times 1, \tag{32}$$

where $\theta = (\lambda', \beta_1', \ldots, \beta_K')'$ is the vector of parameters, $L = K + 1$ and $Z_t = (1, F_t')'$. The GMM estimator requires the solution of

$$\min Q = g_T(\theta)' W_T g_T(\theta), \tag{33}$$

under the constraint (21), where W_T, $NL \times NL$, is a positive definite weighting matrix. Since the constraint is non-linear and the number of parameters, $q = NK + L$, is large, numerical solutions to (33) are difficult to obtain. The problem is exacerbated as numerical solutions may not converge to the global minimum or even converge at all. These difficulties in implementing GMM might be one reason for its infrequent use in estimating the cross-sectional regression model.

While intractable in general, for a special class of weighting matrices, it is possible to solve (33) analytically in terms of λ_0 only. Thus, the multi-dimensional optimization problem, which is often difficult to implement in practice, is reduced to a one-dimensional problem that can easily be solved using one of the many available algorithms. This is particularly important for simulations in which thousands of estimates must be computed. We summarize the key result as

Proposition 5   If the weighting matrix is of the following form,

$$W_T = W_1 \otimes W_2, \qquad W_1 : \; N \times N, \quad W_2 : L \times L, \tag{34}$$

then the GMM estimator of θ is given as a function of λ_0 and the data,

$$
\begin{aligned}
\tilde{\lambda}'_a &= \tilde{A}_1 \tilde{A}_2^{-1}, \\
(\tilde{\beta}_1, \ldots, \tilde{\beta}_K)' &= \tilde{A}_2 (X^{*\prime} P X^*)^{-1} X^{*\prime} P (R - \lambda_0 1_T 1'_N),
\end{aligned}
\tag{35}
$$

where $\tilde{A} = (Z'PZ/T^2)^{-1/2} \tilde{E} = (\tilde{A}'_1, \tilde{A}'_2)'$ with $\tilde{A}_2$ the lower $K \times K$ submatrix of $\tilde{A}$, $P = ZW_2Z'$, $X^* = Z\tilde{A}$, R is a $T \times N$ matrix formed from the R_t's, Z is a $T \times L$ matrix formed from the Z_t's, and $\tilde{E}$ is an $L \times K$ matrix formed from the standardized eigenvectors ($\tilde{E}'\tilde{E} = I_K$) corresponding to the K largest eigenvalues, $\xi_1, \ldots, \xi_K$, of the following $L \times L$ matrix:

$$(Z'PZ/T^2)^{-1/2}[Z'P(R - \lambda_0 1_T 1'_N)/T^2]W_1[Z'P(R - \lambda_0 1_T 1'_N)/T^2]'(Z'PZ/T^2)^{-1/2}. \tag{36}$$

Moreover, the estimator for λ_0 is obtained by minimizing the objective function

$$Q^*(\lambda_0) = \text{tr}\left[W_1 (R - \lambda_0 1_T 1'_N)' P (R - \lambda_0 1_T 1'_N)/T^2 \right] - \xi_1 - \cdots - \xi_K. \tag{37}$$

Proof.   See Appendix A.

In general, the optimal weighting matrix associated with the moment conditions (31) is S_T^{-1}, where S_T is a consistent estimator of

$$S_0 = \sum_{j=-\infty}^{\infty} E[g_t(\theta)g_{t-j}(\theta)']. \tag{38}$$

One example is the well-known estimator of Newey and West (1987). The optimal matrix is of the required form when the data are iid, but otherwise need not satisfy the condition in Proposition 5. In the iid case, the natural consistent estimator of S_0 is

$$\hat{S}_{\text{iid}} = \hat{\Sigma} \otimes \left(\frac{1}{T} \sum_{t=1}^{T} Z_t Z'_t \right). \tag{39}$$

With this choice of the weighting matrix, we refer to the estimator as GMM1.[15] We then have the following interesting result, which has not been noted previously in the literature.

Proposition 6 With weighting matrix $W_T = \hat{S}_{\text{iid}}^{-1}$ as in (39),

$$\lambda^{\text{GMM1}} = \lambda^{\text{ML}}, \tag{40}$$

i.e., the GMM1 estimator is numerically identical to the ML estimator.

Proof. See Appendix A.

This equivalence is surprising, insofar as the objective function that is minimized by the GMM1 estimator appears, at first glance, to be quite different from the likelihood function that is maximized by the ML estimator. We have verified the equivalence numerically as well, however. Although the ML approach is computationally more convenient in the iid normal case, the computational simplification provided by Proposition 5 for GMM estimation may be important in other applications in which these strong assumptions are weakened. Given the surprising Chen and Kan (2005) results on ML estimation noted earlier, it follows immediately from Proposition 6 that the finite sample moments of the GMM1 estimator do not exist. On the other hand, it is not easy to directly derive the asymptotic distribution of the ML estimator in the presence of conditional heteroscedasticity and/or serial correlation. Now this follows straightforwardly from standard GMM results, as outlined below.

With an arbitrary weighting matrix W_T and a given consistent estimator, S_T, of S_0, the asymptotic covariance matrix of the GMM1 estimator is provided by the standard GMM theory as

$$\hat{\Sigma}_\theta = (D_T' W_T D_T)^{-1} D_T' W_T S_T W_T D_T (D_T' W_T D_T)^{-1}, \tag{41}$$

where D_T, $NL \times (NK + L)$, is the matrix of derivatives of $g_T(\theta)$ with respect to the parameters. This formula can be used to obtain standard errors for the risk premium estimates for any weighting matrix W_T. Based on this, the associated "t-ratios" can then be computed and are asymptotically valid under conditional heteroscedasticity and/or serial correlation of the data.

When the optimal weighting matrix is employed, T times the GMM quadratic in (33) follows a central chi-squared distribution with degrees of freedom equal to the number of moment conditions minus the number of parameters, a standard result in the GMM literature. More generally - for example, if an alternative weighting matrix is used to obtain an analytical GMM estimator, the

---

[15]Kan and Zhou (1999) and Jagannathan and Wang (2002) consider a related application of GMM for a model with the (excess) zero-beta rate constrained to equal zero.

distribution will be non-central chi-squared. Nevertheless, a simple analytical chi-squared test of model specification can be obtained as described in Theorem 1 of Zhou (1994). Like the GMM1 estimator, this test is robust to conditional heteroscedasticity and serial correlation.

The estimator GMM1 has been defined in terms of moment conditions that are based on the factor-model regression parameterization of returns. As with ML, the parameters in the joint distribution of the factors play no role in the estimation of λ in this case. These moments become relevant when the factor means are added to the λ estimates to obtain the GMM1 estimates of the γ's.[16] An alternative formulation of the moment conditions provides direct estimates of the γ's. This approach, which builds on Harvey and Kirby (1995), uses the fact that betas can be expressed in terms of moments involving the return/factor means and the covariance matrix of the factors. Alphas need not be separately identified.

For simplicity, consider the case $K = 1$. The multifactor case is treated in Appendix A.7. Note that the asset pricing restriction, equation (2), can be written as

$$E[R_t] = \gamma_0 1_N + \gamma_1 \frac{\text{cov}(R_t, f_{1t})}{\sigma_1^2}, \tag{42}$$

where σ_1^2 is the factor variance. Using this parameterization, the relevant moment conditions are

$$E[h_t(\varphi)] = E \begin{bmatrix} R_t - \mu_r \\ f_{1t} - \mu_1 \\ (f_{1t} - \mu_1)^2 - \sigma_1^2 \\ R_t - \gamma_0 1_N - \gamma_1 \frac{(R_t - \mu_r)(f_{1t} - \mu_1)}{\sigma_1^2} \end{bmatrix} = 0, \tag{43}$$

where μ_r and μ_1 are the population means of the returns and the factor, and φ is the vector of all the parameters, μ_r, μ_1, σ_1^2 and Γ. There are $2(N + 1)$ moment conditions in (43).

We partition $h_t(\varphi)$ into two sub-vectors, $h_{1t}(\varphi_1)$ and $h_{2t}(\varphi_1, \varphi_2)$, where φ_1 is a vector of the first $N + 2$ parameters, and $\varphi_2 = \Gamma$. Since the number of moment conditions in the first set is equal to the number of parameters, this subsystem is exactly identified. Hence the GMM estimator of φ_1 is $\hat{\varphi}_1 = (\bar{R}', \bar{f}_1{}', \hat{\sigma}_1^2)'$, independent of the weighting matrix. Plugging these estimates into the last N moment conditions enables us to identify the risk-return parameters by setting $E[h_{2t}(\hat{\varphi}_1, \varphi_2)] = 0$. Newey (1984) was the first to consider this type of sequential estimator in the GMM context.

---

[16]One can add moment conditions for the factor means to the GMM1 system. Under the assumption that the factor model disturbances have zero mean conditional on the factors, the covariance matrix of the GMM estimator will be block diagonal and so the covariance matrix of the lambdas will be unaffected. In this case, the asymptotic covariance matrix of $\hat{\gamma}_a$ is obtained by adding the asymptotic covariance matrix of the factor sample mean vector (the factor covariance matrix in the iid case) to the asymptotic covariance matrix of $\hat{\lambda}_a$.

To estimate φ_2 in the second-step, the choice of weighting matrix will matter, however. Indeed, given $\hat{\varphi}_1$ and a weighting matrix W_{2T}, we can find the solution to $\min h'_{2T} W_{2T} h_{2T}$ analytically,

$$\hat{\varphi}_2 = (D'_{22} W_{2T} D_{22})^{-1} D'_{22} W_{2T} \bar{R}, \tag{44}$$

where D_{ij}, $i, j = 1, 2$, is the (i, j) block of D_T, the $2(N+1) \times (N+4)$ matrix of the derivatives of $h_T(\varphi)$ with respect to the parameters.[17] Following Ogaki (1993), the optimal weighting matrix is

$$W_{2T} = \left(\begin{bmatrix} -D_{21} D_{11}^{-1} & I_N \end{bmatrix} S_T \begin{bmatrix} -D_{21} D_{11}^{-1} & I_N \end{bmatrix}' \right)^{-1}, \tag{45}$$

where S_T is a consistent estimator of the covariance matrix of moment conditions (43). We call the associated estimator GMM2 in what follows. The asymptotic covariance matrix of GMM2 is $(D'_{22} W_{2T} D_{22})^{-1}$ and the GMM specification test is $J_2 = T \min h'_{2T} W_{2T} h_{2T}$, which has a standard chi-squared distribution in the limit, with degrees of freedom $N - 2$.[18]

As Cochrane emphasizes, the Fama-MacBeth estimators can, like most estimators, be embedded in the general GMM framework. In the context of our sequential GMM formulation, the OLS and GLS second-pass estimators are obtained by letting the weighting matrix used in estimating $\phi_2 = \Gamma$ equal the identity matrix or $\hat{\Sigma}$, respectively. Cochrane's (2001) GMM formulation combines elements of each of our GMM approaches. He starts with the moment conditions (31), but does not impose the asset pricing restrictions on the alphas. Thus, the regression parameters are exactly identified as the usual OLS estimates. For each asset, he then adds a separate moment condition that corresponds to the expected return relation (2), and different weighting matrices are considered in estimating Γ (he excludes the zero-beta rate for simplicity). Although one can show that the matrix, S_0, for this system is singular, the usual formula does deliver valid standard errors for the GLS case that Cochrane considers.[19] We have also verified numerically that the GLS estimator is the optimal GMM estimator for this system in the sequential sense of Ogaki (1993).[20]

---

[17] D_{22} is $N \times 2$, D_{11} is is $(N+2) \times (N+2)$, and D_{21} is $N \times (N+2)$.

[18] Although we have not proved the existence of finite moments for GMM2, we doubt that there is a problem. The moment conditions for GMM1 and GMM2 are quite different and, as we will see, so are their estimates. Indeed, GMM2 tends to behave much more like the GLS estimator in simulations, as might be expected from Eq. (44).

[19] Let v be any nonzero vector orthogonal to both a and β in (12.23) of Cochrane (2001) and consider the vector $(v, 0, -v)$. Pre-multiplying the moment conditions by the transpose of the vector $(v, 0, -v)$ yields 0, so the covariance matrix S must be singular.

[20] Another way of bypassing the numerical difficulty of solving (33), as theoretically justified by Newey (1985), is to obtain the optimal GMM estimator from the scoring algorithm based on a known consistent estimator, such as second-pass OLS. This estimator and the associated tests are not at all reliable, however. For example, with the simulated data later in the paper, the empirical rejection rates are over 90% for a nominal 5% test, even when the sample size is 960. Similar problems have been found in latent variable models by Zhou (1994). Therefore, this method will not be used here.

5 Simulations

In this section, we study the finite sample properties of the various estimation procedures and the associated tests. One-factor and three-factor pricing models are explored. To make our simulations realistic, we calibrate the parameters by using the most recent 40 years, January 1964 - December 2003, of monthly returns on the well-known Fama-French 25 book-to-market and size portfolios, which are available from French's website.[21] In addition, to see how the results vary with the number of assets and over different groups of assets, we also calibrate the parameters using French's 48 industry portfolios. Given the model parameters, returns can be simulated for any sample size T. In the simulations that follow, we draw 10,000 data sets for each scenario considered.

5.1 Under the null that asset pricing restrictions hold

Assume in this subsection that the asset pricing restrictions are true. In this case, the expected returns on the N assets can be obtained from (2), given pre-specified risk premia, with the other parameter values, the betas and Σ obtained from time-series regressions using the actual data. We set $\gamma_0 = 0.0833\%$, or 1% on an annualized basis. This can be viewed as the differential between γ_0 and the riskless rate since the various quantities of interest will be invariant to the level of expected returns and power will depend only on the differential. In this context, the Sharpe-Lintner restriction amounts to the null $\gamma_0 = 0$. To examine the size of the t-ratio test under the null, we let γ_0 equal 0 in the simulations.

We consider two scenarios for the number of factors, $K = 1$ and $K = 3$. When $K = 1$, the excess return on the market index is used to calibrate the parameters with $\gamma_1 = 0.6667\%$ unless indicated otherwise. This value for γ_1 implies an annualized market risk premium of 8%. When $K = 3$, the Fama-French book-to-market and size factors are used to calibrate the parameters with risk premia set to $\gamma_2 = 0.3333\%$ and $\gamma_3 = 0.1667\%$ unless otherwise noted.

