Return Distributions and Improved Tests of Asset Pricing Models

Keith Vorkink
Brigham Young University

We compare and contrast some existing ordinary least squares (OLS)- and generalized method of moments (GMM)-based tests of asset pricing models with a new more general test. This new test is valid under the assumption that returns are elliptically distributed, a necessary and sufficient assumption of the linear capital asset pricing model (CAPM). This new test fails to reject the CAPM on a dataset of stocks sorted by market valuations, whereas similar tests constructed from OLS and GMM estimation methods reject the linear CAPM. We also find that outliers reduce the OLS-estimated mispricing of the linear CAPM on monthly returns sorted by previous performance, that is, momentum. Monte Carlo evidence supports superior size and power properties of the new test relative to OLS- and GMM-based tests.

Quantifying the relation between risk and expected return is a critical and largely unresolved issue in finance despite enormous effort to parse out any meaningful relationships. Although the risk-return relationship remains unresolved, a consensus exists regarding one aspect of returns: they exhibit features in third and fourth moments that are not consistent with a normal distribution. Despite this stylized fact that describes returns, investigations into the risk-return relationship continue to rely on asset pricing tests that are suboptimal in the presence of these non-normalities.

Historically most of the empirical work in asset pricing has used ordinary least squares (OLS) [e.g., see Fama and French (1992, 1993)], a method that efficiently estimates parameters only under the assumption of multivariate normal return distributions. Many have responded to the evidence of nonnormalities by turning to Hansen's (1982) generalized method of moments (GMM), which greatly relaxes distributional assumptions relative to OLS. Investigation of GMM methods is quite extensive in the context of robust asset pricing tests, and the results are mixed. MacKinlay and Richardson (1991) extend OLS test statistics to cases of nonnormality and develop asset pricing tests using GMM estimates.

I would like to thank Greg Bauer, Jim Brau, Emin Dinlersoz, Campbell Harvey (the editor), Douglas Hodgson, Jon Lewellen, Grant McQueen, Bill Schwert, Jay Shanken, Steve Thorley, and an anonymous referee for their comments, as well as seminar participants at the University of Rochester and the University of Utah. Excellent research assistance was provided by Tyler Brough. Address correspondence to Marriott School of Management, Brigham Young University, Provo UT 84602, or e-mail: keith_vorkink@byu.edu.

The Review of Financial Studies Fall 2003 Vol. 16, No. 3, pp. 845–874, DOI: 10.1093/rfs/hhg015

Harvey and Zhou (1993) compare tests using OLS methods against tests using GMM methods and find that, even though in most cases the test statistics differ, the conclusions do not. While GMM provides robustness to asset pricing tests, under nonnormality it generally does not lead to fully efficient (minimum variance) estimates and powerful asset pricing tests. In addition, evidence suggests that asset pricing test statistics based on GMM methods have aberrant properties [see Ferson and Foerster (1994)]. Thus the use of standard estimation (OLS and GMM) and testing techniques may contribute to our confusion about the relationship between risk and return.

Early on, Mandlebrot (1963) and Fama (1965) pointed out that the normality assumption was not accurate. Since then, others have supported their results [see Affleck-Graves and McDonald (1989) and Zhou (1993)]. These rejections of normality have led researchers to investigate models of expected returns, such as the capital asset pricing model (CAPM) under alternative distributions, including the family of elliptical distributions [see Harvey (2002)]. Elliptical distributions are a semiparametric family of distributions that include the normal distribution, distributions having conditional dependence among higher moments (GARCH), infinite variance (Cauchy), symmetric Pareto–Levy, central Student's t, and others. In particular, this semiparametric family allows for flexibility in the fourth moment, enabling kurtosis levels to differ from the case of normality. In addition, elliptical distributions are shown to play a role in robust derivation of the CAPM. For example, Berk (1997) shows that elliptical distributions are a necessary assumption for obtaining the linear CAPM when agents maximize well-behaved expected utility functions [see also Chamberlain (1983) and Owen and Rabinovitch (1983)]. Although elliptical distributions cannot model skewness often found in return distributions, they do have full flexibility to account for kurtosis, a feature more prevalent in returns than skewness.

Recently Hodgson, Linton, and Vorkink (2002; hereafter HLV) have developed an estimation methodology that enables linear asset pricing models, including the CAPM, to be estimated and tested efficiently assuming elliptical distributions. Contrasted with GMM methods, which adjust only the standard errors of OLS estimates, the HLV method incorporates the elliptical assumption when constructing coefficients and corresponding standard errors. The resulting estimates are asymptotically efficient, and consequently, resulting tests have attractive size and power properties. Not only does the HLV methodology allow one to estimate linear asset pricing models under very general distributional assumptions, but it should also provide better-behaved test statistics, given the preponderance of kurtosis in returns.

Our first contribution is to provide tests of ellipticity on datasets of monthly equity returns. Zhou (1993) performed tests of particular distributions (t-distributions and mixture of normals) within the family of elliptical distributions but did not test the general hypothesis that returns are elliptically distributed. We employ the test of Beran (1979), which specifically tests the general hypothesis of ellipticity on a number of monthly stock return datasets and fail to reject the elliptical assumption.

Our second, and primary, contribution results from an investigation into the risk-return relationship of two well-known datasets on which traditional models of expected return have fallen short: a dataset of monthly portfolio returns sorted by market valuation (firm size), and a dataset of monthly portfolio returns sorted by previous performance (momentum). We find that using the more general HLV method, the CAPM is not rejected on the size-sorted dataset, whereas OLS and GMM both strongly reject the CAPM on the same dataset. We find that the OLS and GMM spurious rejections of the CAPM are primarily caused by a few extreme returns whose influence is largely felt in the smaller size deciles. In contrast, the HLV method appears to be robust to the presence of these outliers.

We also find striking differences between OLS and HLV estimates of asset pricing models on the momentum-sorted dataset. These differences are again driven by the presence of outliers, and these outliers are primarily tied to January returns.[1] However, in contrast to the size-sorted data, outliers cause OLS/GMM to understate the mispricing of the linear CAPM on this dataset. We also find remarkably large differences between estimated slope coefficients of additional factors (Fama–French three-factor model, quadratic and cubic CAPM) on this dataset. One extreme case in the differences between OLS/GMM and HLV estimates leads to a difference in expected annual returns of 5%.

Our third contribution is a Monte Carlo study of asset pricing tests based on the HLV estimator as well as OLS and GMM. We find that the tests generated by the HLV estimator exhibit substantially more power relative to the OLS- and GMM-based tests in the presence of kurtosis and are well sized, even in the presence of skewness. We also find that the HLV method has superior size properties relative to GMM methods in the presence of kurtosis. We document the poor performance of the GMM-based tests when either the number of estimated and tested portfolios increases or when kurtosis increases, calling into question the robustness properties thought to exist with GMM.

The article is organized as follows. Section 1 provides a brief discussion of existing OLS and GMM methods for testing mean-variance efficiency

[1] Knez and Ready (1997), in a related work, find that cross-sectional regressions similar to those found in Fama and French (1992, 1993) are quite senstive to outliers using a least trimmed squares estimation method.

and introduces the test based on the HLV estimation method. Section 2 conducts distributional tests on our dataset of monthly returns. Section 3 applies the HLV method, along with OLS and GMM, in testing the CAPM on two datasets of monthly stock returns. Section 4 discusses a Monte Carlo exercise that analyzes the size and power properties of OLS-, GMM-, and HLV-generated asset pricing tests. Section 5 concludes.