The number of assets N is either 25 or 48. When $N = 25$, we consider sample sizes $T = 60$, 120, 240, 360, 480 and 960. When $N = 48$, we use the same T's except 60 to avoid near singularity of the simulated sample covariance matrices. Studies such as Fama and French (1992) use sample sizes close to $T = 360$. $T = 960$ approximates the sample size of a study that uses all data going back to the 1920's. Varying T is useful in understanding the small-sample behavior of the tests and the validity of asymptotic approximations. For now, the data-generating process is the standard

---

[21]We are grateful to Ken French for making this data available on his website.

multivariate normal distribution, while a t distribution will be used later.

Consider first the case in which there is only one factor, $K = 1$. Table 1 provides the estimation results for the factor risk premium, γ_1. For each of the estimation methods, we report the average estimate, $\bar{\gamma}_1$, then the percentage error, $(\bar{\gamma}_1 - \gamma_1)/\gamma_1$, and finally the root-mean-square error (RMSE), the square root of $\sum_{m=1}^{M} (\hat{\gamma}_1^{(m)} - \gamma_1)^2/M$ with $M = 10,000$. We begin by discussing the results for $N = 25$ shown in the upper portion of Table 1.

Amsler and Schmidt (1985) note that when T is small, the maximum likelihood estimator is extremely volatile across simulations. We find this as well for the different test portfolios we consider. The ML results reported in the tables reflect truncation, as discussed earlier. Without truncation, the RMSE for the estimator of γ_1 is theoretically unreliable. Nevertheless, it is of interest to examine its value in the given simulation, which is 6.4 (13% bias) when $T = 60$, as compared to 1.12 (−8% bias) in Table 1. Truncation has little impact when $T = 120$ and no effect for larger values of T. These are much longer samples than the 6 years examined by Amsler and Schmidt, but modest in relation to typical applications. GLS and GMM2 perform the best in terms of RMSE, with truncated ML (henceforth simply ML) not far behind when T is at least 240. With a sample size as large as $T = 960$, all five of these estimators have about the same RMSE.

When $N = 25$ and T is at least 360, all methods have negative but fairly small percentage errors, implying that the estimators are slightly biased downward. This is expected for the two-pass estimators, given the well-known errors-in-variables (EIV) problem relating to estimation of the betas. Consistent with Proposition 4 and the positive value of γ_1, the average ML estimate across simulations exceeds the average GLS estimate. For example, when $T = 360$ the average is 0.664 for ML and 0.615 for GLS. As expected, the magnitude of the difference decreases as the sample size increases. When $T = 120$ or greater, ML has percentage errors of just 1% or less, the least among all the methods. The finding for ML is in keeping with Shanken's demonstration that the ML estimator has a desirable N-consistency property, a sort of asymptotic unbiasedness, for fixed T. Panel B of Table 1 provides the results when $N = 48$. The qualitative conclusions are essentially the same as those for Panel A. Although the biases are minimal for ML, they are larger for the other estimators. This may be due to higher residual variances for the 48 industry portfolios, exacerbating the EIV problem.

Table 2 provides the estimation results on γ_0, the intercept in the expected return relation. As was the case in estimating γ_1, GLS and GMM2 are the best estimators of γ_0 in terms of RMSE, with ML close behind. Consistent with the EIV perspective, the biases in the intercept are all

positive for the two-pass estimators and for GMM2 as well. The percentage biases are quite large, mainly due to the small magnitude of γ_0, as we discuss below. As earlier, ML has the least bias, at most 5% in magnitude when $T \geq 240$. To conserve space, we do not report the simulation standard deviations of the estimates in Tables 1 and 2. These are generally close to the RMSE's, however. For example, with $N = 48$ and $T = 240$ the difference is about 0.015 for the GLS estimator of γ_0 despite the large percentage bias of 110%.

Next, we examine the standard errors of the estimates. For each estimator, Table 3 provides the RMSE followed by the asymptotic standard error evaluated at the true parameters, the average of the estimated asymptotic standard errors and, finally, the root-mean-square error of these estimated standard errors in parentheses. Results are given for $N = 25$ in Table 3a and $N = 48$ in Table 3b. Notice that the standard errors for γ_0 and γ_1 happen to be similar in magnitude. These standard errors would be unaffected by changing the value of γ_0, as the γ_0 estimates would be shifted by a fixed amount and the risk premium estimates would be unchanged. It also follows that the percentage bias in estimating γ_0 would decline if its true value were increased in the simulations.

There are several important observations to be made about the standard errors. First, there is a tendency for the asymptotic standard errors of the OLS and WLS estimators to be a bit greater than RMSE when evaluated at the true parameter values, particularly in small samples. However, the asymptotic standard errors decline when evaluated at the second-pass estimates. In fact, this decline is observed for all estimators, though both standard errors are consistently below the RMSE for ML. For example, with $N = 48$ and $T = 120$, the OLS standard error for γ_0 drops from 0.57 to 0.50, which is less than the RMSE of 0.53, whereas the standard errors for ML are 0.38 and 0.26, and RMSE is 0.58.

When $T \geq 360$, the two asymptotic standard errors are fairly close to each other and to the RMSE's of the estimates for the second-pass GLS/GMM2 estimators of the risk premium. The estimated second-pass GLS/GMM2 standard errors for the pricing intercept display more downward bias, however. The estimated ML standard errors have the worst bias; when $T = 360$, the RMSE is understated by as much as 10% for γ_1 and 20% for γ_0, with $N = 48$. The problem diminishes as T increases but would be of particular concern if, contrary to typical practice, relatively short subperiods were to be used. Note that GLS, ML, and GMM2 all have the same asymptotic standard errors when evaluated at the true parameter values. Asymptotic equivalence was proven analytically for GLS and MLE in Shanken (1992), but apparently extends to our sequential GMM estimator as well when returns are iid normal.

Now we consider "t-ratio" tests of the null hypothesis that a given parameter, either γ_0 or γ_1, equals zero. All tests are two-sided, although one could argue for one-sided tests in a CAPM framework. The nominal size of the tests is set at 5% based on the asymptotic chi-squared distribution of the squared ratios. Table 4a (4b) provides both the actual rejection rates under the null and the empirical power of the test when $N = 25$ ($N = 48$). The traditional OLS test and the WLS test are well-specified under the null in most cases.

In testing $\gamma_0 = 0$, the rejection rates are often large relative to 5% for the efficient estimators. Misspecification is greatest for the ML test and $N = 48$, with rejection probability 10% even when $T = 480$. This is due, in part, to the high kurtosis of ML (despite truncation) in the simulated data sets. Since the actual size of the tests often exceeds 5%, rejection rates under the alternative overstate the power of a 5% test. To overcome this problem, we compute the empirical power by using the upper 95th percentile of the 10,000 squared ratios simulated under the null. Although this is really an estimate of the population percentile, it should result in a better indication of the actual power of the test. In the limit, as T approaches infinity, power must approach one. However, even with $T = 960$, power is always less than 10% in the second panels of Tables 4a and 4b. Thus, we conclude that the test would have virtually no power in typical applications when the annualized value of γ_0 is only 1%, a plausible differential a priori in the context of the zero-beta CAPM.

When testing the risk premium hypothesis $\gamma_1 = 0$, all of the tests except ML have rejection rates close to 5% under the null. Again, ML tends to reject too often, though the problem is not severe (rate $< 8\%$) when $T \geq 360$. The power of the tests is substantial when the annualized risk premium equals 8%, but does not exceed 0.5 until $T \geq 480$. There are some power differences across estimation methods, with tests based on the OLS and WLS estimators exhibiting the lowest power and ML tests the highest.

It is interesting that the OLS version of the Fama-MacBeth method, used so extensively in the literature, is clearly dominated in terms of precision and power by the estimation methods that incorporate information about the covariance matrix of asset returns. A common view, generally unstated, is that while GLS-type estimators surely dominate in terms of asymptotic properties, the benefits will be lost when covariances have to be estimated. Our results for the one-factor model do not support this view. Moreover, it is surprising that adjusting the cross-sectional regressions for heteroskedasticity barely makes any difference in terms of bias, precision or power.[22] Next, we

---

[22]An advantage for WLS could emerge when a large set of individual stocks or less-diversified portfolios is employed since heteroskedasticity is likely to be greater in that case. Typical applications employ a relatively small number of portfolios, however.

see whether these conclusions continue to hold in the context of a multifactor model.

Consider now what happens when there are $K = 3$ factors. As earlier, we set the (truncated) ML estimates for all parameters equal to the GLS estimates when the magnitude of the ML estimate of any risk premium is more than twice that of the GLS estimate. Truncation is observed more often now, occurring a few times even when $T = 960$. Conclusions for the market risk premium estimators (results not shown) are qualitatively similar to those based on the one-factor model, except for a decline in the relative precision of ML. The variation in RMSE across estimators is less with the Fama-French factors, however, and biases are generally larger (except ML).

Table 5 reports estimation results for the book-to-market premium. Once again, ML has very little, if any, bias. GMM2 has the lowest RMSE, despite the fact that it is the most biased estimator. Differences in RMSE are minor when $N = 25$, except for the smallest values of T. A bit more variation in RMSE is observed when $N = 48$, ranging from 0.18 for GMM2 to 0.22 for OLS when $T = 360$. Results for the size premium and the zero-beta intercept (not shown) are similar to those for book-to-market.

We have also examined simulation results for standard errors and t-ratios in the three-factor model. All of the findings are summarized, though only results for the market and book-to-market premia are shown (Tables 6a and 6b). The rest are available on request. As in Tables 3a and 3b, we find downward bias in the estimated standard errors for γ_0 using GLS and GMM2, with the worst bias using ML. The estimated ML standard errors for γ_1 are too low as well, by as much as 15% when $T = 360$, and the ML RMSE's are particularly high with $N = 25$. The behavior of the t-ratios for γ_0 and γ_1 is also fairly similar to that observed earlier. The estimated standard errors for the size and book-to-market premia are generally quite close to the RMSE's and to the true asymptotic values for all estimators. Likewise, the tests for size and book-to-market premia are typically well specified under the null, with GMM2 sometimes displaying a slight tendency to underreject.

We now turn to tests of the maintained assumption of an exact linear relation between expected return and beta, as given in (2). Tables 7 and 8 examine the actual size of the various (nominal) 5% level tests of this linear specification for the one and three-factor models, respectively. The rejection rates are based on simulations in which the expected return relation holds, with the positive parameter values described earlier. The cross-sectional tests both perform quite well for all models except when T is very small. GMM2 over-rejects in tests of the three-factor model. LRT displays a more systematic tendency to over-reject somewhat, with probabilities ranging from 0.06

to 0.09 when $T = 360$. Finally, our univariate specification test based on the difference between OLS and GLS estimators is fairly well specified for samples this large.

In addition to the multivariate tests, we have examined the traditional Fama-MacBeth univariate test based on the t-ratio for an additional cross-sectional variable. This experiment is conducted for $N = 25$ under the CAPM null hypothesis. The additional independent variable is taken to be the (time-series) average book-to-market ratio for each portfolio, as reported in Fama and French (1993). The rejection rates (not shown) are close to 5% for OLS and WLS at all sample sizes. The GLS test rejects too often in small samples. The rejection rate declines to 6.5% for $T = 360$, but is still 6% for T as large as 960.

5.2 Results when the linear expected return model is misspecified

Like the asymptotic results on estimation in Shanken (1992) and Jagannathan and Wang (1998), our simulations thus far have assumed that the linear expected return model is correctly specified. Now we allow for the possibility of deviations from the model. Of course, there are many ways in which the expected return restriction could be violated. We consider two cases of interest, both of which entail estimation and testing of a misspecified zero-beta CAPM. In each case, pricing errors are added to the CAPM expected returns with $\gamma_0 = 0.0833\%$ and $\gamma_1 = 0.6667\%$, as earlier. When $N = 25$, the pricing errors are taken to be the Fama-French estimates of the excess-return alphas for the size/book-to-market portfolios. When $N = 48$, we randomly draw a CAPM deviation from a normal distribution with mean zero and standard deviation 0.1667% or 2% annualized. In each case, the deviations are taken to be fixed over time and thus betas and covariance parameters are unaffected.

As discussed in Section 2, when using the Fama-MacBeth method, the implied coefficients in the single-factor expected return relation are determined by projecting the true expected return vector on the univariate beta vector and a constant vector. The projection varies with the weighting matrix employed. For $N = 25$, the true parameters under OLS, WLS, and GLS are 1.03%, 0.76%, and 1.01%, respectively, for γ_0 and -0.06%, 0.14%, and -0.25% for γ_1. Thus, introducing the Fama-French alphas largely eliminates the correlation between expected returns and market betas. This is presumably driven by the relatively high (low) betas and negative (positive) alphas of low (high) book-to-market stocks. When $N = 48$, the projection coefficients are $\gamma_0 = 0.22\%$, 0.28%, 0.52%, and $\gamma_1 = 0.55\%$, 0.48%, 0.20%, for OLS, WLS, and GLS, respectively. In this case, the OLS and WLS parameters are closer to the values used for the baseline CAPM model due to the

random nature of the pricing errors.

As expected, when we simulate the misspecified models, the average estimates (not shown) approach the corresponding theoretical projection parameters as T increases. Less clear is the impact of misspecification on the standard errors of the estimates. Table 9 reports, for $N = 25$, the RMSE of each estimator followed by two sets of asymptotic standard errors. The first set is evaluated at the true parameter values, ignoring model misspecification and then taking it into account. The second pair consists of the corresponding simulation averages of the estimated asymptotic standard errors. Interestingly, we find that the asymptotic adjustment for model misspecification increases the standard errors, but minimally in this context. For example, when $T = 360$, the estimated standard error for the GLS estimator of γ_1 increases from 0.34 to 0.37. Similar results (not shown) are obtained when $N = 48$.

Tests of the linear expected return relation for $N = 25$ and $N = 48$ are reported in Table 10. Since the CST's appear to have the appropriate size based on the earlier analysis, the rejection rates for these tests can be viewed as estimates of the power of a 5% test. We see that the multivariate CST's have substantial power against both alternatives when $T \geq 240$. The power of the coefficient-based test (which may be overstated slightly in small samples) is lower, particularly for $N = 48$.

5.3   Conditional heteroscedasticity

Thus far, returns have been assumed to be distributed iid normal in our simulations. Now we examine the sensitivity of estimation and test results to the joint normality assumption. As Kan and Zhou (2003) show, a multivariate t-distribution with 8 degrees of freedom fits the data well.[23] In addition, under the t assumption, the residual variance depends on the factors, and so contemporaneous conditional heteroscedasticity is introduced. We generate the data as before, except that the joint normality assumption for the factors and returns is replaced by the joint t assumption.