1. Asset Pricing Tests Background

1.1 OLS-based tests

We restrict our attention to mean-variance expected return models where the excess return on an asset (i) is linear in its covariances ($\boldsymbol{\beta}_i$), with a list of factors (**F**) as follows:[2]

$$r_{i,t} = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i \mathbf{F}_t + u_{i,t}, \quad i = 1 \ldots N, \;\; t = 1 \ldots T,$$

where $r_{i,t}$ corresponds to the excess return on, portfolio i in period t, N is the number of portfolios, T is the length of the time series, $\mathbf{F}_t$ is a $k \times 1$ vector of factor portfolio excess returns, β_i is a $1 \times k$ vector of scaled covariances (or betas), and $u_{i,t}$ is a mean zero innovation. We restate the model in vector form below:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} \mathbf{F}_t + \mathbf{u}_t, \tag{2}$$

with vectors $\mathbf{r}_t$, $\mathbf{u}_t$, and α being $N \times 1$ and $\boldsymbol{\beta}$ being $N \times k$. Historically the residuals have been assumed to be jointly normally distributed either directly or as a result of assuming that the returns and factors are jointly multivariate normal.

The mean-variance efficiency of **F** implies that estimates of α should not statistically differ from zero if the return model is the "true" model. Given a time series of returns and factors $\{\mathbf{r}_t, \mathbf{F}_t\}_{t=1}^{T}$, Equation (2) can be estimated using OLS, followed by the construction of a standard Wald test of mean-variance efficiency based on the null hypothesis:

$$H_0\colon \alpha_i = 0 \quad i = 1, \ldots, N.$$

The standard OLS-constructed Wald statistic that jointly tests the intercepts' significance is

$$J_{\mathrm{OLS}} = \hat{\boldsymbol{\alpha}}^T \mathrm{var}(\hat{\boldsymbol{\alpha}})^{-1} \hat{\boldsymbol{\alpha}} \overset{a}{\sim} \chi^2_N. \tag{3}$$

[2] Examples of such models include Fama and French's (1993) multifactor models, Ross' (1976) APT model, and Sharpe's (1964) and Lintner's (1965) CAPM. The model can also accommodate conditional pricing models such as Shanken (1990).

Under the null, the portfolio vector **F** spans the mean-variance frontier and the given asset pricing model explains expected returns, leading to statistically insignificant values of J_{OLS}.[3]

1.2 GMM-based tests

Hansen's (1982) GMM allows the researcher to estimate and test a particular asset pricing model with minimal restrictions on the residual density. Parameter estimates result from a quadratic minimization of moment conditions implied by the particular asset pricing model. For asset pricing models in the form of Equation (2), the sample moment conditions are

$$s(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{t=1}^{T} \begin{pmatrix} \mathbf{u}_t \\ \mathbf{u}_t \mathbf{F}_t \end{pmatrix}, \tag{4}$$

based on the population moments $E(\mathbf{u}_t) = E(\mathbf{u}_t \mathbf{F}_t) = 0$. The GMM objective function can be written as

$$s(\boldsymbol{\alpha}, \boldsymbol{\beta})^{\mathrm{T}} W^{-1} s(\boldsymbol{\alpha}, \boldsymbol{\beta}),$$

with W representing a weighting matrix, optimally chosen to consistently estimate the variance of the score vectors $s(\boldsymbol{\alpha}, \boldsymbol{\beta})$ [see Andrews (1991)].

We analyze two statistics available from GMM estimation methods. For a more complete discussion of possible GMM asset pricing tests, see MacKinlay and Richardson (1991) and Harvey and Zhou (1993). Our first statistic results from estimating the candidate asset pricing model while imposing the null H_0: $\alpha_i = 0, i = 1, \ldots, N$. We construct the test statistic as

$$J_{\mathrm{GMM1}} = s_r^{\mathrm{T}} W_r^{-1} s_r \overset{a}{\sim} \chi_N^2, \tag{5}$$

where the subscript r denotes the restricted model and the dependence of the scores on the parameters is suppressed for simplicity.

Our second statistic follows Harvey and Zhou (1993) and is constructed as follows:

$$J_{\mathrm{GMM2}} = s_r^{\mathrm{T}} W_{ur}^{-1} s_r - s_{ur}^{\mathrm{T}} W_{ur}^{-1} s_{ur} \overset{a}{\sim} \chi_N^2, \tag{6}$$

where the subscript ur represents the unrestricted model where $\boldsymbol{\alpha}$ is estimated [see Gallant and Jorgensen (1979)]. Although the asymptotic distribution of these two statistics is the same, Harvey and Zhou (1993) show that finite sample performance differs.

---

[3] Gibbons, Ross, and Shanken (1989) show that under multivariate normality, a correction to J_{OLS} will lead to a finite F-distribution test statistic. Campbell, Lo, and MacKinlay (1997) review the literature of asset pricing tests of this type.

1.3 Existing tests under elliptical returns

A random variable is said to be elliptical if and only if its density $p(u)$ can be represented by the function $g(u^{\mathrm{T}}\Sigma^{-1}u)$. This semiparametric family of distributions has an appeal in asset pricing because of its ability to mimic the kurtosis found in return data as well as its theoretical connection to robust mean-variance optimization, as noted by Berk (1997).

The performance of traditional asset pricing tests, such as J_{OLS}, may break down in the presence of elliptical returns. This problem arises because the kurtosis present in elliptical returns may be driven by conditional heteroscedasticity. If conditional heteroscedasticity is such that the variance of u_t is related to regressors (F_t in our case), the OLS-estimated standard errors are biased and resulting statistical tests will have problems with size [see Van Praag and Wessleman (1989)]. One solution to the problem, as shown in Zhou (1993), is to use information from the unconditional distribution to correct for the conditional heteroscedasticity-induced bias in estimated second moments [see also Géczy (1998)].[4] Zhou's correction is accomplished by a simple multiplicative correction to the J_{OLS} statistic, which generalizes the statistic to the assumption of elliptical return distributions. The correction is a follows:

$$J_{\mathrm{C-OLS}} = J_{\mathrm{OLS}} * \eta^{-1} \overset{a}{\sim} \chi^2_N, \tag{7}$$

with $\eta = 1 + \vartheta$, $\vartheta = \kappa_x / N(N+2)$, and κ_x is Mardia's (1970) multivariate measure of kurtosis. Under multivariate normality $\vartheta = 0$, and $J_{\mathrm{C-OLS}} = J_{\mathrm{OLS}}$. However, given excess kurtosis, $\vartheta > 0$, and $J_{\mathrm{C-OLS}} < J_{\mathrm{OLS}}$.

Zhou's correction accounts for the presence of kurtosis in the distributions of returns, but ignores the potential information contained in the higher moments of the true distribution when estimating the parameters and their covariance matrix. Incorporating this information in the estimation procedure can lead to efficiency gains. The challenge, however, is to place structure on the model consistent with the data and to allow for generality where the data departs from standard assumptions. If this trade-off is optimally chosen, efficiency should increase with minimal effect on robustness, which will lead to more powerful asset pricing tests. Below we briefly discuss the methodology of HLV, which attempts to improve this trade-off.

1.4 HLV-based tests

We refer the interested reader to Appendix A for a discussion of the HLV estimation procedure; here we simply state the asymptotic distribution of

[4] Another correction for this problem is to first weight returns with their estimated conditional standard deviations, and then estimate and test candidate asset pricing models, as seen in White (1980).

the HLV estimator. Let $\theta = [\boldsymbol{\alpha}^{\mathrm{T}}, \mathrm{vec}(\boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}]^{\mathrm{T}}$ be the $N(k+1)$ vector of parameters in the asset pricing model defined in Equation (2). Assuming that the joint distribution of $\{\mathbf{r}_t, \mathbf{F}_t\}_{t=1}^{T}$ is elliptical, the HLV estimate ($\tilde{\theta}$) has the following asymptotic distribution:

$$\sqrt{T}(\tilde{\theta} - \theta_0) \Longrightarrow N(0,\ \mathcal{I}^{-1}), \tag{8}$$

where $\Longrightarrow$ denotes weak convergence of probability measures, θ_0 is the true value of θ, and $\mathcal{I}$ is Fisher's asymptotic information matrix.[5] Equation (8) shows that the variance of $\tilde{\theta}$ will be asymptotically equivalent to the variance of the maximum-likelihood estimator ($\mathcal{I}^{-1}$), demonstrating $\tilde{\theta}$'s adaptive, or asymptotically efficient, property. Intuitively $\tilde{\theta}$ is estimated semiparametrically, a procedure that contains both nonparametric and parametric elements. The nonparametric features allow the data to determine the kurtosis of the distribution, while the parametric features of the method impose the elliptical assumption.

Given our estimates of $\tilde{\theta}$, Wald tests can be constructed in the usual manner. The distribution of the test statistic result hinges on the asymptotic distribution of our estimator given in Equation (8). Following White (1994), a Wald test similar to J_{OLS}, defined in Equation (3), will hold asymptotically. We denote the Wald test generated using HLV estimates, denoted as J_{HLV}, to distinguish it from the OLS-based Wald test:

$$J_{\mathrm{HLV}} = \tilde{\boldsymbol{\alpha}}^{\mathrm{T}} \mathrm{var}(\tilde{\boldsymbol{\alpha}})^{-1} \tilde{\boldsymbol{\alpha}} \overset{a}{\sim} \chi^2_N. \tag{9}$$

2. Distributional Tests

2.1 Data description

Our dataset consists of CRSP monthly returns from July 1963 through December 1995. Firms traded on the NYSE, NASDAQ, and AMEX are included in the sample. To compare OLS, GMM, and HLV estimations and asset pricing tests, we sort returns into two sets of portfolios: firm size and previous performance (momentum). The size-sorted dataset is well known from its abundant use in empirical studies. The momentum-sorted

[5] The HLV method requires a distributional assumption regarding $\mathbf{u}_t$ which may be obtained either directly or indirectly through assumptions describing $\{\mathbf{r}_t, \mathbf{F}_t\}_{t=1}^{T}$. By making the distributional assumption on $\{\mathbf{r}_t, \mathbf{F}_t\}_{t=1}^{T}$, a potential high-order dependence between $\mathbf{F}_t$ and $\mathbf{u}_t$ may cause the adaptive properties of the HLV estimator to break down. As discussed earlier, White's (1980) method can be used to remove this dependence. In a previous version of the article we used various methods to remove this dependence according to White (1980), such as nonparametric estimations of $h(Z_t)$ as well as GARCH models. These results are more difficult to interpret, do not deviate materially from the results reported, and consequently they are not included, but are available on request. In the current version of the article, we estimate $\tilde{\theta}$ assuming that returns follow some unconditional leptokurtokic distribution and ignore any possible conditional heteroscedasticity.

dataset is based on the results of Jegadeesh and Titman (1993), who found that abnormal returns are possible by buying recent winners and selling recent losers. We constructed this dataset of portfolio returns by placing firms into deciles based on their returns over the $t-12$ to $t-2$ prior months and holding those portfolios for one month.

We examine the ability of a number of linear asset pricing models to explain the returns of these two datasets: the linear CAPM, a quadratic CAPM, a cubic CAPM, and the Fama and French (1993) three-factor asset pricing model. The linear CAPM, the Fama and French model, and the factors these models use to explain returns are well known. The quadratic and cubic CAPM have recently been used to explain returns in cases where other models have fallen short. Harvey and Siddique (2000) find that quadratic CAPM factors add substantially to the explanation of returns on a number of datasets, including momentum-sorted returns. Dittmar (2002) shows that cubic CAPM factors add value in explaining portfolio returns sorted by size and book-to-market. The quadratic CAPM postulates that in addition to covariance with the market portfolio, a firm's return is determined by its coskewness with the market portfolio. Similarly a cubic CAPM allows covariance, coskewness, and cokurtosis with the market portfolio to influence expected returns. We form two portfolio factors to proxy for coskewness and cokurtosis with the market portfolio and use these to estimate and test the quadratic and cubic CAPM. We construct these portfolios in a manner similar to that of Harvey and Siddique (2000), where portfolios are formed by sorting firms based on measures of coskewness ($\beta_{s,i}$), and cokurtosis ($\beta_{k,i}$), defined as

$$\beta_{s,i} = \frac{E(\varepsilon_{i,t+1}\varepsilon^2_{m,t+1})}{\sqrt{E(\varepsilon^2_{i,t+1})}E(\varepsilon^2_{m,t+1})} \tag{10}$$

$$\beta_{k,i} = \frac{E(\varepsilon^2_{i,t+1}\varepsilon^2_{m,t+1})}{E(\varepsilon^2_{i,t+1})E(\varepsilon^2_{m,t+1})} \tag{11}$$

with $\varepsilon_{i,t+1} = r_{i,t+1} - \alpha_i - \beta_{i,m} r_{m,t+1}$, where $r_{m,t+1}$ is the excess return on the market portfolio. To obtain the returns for these factors in month t we first use months $t-60$ through $t-1$ to estimate Equations (10) and (11) for all firms in our sample. We then form six portfolios by ranking firms based on coskewness or cokurtosis: the 30% of firms with the lowest values of coskewness (cokurtosis) form the portfolio entitled S^- (K^-); the middle 40% is called S^0 (K^0); and the final 30% form a portfolio called S^+ (K^+). We retain the value-weighted return for each portfolio in month t and repeat this process for all months in the sample. We use the excess return on the S^- portfolio to proxy for our coskewness factor, and use

Table 1
Univariate summary statistics

	Factor	Mean	Median	Std. dev.	Skewness	Kurtosis	J–B
Panel A: Factors							
	r_m	0.41	0.54	4.41	−0.34**	0.81**	101.67**
	SMB	0.29	0.13	2.88	0.19	0.40**	25.46**
	HML	0.45	0.39	2.59	0.01	0.37**	19.86**
	S^-	0.56	0.68	4.68	0.18	0.73**	84.26**
	K^+	0.50	0.55	4.41	0.17	1.02**	162.03**
	Size decile	Mean	Median	Std. dev.	Skewness	Kurtosis	J–B
Panel B: Size-sorted portfolios							
Small	1	1.48	0.91	7.26	1.21**	3.67**	1995.25**
	2	1.30	1.20	6.31	0.43**	2.40**	824.57**
	3	1.20	1.10	5.93	0.08	1.92**	524.84**
	4	1.26	1.36	5.64	−0.19	1.41**	285.69**
	5	1.17	1.24	5.40	−0.37**	1.21**	216.64**
	6	1.19	1.36	5.27	−0.38**	1.21**	216.64**
	7	1.10	1.07	5.01	−0.34**	0.95**	137.15**
	8	1.10	1.09	4.96	−0.30*	0.99**	145.62**
	9	0.99	1.29	4.70	−0.22	0.63**	60.19**
Large	10	0.83	0.82	4.14	−0.12	0.78**	88.46**
	Decile	Mean	Median	Std. dev.	Skewness	Kurtosis	J–B
Panel C: Momentum-sorted portfolios, one-month rebalancing, equal-weighted momentum							
	1(losers)	0.65	0.19	8.28	1.10**	1.78**	581.41**
	2	0.77	0.59	6.22	0.42**	1.37**	306.86**
	3	1.04	1.09	5.58	0.21	1.35**	293.00**
	4	1.15	1.10	5.14	0.14	1.52**	405.45**
	5	1.22	1.28	4.90	−0.16	1.71**	462.11**
	6	1.35	1.60	4.85	−0.27*	1.61**	415.34**
	7	1.48	1.86	4.87	−0.52**	1.72**	484.32**
	8	1.56	1.93	5.08	−0.55**	1.47**	364.59**
	9	1.77	2.04	5.49	−0.62**	1.37**	324.19**
	10(winners)	2.04	2.38	6.57	−0.58**	0.91**	155.82**

The above statistics are based on monthly portfolio returns, in percentage points, for the period of July 1963 through December 1995. J–B refers to the Jarque–Bera test statistic. r_m is the excess market return, and SMB and HML are portfolio returns from Fama and French three-factor model. S^- (K^+) corresponds to our coskewness (cokurtosis) factor proxy. ** (*) refers to a rejection of the hypothesis that a given moment condition is consistent with the normal distribution at the .01 (.05) level.

factors S^- and r_m to estimate and test the quadratic CAPM.[6] We use the excess return on K^+ to proxy for the cokurtosis factor and use factors K^+, S^-, and r_m to test the cubic CAPM.