Table 11 reports estimation results for γ_1. The results are fairly robust to the assumed conditional heteroscedasticity when $T \geq 360$, but some small effects are observed. For example, compared to Table 1, the GLS estimator is slightly more biased and its RMSE slightly higher in Table 11, while the OLS RMSE declines a bit for $N = 25$. Also, MLE now has a negative bias,

---

[23]Here, we refer to the unconditional distribution of the data. Exploring the behavior of conditional tests when the variance of returns is allowed to change is beyond the scope of this paper, but an important topic that we hope to address in future work.

but it is less than 5% of the true value for reasonable sample sizes. The cross-sectional F tests of expected return linearity, which performed quite well under homoskedasticity, continue to display the proper size (close to 5%) under conditional heteroskedasticity (results not shown).

6   Application

The standard excess-return time-series formulation of the Fama and French (1993) three-factor model constrains the alphas to equal zero, implicitly assuming that the zero-beta rate equals the riskless rate and the factor risk premia equal the corresponding factor means. As noted in Shanken and Weinstein (1990), the latter restriction is implied whenever the factors are spread portfolio returns. It is well-known that the constrained model can be evaluated by the Gibbons, Ross and Shanken (1989) test or, if the normality assumption is a concern, by methods that allow for conditional heteroskedasticity, as in MacKinlay and Richardson (1991) and Shanken (1990). In this section, we relax the usual pricing restrictions and apply the earlier cross-sectional regression methods in an analysis of the Fama-French model. This can be likened, in some respects, to Fama and MacBeth's (1973) analysis of the original Sharpe-Lintner CAPM.

The empirical factor model is:

$$R_{it} - r_{ft} = \alpha_i + \beta_{i1}(f_{M,t} - r_{ft}) + \beta_{i2}f_{SMB,t} + \beta_{i3}f_{HML,t} + \epsilon_{it}, \tag{46}$$

where f_M is the market return factor, f_{SMB} is the small-big return spread, f_{HML} is the high-low return spread, and r_{ft} is the 30-day T-bill rate. The R_{it}'s are the test asset returns on the 25 stock portfolios formed on size and book-to-market. The following asset pricing relationship is of interest,

$$H_0 : \qquad E(R_{it} - r_{ft}) = \gamma_0 + \gamma_1\beta_{i1} + \gamma_2\beta_{i2} + \gamma_3\beta_{i3}, \tag{47}$$

where γ_1, γ_2, and γ_3 are the risk premium parameters, and γ_0 is the excess zero-beta rate.

Table 12 reports the results. The first column indicates which of the five estimation methods is used. The next four columns provide estimation results for each of the risk-return parameters. The point estimate is given with its standard error in parentheses, both in percent. The last column contains, for each second-pass estimation method, a chi-squared test of significance comparing the cross-sectional regression estimates of the risk-return parameters to the usual time-series means (and $\gamma_0 = 0$).

The first portion of Table 12 shows the sample mean of each factor over the period 1964-

2003. The market and HML risk premia are significantly positive and the SMB risk premium is marginally significant with a (2-sided) p-value of 7%. The rest of the table shows the cross-sectional estimates when the 25 size and book-to-market portfolios are used as the test assets and the $\gamma's$ are unconstrained. All of the estimates of the market risk premium, γ_1, are close to -1% per month or lower, while the zero-beta rates are about $1\frac{1}{2}\%$ or higher, both far from the constrained values. On the other hand, the size and book-to-market premia are close to the sample means. A chi-squared test given in the last column indicates that each set of second-pass estimates is jointly significantly different from the constrained values at the 0.01 level or lower.

The risk premia estimates are fairly similar across estimation methods, as might be expected if expected returns were indeed linear in the betas. However, the general specification tests reported in Table 13 easily reject the null at the 1% level and the test for equality of OLS and GLS projection coefficients is marginally significant. The numbers in parentheses in Table 12 are the usual asymptotic standard errors, while those in brackets take into account model misspecification. The differences are a little larger than we saw earlier with K=1, but not big enough to materially change any of the inferences. For example, when $N = 20$, the OLS standard error for γ_1 increases from 0.38 to 0.42, but the estimate is still significantly negative.

That the three-factor model can be rejected is not new. In particular, it is well-known, e.g., Fama and French (1996), that the small growth portfolio is a problem for the model. Its alpha is about -45 basis points (bps) per month, contributing to the strong rejection by the GRS test in Table 13. Although our less restrictive cross-sectional model is rejected as well, it is of interest to assess the extent to which the model is improved by allowing for a large zero-beta rate, as the data apparently prefer. The impact need not be large since the (multiple regression) market betas of the portfolios tend to be close to one. Thus, the impact of the low value of γ_1 can offset that of the high γ_0 estimate. The counterpart to alpha, in the CSR context, is the difference between the time-series average of the portfolio excess returns and the model fitted expected return. The average absolute value of this measure, across the 25 portfolios, is 8.9 bps for each of the second-pass estimators, as compared to the average absolute alpha of 10.5 bps, a 15% reduction. The small growth portfolio is still the biggest challenge, however, with estimated deviations ranging from -32 to -37 bps.

7 Conclusion

The Fama-MacBeth two-pass estimation method has been used time and again in testing asset pricing models and analyzing the cross-section of expected returns. Yet, there has been surprisingly little analysis of its finite-sample distribution and performance relative to other estimation procedures that have occasionally been used. These include maximum likelihood estimation and various forms of the generalized method of moments. We provide analytic results that simplify the computation of these alternative estimators and explore their properties, along with those of the two-pass procedure, through extensive simulations. We also show that one natural formulation of the GMM method turns out to be identical to ML estimation.

The ML estimator was originally proposed by Gibbons (1982) as a solution to the errors-in-variables problem inherent in two-pass estimation. Shanken (1992) provided some formal justification for this conclusion, though the early simulation results of Amsler and Schmidt (1985) raised doubts about the finite-sample behavior of the ML estimator. The recent insightful analysis of Chen and Kan (2005) reveals the analytic basis for the erratic behavior of the ML estimator. In particular, they show that its mean does not exist. Given this, we explore the properties of a truncated ML estimator, which involves switching to the GLS version of the cross-sectional regression estimator when the magnitude of the ML estimator of any risk premium gets too big, say more than twice that of the GLS estimator.

A striking result is that, for sample sizes of 30 years or more, typical in applications, the ML risk premia estimators are virtually unbiased in our simulations. This is true not just for the single-factor CAPM simulations, but for the Fama-French (1993) three-factor model as well, whether the assets are the 25 Fama-French size and book-to-market portfolios or 48 industry portfolios. The precision of the ML estimator, as measured by root mean-squared error, is typically close to, or a bit lower than that of GLS, and both are much more precise than OLS or WLS in CAPM simulations. However, OLS and WLS tend to be less biased than GLS. The differences in estimation performance across methods are more minor for the size and book-to-market factor premia.

When evaluated at the true parameter values, the average asymptotic standard errors for the second pass estimators are often greater than their RMSE's. However, the standard errors always decline when the parameter estimates are used, as would be the case in practice. The behavior of *estimated* standard errors has not, to our knowledge, been examined previously, and that of the *t*-ratio for the hypothesis that the (excess) zero-beta rate or risk premium is zero has received little

attention.[24] Here, ML does not perform very well, sometimes overstating precision by perhaps 10-15% and rejecting true null hypotheses at twice the nominal 5% level when, for example, 40 years of data are used. Though the estimators are less precise, inference with OLS/WLS is fairly reliable in all scenarios studied here.

We also simulate tests of model specification and find that Shanken's (1985) cross-sectional F-test for expected return linearity consistently exhibits size close to the nominal 5% level considered, whereas the Bartlett-corrected likelihood ratio test sometimes rejects a bit too much. With regard to misspecification, an innovative aspect of our paper is the analysis of asymptotic and small-sample properties of Fama-MacBeth estimators when exact linearity is violated. We find that standard errors are understated when misspecification is ignored, by as much as 10% in one case, but generally less in the examples examined.

Although our comparison of estimation methods has suggested some important regularities, different patterns might, of course, be observed with other factors or test portfolios. Therefore, further analysis of a range of scenarios would be worthwhile. Since no single estimation procedure dominates in all respects, it might be wise to explore robustness of results to several estimation approaches in applied work. As we have shown, comparison of OLS and GLS results can also serve as the basis for a formal test of model specification. In our cross-sectional estimation of the Fama-French model, a very high zero-beta rate and a significantly negative market risk premium are obtained regardless of the method used.

Our limited simulations based on the multivariate t distribution allow for residual return heteroskedasticity conditional on the factors. We observe some, but not too much sensitivity to this form of departure from normality. Exploring the impact of time-varying ex ante volatility would be one natural direction for further research. Given that the ML estimator tends to perform well overall, in terms of bias and precision, it would be interesting to see whether more accurate standard errors and inferences can be obtained by some sort of bootstrap procedure. The computational simplifications developed here should be valuable in this context. Finally, although it will pose some computational challenges in simulations, we also plan to explore the merits of incorporating a large number of test portfolios or individual securities in cross-sectional analysis. Although this is sometimes done in practice, little is known about the small-sample properties of the estimators in this context, with the exception of the recent work by Petersen (2005), who also considers the serial correlation that arises in many corporate finance applications.

---

[24] An exception is Chen, Kan and Zhang (1999) who provide some simulation results on the t-ratios when the test assets are ten size portfolios.

References

Amsler, C. E., Schmidt, P., 1985. A Monte Carlo investigation of the accuracy of multivariate CAPM tests. Journal of Financial Economics 14, 359–375.

Anderson, T.W., 1984. An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Balduzzi, P., Robotti, C., 2004. Mimicking portfolios, economic risk premia, and tests of multi-beta models, working paper, Federal Reserve Bank of Atlanta.

Bartlett, M. S., 1947. Multivariate analysis. Journal of the Royal Statistical Society (Suppl.) 9, 176–190.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–454.

Black, F., Jensen, M. C., Scholes, M., 1972. The capital asset pricing model: some empirical findings. In: M. C. Jensen (Ed.), Studies in the Theory of Capital Markets, Praeger, New York.

Breeden, D. T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 265–296.

Chen, N., Kan, R., Zhang, C., 1999. A critique of the use of t-ratios in model selection, working paper, University of Toronto.

Chen, R., Kan, R., 2004. Finite sample analysis of two-pass cross-sectional regressions, working paper, University of Toronto.

Cochrane, J. H., 2001. Asset Pricing. Princeton University Press, New Jersey.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1–33.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? Journal of Finance 57, 2185–2221.

Fama, E. F., MacBeth, J., 1973. Risk, returns and Equilibrium: Empirical tests. Journal of Political Economy 71, 607–636.

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

Fama E.F., French, K.R., 1998. Taxes, financing decisions, and firm value. Journal of Finance 53, 819–843.

Fama E.F., French, K.R., 2002. Testing tradeoff and pecking order predictions about dividends and debt. Review of Financial Studies 15, 1–33.

Gibbons, M. R., 1982. Multivariate tests of financial models: A new approach. Journal of Financial Economics 10, 3–27.

Gibbons, M.R., Ross, S.A., Shanken J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Grinstein, Y., Michaely, R., 2002. Institutional holdings and payout policy. Journal of Finance, forthcoming.

Harvey, C. R., Kirby, C. M., 1995. Analytic tests of factor pricing models, working paper, Duke University.

Harvey, C.R., Zhou, G., 1993. International asset pricing with alternative distributional specifications. Journal of Empirical Finance 1, 107–131

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

Jagannathan, R., Wang, Z., 1998. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. Journal of Finance 53, 1285–1309.

Jagannathan, R., Wang, Z., 2002. Empirical evaluation of asset pricing models: a comparison of the SDF and beta methods. Journal of Finance 57, 2337–2367.

Jobson, J. D., Korkie, B. M., 1982. Potential performance and tests of portfolio efficiency. Journal of Financial Economics 10, 433–466.

Kan, R., Zhang, C., 1999. Two-pass tests of asset pricing models with useless factors. Journal of Finance 54, 203–235.

Kan, R., Zhou, G., 1999. A critique of the stochastic discount factor methodology. Journal of Finance 54, 1221–1248.

Kan, R., Zhou, G., 2003. Modeling non-normality using multivariate t: implications for asset pricing, working paper, Washington University in St.Louis.

Kandel, S., 1984. The likelihood ratio test statistic of mean-variance efficiency without a riskless asset. Journal of Financial Economics 13, 575–592.

Kandel, S., Stambaugh, R., 1995. Portfolio inefficiency and the cross-section of expected returns. Journal of Finance 50, 157–184.

Kimmel, R., 2003. Risk premia in linear factor models: theoretical and econometric issues. working paper, Department of Economics, Princeton University.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Litzenberger, R.H., Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. Journal of Financial Economics 7, 163–195.

Lo, A., MacKinlay, C., 1990. Data snooping biases in tests of financial asset pricing models. Review of Financial Studies 3, 431–468.

MacKinlay, A.C., 1995. Multifactor models do not explain the CAPM. Journal of Financial Economics 38, 3–28.

MacKinlay, A.C., Richardson, M.P., 1991. Using generalized method of moments to test mean-variance efficiency. Journal of Finance 46, 511–527.

Merton, R. C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.

Muirhead, R. J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Newey, W. K., 1984. A method of moments interpretation of sequential estimators. Economic Letters 14, 201–206.

Newey, W. K., 1985. Generalized method of moments specification testing. Journal of Econometrics 29, 229–256.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Ogaki, M., 1993. Generalized method of moments: econometric applications. In: Maddala, et al (Ed.), Handbook of Statistics, Vol. 11. North-Holland, Amsterdam, pp. 455–488.

Petersen, M., 2005. Estimating standard errors in finance panel data sets: comparing approaches, NBER working paper 11280, Cambridge, MA.

Roll, R., 1985. A note on the geometry of Shankens CSR T^2 test for mean/variance efficiency. Journal of Financial Economics 14, 349–357.

Roll, R., Ross, S.A., 1994. On the cross-sectional relation between expected returns and betas. Journal of Finance 49, 101–121.

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

Rubinstein, M., 1976. The valuation of uncertain income streams and the pricing of options. The Bell Journal of Economics 7, 407–425.

Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. Journal of Financial Economics 14, 327–348.

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., Weinstein, M., 1990. Macroeconomic variables and asset pricing,: further results, working paper, University of Rochester.

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

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

Siskind, V., 1972. Second moments of inverse Wishart-matrix elements. Biometrika 59, 690–691.

Stambaugh, R.F., 1982. On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. Journal of Financial Economics 10, 237–268.

Velu, R., Zhou, G., 1999. Testing multi-beta asset pricing models. Journal of Empirical Finance 6, 219–241.