2.2 What distribution do returns follow?

Univariate summary statistics for the factors and the two datasets (size-sorted and momentum-sorted) are listed in Table 1. Included in this table

[6] HLV estimates of a linear asset pricing model including coskewness as a factor are no longer adaptive except under certain circumstances (e.g., coskewness coefficients across portfolios must sum to zero). However, Section 4 shows that HLV estimates are likely to be well behaved in the presence of asymmetry.

are measures of skewness ($\hat{S}$) and excess kurtosis ($\hat{K}$), which are calculated in the following manner:

$$\hat{S} = \frac{1}{T}\sum_{t=1}^{T}\frac{(r_t - \bar{r})^3}{\hat{\sigma}^3} \tag{12}$$

$$\hat{K} = \frac{1}{T}\sum_{t=1}^{T}\frac{(r_t - \bar{r})^4}{3\hat{\sigma}^4} - 1. \tag{13}$$

Under the assumption of normality, $\hat{S}$ and $\hat{K}$ would have mean zero asymptotic normal distributions with variances $6/T$ and $8/3T$, respectively.

Panel A contains statistics on the five factor returns: the excess market return (r_m), the Fama and French size factor return (SMB), the Fama and French value factor return (HML), the excess coskewness factor return (S^-), and the excess cokurtosis factor return (K^+).

Results for the 10 size-sorted portfolios are found in panel B of Table 1. We observe a relationship between $\hat{K}$ and the size decile classification similar to the means and standard deviations. In particular, the smallest decile portfolio has the largest $\hat{K}$ and these measures decrease almost monotonically to the largest decile portfolio. The hypothesis that the kurtosis levels are equivalent to the normal distribution is rejected for all the portfolios at the .01 level, while only five portfolios reject normality based on the $\hat{S}$ measure at the .01 level. We do observe a pattern of $\hat{S}$ across the portfolios, where the small firm deciles have positive $\hat{S}$ measures and the larger firm deciles have negative $\hat{S}$ measures, similar to Harvey and Siddique (2000).

Panel C reports the univariate statistics for the momentum dataset. We find a clear pattern for $\hat{S}$ across the momentum portfolios, where $\hat{S}$ is large and positive for the decile 1 portfolio and decrease monotonically to the large negative $\hat{S}$ of the decile 10 portfolio, also similar to Harvey and Siddique (2000). The $\hat{K}$ measures on the momentum dataset are largest for the decile 1 momentum (recent losers) portfolio, although no significant pattern exists across this sorting.

We also conduct multivariate distribution tests, reported in Table 2. Panel A reports estimates of Mardia's (1970) measure of multivariate skewness while Panels B and C report Mardia's (1970) measure of multivariate kurtosis. We construct Mardia's skewness and kurtosis measures using the returns of size- and momentum-sorted portfolios, as well as the residuals from estimating the linear CAPM on these datasets.[7] Testing the hypothesis that either multivariate skewness or excess multivariate kurtosis is zero, necessary conditions of the normal distribution, is

[7] We are primarily interested in the return distributional tests; however, the HLV methodology can also be motivated by the distribution of the residuals.

Table 2
Multivariate distribution tests

	Size sorted	Momentum sorted
Panel A: Multivariate skewness measures		
Returns	17.86 (.00*)	17.02 (.00*)
Residuals (CAPM)	16.06 (.00*)	15.35 (.00*)
Panel B: Multivariate kurtosis measures		
Returns	31.29 (.00*)	48.99 (.00*)
Residuals (CAPM)	28.54 (.00*)	37.67 (.00*)
Panel C: Multivariate kurtosis measures (EGARCH weighted)		
Returns	35.54 (.00*)	56.65 (.00*)
Residuals (CAPM)	30.33 (.00*)	55.92 (.00*)
Panel D: Elliptical symmetric measures (ES_T)		
Returns	−1.71 (.09)	0.58 (.56)
Residuals (CAPM)	−1.22 (.22)	1.69 (.09)

The test statistic in panel A is Mardia's (1970) multivariate skewness measure, while the test statistics in panels B and C are Mardia's (1970) multivariate kurtosis measure. The test statistic in panel D is Beran's (1979) elliptically symmetric statistic, also discussed in Appendix B. The multivariate skewness measure follows a chi-square distribution, while all other statistics reported in this table have an asymptotic standard normal distribution. *P*-values corresponding to each statistic are located in the parenthesis.

*Indicates a *p*-value less than .005.

rejected at the .01 level in all cases. These results are similar to those in Zhou (1993), where multivariate normality is rejected strongly.

To assess the impact of conditional heteroscedasticity on the kurtosis of unconditional distributions, we construct Mardia's multivariate kurtosis statistic using returns that are weighted by their EGARCH(1,1) estimated conditional standard deviations. We find that sufficient kurtosis is present in the conditional return and residual distributions to reject normality at the .01 level [see also Bollerslev, Chou, and Kroner (1992)].[8]

Some feel inclined to use the above rejections of multivariate normality based on the kurtosis estimates to motivate the claim that returns are distributed elliptically [see Zhou (1993)]. However, these tests cannot make that claim, because the alternative hypothesis does not rule out all elliptical distributions. To test specifically the elliptical assumption we make use of a better specified test.

Beran (1979) proposes a test for multivariate ellipticity based on two properties of elliptical random variables: the elliptical random variable's length is independent of its direction, and the standardized elliptical random variable follows a uniform distribution on the $N-1$-dimensional unit hypersphere. The construction and properties of this test statistic are discussed in Appendix B. The null of this distributional test statistic assumes that the multivariate series follows an elliptical distribution.

[8] We also find kurtosis present in standardized residuals when nonparametric models of conditional volatility in the form of $h(Z_t)$ are estimated.

Consequently the alternative of this test statistic is that the series does not follow an elliptical distribution. The Beran (1979) statistic, denoted as ES_T, has an asymptotic standard normal distribution for the purposes of hypothesis testing.[9]

The Beran statistic may also shed light on the strong rejections of normality based on the multivariate skewness estimates in panel A. The distribution of the two Mardia statistics, under the null, are based on an obviously incorrect assumption of multivariate normality, and consequently may behave different than expected. The Beran statistic, on the other hand, is based on an assumption of ellipticity, an assumption that not only encompasses the normal distribution but is much less restrictive, and consequently allows testing under a much more general "null" hypothesis.