Zhou, G., 1994. Analytical GMM tests: Asset pricing with time-varying risk premiums. Review of Financial Studies 7, 687–709.

Zhou, G., 1995. Small sample rank tests with applications to asset pricing. Journal of Empirical Finance 2, 71–93.

A Appendix

A.1 *Proof of Proposition 1*

Our proof follows Shanken (1992) closely, but accommodates model mispricing. Inclusion of a vector of constant characteristics, as in section 3.3 of that paper, poses no additional complications. We derive the asymptotic distribution of $\hat{\Gamma} - \bar{\Gamma}$. The implied asymptotic distribution for $\hat{\Gamma} - \Gamma$ can then be obtained as in (iii) of Shanken's Theorem 1. For notational brevity, in what follows we omit the dependence of Γ, etc., on W. Averaging the factor regression model, equation (1), and using $\alpha = E(R_t) - \beta E(F_t)$, we have

$$\bar{R} = X\bar{\Gamma} + \eta + \bar{\epsilon}, \tag{A1}$$

where

$$X = [1_N, \beta] \quad \text{and} \quad \bar{\Gamma} = [\gamma_0, \bar{\gamma}_a']', \tag{A2}$$

with $\bar{\gamma}_a = \gamma_a + \bar{F}_t - E(F_t)$. Letting $U = \hat{\beta} - \beta$, we can rewrite (A1) as

$$\bar{R} = \hat{X}\bar{\Gamma} + (\bar{\epsilon} - U\bar{\gamma}_a) + \eta. \tag{A3}$$

Premultiplying by $\hat{A} = (\hat{X}'\hat{W}\hat{X})^{-1}\hat{X}'\hat{W}$ and noting that $\hat{A}\hat{X} = I_{K+1}$, we get

$$\hat{\Gamma} - \bar{\Gamma} = \hat{A}(\bar{\epsilon} - U\bar{\gamma}_a) + \hat{A}\eta. \tag{A4}$$

As $A\eta = 0$ by construction and $\hat{A}$ converges to A, so $\hat{A}\eta$ converges in probability to zero as $T \to \infty$. Consistency then follows easily from Shanken (1992). Moreover, the asymptotic covariance matrix of $\hat{\Gamma} - \bar{\Gamma}$ is determined by the covariance matrix of the righthand side of (A4). There are three components. The first component, $\Upsilon_w - \Sigma_f^*$, is the asymptotic covariance matrix of $\hat{A}(\bar{\epsilon} - U\bar{\gamma}_a)$, obtained earlier by Shanken. The second component, $\Upsilon_{w1} + \Upsilon_{w1}'$, is the asymptotic covariance matrix between $\hat{A}(\bar{\epsilon} - U\bar{\gamma}_a)$ and $\hat{A}\eta$ and the third component, Υ_{w2}, is the asymptotic covariance matrix of $\hat{A}\eta$.

When $\eta = 0$, as in the earlier analysis by Shanken, error in estimating the covariance matrix does not affect the asymptotic distributions of the second-pass WLS/GLS estimators. To see this, note that the expression in parentheses in (A4) has mean zero and converges in distribution after multiplication by $\sqrt{T}$. In this context, the effect of $\hat{A}$ on the asymptotic distribution amounts to multiplication by a constant matrix. The role of the covariance estimator is more complicated when the model is misspecified. Using $A\eta = 0$, we have $X'W\eta = 0$, and hence

$$\hat{X}'\hat{W}\eta = [0, U]'W\eta + \hat{X}'(\hat{W} - W)\eta. \tag{A5}$$

Note that $\hat{A}\eta$ equals the inverse of $\hat{X}'\hat{W}\hat{X}$ times the expression in (A5). Now, this inverse acts as a constant matrix in the limit, as does $\hat{X}$ in premultiplying the estimation error in $\hat{W}$.

Given the usual regression assumptions on the factor model, $\bar{\epsilon}$ and U are orthogonal. Imposing the conditional joint normality assumption on the disturbances further ensures, by Lemma 1 of Shanken (1992), that $\bar{\epsilon}$, U and $\hat{W}$ are mutually independent. In the OLS case, $\hat{W} = W = I$ and hence the normality assumption is not required. Although normality is sufficient to guarantee orthogonality between $\hat{W}$ and the other components, this condition may be (approximately) true more generally. While it is beyond the scope of this paper, it should be possible to derive the asymptotic distribution under more general assumptions that do not require such orthogonality.

It follows that

$$\Upsilon_{w1} = -(X'WX)^{-1} \begin{pmatrix} 0 \\ \lim E(TU'W\eta\gamma_a'U') \end{pmatrix} WX(X'WX)^{-1}, \tag{A6}$$

where the limit is taken as $T \to \infty$. Noting that

$$\hat{X}'(\hat{W} - W)\eta = \text{vec}[\hat{X}'(\hat{W} - W)\eta] = (\eta' \otimes \hat{X}')\text{vec}(\hat{W} - W), \tag{A7}$$

we have

$$\Upsilon_{w2} = (X'WX)^{-1} \left[\begin{pmatrix} 0 & 0 \\ 0 & \lim E(TU'W\eta\eta'WU) \end{pmatrix} + (\eta' \otimes X')V_w(\eta \otimes X) \right] (X'WX)^{-1}, \tag{A8}$$

where V_w is the asymptotic covariance matrix of $\text{vec}(\hat{W} - W)$, which is zero in the OLS case. In the GLS case, V_w is given by (17) based on Siskind (1972). Interestingly, the expression is similar to that for $\hat{\Sigma}$ as given by Muirhead (1982, p. 20). In the WLS case, since $\hat{W}$ and W are diagonal, the nonzero elements of V_w must be of the form given on the left-hand side of Eq. (18). The right-hand side then follows from the equation

$$\left(\frac{1}{\hat{\sigma}_{ii}} - \frac{1}{\sigma_{ii}} \right) \left(\frac{1}{\hat{\sigma}_{jj}} - \frac{1}{\sigma_{jj}} \right) = \frac{(\hat{\sigma}_{ii} - \sigma_{ii})(\hat{\sigma}_{jj} - \sigma_{jj})}{\hat{\sigma}_{ii}\sigma_{ii}\hat{\sigma}_{jj}\sigma_{jj}} \tag{A9}$$

and the aforementioned formula for the covariance matrix of $\hat{\Sigma}$.

Finally, there are two useful identities,[25]

$$E[Z'CZ'] = \Sigma_f^{-1}C'\Sigma, \tag{A10}$$

$$E[Z'BZ] = \text{tr}(B\Sigma)\Sigma_f^{-1}, \tag{A11}$$

---

[25]We are grateful to an anonymous referee for pointing them out, which has substantially simplified our earlier expressions.

where Z, $N \times K$, is a random matrix with $E[Z] = 0$ and $\text{Cov}[\text{vec}(Z)] = \Sigma_f^{-1} \otimes \Sigma$, and C and B are compatible matrices. The simple expressions for the limiting values of Υ_{w1} and Υ_{w2}, as given in (15) and (16), follow directly from these two identities. In the GLS case, the orthogonality condition, $X'W\eta = 0$, implies that $\Upsilon_{w1} = 0$.

To see why the identities hold, consider the first one. Let Z_i be the i-th column of Z and Z^n be its n-th row. Then, $E(Z_i Z_j') = g_{ij}\Sigma$, where g_{ij} is the (i, j) element of Σ_f^{-1}. Hence, the (k, n) element of $Z'CZ'$ is

$$E(Z_k'CZ^{n\prime}) = E \sum_{j=1}^{K} \sum_{m=1}^{N} z_{mk} c_{mj} z_{nj} = \sum_{j=1}^{K} \sum_{m=1}^{N} g_{kj} c_{mj} \sigma_{mn}, \tag{A12}$$

where σ_{mn} is the (m, n) element of Σ. One can then verify directly that the right-hand side of (A12) is the (k, n) element of the right-hand side of (A10). The other identity can be proved similarly. Q.E.D

A.2   *Proof of Proposition 2*

The key is to note that equation (A4) holds for both the OLS and GLS estimators. Taking the difference of the two equations yields the desired result under the null that $\eta = 0$. Q.E.D

A.3   *Proof of Proposition 3*

Denote by $\mathcal{L}_c$ the log-likelihood function under the constraints (2),

$$\begin{aligned}
\mathcal{L}_c(\lambda, \beta, \Sigma) = {}& -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| \\
& -\frac{1}{2}\sum_{t=1}^{T}[R_t - \lambda_0 1_N - \beta_1(f_{1t} + \lambda_1) - \cdots - \beta_K(f_{Kt} + \lambda_K)]'\Sigma^{-1} \\
& \times [R_t - \lambda_0 1_N - \beta_1(f_{1t} + \lambda_1) - \cdots - \beta_K(f_{Kt} + \lambda_K)].
\end{aligned} \tag{A13}$$

The usual unconstrained log-likelihood function is

$$\mathcal{L}_u = -\frac{NT}{2}\log(2\pi) - \frac{T}{2}\log|\widehat{\Sigma}| - \frac{NT}{2}. \tag{A14}$$

Following Shanken (1986), we can write the likelihood ratio for testing (21) as

$$\text{LR} = T\log[1 + Q(\lambda)], \qquad Q(\lambda) = \frac{\tilde{a}'\widehat{\Sigma}^{-1}\tilde{a}}{1 + (\bar{F} + \lambda_a)'\widehat{\Delta}^{-1}(\bar{F} + \lambda_a)}, \tag{A15}$$

where $\lambda_a = (\lambda_1, \ldots, \lambda_K)'$, $\tilde{a} = \hat{\alpha} - \lambda_0 1_N - \lambda_1 \hat{\beta}_1 - \cdots - \lambda_K \hat{\beta}_K$, and $\bar{F}$ and Δ are the sample mean and covariance matrix of the factors. Moreover, maximizing the likelihood function under the null is equivalent to minimizing LR, and hence the maximum likelihood estimator of $\lambda_0, \lambda_1, \ldots, \lambda_K$ is the solution to minimizing $Q(\lambda)$.

We begin by minimizing Q with respect to λ_0, since this parameter only enters the numerator of Q. Conditional on λ_a, the solution to this minimization amounts to a GLS regression of $\tilde{a}$ on 1_N:

$$\lambda_0 = (1_N' \widehat{\Sigma}^{-1} 1_N)^{-1} 1_N' \widehat{\Sigma}^{-1} \tilde{a}. \tag{A16}$$

Therefore, we only need to minimize

$$Q^*(\lambda_a) = \frac{(\hat{\alpha}^* - \hat{\beta}^* \lambda_a)' \widehat{\Sigma}^{-1} (\hat{\alpha}^* - \hat{\beta}^* \lambda_a)}{1 + (\bar{F} + \lambda_a)' \widehat{\Delta}^{-1} (\bar{F} + \lambda_a)}, \tag{A17}$$

where $\hat{\alpha}^* = \hat{\alpha} - (1_N' \widehat{\Sigma}^{-1} 1_N)^{-1} (1_N' \widehat{\Sigma}^{-1} \hat{\alpha}) 1_N$ and $\hat{\beta}^* = (\hat{\beta}_1^*, \ldots, \hat{\beta}_K^*)$ with $\hat{\beta}_j^* = \hat{\beta}_j - (1_N' \widehat{\Sigma}^{-1} 1_N)^{-1} (1_N' \widehat{\Sigma}^{-1} \hat{\beta}_j) 1_N$ for $j = 1, \ldots, K$. To solve (A17) analytically, we express $Q^*(\lambda_a)$ as

$$Q^*(\lambda_a) = \frac{w' A w}{w' B w}, \qquad w = \begin{pmatrix} 1 \\ \lambda_a \end{pmatrix}, \tag{A18}$$

where A and B are defined in (25). The minimum of Q^* is given by the smallest eigenvalue of $|A - \zeta B| = 0$ (see, e.g., Anderson, 1984, p. 590). The rest follows from a transformation of this equation. Q.E.D.

A.4 *Proof of Proposition 4*

With $K = 1$, we can simply write $Q(\lambda) = h(\gamma_0, \gamma_1)$. By Proposition 3, the ML estimator of γ_0 is given by

$$\tilde{\gamma}_0^{\text{ML}} = (1_N' \widehat{\Sigma}^{-1} \bar{R} - \tilde{\gamma}_1^{\text{ML}} 1_N' \widehat{\Sigma}^{-1} \hat{\beta}_1)/(1_N' \widehat{\Sigma}^{-1} 1_N). \tag{A19}$$

We want to show first that the ML estimator of γ_1, $\tilde{\gamma}_1^{\text{ML}}$, is one of the two roots of the following quadratic equation

$$H(\gamma_1) = \hat{\gamma}_1^{\text{GLS}} \gamma_1^2 - (\hat{\sigma}_f^2 - d_1/d) \gamma_1 - \hat{\sigma}_f^2 \hat{\gamma}_1^{\text{GLS}} = 0, \tag{A20}$$

where

$$d_1 = \det \begin{pmatrix} 1_N' \widehat{\Sigma}^{-1} 1_N & 1_N' \widehat{\Sigma}^{-1} \bar{R} \\ \bar{R}' \widehat{\Sigma}^{-1} 1_N & \bar{R}' \widehat{\Sigma}^{-1} \bar{R} \end{pmatrix}, \qquad d = \det \begin{pmatrix} 1_N' \widehat{\Sigma}^{-1} 1_N & 1_N' \widehat{\Sigma}^{-1} \hat{\beta}_1 \\ \hat{\beta}_1' \widehat{\Sigma}^{-1} 1_N & \hat{\beta}_1' \widehat{\Sigma}^{-1} \hat{\beta}_1 \end{pmatrix}. \tag{A21}$$

Plugging $\tilde{\gamma}_0^{\text{ML}}$ into $h(\gamma_0, \gamma_1)$ we need only minimize

$$
\begin{aligned}
h(\gamma_1) &= \left[(\bar{R} - \frac{a}{b}1_N) + (\frac{c}{b}1_N - \hat{\beta}_1)\gamma_1\right]' \widehat{\Sigma}^{-1} \left[(\bar{R} - \frac{a}{b}1_N) + (\frac{c}{b}1_N - \hat{\beta}_1)\gamma_1\right] /(1 + \gamma_1^2/\hat{\sigma}_f^2) \\
&= (d\gamma_1^2 - 2\hat{\gamma}_1^{\text{GLS}}\gamma_1 + d_1)/[b(1 + \gamma_1^2/\hat{\sigma}_f^2)],
\end{aligned}
\tag{A22}
$$

where $a = 1_N'\widehat{\Sigma}^{-1}\bar{R}$, $b = 1_N'\widehat{\Sigma}^{-1}1_N$ and $c = 1_N'\widehat{\Sigma}^{-1}\hat{\beta}_1$. Then, taking the first order derivative of $h(\gamma_1)$, it can be verified (though tedious) that $h'(\gamma_1)$ has the same sign as $H(\gamma_1)$. It follows that if $\hat{\gamma}_1^{\text{GLS}} > 0$, then h is initially increasing, then decreasing, and then increasing again, with the same finite asymptotic values at plus or minus infinity. Hence, h achieves a global minimum at the larger root of H in this case.