Panel C of Table 2 includes the estimates of Beran's (1979) test statistic. In all but two cases, ellipticity cannot be rejected at the 5% level. The size-sorted returns and momentum-sorted residuals are rejected at the 10% level. By definition, the elliptical assumption will be at least as good in describing returns as the normality assumption; however, this generalization appears to help quite dramatically. On the other hand, if the rejections of normality in Tables 1 and 2 are truly driven by skewness, the Beran (1979) test may suffer from a lack of power. Yet, even if skewness is present in our data, the HLV estimator may dominate OLS/GMM. Excess kurtosis is more prevalent in portfolio returns than skewness (see the respective skewness and kurtosis columns in Table 1), and the HLV methodology should model kurtosis well. Also, Monte Carlo simulations in Section 4 show that the HLV estimator can provide consistent estimates that lead to well-sized test statistics in the presence of skewness. Consequently we interpret these distributional tests as substantial motivation for the HLV estimation methodology.

3. Model Investigation

3.1 Estimation results

This section describes our estimates of various asset pricing models using three different estimation procedures: OLS, GMM, and HLV.[10] We estimate a linear CAPM, quadratic CAPM, cubic CAPM, and the Fama and French three-factor model on our two candidate datasets.

[9] The ES_T statistic is short for elliptical symmetry, a phrase common in the statistics literature and synonymous with ellipticity; see Beran (1979).

[10] The HLV estimation procedure is used to estimate a particular asset pricing model on each portfolio as opposed to estimating the system as a whole. Estimating the parameters using individual equations comes at the cost of ignoring potential higher moment dependencies between portfolios. Preliminary estimations found that the costs of estimating a large system largely dominate the benefits, leading us to our estimation choices.

Table 3
Coefficients and standard errors from asset-pricing model estimations

	Size decile	Linear CAPM								Quadratic CAPM		Cubic CAPM	
		OLS		GMM		HLV		LAD		S^-		K^+	
		α	r_m	α	r_m	α	r_m	α	r_m	OLS	HLV	OLS	HLV
Panel A: size-sorted portfolios													
Point estimate	1	0.463	1.216	0.463	1.216	0.012	1.179	0.104	1.224	0.019	0.053	−0.438	−0.567
Std. error		0.256	0.058	0.278	0.084	0.195	0.044	0.256	0.058	0.219	0.201	0.285	0.263
	2	0.289	1.188	0.289	1.188	0.028	1.155	0.050	1.133	−0.003	−0.129	−0.081	−0.107
		0.185	0.042	0.244	0.066	0.160	0.037	0.185	0.042	0.158	0.141	0.206	0.184
	3	0.211	1.153	0.211	1.153	0.112	1.129	0.031	1.151	0.044	−0.128	−0.128	−0.383
		0.161	0.036	0.221	0.057	0.139	0.032	0.161	0.036	0.138	0.125	0.179	0.162
	4	0.273	1.144	0.273	1.144	0.099	1.139	0.041	1.166	0.036	−0.013	−0.190	−0.222
		0.133	0.030	0.186	0.042	0.121	0.028	0.134	0.030	0.114	0.101	0.148	0.132
	5	0.193	1.115	0.193	1.115	0.137	1.126	0.169	1.146	−0.004	0.050	−0.113	0.002
		0.119	0.027	0.152	0.037	0.107	0.025	0.119	0.027	0.101	0.091	0.132	0.117
	6	0.215	1.111	0.215	1.111	0.140	1.117	0.185	1.121	−0.031	0.029	0.008	−0.050
		0.104	0.024	0.133	0.035	0.095	0.022	0.104	0.024	0.089	0.075	0.116	0.098
	7	0.131	1.101	0.131	1.101	0.114	1.099	0.022	1.104	0.057	0.059	0.019	−0.046
		0.084	0.019	0.111	0.032	0.080	0.018	0.084	0.019	0.072	0.061	0.093	0.079
	8	0.140	1.108	0.140	1.108	0.073	1.079	0.078	1.070	−0.003	−0.072	0.090	0.122
		0.074	0.017	0.081	0.021	0.069	0.016	0.074	0.017	0.063	0.055	0.082	0.071
	9	0.042	1.104	0.042	1.104	0.079	1.046	0.041	1.043	0.142	0.164	0.105	0.119
		0.054	0.012	0.060	0.014	0.051	0.011	0.054	0.012	0.046	0.038	0.059	0.049
	10	−0.073	0.925	−0.073	0.925	−0.048	0.917	−0.067	0.918	0.020	0.007	0.007	0.003
		0.043	0.010	0.055	0.012	0.040	0.009	0.043	0.010	0.037	0.032	0.048	0.042

Table 3
Continued.

	Linear CAPM				Fama and French				Quadratic CAPM		Cubic CAPM	
	OLS		HLV		OLS		HLV		S^-		K^+	
Momentum decile	α	r_m	α	r_m	SMB	HML	SMB	HML	OLS	HLV	OLS	HLV
Panel B: Momentum-sorted portfolios, one-month rebalancing, equal-weighted												
1	−0.518	1.285	−0.858	1.186	1.629	0.464	1.524	0.253	−0.103	−0.842	−1.366	−1.065
	0.306	0.070	0.274	0.063	0.077	0.089	0.073	0.084	0.266	0.238	0.333	0.297
2	−0.310	1.114	−0.268	1.086	1.136	0.353	1.103	0.296	−0.075	−0.328	−0.692	−0.547
	0.197	0.045	0.178	0.041	0.044	0.051	0.040	0.046	0.172	0.156	0.216	0.198
3	−0.012	1.045	−0.190	1.026	0.988	0.341	0.998	0.340	−0.057	−0.196	−0.522	−0.580
	0.164	0.038	0.145	0.033	0.033	0.038	0.029	0.033	0.143	0.127	0.181	0.161
4	0.123	0.999	0.025	0.955	0.865	0.340	0.851	0.337	−0.121	−0.260	−0.405	−0.307
	0.140	0.032	0.124	0.029	0.026	0.030	0.023	0.026	0.122	0.109	0.155	0.138
5	0.195	0.996	0.167	0.994	0.740	0.333	0.715	0.326	−0.081	−0.154	−0.421	−0.429
	0.119	0.027	0.105	0.024	0.020	0.023	0.017	0.020	0.103	0.091	0.130	0.116
6	0.329	0.994	0.367	0.981	0.736	0.320	0.731	0.319	−0.077	−0.159	−0.422	−0.387
	0.115	0.026	0.101	0.023	0.018	0.020	0.015	0.018	0.100	0.088	0.126	0.109
7	0.455	1.008	0.492	1.015	0.703	0.295	0.704	0.309	−0.113	−0.129	−0.453	−0.358
	0.110	0.025	0.097	0.022	0.018	0.020	0.016	0.018	0.096	0.084	0.120	0.106
8	0.516	1.047	0.459	1.061	0.744	0.275	0.730	0.276	−0.092	−0.031	−0.487	−0.356
	0.118	0.027	0.105	0.024	0.020	0.024	0.018	0.020	0.102	0.092	0.128	0.118
9	0.687	1.116	0.719	1.106	0.787	0.204	0.797	0.186	−0.107	−0.107	−0.687	−0.539
	0.132	0.030	0.115	0.027	0.028	0.033	0.026	0.030	0.115	0.101	0.143	0.124
10	0.881	1.271	0.700	1.270	1.036	−0.003	1.096	−0.013	−0.120	−0.151	−0.905	−0.804
	0.181	0.042	0.159	0.037	0.043	0.049	0.037	0.043	0.158	0.140	0.196	0.173

Above are results of estimating various asset pricing models on two datasets of monthly U.S. equity returns. We report intercepts (α) and various slope coefficients, though not always all, of a given asset pricing model in columns below the given model. For example, under the columns listed "Fama and French," we report the coefficients on the SMB and HML factors but exclude the coefficients on the r_m factor and the intercept. Standard errors are listed directly below the point estimates.

Table 3 provides the results from these estimations. In general, we find some striking differences between the point estimates of OLS and GMM versus the HLV method, and we find that the HLV estimation results in more efficient estimates than OLS/GMM in almost all cases.