Now, to establish the inequality, let $h = f/g$. Then, since $h'(\gamma_1) = 0$ at the ML estimator, we have $f/f' = g/g'$. This implies that f' and g' must have the same sign at the ML estimator. Because $d > 0$, f' must also be an increasing function of γ_1. Now if $\hat{\gamma}_1^{\text{GLS}} > 0$, it follows from (A20) that $\tilde{\gamma}_1^{\text{ML}} \neq 0$. If $\tilde{\gamma}_1^{\text{ML}} < 0$, we have $f(-\tilde{\gamma}_1^{\text{ML}}) = f(\tilde{\gamma}_1^{\text{ML}}) + 4\hat{\gamma}_1^{\text{GLS}}\tilde{\gamma}_1^{\text{ML}} < f(\tilde{\gamma}_1^{\text{ML}})$, a contradiction since $\tilde{\gamma}_1^{\text{ML}}$ minimizes (A22). Therefore, it must be the case that $\tilde{\gamma}_1^{\text{ML}} > 0$ whenever $\hat{\gamma}_1^{\text{GLS}} > 0$. Since $g' = 2b\gamma_1/\hat{\sigma}_f^2$, we have $g' > 0$ and hence $f' > 0$ at the ML estimator if $\hat{\gamma}_1^{\text{GLS}} > 0$. The proposition then follows by noting that $f'(\hat{\gamma}_1^{\text{GLS}}) = 0$. Q.E.D

A.5 *Proof of Proposition 5*

First, we rewrite the model in a more tractable form. Let

$$
\Theta = \begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix} = \begin{bmatrix} \lambda_1\beta_1' + & \cdots & +\lambda_K\beta_K' \\ (\beta_1 & \ldots & \beta_K)' \end{bmatrix},
\tag{A23}
$$

a $(K+1) \times N$ matrix. The model residuals can then be written as a $T \times N$ matrix,

$$
U = R - \lambda_0 1_T 1_N' - Z\Theta, \qquad Z = [1_T, F_1, \ldots, F_K],
\tag{A24}
$$

where Z is a $T \times (K+1)$ matrix formed by the unit vector and T observations on the factors.

The key is to note that, conditional on γ_0, the null hypothesis imposes a rank K restriction on Θ, i.e.,

$$
\Theta = AB, \qquad A = \begin{bmatrix} \lambda_a' \\ I_K \end{bmatrix}, \quad B = \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_K' \end{bmatrix},
\tag{A25}
$$

where A is $(K+1) \times K$, B is $K \times N$, and $\lambda_a = (\lambda_1, \ldots, \lambda_K)'$. Hence, the technique of Zhou (1994) can used to solve for A and B. Q.E.D.

A.6 *Proof of Proposition 6*

To simplify the proof, we show first that adding a constant vector to the factor observations will reduce the GMM1 estimator of λ_a by the same amount. This is clearly true for the maximum likelihood method as the constrained likelihood function will be unchanged by the two offsetting shifts. But this is not obvious for GMM1 because both the weighting matrix and the instruments (factors) change with the shifts, and so the impact on the objective function requires further analysis.

To prove the claim, consider the following re-ordered GMM moment conditions

$$g_T^*(\theta) = \frac{1}{T} \sum_{t=1}^{T} Z_t \otimes \mathcal{E}_t = \frac{1}{T} \sum_{t=1}^{T} \left(Z_t \otimes [(R_t - \lambda_0 1_N) - \beta(F_t + \lambda_a)] \right), \tag{A26}$$

and the associated weighting matrix $W = W_T^* = W_2 \otimes W_1$ with $W_2 = (Z'Z/T)^{-1}$. Using the identity

$$z_t \otimes \beta(F_t + \lambda_a) = [z_t(F_t + \lambda_a)' \otimes I_N]b, \tag{A27}$$

where $b = \text{vec}(\beta)$, we can write the moment conditions as a quadratic function of b conditional on λ_0 and λ_a,

$$g_T^*(\theta) = \frac{1}{T} \sum_{t=1}^{T} Z_t \otimes \mathcal{E}_t(\theta) = y - xb, \tag{A28}$$

where

$$y = \begin{pmatrix} \bar{R} - \lambda_0 1_N \\ \frac{1}{T} \sum_{t=1}^{T} F_t \otimes (R_t - \lambda_0 1_N) \end{pmatrix}, \qquad x = \begin{pmatrix} (\bar{F} + \lambda_a)' \\ \widehat{\Delta} + \bar{F}\bar{F}' + \bar{F}\lambda_a' \end{pmatrix} \otimes I_N. \tag{A29}$$

To carry out the matrix multiplications in what follows, it is necessary to write y in a more tractable form. Because

$$F_t \otimes (R_t - \lambda_0 1_N) = F_t \otimes (R_t - \bar{R}) + F_t \otimes (\bar{R} - \lambda_0 1_N) \tag{A30}$$

$$= (F_t - \bar{F}) \otimes (R_t - \bar{R}) + \bar{F} \otimes (R_t - \bar{R}) + F_t \otimes (\bar{R} - \lambda_0 1_N), \tag{A31}$$

we have

$$\frac{1}{T} \sum_{t=1}^{T} F_t \otimes (R_t - \lambda_0 1_N) = \frac{1}{T} \sum_{t=1}^{T} (F_t - \bar{F}) \otimes (R_t - \bar{R}) + \bar{F} \otimes (\bar{R} - \lambda_0 1_N). \tag{A32}$$

Hence, using the identity $(F_t - \bar{F}) \otimes (R_t - \bar{R}) = \text{vec}[(R_t - \bar{R})(F_t - \bar{F})']$ and the definition for $\hat{\beta}$, we obtain

$$\frac{1}{T} \sum_{t=1}^{T} (F_t - \bar{F}) \otimes (R_t - \bar{R}) = \frac{1}{T} \sum_{t=1}^{T} \text{vec}[(R_t - \bar{R})(F_t - \bar{F})'] = \text{vec}(\hat{\beta}\widehat{\Delta}), \tag{A33}$$

which implies that

$$y = \begin{pmatrix} \bar{R} - \lambda_0 1_N \\ (\widehat{\Delta} \otimes I_N)\hat{b} + \bar{F} \otimes (\bar{R} - \lambda_0 1_N) \end{pmatrix}, \quad N(K+1) \times 1. \tag{A34}$$

Now, it is easy to verify that

$$W_2 = \left(\frac{1}{T} Z'Z \right)^{-1} = \begin{pmatrix} 1 + \bar{F}'\widehat{\Delta}^{-1}\bar{F} & -\bar{F}'\widehat{\Delta}^{-1} \\ -\widehat{\Delta}^{-1}\bar{F} & \widehat{\Delta}^{-1} \end{pmatrix}, \tag{A35}$$

and hence

$$x'W = x'(W_2 \otimes W_1) = \begin{pmatrix} \lambda_a & I_K \end{pmatrix} \otimes W_1, \tag{A36}$$

and

$$x'Wx = [\widehat{\Delta} + (\bar{F} + \lambda_a)(\bar{F} + \lambda_a)'] \otimes W_1. \tag{A37}$$

As a function of the observed data, the GMM objective function can be decomposed into three terms,

$$Q_{\text{GMM}}(\theta) = (y - xb)'W(y - xb) = y'Wy - 2y'Wxb + b'x'Wxb. \tag{A38}$$

Equation (A37) implies that the last term will be invariant to the offsetting shifts. By (A34), we can write y as

$$y = y_1 + y_2 = \begin{pmatrix} 0_{N \times 1} \\ (\widehat{\Delta} \otimes I_N)\hat{b} \end{pmatrix} + \begin{pmatrix} 1 \\ \bar{F} \end{pmatrix} \otimes (\bar{R} - \lambda_0 1_N). \tag{A39}$$

Then, from (A37) and (A39), we have

$$y'Wx = \hat{b}'(\widehat{\Delta} \otimes W_1) + (\bar{F} + \lambda_a)' \otimes \left[(\bar{R} - \lambda_0 1_N)'W_1 \right], \tag{A40}$$

i.e., the second term of (A38) is also invariant to the shifts. Similarly, one can verify that $y'Wy = y_1'Wy_1 + 2y_1'Wy_2 + y_2'Wy_2$ is invariant too, and so the claim follows.

Given the claim, we can assume $\bar{F} = 0$ in the remainder of the proof. Maximizing the objective function over b, we obtain, from standard regression theory, that

$$Q_{\text{GMM}}(\theta) = Q(\lambda_0, \lambda_a) = (y - xb^*)'W(y - xb^*) = y'[W - Wx(x'Wx)^{-1}x'W]y, \tag{A41}$$

where $b^* = (x'Wx)^{-1}x'Wy$. To finish the proof, we need only show that $Q(\lambda_0, \lambda_a)$ is the same as the objective function of the maximum likelihood method, equation (22), when $W_1 = \hat{\Sigma}^{-1}$. Since $\bar{F} = 0$, the inverse of (A37) is

$$W - Wx(x'Wx)^{-1}x'W = \frac{[1, -\lambda_a'\widehat{\Delta}^{-1}]'[1, -\lambda_a'\widehat{\Delta}^{-1}] \otimes W_1}{1 + \lambda_a'\widehat{\Delta}^{-1}\lambda_a}. \tag{A42}$$

As $y = \text{vec}[\bar{R} - \lambda_0 1_N, \hat{\beta}\hat{\Delta}]$ from (A33), we have, using identity $\text{vec}(ABC) = (C' \otimes A)\text{vec}B$,

$$
\left([1, -\lambda_a'\hat{\Delta}^{-1}] \otimes W_1^{1/2}\right)y = \text{vec}\left(W_1^{1/2}[\bar{R} - \lambda_0 1_N, \hat{\beta}\hat{\Delta}][1, -\lambda_a'\hat{\Delta}^{-1}]'\right) \tag{A43}
$$

$$
= W_1^{1/2}(\bar{R} - \lambda_0 1_N - \hat{\beta}\lambda_a). \tag{A44}
$$

Therefore,

$$
Q(\lambda_0, \lambda_a) = y'[W - Wx(x'Wx)^{-1}x'W]y = \frac{(\bar{R} - \lambda_0 1_N - \hat{\beta}\lambda_a)'W_1(\bar{R} - \lambda_0 1_N - \hat{\beta}\lambda_a)}{1 + \lambda_a'\hat{\Delta}^{-1}\lambda_a}. \tag{A45}
$$

This proves the Proposition. It also shows that, for an arbitrary W_1, the GMM1 estimator can be solved analytically similar to the approach for ML in appendix A.3, as long as $W_2 = \left(\frac{1}{T}Z'Z\right)^{-1}$.

A.7 GMM2: the multifactor case

When there are multiple factors, there are two cases of interest. The first case is simpler and assumes that the asset pricing restrictions are expressed in terms of *univariate* betas, that is,

$$
E[R_t] = \gamma_0 1_N + \gamma_1 \frac{\text{cov}(R_t, f_1)}{\sigma_1^2} + \cdots + \gamma_K \frac{\text{cov}(R_t, f_K)}{\sigma_K^2}, \tag{A46}
$$

where σ_j^2 is the variance of the j-th factor and $\beta_j = \text{cov}(R_t, f_j)/\sigma_j^2$ is the vector of betas from the univariate regressions of the components of R_t on f_j. Then, as in (43), the sample moment conditions are

$$
E[h_t(\varphi)] = E\begin{bmatrix} R_t - \mu_r \\ F_t - \mu_f \\ (f_{1t} - \mu_1)^2 - \sigma_1^2 \\ \vdots \\ (f_{Kt} - \mu_K)^2 - \sigma_K^2 \\ R_t - \gamma_0 1_N - \sum_{j=1}^{K} \gamma_j \frac{(R_t - \mu_r)(f_{jt} - \mu_j)}{\sigma_j^2} \end{bmatrix} = 0. \tag{A47}
$$

The sequential GMM estimator can be solved analytically in the same manner as before, and the associated GMM theory provides asymptotic standard errors for the risk premium estimators under general conditions. Jagannathan and Wang (1998) consider the case of two-pass estimation with univariate betas and derive asymptotic standard errors under conditional heteroscedasticity.

Typical multifactor applications of the Fama-MacBeth methodology employ *multiple* regression betas from time-series regressions of returns on the K factors. Sequential GMM estimation in this

case is a bit more complex. Recalling that $\gamma_a = (\gamma_1, \ldots, \gamma_K)'$, the asset pricing relation (2) can be written as

$$E[R_t] = \gamma_0 1_N + \Sigma_{rf} \Sigma_f^{-1} \gamma_a, \tag{A48}$$

where the multivariate-regression β matrix equals Σ_{rf}, the covariance matrix between the returns and the factors, times the inverse of Σ_f, the covariance matrix of the factors. To obtain the sequential GMM estimator of Γ from (A48), we modify the univariate procedure in two ways. First, we add $K(K+1)/2 - K$ moment conditions,

$$E[(f_{it} - \mu_i)(f_{jt} - \mu_j) - \text{cov}(f_{it}, f_{jt})] = 0 \tag{A49}$$

to estimate the off-diagonal elements of Σ_f. Then, the last N moment conditions of (A47) are replaced by a sample analogue of (A48),

$$E\left[R_t - \gamma_0 1_N - \Sigma_{rft} \Sigma_f^{-1} \gamma_a\right] = 0, \tag{A50}$$

where Σ_{rft} is the $N \times K$ matrix whose j-th column is given by $(R_t - \mu_r)(f_{jt} - \mu_j)$ for $j = 1, \ldots, K$. Given this set of moment conditions, it is readily seen that the sequential GMM estimators of all the parameters except the gammas are given by their sample analogues, while the gammas are estimated from (A50), which can be explicitly solved, yielding a formula similar to (44).

Table 1: Market Risk Premium Estimates in a One-factor Model

The table reports the average estimate, its percentage error, and root-mean-square error (all in percent) over 10,000 simulated data sets. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1}f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the asset pricing restrictions are

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length and N is the number of assets. The true value for the factor risk premium is $\gamma_1 = 0.6667\%$ and the zero-beta intercept is $\gamma_0 = 0.0833\%$. The estimation methods are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (truncated maximum likelihood) and GMM2 (generalized method of moments).

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			N=25			
OLS	0.6189	0.6562	0.6404	0.6554	0.6574	0.6600
	-7%	-2%	-4%	-2%	-1%	-1%
	(1.1830)	(0.8629)	(0.6295)	(0.5041)	(0.4490)	(0.3114)
WLS	0.5920	0.6415	0.6340	0.6511	0.6530	0.6585
	-11%	-4%	-5%	-2%	-2%	-1%
	(1.1445)	(0.8427)	(0.6190)	(0.4953)	(0.4423)	(0.3070)
GLS	0.4508	0.5421	0.5905	0.6153	0.6282	0.6448
	-32%	-19%	-11%	-8%	-6%	-3%
	(0.8332)	(0.5867)	(0.4300)	(0.3492)	(0.3056)	(0.2153)
GMM2	0.4506	0.5424	0.5907	0.6152	0.6280	0.6449
	-32%	-19%	-11%	-8%	-6%	-3%
	(0.8393)	(0.5906)	(0.4311)	(0.3498)	(0.3061)	(0.2155)
ML	0.6148	0.6743	0.6610	0.6639	0.6652	0.6637
	-8%	1%	-1%	-0%	-0%	-0%
	(1.1241)	(0.7206)	(0.4753)	(0.3730)	(0.3212)	(0.2206)
			N=48			
OLS		0.5818	0.6139	0.6372	0.6367	0.6553
		-13%	-8%	-4%	-4%	-2%
		(0.6173)	(0.4595)	(0.3865)	(0.3369)	(0.2409)
WLS		0.5954	0.6218	0.6423	0.6419	0.6586
		-11%	-7%	-4%	-4%	-1%
		(0.6058)	(0.4516)	(0.3773)	(0.3288)	(0.2348)
GLS		0.5087	0.5705	0.6042	0.6148	0.6433
		-24%	-14%	-9%	-8%	-4%
		(0.5492)	(0.3862)	(0.3185)	(0.2784)	(0.1961)
GMM2		0.5088	0.5703	0.6043	0.6148	0.6434
		-24%	-14%	-9%	-8%	-3%
		(0.5528)	(0.3878)	(0.3193)	(0.2790)	(0.1964)
ML		0.6793	0.6658	0.6716	0.6657	0.6697
		2%	-0%	1%	-0%	0%
		(0.7130)	(0.4374)	(0.3480)	(0.2964)	(0.2027)

Table 2: Pricing Intercept Estimates in a One-factor Model

The table reports the average estimate, its percentage error, and root-mean-square error (all in percent) over 10,000 simulated data sets. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the asset pricing restrictions are

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length and N is the number of assets. The true value for the factor risk premium is $\gamma_1 = 0.6667\%$ and the zero-beta intercept is $\gamma_0 = 0.0833\%$. The estimation methods are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (truncated maximum likelihood) and GMM2 (generalized method of moments).

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			N=25			
OLS	0.1339	0.0927	0.1075	0.0903	0.0939	0.0882
	61%	11%	29%	8%	13%	6%
	(1.0437)	(0.7575)	(0.5489)	(0.4457)	(0.3956)	(0.2739)
WLS	0.1577	0.1053	0.1136	0.0945	0.0979	0.0896
	89%	26%	36%	13%	18%	8%
	(1.0241)	(0.7465)	(0.5428)	(0.4393)	(0.3916)	(0.2715)
GLS	0.2973	0.2037	0.1572	0.1304	0.1220	0.1037
	257%	144%	89%	56%	46%	24%
	(0.7586)	(0.4878)	(0.3351)	(0.2698)	(0.2334)	(0.1618)
GMM2	0.2971	0.2034	0.1569	0.1305	0.1220	0.1037
	256%	144%	88%	57%	46%	24%
	(0.7624)	(0.4907)	(0.3363)	(0.2704)	(0.2338)	(0.1619)
ML	0.1338	0.0719	0.0869	0.0819	0.0850	0.0849
	61%	-14%	4%	-2%	2%	2%
	(0.9888)	(0.5891)	(0.3663)	(0.2866)	(0.2441)	(0.1652)
			N=48			
OLS		0.1765	0.1344	0.1134	0.1121	0.0953
		112%	61%	36%	34%	14%
		(0.5312)	(0.3865)	(0.3221)	(0.2793)	(0.2001)
WLS		0.1614	0.1261	0.1082	0.1065	0.0918
		94%	51%	30%	28%	10%
		(0.5092)	(0.3741)	(0.3104)	(0.2694)	(0.1926)
GLS		0.2443	0.1754	0.1453	0.1325	0.1067
		193%	110%	74%	59%	28%
		(0.4687)	(0.2966)	(0.2360)	(0.1997)	(0.1380)
GMM2		0.2435	0.1754	0.1450	0.1324	0.1066
		192%	111%	74%	59%	28%
		(0.4704)	(0.2976)	(0.2365)	(0.2002)	(0.1382)
ML		0.0764	0.0816	0.0789	0.0824	0.0807
		-8%	-2%	-5%	-1%	-3%
		(0.5826)	(0.3262)	(0.2517)	(0.2092)	(0.1414)

Table 3a: Standard Error Estimates in a One-factor Model (N=25)

The table first reports the root-mean square error of the γ_0 or γ_1 estimates over 10,000 independent simulations, then the asymptotic standard error evaluated at the true parameter values, the simulation average of the estimated asymptotic standard errors, and finally the root-mean square error of the estimated standard errors in parentheses. The data-generating process is the same as in Tables 1 and 2. The estimation methods are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (truncated maximum likelihood) and GMM2 (generalized method of moments).

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			Standard Error on γ_0			
OLS	1.0437	0.7575	0.5489	0.4457	0.3956	0.2739
	1.1099	0.7848	0.5549	0.4531	0.3924	0.2775
	0.9955	0.7428	0.5374	0.4424	0.3851	0.2736
	(0.2253)	(0.1183)	(0.0614)	(0.0410)	(0.0313)	(0.0158)
WLS	1.0241	0.7465	0.5428	0.4393	0.3916	0.2715
	1.1022	0.7793	0.5511	0.4500	0.3897	0.2755
	0.9998	0.7466	0.5397	0.4441	0.3867	0.2746
	(0.2291)	(0.1195)	(0.0623)	(0.0412)	(0.0316)	(0.0158)
GLS	0.7586	0.4878	0.3351	0.2698	0.2334	0.1618
	0.6482	0.4583	0.3241	0.2646	0.2292	0.1620
	0.4121	0.3703	0.2917	0.2468	0.2179	0.1580
	(0.2462)	(0.0983)	(0.0408)	(0.0247)	(0.0173)	(0.0079)
GMM2	0.7624	0.4907	0.3363	0.2704	0.2338	0.1619
	0.6482	0.4583	0.3241	0.2646	0.2292	0.1620
	0.4003	0.3663	0.2903	0.2460	0.2173	0.1578
	(0.2569)	(0.1017)	(0.0418)	(0.0253)	(0.0177)	(0.0080)
ML	0.9888	0.5891	0.3663	0.2866	0.2441	0.1652
	0.6482	0.4583	0.3241	0.2646	0.2292	0.1620
	0.4198	0.3735	0.2927	0.2472	0.2182	0.1581
	(0.2407)	(0.0966)	(0.0404)	(0.0245)	(0.0172)	(0.0079)
			Standard Error on γ_1			
OLS	1.1830	0.8629	0.6295	0.5041	0.4490	0.3114
	1.2584	0.8898	0.6292	0.5137	0.4449	0.3146
	1.1554	0.8521	0.6135	0.5042	0.4384	0.3111
	(0.2013)	(0.1051)	(0.0545)	(0.0363)	(0.0277)	(0.0140)
WLS	1.1445	0.8427	0.6190	0.4953	0.4423	0.3070
	1.2446	0.8801	0.6223	0.5081	0.4400	0.3112
	1.1522	0.8503	0.6119	0.5029	0.4373	0.3103
	(0.2056)	(0.1063)	(0.0553)	(0.0365)	(0.0281)	(0.0140)
GLS	0.8332	0.5867	0.4300	0.3492	0.3056	0.2153
	0.8703	0.6154	0.4351	0.3553	0.3077	0.2176
	0.7087	0.5522	0.4114	0.3421	0.2993	0.2146
	(0.1738)	(0.0732)	(0.0312)	(0.0191)	(0.0136)	(0.0063)
GMM2	0.8393	0.5906	0.4311	0.3498	0.3061	0.2155
	0.8703	0.6154	0.4351	0.3553	0.3077	0.2176
	0.6949	0.5478	0.4100	0.3413	0.2988	0.2144
	(0.1871)	(0.0771)	(0.0323)	(0.0197)	(0.0140)	(0.0064)
ML	1.1241	0.7206	0.4753	0.3730	0.3212	0.2206
	0.8703	0.6154	0.4351	0.3553	0.3077	0.2176
	0.7135	0.5545	0.4121	0.3425	0.2996	0.2146
	(0.1703)	(0.0719)	(0.0309)	(0.0189)	(0.0135)	(0.0062)

Table 3b: Standard Error Estimates in a One-factor Model (N=48)

The table first reports the root-mean square error of the γ_0 or γ_1 estimates over 10,000 independent simulations, then the asymptotic standard error evaluated at the true parameter values, the simulation average of the estimated asymptotic standard errors, and finally the root-mean square error of the estimated standard errors in parentheses. The data-generating process is the same as in Tables 1 and 2. The estimation methods are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (truncated maximum likelihood) and GMM2 (generalized method of moments).

Methods	T=120	T=240	T=360	T=480	T=960
		Standard Error on γ_0			
OLS	0.5312	0.3865	0.3221	0.2793	0.2001
	0.5699	0.4030	0.3291	0.2850	0.2015
	0.5035	0.3754	0.3129	0.2738	0.1965
	(0.0851)	(0.0402)	(0.0261)	(0.0192)	(0.0095)
WLS	0.5092	0.3741	0.3104	0.2694	0.1926
	0.5500	0.3889	0.3175	0.2750	0.1944
	0.4915	0.3666	0.3056	0.2673	0.1918
	(0.0776)	(0.0357)	(0.0229)	(0.0167)	(0.0081)
GLS	0.4687	0.2966	0.2360	0.1997	0.1380
	0.3822	0.2703	0.2207	0.1911	0.1351
	0.2581	0.2241	0.1955	0.1746	0.1293
	(0.1274)	(0.0492)	(0.0280)	(0.0190)	(0.0076)
GMM2	0.4704	0.2976	0.2365	0.2002	0.1382
	0.3822	0.2703	0.2207	0.1911	0.1351
	0.2532	0.2223	0.1944	0.1738	0.1290
	(0.1321)	(0.0509)	(0.0289)	(0.0196)	(0.0079)
ML	0.5826	0.3262	0.2517	0.2092	0.1414
	0.3822	0.2703	0.2207	0.1911	0.1351
	0.2609	0.2251	0.1960	0.1749	0.1294
	(0.1251)	(0.0484)	(0.0276)	(0.0187)	(0.0076)
		Standard Error on γ_1			
OLS	0.6173	0.4595	0.3865	0.3369	0.2409
	0.6916	0.4890	0.3993	0.3458	0.2445
	0.6390	0.4670	0.3865	0.3370	0.2405
	(0.0703)	(0.0335)	(0.0218)	(0.0160)	(0.0079)
WLS	0.6058	0.4516	0.3773	0.3288	0.2348
	0.6742	0.4767	0.3892	0.3371	0.2384
	0.6287	0.4591	0.3799	0.3311	0.2362
	(0.0637)	(0.0297)	(0.0191)	(0.0139)	(0.0068)
GLS	0.5492	0.3862	0.3185	0.2784	0.1961
	0.5636	0.3985	0.3254	0.2818	0.1993
	0.4853	0.3677	0.3084	0.2706	0.1952
	(0.0835)	(0.0346)	(0.0201)	(0.0138)	(0.0058)
GMM2	0.5528	0.3878	0.3193	0.2790	0.1964
	0.5636	0.3985	0.3254	0.2818	0.1993
	0.4792	0.3655	0.3071	0.2698	0.1949
	(0.0895)	(0.0365)	(0.0212)	(0.0145)	(0.0061)
ML	0.7130	0.4374	0.3480	0.2964	0.2027
	0.5636	0.3985	0.3254	0.2818	0.1993
	0.4869	0.3683	0.3087	0.2708	0.1953
	(0.0822)	(0.0341)	(0.0198)	(0.0137)	(0.0058)

Table 4a: T-ratio Tests in a One-factor Model (N=25)

The table reports the empirical rejection rates over 10,000 simulated data sets of the t-ratio tests of the hypothesis that $\gamma_0 = 0$ or $\gamma_1 = 0$. The nominal size of the test is set at 5% and empirical power is adjusted for size distortions. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the asset pricing restrictions are

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length and N is the number of assets. The gamma values are set at $\gamma_0 = 0.0833\%$ and $\gamma_1 = 0.6667\%$ except when the null value of 0 is imposed.

Tests	T=60	T=120	T=240	T=360	T=480	T=960
		Testing the null: $\gamma_0 = 0$	when $\gamma_0 = 0$			
OLS	0.0587	0.0546	0.0535	0.0490	0.0564	0.0494
WLS	0.0429	0.0458	0.0468	0.0428	0.0518	0.0457
GLS	0.2842	0.1374	0.0879	0.0723	0.0684	0.0552
GMM2	0.2983	0.1419	0.0910	0.0744	0.0693	0.0547
ML	0.3808	0.2000	0.1159	0.0896	0.0792	0.0586
		Power of testing the null: $\gamma_0 = 0$	when $\gamma_0 = 0.0833\%$			
OLS	0.0509	0.0522	0.0547	0.0553	0.0546	0.0623
WLS	0.0508	0.0516	0.0519	0.0556	0.0548	0.0607
GLS	0.0510	0.0537	0.0594	0.0606	0.0658	0.0812
GMM2	0.0529	0.0561	0.0590	0.0614	0.0685	0.0811
ML	0.3902	0.2049	0.1280	0.1068	0.1003	0.0950
		Testing the null: $\gamma_1 = 0$	when $\gamma_1 = 0$			
OLS	0.0483	0.0492	0.0534	0.0479	0.0569	0.0473
WLS	0.0277	0.0383	0.0462	0.0426	0.0531	0.0442
GLS	0.0769	0.0535	0.0550	0.0508	0.0497	0.0505
GMM2	0.0886	0.0578	0.0568	0.0505	0.0506	0.0503
ML	0.1840	0.1245	0.0874	0.0674	0.0641	0.0573
		Power of testing the null: $\gamma_1 = 0$	when $\gamma_1 = 0.6667\%$			
OLS	0.0770	0.1186	0.1710	0.2526	0.2864	0.5777
WLS	0.0777	0.1220	0.1740	0.2661	0.2941	0.5899
GLS	0.0910	0.1627	0.2939	0.4248	0.5545	0.8605
GMM2	0.0826	0.1564	0.2878	0.4343	0.5472	0.8602
ML	0.2497	0.2776	0.3761	0.4872	0.5983	0.8683

Table 4b: T-ratio Tests in a One-factor Model (N=48)

The table reports the empirical rejection rates over 10,000 simulated data sets of the t-ratio tests of the hypothesis that $\gamma_0 = 0$ or $\gamma_1 = 0$. The nominal size of the test is set at 5% and empirical power is adjusted for size distortions. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the asset pricing restrictions are

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length and N is the number of assets. The gamma values are set at $\gamma_0 = 0.0833\%$ and $\gamma_1 = 0.6667\%$ except when the null value of 0 is imposed.