Estimation results of the linear CAPM on the two datasets lead to some of the greatest differences between parameter values. For example, the size-sorted decile 1 CAPM intercept (α) ranges from an OLS value of .463 to .012 for the HLV estimator, as seen in panel A, a difference in expected return of more than 5.5% annually. In another case, the momentum-sorted decile 1 estimated intercept ranges from an OLS value of $-.518$ to an HLV value of $-.858$, as shown in panel B, leading to a difference in expected return of 4.4% annually. Both of these portfolios, the small firm portfolio and the recent losers portfolio, exhibit substantial nonnormality, as evidenced by their estimates of $\hat{S}$ and $\hat{K}$ in Table 1.

The most plausible explanation for the difference in estimated α is the HLV estimator's "better" treatment of outliers. Intuitively the HLV method first constructs a nonparametric estimate of the residual density, and then uses this density estimate to form the parameter estimates. Because the HLV method allows the data, and not a parametric density, to assign the probability in the tails of the residual distribution, its parameter estimates have a reduced exposure to density misspecification and sensitivity to outliers, unlike OLS.[11] For example, in the size-sorted portfolios we find that one monthly return (January 1975) for the decile 1 portfolio was 57%, while the OLS predicted return was only 17%. We reestimated the linear CAPM using OLS after removing this month and found that $\hat{\alpha}$ declined from .463 to .381, and $\hat{\beta}_m$ also declined from 1.216 to 1.153.[12]

Similarly, for the momentum-sorted portfolios, we found one monthly excess return (January 1992) in decile 1 that was 42%, while the OLS predicted return was -1%. Estimation of the linear CAPM without this observation decreased $\hat{\alpha}$ from $-.518$ to $-.625$. We also observe that the three largest residual outliers of the linear CAPM estimation of the decile 1 portfolio are all January returns (January 1992, January 1975, and January 1967), which also happen to all be positive outliers. These three January returns have the effect of increasing the intercept, and although the linear CAPM is rejected on this portfolio, its performance in greatly overstated by OLS as a result of these outliers. The puzzle associated with

[11] The fact that, in our case, the HLV method places less weight on outliers than OLS in constructing parameter estimates is purely a function of the data and the estimated residual density. The HLV method is agnostic regarding the economic importance of these outliers or "extreme" observations and only suggests OLS places too much statistical weight on these observations in constructing parameter estimates.

[12] We also estimate the CAPM after removing four dates we defined to be outliers (January 1974, January 1975, February 1976, and January 1992) and find that the intercepts decline such that none of the OLS estimated $\hat{\alpha}$ are significantly different from zero for the size-sorted portfolios.

the decile 1 momentum portfolio is larger than one would conclude using OLS methods.

More evidence that tail behavior is influencing the OLS and GMM estimates can be found in our least absolute deviations (LAD) estimation of the linear CAPM on the size-sorted portfolios. LAD is a more robust estimation method than OLS, and we find that the LAD estimates are much closer to the HLV estimates than the OLS counterparts. This is particularly true for the small-firm intercepts, where OLS intercept estimates are large and statistically significant, whereas LAD- and HLV-estimated intercepts are much closer to zero.[13]

We find interesting results for the multifactor model estimations, where the HLV method again leads to markedly different point estimates than OLS. These results are also found in Table 3, where we report the coefficients on SMB and HML from estimation of the Fama and French model; the coefficient on S^- from estimating the quadratic CAPM; and the coefficient on K^+ from our estimation of the cubic CAPM.

We find the largest differences in the point estimates between HLV and OLS for the Fama and French three-factor model to be on the portfolios that are the most nonnormal. The most extreme case is found for the decile 1 momentum portfolio. The estimates of the SMB coefficient range from an OLS value of 1.629 to an HLV value of 1.524, while the HML coefficient ranges from an OLS value of .464 to an HLV value of .253, or a 46% decline in the parameter estimate.

For the quadratic and cubic CAPM estimations, some of the differences again are remarkably large. In the case of size-sorted portfolios, the results of decile 3 change dramatically between OLS and HLV. The S^- coefficient goes from .044 using OLS to $-.128$ for HLV when estimating the quadratic CAPM. Given the average return on the S^- factor, this results in a reduction in the monthly expected return of almost 10 basis points. In addition, the K^+ coefficient declines by almost 200% to the HLV estimate of $-.383$ from the OLS value of $-.128$ when estimating the cubic CAPM, a difference in annual expected returns of almost 2%.

Perhaps the most striking differences between OLS and HLV estimates of multifactor models are found in the results of the momentum portfolios. In particular, the decile 1 portfolio has dramatic differences between point estimates for almost all models tested. In the most extreme case, the coefficient on the quadratic CAPM's S^- declines by more than 700% from $-.103$ for OLS to $-.842$ for HLV, leading to a change in monthly excess returns of more than 40 basis points for the quadratic CAPM. This translates into a 5% difference in annual expected returns between OLS and HLV.

[13] LAD estimation of the linear CAPM on the momentum-sorted data are available on request.

One other interesting result on the momentum-sorted dataset is that the HLV-estimated S^{-} coefficient tends to be greater in magnitude than the OLS-estimated counterpart, while the HLV-estimated K^{+} tends to be smaller in magnitude than the OLS-estimated counterpart. While the results for both estimation methods show that the quadratic and cubic CAPM are valuable factors, this systematic movement in coefficients suggests that the cubic CAPM's marginal value may lie in fitting the outlier returns relative to the quadratic CAPM, whose marginal value increases when these outliers are downweighted.

Broadly speaking, the differences in estimated coefficients seem to be driven by the opposing importance placed on extreme return observations; OLS using outliers to construct parameter estimates more than HLV. Even if one felt particularly tied to OLS, or skeptical of the HLV method, these results indicate how sensitive the estimated coefficients are to the relative weight placed on extreme observations.

In terms of efficiency, we find that the average standard error reduction across the two datasets is 13%. The largest reduction is found in the size-sorted decile 1 (smallest) portfolio, where the standard errors of the HLV-estimated coefficients are 24% (29%) less than their OLS (GMM) counterparts. Although no formal test is conducted, the degree of efficiency gain appears to be proportional to the deviation in the fourth moment from the normality assumption.

3.2 Asset pricing tests

We next test the mean-variance efficiency of our candidate models (linear CAPM, quadratic CAPM, cubic CAPM, and Fama and French three-factor model) on the size and momentum datasets and report these results in Table 4. Using the results from our estimations in Table 3, we test whether the intercepts of all portfolios in a given group are jointly zero. The OLS-constructed J_{OLS} test, from Equation (3), and both of the GMM statistics J_{GMM1} and J_{GMM2}, from Equations (5) and (6), respectively, strongly reject mean-variance efficiency of the market portfolio on the size-sorted portfolios. However, the HLV-constructed J_{HLV}, from Equation (9), fails to reject the CAPM on the size-sorted portfolios. The conflicting conclusions appear to be a result of the influence that a few monthly returns (outliers) have on the intercept coefficients of portfolios. To see how sensitive the results are to these outliers we perform mean-variance tests on the dataset where the returns on the four previously mentioned dates are removed.[14] The changes in conclusions are dramatic. On the revised dataset, OLS fails to reject the CAPM with a J_{OLS} value of

[14] See Footnote 12. We can actually remove just three of these outlier observations (Jan. 1974, Jan. 1975, and Jan. 1992) and still fail to reject the CAPM using the J_{OLS} statistic. In addition, the J statistic for the complete data set using the LAD estimates is 9.43 with a p-value of 0.49.