Tests	T=120	T=240	T=360	T=480	T=960
		Testing the null: $\gamma_0 = 0$ when $\gamma_0 = 0$			
OLS	0.0619	0.0555	0.0595	0.0560	0.0539
WLS	0.0550	0.0547	0.0547	0.0514	0.0510
GLS	0.2795	0.1410	0.1016	0.0889	0.0683
GMM2	0.2897	0.1458	0.1041	0.0891	0.0687
ML	0.3662	0.1715	0.1211	0.1005	0.0748
		Testing the null: $\gamma_0 = 0$ when $\gamma_0 = 0.0833\%$			
OLS	0.0515	0.0543	0.0599	0.0573	0.0705
WLS	0.0518	0.0550	0.0613	0.0603	0.0699
GLS	0.0559	0.0610	0.0659	0.0716	0.0925
GMM2	0.0549	0.0618	0.0652	0.0708	0.0947
ML	0.0525	0.0608	0.0595	0.0694	0.0896
		Testing the null: $\gamma_1 = 0$ when $\gamma_1 = 0$			
OLS	0.0492	0.0534	0.0479	0.0569	0.0473
WLS	0.0383	0.0462	0.0426	0.0531	0.0442
GLS	0.0535	0.0550	0.0508	0.0497	0.0505
GMM2	0.0578	0.0568	0.0505	0.0506	0.0503
ML	0.1700	0.0960	0.0783	0.0723	0.0592
		Testing the null: $\gamma_1 = 0$ when $\gamma_1 = 0.6667\%$			
OLS	0.1625	0.2694	0.3750	0.4982	0.7827
WLS	0.1753	0.2800	0.4023	0.5219	0.8002
GLS	0.1555	0.3360	0.4843	0.6126	0.9069
GMM2	0.1484	0.3348	0.4801	0.6166	0.9059
ML	0.3479	0.4446	0.5676	0.6739	0.9202

Table 5: B/M Factor Premium Estimates in a Three-factor Model

The table reports the average estimate, its percentage error, and root-mean-square error (all in percent) over 10,000 simulated data sets. The data-generating process is the Fama-French three-factor (market, book-to-market, size) regression model:

$$R_{it} = \alpha_i + \beta_{i1}f_{1t} + \beta_{i2}f_{2t} + \beta_{i3}f_{3t} + \epsilon_{it}, \quad i = 1,\dots,N, \quad t = 1,\dots,T;$$

with normally distributed residuals and factors, and the asset pricing restrictions are

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1\beta_1 + \gamma_2\beta_2 + \gamma_3\beta_3,$$

where R_{it} is the return on asset i in period t, f_{jt} is the realization of the j-th factor in period t, T is the time-series length and N is the number of assets. The true values of the factor premia are $\gamma_1 = 0.6667\%$, $\gamma_2 = 0.3333\%$, and $\gamma_3 = 0.1667\%$ and the zero-beta intercept is $\gamma_0 = 0.0833\%$. The estimation methods are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (truncated maximum likelihood) and GMM2 (generalized method of moments).

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			N=25			
OLS	0.3226	0.3259	0.3306	0.3312	0.3326	0.3328
	-3%	-2%	-1%	-1%	-0%	-0%
	(0.4291)	(0.3041)	(0.2143)	(0.1755)	(0.1546)	(0.1093)
WLS	0.3248	0.3267	0.3311	0.3315	0.3329	0.3330
	-3%	-2%	-1%	-1%	-0%	-0%
	(0.4272)	(0.3033)	(0.2135)	(0.1745)	(0.1538)	(0.1087)
GLS	0.3332	0.3305	0.3328	0.3333	0.3340	0.3335
	-0%	-1%	-0%	-0%	0%	0%
	(0.4270)	(0.3013)	(0.2113)	(0.1726)	(0.1521)	(0.1075)
GMM2	0.2488	0.2919	0.3142	0.3211	0.3249	0.3290
	-25%	-12%	-6%	-4%	-3%	-1%
	(0.3391)	(0.2705)	(0.2003)	(0.1665)	(0.1480)	(0.1061)
ML	0.3336	0.3309	0.3330	0.3335	0.3341	0.3335
	0%	-1%	-0%	0%	0%	0%
	(0.4278)	(0.3017)	(0.2114)	(0.1727)	(0.1522)	(0.1075)
			N=48			
OLS		0.3400	0.3381	0.3378	0.3398	0.3349
		2%	1%	1%	2%	0%
		(0.3560)	(0.2701)	(0.2237)	(0.1975)	(0.1408)
WLS		0.3413	0.3382	0.3394	0.3390	0.3345
		2%	1%	2%	2%	0%
		(0.3317)	(0.2472)	(0.2044)	(0.1781)	(0.1261)
GLS		0.3214	0.3293	0.3319	0.3339	0.3320
		-4%	-1%	-0%	0%	-0%
		(0.3368)	(0.2404)	(0.1975)	(0.1714)	(0.1213)
GMM2		0.2452	0.2915	0.3065	0.3151	0.3228
		-26%	-13%	-8%	-5%	-3%
		(0.2698)	(0.2155)	(0.1836)	(0.1622)	(0.1181)
ML		0.3326	0.3324	0.3331	0.3344	0.3319
		-0%	-0%	-0%	0%	-0%
		(0.3659)	(0.2550)	(0.2060)	(0.1771)	(0.1234)

Table 6a: Market Risk and B/M Premium Standard Error Estimates (N=25, K=3)

The table reports the root-mean square error of the risk premium estimates over 10,000 independent simulations, the asymptotic standard error evaluated at the true parameter values, the simulation average of the estimated asymptotic standard errors, and the root-mean square error of the estimated standard errors in parentheses. The data-generating process is the same Fama-French three-factor (market, book-to-market, size) regression model as in Table 5.

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			Standard Error on γ_1			
OLS	0.8932	0.6954	0.5116	0.4225	0.3740	0.2709
	1.0874	0.7689	0.5437	0.4439	0.3845	0.2719
	0.8928	0.6857	0.5081	0.4219	0.3686	0.2643
	(0.2214)	(0.1061)	(0.0513)	(0.0340)	(0.0253)	(0.0129)
WLS	0.9145	0.7186	0.5316	0.4382	0.3890	0.2807
	1.1330	0.8012	0.5665	0.4626	0.4006	0.2833
	0.9271	0.7184	0.5352	0.4452	0.3893	0.2795
	(0.2379)	(0.1121)	(0.0529)	(0.0346)	(0.0253)	(0.0126)
GLS	0.9291	0.6524	0.4701	0.3867	0.3441	0.2446
	0.9821	0.6944	0.4910	0.4009	0.3472	0.2455
	0.7413	0.5943	0.4521	0.3792	0.3329	0.2406
	(0.2525)	(0.1118)	(0.0487)	(0.0302)	(0.0216)	(0.0099)
GMM2	0.7737	0.6061	0.4538	0.3780	0.3382	0.2423
	0.9821	0.6944	0.4910	0.4009	0.3472	0.2455
	0.7412	0.5918	0.4497	0.3772	0.3311	0.2393
	(0.2533)	(0.1136)	(0.0500)	(0.0312)	(0.0224)	(0.0104)
ML	1.2169	0.8425	0.5488	0.4284	0.3712	0.2549
	0.9821	0.6944	0.4910	0.4009	0.3472	0.2455
	0.7477	0.5992	0.4540	0.3802	0.3335	0.2408
	(0.2479)	(0.1093)	(0.0480)	(0.0299)	(0.0214)	(0.0099)
			Standard Error on γ_2			
OLS	0.4291	0.3041	0.2143	0.1755	0.1546	0.1093
	0.4333	0.3064	0.2167	0.1769	0.1532	0.1083
	0.4266	0.3039	0.2158	0.1764	0.1528	0.1081
	(0.0385)	(0.0190)	(0.0096)	(0.0064)	(0.0048)	(0.0024)
WLS	0.4272	0.3033	0.2135	0.1745	0.1538	0.1087
	0.4310	0.3048	0.2155	0.1759	0.1524	0.1077
	0.4254	0.3029	0.2149	0.1756	0.1522	0.1076
	(0.0384)	(0.0190)	(0.0096)	(0.0064)	(0.0048)	(0.0024)
GLS	0.4270	0.3013	0.2113	0.1726	0.1521	0.1075
	0.4259	0.3011	0.2129	0.1739	0.1506	0.1065
	0.4197	0.2990	0.2123	0.1735	0.1503	0.1063
	(0.0389)	(0.0192)	(0.0097)	(0.0065)	(0.0048)	(0.0024)
GMM2	0.3391	0.2705	0.2003	0.1665	0.1480	0.1061
	0.4259	0.3011	0.2129	0.1739	0.1506	0.1065
	0.4316	0.3059	0.2164	0.1768	0.1530	0.1082
	(0.0452)	(0.0223)	(0.0114)	(0.0078)	(0.0059)	(0.0032)
ML	0.4278	0.3017	0.2114	0.1727	0.1522	0.1075
	0.4259	0.3011	0.2129	0.1739	0.1506	0.1065
	0.4197	0.2990	0.2123	0.1735	0.1503	0.1063
	(0.0389)	(0.0192)	(0.0097)	(0.0065)	(0.0048)	(0.0024)

Table 6b: Market Risk and B/M Premium Standard Error Estimates (N=48, K=3)

The table reports the root-mean square error of the risk premium estimates over 10,000 independent simulations, the asymptotic standard error evaluated at the true parameter values, the simulation average of the estimated asymptotic standard errors, and the root-mean square error of the estimated standard errors in parentheses. The data-generating process is the same Fama-French three-factor (market, book-to-market, size) regression model as in Table 5.

Methods	T=120	T=240	T=360	T=480	T=960
		Standard Error on γ_1			
OLS	0.6542	0.5054	0.4246	0.3826	0.2762
	0.8064	0.5702	0.4656	0.4032	0.2851
	0.6611	0.5040	0.4244	0.3745	0.2720
	(0.1562)	(0.0748)	(0.0483)	(0.0353)	(0.0172)
WLS	0.5820	0.4369	0.3681	0.3269	0.2348
	0.6782	0.4796	0.3916	0.3391	0.2398
	0.6047	0.4501	0.3750	0.3283	0.2357
	(0.0863)	(0.0391)	(0.0246)	(0.0179)	(0.0086)
GLS	0.5891	0.4130	0.3400	0.2973	0.2112
	0.6051	0.4279	0.3494	0.3026	0.2139
	0.4966	0.3837	0.3241	0.2859	0.2076
	(0.1129)	(0.0477)	(0.0283)	(0.0195)	(0.0084)
GMM2	0.5276	0.3966	0.3320	0.2914	0.2092
	0.6051	0.4279	0.3494	0.3026	0.2139
	0.4986	0.3841	0.3243	0.2860	0.2076
	(0.1115)	(0.0474)	(0.0282)	(0.0195)	(0.0084)
ML	0.7517	0.4866	0.3808	0.3247	0.2211
	0.6051	0.4279	0.3494	0.3026	0.2139
	0.4993	0.3852	0.3250	0.2865	0.2079
	(0.1106)	(0.0465)	(0.0276)	(0.0191)	(0.0082)
		Standard Error on γ_2			
OLS	0.3560	0.2701	0.2237	0.1975	0.1408
	0.4074	0.2881	0.2352	0.2037	0.1441
	0.3761	0.2735	0.2259	0.1970	0.1408
	(0.0387)	(0.0193)	(0.0129)	(0.0096)	(0.0048)
WLS	0.3317	0.2472	0.2044	0.1781	0.1261
	0.3619	0.2559	0.2089	0.1809	0.1279
	0.3549	0.2536	0.2077	0.1802	0.1277
	(0.0211)	(0.0103)	(0.0069)	(0.0051)	(0.0025)
GLS	0.3368	0.2404	0.1975	0.1714	0.1213
	0.3448	0.2438	0.1991	0.1724	0.1219
	0.3226	0.2356	0.1945	0.1694	0.1208
	(0.0291)	(0.0125)	(0.0078)	(0.0056)	(0.0026)
GMM2	0.2698	0.2155	0.1836	0.1622	0.1181
	0.3448	0.2438	0.1991	0.1724	0.1219
	0.3263	0.2379	0.1962	0.1709	0.1218
	(0.0279)	(0.0119)	(0.0074)	(0.0054)	(0.0025)
ML	0.3659	0.2550	0.2060	0.1771	0.1234
	0.3448	0.2438	0.1991	0.1724	0.1219
	0.3233	0.2360	0.1947	0.1696	0.1209
	(0.0285)	(0.0122)	(0.0076)	(0.0055)	(0.0025)

Table 7: Specification Tests in a One-factor Model

The table reports the empirical rejection rates of the model specification tests over 10,000 simulated data sets. The nominal size of the test is set at 5%. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the null hypothesis is the asset pricing restriction

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length and N is the number of assets. CST is the cross-sectional F test of Shanken (1985), LRT is the likelihood ratio test with Bartlett correction, and GMM2 is the generalized method of moments chi-squared test.

Tests	T=60	T=120	T=240	T=360	T=480	T=960
			N=25			
$\gamma^{\text{OLS}} = \gamma^{\text{GLS}}$	0.0724	0.0523	0.0498	0.0529	0.0549	0.0517
CST-GLS	0.0450	0.0445	0.0513	0.0487	0.0491	0.0507
GMM2	0.0252	0.0390	0.0479	0.0470	0.0481	0.0499
CST-ML	0.0428	0.0419	0.0500	0.0473	0.0483	0.0500
LRT	0.0678	0.0595	0.0608	0.0605	0.0586	0.0594
			N=48			
$\gamma^{\text{OLS}} = \gamma^{\text{GLS}}$		0.0804	0.0580	0.0560	0.0540	0.0533
CST-GLS		0.0484	0.0493	0.0551	0.0499	0.0495
GMM2		0.0264	0.0403	0.0512	0.0468	0.0497
CST-ML		0.0443	0.0479	0.0532	0.0486	0.0491
LRT		0.0791	0.0649	0.0681	0.0633	0.0634

Table 8: Specification Tests in a Three-factor Model

The table reports the empirical rejection rates of the model specification tests over 10,000 simulated data sets. The nominal size of the test is set at 5%. The data-generating process is the Fama-French three-factor (market, book-to-market, size) regression model:

$$R_{it} = \alpha_i + \beta_{i1}f_{1t} + \beta_{i2}f_{2t} + \beta_{i3}f_{3t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the null hypothesis is the asset pricing restriction

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1\beta_1 + \gamma_2\beta_2 + \gamma_3\beta_3,$$

where R_{it} is the return on asset i in period t, f_{jt} is the realization of the j-th factor in period t, T is the time-series length and N is the number of assets. CST is the cross-sectional F test of Shanken (1985), LRT is the likelihood ratio test with Bartlett correction, and GMM2 is the generalized method of moments chi-squared test.

Tests	T=60	T=120	T=240	T=360	T=480	T=960
			$N = 25$			
$\gamma^{\mathrm{OLS}} = \gamma^{\mathrm{GLS}}$	0.1527	0.0754	0.0631	0.0561	0.0578	0.0484
CST-GLS	0.0421	0.0414	0.0475	0.0499	0.0485	0.0498
GMM2	0.0472	0.0556	0.0654	0.0671	0.0665	0.0659
CST-ML	0.0410	0.0394	0.0446	0.0479	0.0473	0.0492
LRT	0.0851	0.0760	0.0743	0.0750	0.0726	0.0695
			$N = 48$			
$\gamma^{\mathrm{OLS}} = \gamma^{\mathrm{GLS}}$		0.1058	0.0655	0.0570	0.0558	0.0551
CST-GLS		0.0481	0.0491	0.0523	0.0519	0.0498
GMM2		0.0406	0.0630	0.0705	0.0705	0.0727
CST-ML		0.0450	0.0446	0.0486	0.0491	0.0479
LRT		0.1014	0.0866	0.0885	0.0835	0.0803

Table 9: Standard Error Estimates in an Alpha-based Pricing Error Model

The table reports the root-mean square error of the second-pass estimates for the single-factor model over 10,000 independent simulations, the asymptotic standard errors evaluated at the true parameter values, first ignoring model misspecification and then taking it into account, and the corresponding simulation averages of the estimated asymptotic standard errors. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1} f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with normally distributed residuals and factors, and the expected returns satisfy

$$E[R_t] = a + \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length, N is the number of assets and a is an N-vector of pricing errors taken from Table 9a of Fama and French (1993) with $N = 25$. The risk premium parameters are fixed at $\gamma_0 = 0.0833\%$, $\gamma_1 = 0.6667\%$.

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			Standard Error on γ_0			
OLS	1.0655	0.7695	0.5560	0.4517	0.3995	0.2780
	1.1022	0.7794	0.5511	0.4500	0.3897	0.2755
	1.1221	0.7935	0.5611	0.4581	0.3967	0.2805
	0.9993	0.7456	0.5394	0.4440	0.3866	0.2746
	1.0775	0.7799	0.5565	0.4559	0.3963	0.2805
WLS	1.0682	0.7761	0.5624	0.4556	0.4043	0.2821
	1.0955	0.7746	0.5477	0.4472	0.3873	0.2739
	1.1572	0.8183	0.5786	0.4724	0.4091	0.2893
	0.9983	0.7434	0.5371	0.4418	0.3845	0.2730
	1.1333	0.8121	0.5761	0.4713	0.4095	0.2895
GLS	0.7810	0.5155	0.3592	0.2925	0.2531	0.1773
	0.6459	0.4567	0.3230	0.2637	0.2284	0.1615
	0.7159	0.5062	0.3579	0.2923	0.2531	0.1790
	0.4133	0.3704	0.2914	0.2463	0.2174	0.1576
	0.6192	0.4783	0.3497	0.2881	0.2510	0.1783
			Standard Error on γ_1			
OLS	1.2002	0.8702	0.6321	0.5070	0.4507	0.3144
	1.2479	0.8824	0.6239	0.5094	0.4412	0.3120
	1.2641	0.8939	0.6321	0.5161	0.4469	0.3160
	1.1554	0.8521	0.6135	0.5042	0.4384	0.3111
	1.2210	0.8806	0.6276	0.5140	0.4464	0.3160
WLS	1.1817	0.8651	0.6321	0.5068	0.4512	0.3156
	1.2350	0.8733	0.6175	0.5042	0.4366	0.3088
	1.2882	0.9109	0.6441	0.5259	0.4554	0.3220
	1.1475	0.8450	0.6078	0.4993	0.4341	0.3080
	1.2666	0.9052	0.6417	0.5249	0.4557	0.3222
GLS	0.8518	0.6058	0.4454	0.3647	0.3204	0.2259
	0.8659	0.6123	0.4329	0.3535	0.3061	0.2165
	0.9186	0.6496	0.4593	0.3750	0.3248	0.2297
	0.7081	0.5509	0.4100	0.3408	0.2981	0.2136
	0.8462	0.6285	0.4530	0.3719	0.3232	0.2292

Table 10: Specification Tests in Misspecified Models

The table reports the empirical rejection rates of the model specification tests over 10,000 simulated data sets. The nominal size of the test is set at 5%. The data-generating process is the same as in Table 9 except there are now two specification for a. The first ($N = 25$) is that in Table 9, referred to here as Fama-French (1993) pricing errors. The second ($N = 48$) allows a to be randomly drawn from a normal distribution with mean 0 and standard deviation 2% (annualized). The null hypothesis is the linear pricing restriction

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1.$$

CST is the cross-sectional F test of Shanken (1985), LRT is the likelihood ratio test with Bartlett correction, and GMM2 is the generalized method of moments chi-squared test.

Tests	T=60	T=120	T=240	T=360	T=480	T=960
		Fama-French (1993) pricing errors				
$\gamma^{\mathrm{OLS}} = \gamma^{\mathrm{GLS}}$	0.1698	0.2504	0.4575	0.6324	0.7636	0.9714
CST-GLS	0.1342	0.3758	0.8149	0.9660	0.9968	1.0000
GMM2	0.0765	0.3308	0.7958	0.9643	0.9959	1.0000
CST-ML	0.1303	0.3661	0.8123	0.9656	0.9968	1.0000
LRT	0.1770	0.4110	0.8242	0.9685	0.9971	1.0000
		Randomly simulated pricing errors				
$\gamma^{\mathrm{OLS}} = \gamma^{\mathrm{GLS}}$		0.1524	0.2232	0.3146	0.3958	0.6832
CST-GLS		0.2458	0.6879	0.9215	0.9888	1.0000
GMM2		0.1417	0.6331	0.9071	0.9868	1.0000
CST-ML		0.2354	0.6847	0.9209	0.9886	1.0000
LRT		0.3048	0.7044	0.9254	0.9890	1.0000

Table 11: Market Risk Premium Estimates Under Conditional Heteroscedasticity

The table reports the average estimate, its percentage error, and root-mean-square error (all in percent) over 10,000 simulated data sets. The data-generating process is the standard market model:

$$R_{it} = \alpha_i + \beta_{i1}f_{1t} + \epsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T;$$

with t-distributed residuals from a joint multivariate t-distribution of the returns and factor with 8 degrees of freedom, and the asset pricing restrictions are

$$H_0: \qquad E[R_t] = \gamma_0 1_N + \gamma_1 \beta_1,$$

where R_{it} is the return on asset i in period t, f_{1t} is the realization of the market factor in period t, T is the time-series length and N is the number of assets. The true value for the factor risk premium is $\gamma_1 = 0.6667\%$ and the zero-beta intercept is $\gamma_0 = 0.0833\%$. The estimation methods are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (truncated maximum likelihood) and GMM2 (generalized method of moments).

Methods	T=60	T=120	T=240	T=360	T=480	T=960
			N=25			
OLS	0.5789	0.6090	0.6403	0.6582	0.6497	0.6600
	-13%	-9%	-4%	-1%	-3%	-1%
	(1.0843)	(0.8100)	(0.5932)	(0.4875)	(0.4263)	(0.3052)
WLS	0.5495	0.5901	0.6317	0.6497	0.6437	0.6575
	-18%	-11%	-5%	-3%	-3%	-1%
	(1.0602)	(0.8034)	(0.5922)	(0.4884)	(0.4274)	(0.3074)
GLS	0.4114	0.4914	0.5637	0.5961	0.6072	0.6354
	-38%	-26%	-15%	-11%	-9%	-5%
	(0.8318)	(0.5948)	(0.4351)	(0.3559)	(0.3173)	(0.2282)
GMM2	0.4351	0.5170	0.5842	0.6115	0.6198	0.6423
	-35%	-22%	-12%	-8%	-7%	-4%
	(0.8715)	(0.6209)	(0.4458)	(0.3618)	(0.3213)	(0.2298)
ML	0.5522	0.6105	0.6358	0.6477	0.6466	0.6562
	-17%	-8%	-5%	-3%	-3%	-2%
	(1.0906)	(0.7166)	(0.4791)	(0.3805)	(0.3330)	(0.2339)
			N=48			
OLS		0.5335	0.5898	0.6147	0.6273	0.6470
		-20%	-12%	-8%	-6%	-3%
		(0.6176)	(0.4586)	(0.3834)	(0.3387)	(0.2405)
WLS		0.5535	0.6051	0.6220	0.6343	0.6508
		-17%	-9%	-7%	-5%	-2%
		(0.6040)	(0.4521)	(0.3752)	(0.3306)	(0.2348)
GLS		0.4719	0.5466	0.5768	0.5963	0.6288
		-29%	-18%	-13%	-11%	-6%
		(0.5532)	(0.3951)	(0.3228)	(0.2830)	(0.2004)
GMM2		0.5036	0.5769	0.6033	0.6185	0.6417
		-24%	-13%	-9%	-7%	-4%
		(0.5773)	(0.4062)	(0.3296)	(0.2878)	(0.2022)
ML		0.6115	0.6354	0.6405	0.6463	0.6558
		-8%	-5%	-4%	-3%	-2%
		(0.6787)	(0.4404)	(0.3459)	(0.2982)	(0.2057)

Table 12: Risk Premium Estimates in the Fama-French Three-Factor Model

The table reports parameter estimates for the model:

$$R_{it} - r_{ft} = \alpha_i + \beta_{i1}(f_{M,t} - r_{ft}) + \beta_{i2}f_{SMB,t} + \beta_{i3}f_{HML,t} + \epsilon_{it},$$

with

$$H_0: \qquad E(R_{it} - r_{ft}) = \gamma_0 + \gamma_1\beta_{i1} + \gamma_2\beta_{i2} + \gamma_3\beta_{i3},$$

where f_M is the return on the market factor, f_{SMB} is the SMB spread return, f_{HML} is the HML spread return, and r_{ft} is the 30-day T-bill rate. The R_{it}'s are the test asset returns on the 25 stock portfolios formed on size and book-to-market. The data is monthly from January 1964 to December 2003. The table first reports the restricted estimates under the assumption that $\gamma_0 = 0$. These are followed by cross-sectional estimates of the linear expected return relation. The estimation methods, which do not impose factor portfolio constraints, are the OLS, WLS, and GLS versions of the (Fama-MacBeth) two-pass regression methodology, ML (maximum likelihood) and GMM2 (generalized method of moments). Standard errors of the estimates under H_0 are given in parentheses and those under model misspecification in brackets. All estimates and standard errors are in percent. The last column reports tests of significance for the difference between the second-pass estimates and the restricted estimates, with p-values in parentheses.

Methods	γ_0	γ_1	γ_2	γ_3	Equality Test
Imposing $\gamma_0 = 0$ and the factor portfolio constraints					
Sample Averaging		0.45*	0.27	0.43*	
		(0.21)	(0.15)	(0.14)	
Linear expected return model: unrestricted					
OLS	1.42*	-0.93*	0.23	0.46*	23.38
	(0.32)	(0.38)	(0.15)	(0.14)	(0.01)
	[0.36]	[0.42]	[0.15]	[0.14]	
WLS	1.44*	-0.93*	0.24	0.43*	27.14
	(0.34)	(0.40)	(0.15)	(0.14)	(0.00)
	[0.36]	[0.42]	[0.15]	[0.14]	
GLS	1.60*	-1.09*	0.26	0.43*	33.93
	(0.29)	(0.35)	(0.15)	(0.14)	(0.00)
	[0.32]	[0.38]	[0.15]	[0.14]	
GMM2	1.57*	-0.99*	0.22	0.40*	
	(0.28)	(0.35)	(0.17)	(0.15)	
ML	1.85*	-1.35*	0.26	0.43*	
	(0.29)	(0.36)	(0.15)	(0.14)	

Table 13: Specification Tests in the Fama-French Three-Factor Model

The table reports the specification tests and their p-values in parentheses for the model:

$$R_{it} - r_{ft} = \alpha_i + \beta_{i1}(f_{M,t} - r_{ft}) + \beta_{i2}f_{SMB,t} + \beta_{i3}f_{HML,t} + \epsilon_{it},$$

with

$$H_0: \qquad E(R_{it} - r_{ft}) = \gamma_0 + \gamma_1\beta_{i1} + \gamma_2\beta_{i2} + \gamma_3\beta_{i3},$$

where f_M is the return on the market factor, f_{SMB} is the SMB spread return, f_{HML} is the HML spread return, and r_{ft} is the 30-day T-bill rate. The R_{it}'s are the test asset returns on the 25 stock portfolios formed on size and book-to-market. The data is monthly from January 1964 to December 2003. The first column reports a test of significance for the difference between the OLS and GLS second-pass estimates. CST is the cross-sectional F test of Shanken (1985), LRT is the likelihood ratio test with Bartlett correction, and GMM2 is the generalized method of moments chi-squared test. GRS is the Gibbons, Ross and Shanken (1989) joint test that the alphas are all zero.

$\gamma^{\mathrm{OLS}} = \gamma^{\mathrm{GLS}}$	CST-GLS	CST-ML	LRT	GMM2	GRS
8.80	2.15	2.12	48.33	45.78	3.15
(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)