Table 4
Asset pricing tests

Model	Test statistic	Size sorted	Momentum sorted
r_m	J_{OLS}	27.94 (.00*)	104.15 (.00*)
	$J_{\text{C-OLS}}$	25.74 (.00*)	95.74 (.00*)
	J_{GMM1}	18.95 (.04)	50.81 (.00*)
	J_{GMM2}	40.77 (.00*)	73.75 (.00*)
	J_{HLV}	12.13 (.28)	132.39 (.00*)
r_m, S^-	J_{OLS}	27.94 (.00*)	101.57 (.00*)
	J_{HLV}	36.50 (.00*)	113.45 (.00*)
r_m, S^-,	J_{OLS}	26.47 (.00*)	99.26 (.00*)
K^+	J_{HLV}	40.41 (.00*)	142.33 (.00*)
FF	J_{OLS}	23.90 (.01)	185.60 (.00*)
	J_{HLV}	27.68 (.00*)	228.30 (.00*)

Test statistics above refer to multivariate intercept tests from time-series estimations of candidate asset pricing models listed in the first column. The linear CAPM includes the factor r_m; the quadratic CAPM includes factors r_m and S^-; the cubic CAPM includes factors r_m, S^-, and K^+; and FF stands for the Fama and French three-factor model. Under the null H_0: $\alpha = 0$, the test statistics follow a chi-square distribution with 10 degrees of freedom, and a statistic equal to 18.3 would have a p-value of 0.05.

*Indicates a p-value less than .005.

11.87 and an associated p-value of .29. Conversely the J_{HLV} statistic using this dataset was 14.74 with a p-value of .14, quite similar its value for the original dataset.

These test statistic values support our claim that the HLV procedure will be robust to the distribution-generating returns and less influenced by outliers than OLS and GMM procedures. Given the OLS results, the well-documented January effect appears to be driving the rejection of the CAPM on the size-sorted portfolios [see Keim (1983)]. However, the more general estimation procedures (HLV and LAD) indicate that the January effect isn't large enough to merit a mispricing of the CAPM on the size-sorted portfolios; consequently the robustness of this anomaly appears questionable.

The small value of J_{HLV} on the size-sorted dataset is not intended to resurrect the linear CAPM, which is later rejected on the same dataset, but to highlight problems with conventional asset pricing tests and their sensitivity to distributional assumptions not required by the asset pricing models themselves.

We find on the momentum-sorted datasets that all candidate models are rejected. In all cases, the J_{HLV} statistic is greater than the J_{OLS} statistic, driven primarily by the HLV's efficient estimation, leading to greater power. While the few January returns, discussed in the previous section lead to smaller OLS/GMM-estimated intercepts (α), all three estimation methods still strongly reject the linear CAPM. Even when the linear CAPM is extended to include our proxies for quadratic or cubic factors, or the Fama and French factors, these additional explanatory variables fail to add sufficient explanation to the momentum-sorted portfolio returns for all three estimation methods. We do find, consistent with Harvey and Siddique (2000), that quadratic and cubic CAPM factors perform much better on the momentum portfolios than does the Fama and French three-factor model, as evidenced by their much smaller test statistics. Unlike Zhou's (1993) standard error corrections, which always lead to a smaller test statistic in the presence of leptokurtosis, J_{HLV} in some cases strengthens conclusions made using J_{OLS}, J_{GMM1}, and J_{GMM2}, and in other cases contradicts them.

4. Monte Carlo Results

While J_{HLV} appears to have better properties than the other test statistics (J_{OLS}, $J_{\mathrm{C-OLS}}$, J_{GMM1}, J_{GMM2}), this could be an artifact of our particular empirical investigation. In this section we present a Monte Carlo exercise intended to better assess the properties of the test statistics. We find that the J_{HLV} has substantially greater power than J_{OLS}, $J_{\mathrm{C-OLS}}$, J_{GMM1}, and J_{GMM2} when leptokurtosis is present. We also find that the J_{HLV} has appropriate size in all cases that we investigate, even when the simulated distribution is skewed. Surprisingly we find that J_{GMM1} and J_{GMM2} have quite poor size properties when either kurtosis or the number of portfolios being tested is increased.

We adopt the Davidson and MacKinnon (1998; hereafter DM) graphical method of comparing simulation results through the use of p-value plots and power-size plots. The DM method focuses on the empirical p-value distribution a test statistic generates, which, independent of the particular statistic, should follow a uniform distribution. The DM method also clearly indicates how inputs such as sample size, distribution characteristics, and other factors affect the performance of a given test statistic over a range of p-values. Along with the graphs, we provide two tables with key simulation results.

To conduct the simulation, we first construct a dataset of artificial returns, $\check{\mathbf{r}}$, a $T \times N$ matrix, using the following model: $\check{\mathbf{r}} = \boldsymbol{\alpha} + \hat{\boldsymbol{\beta}} \boldsymbol{r}_m + \check{\boldsymbol{\varepsilon}}$, where T represents the sample size and N represents the number of portfolios. For each simulation we construct both a series of null returns, $\check{\mathbf{r}}^n$, and a series of alternative returns, $\check{\mathbf{r}}^a$, in the following manner. First, we sum the product of r_m, the series of excess returns on the CRSP value-weighted index, and $\hat{\boldsymbol{\beta}}$, an $N \times 1$ vector of parameters taken from portfolio betas in our empirical exercise, where $N = 1$ or 13.

We then add to $\hat{\boldsymbol{\beta}} r_m$ a series of randomly selected residuals, $\check{\boldsymbol{\varepsilon}}$, drawn from some prespecified distribution. We use four residual distributions from which to randomly draw $\check{\boldsymbol{\varepsilon}}$: central $t_{(3)}$, central $t_{(5)}$, normal distribution, and $\chi^2_{(4)}$. The first three are all elliptical distributions with varying degrees of kurtosis; the last distribution is asymmetric, which we include to observe the robustness properties of the J_{HLV} statistic.[15]

Last, we add an intercept, $\boldsymbol{\alpha}$, to the artificial returns, where $\boldsymbol{\alpha} = 0$ to construct $\check{\mathbf{r}}^n$, and $\boldsymbol{\alpha} = .15$ (.075) to construct $\check{\mathbf{r}}^a$ for $N = 1(13)$.[16] We now have two sets of returns, $\check{\mathbf{r}}^n$ and $\check{\mathbf{r}}^a$, and proceed to estimate the linear CAPM using OLS, GMM, and the HLV method on each set of returns. Using the coefficients from a given estimation, we construct the test statistics J_{OLS}, $J_{\mathrm{C-OLS}}$, J_{GMM1}, J_{GMM2}, and J_{HLV}, as well as the p-value under the null for each statistic. We store the various p-values for each simulation and repeat this process 10,000 times.[17] Given that we are investigating five test statistics (J_{OLS}, $J_{\mathrm{C-OLS}}$, J_{GMM1}, J_{GMM2}, and J_{HLV}), this means we store 10 unique p-values for each simulation.

With a series of p-values for a particular test statistic, we can construct the statistic's empirical p-value cdf. Let $\hat{F}(x_i)$ represent the estimated cdf of the p-values generated by a test statistic at the point $x_i \in [0, 1]$. Define $p(\varpi_j)$ to be the p-value generated by the test statistic for simulation j. We calculate $\hat{F}(x_i)$ by using the following formula:

$$\check{F}(x_i) = \frac{1}{W}\sum_{j=1}^{W}\mathbb{I}(p(\varpi_j) > x_i),$$

where W is the number of simulations and $\mathbb{I}(\cdot)$ is an indicator function. We actually form a grid of values, denoted as $\mathbf{X} = \{.001, .002, .003, \ldots, 1\}$, and construct the associated $\hat{F}(\mathbf{X})$ for all 10 different combinations of test statistics and return series ($\check{\mathbf{r}}^n$ and $\check{\mathbf{r}}^a$) over this grid.

To compare size properties, or performance of the statistics under the null, we construct the following graph, entitled a p-value plot. The plot is constructed by graphing our grid, $\mathbf{X}$, on the x-axis versus $\hat{F}(\mathbf{X})$ on the y-axis for the given statistic. A test statistic with appropriate size follows the 45-degree line, since the distribution of p-values from any test statistic is uniform under the null, and the cdf of the uniform distribution is the 45-degree line. When the p-value plot of a statistic lies above (below) the 45-degree line, the associated statistic over- (under-)rejects the null hypothesis.

[15] We use the same residual in constructing both the $\check{\mathbf{r}}^n$ and $\check{\mathbf{r}}^a$ series to isolate the results from randomness introduced by the simulations. Also, a Monte Carlo experiment on elliptically symmetric returns with kurtosis driven by conditional heteroscedasticity is conducted in Hodgson and Vorkink (2002).

[16] The value $\alpha_h = .15$ was chosen by finding the average absolute intercept from the CAPM regressions and rounding to the second digit. The choice for simulating a 13 portfolio series was to match industry portfolio estimations that are not included in this version of the article. The decision to decrease α_h to .075 when $N = 13$ is because all estimation methods reject the alternative using $\alpha_h = .15$ and $N = 13$ more than 95% of the time.

[17] By repeating this process 10,000 times, we are able to obtain two- or three-digit accuracy in the p-values we report.

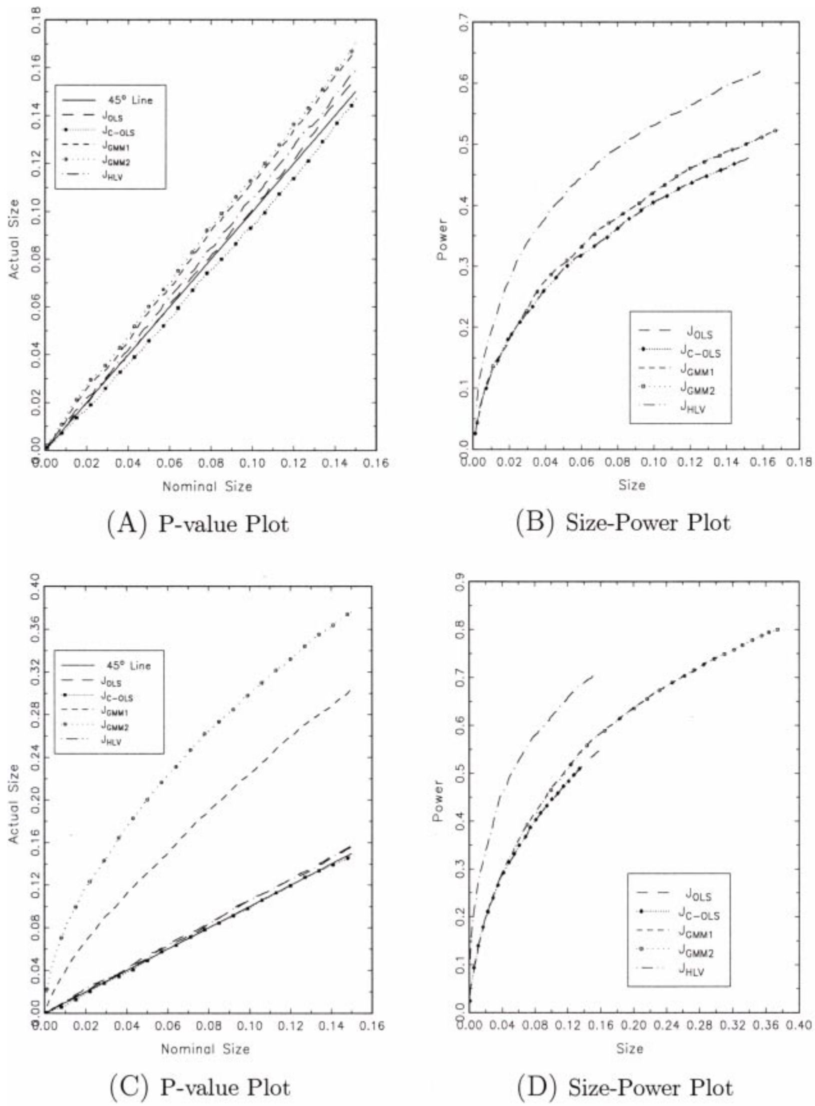

Figure 1
P-value and size-power plots from Monte Carlo simulations with $\varepsilon_t \sim t_{(3)}$, and $T = 250$. (a and b) report simulations with $N = 1$. (c and d) report simulations with $N = 13$.

Figure 1 (a and c) shows the *p*-value plots for all of our candidate test statistics: J_{OLS}, $J_{\mathrm{C-OLS}}$, J_{GMM1}, J_{GMM2}, and J_{HLV}. Five percent and 10% size values are also found in panel A of Table 5. Figure 1a reports the results of simulated data using $T = 250$, a $t_{(3)}$ residual distribution, and $N = 1$, while Figure 1b reports an identical simulation except we

Table 5
Monte carlo results

			5%					10%				
T	N	Dist	J_{OLS}	$J_{\mathrm{C-OLS}}$	J_{GMM1}	J_{GMM2}	J_{HLV}	J_{OLS}	$J_{\mathrm{C-OLS}}$	J_{GMM1}	J_{GMM2}	J_{HLV}
Panel A: Size												
250	1	$t_{(3)}$	0.058	0.049	0.058	0.060	0.050	0.122	0.097	0.112	0.114	0.102
250	1	$t_{(5)}$	0.058	0.050	0.051	0.053	0.050	0.116	0.097	0.105	0.107	0.097
250	1	Normal	0.052	0.051	0.050	0.052	0.051	0.105	0.103	0.103	0.106	0.102
250	1	$\chi^2_{(4)}$	0.057	0.018	0.054	0.056	0.058	0.108	0.044	0.103	0.106	0.115
372	1	$t_{(3)}$	0.057	0.047	0.058	0.060	0.051	0.121	0.097	0.115	0.116	0.106
372	1	$t_{(5)}$	0.058	0.046	0.050	0.052	0.046	0.113	0.097	0.104	0.105	0.098
372	1	Normal	0.051	0.050	0.052	0.054	0.049	0.106	0.099	0.105	0.107	0.099
250	13	$t_{(3)}$	0.060	0.049	0.131	0.200	0.053	0.122	0.086	0.224	0.299	0.105
250	13	$t_{(5)}$	0.059	0.052	0.079	0.130	0.049	0.110	0.103	0.149	0.212	0.098
250	13	Normal	0.052	0.050	0.073	0.114	0.051	0.106	0.104	0.113	0.193	0.102
250	13	$\chi^2_{(4)}$	0.065	0.015	0.092	0.130	0.054	0.127	0.058	0.161	0.212	0.102
Panel B: Power												
250	1	$t_{(3)}$	0.292	0.288	0.303	0.303	0.444	0.400	0.400	0.421	0.421	0.563
250	1	$t_{(5)}$	0.431	0.428	0.456	0.456	0.484	0.561	0.561	0.572	0.571	0.620
250	1	Normal	0.636	0.635	0.657	0.658	0.592	0.749	0.750	0.764	0.764	0.707
250	1	$\chi^2_{(4)}$	0.629	0.627	0.649	0.652	0.619	0.752	0.750	0.774	0.775	0.726
372	1	$t_{(3)}$	0.402	0.399	0.412	0.412	0.601	0.520	0.520	0.523	0.523	0.714
372	1	$t_{(5)}$	0.587	0.586	0.610	0.611	0.655	0.696	0.697	0.717	0.717	0.755
372	1	Normal	0.809	0.808	0.822	0.823	0.783	0.883	0.883	0.890	0.890	0.868
250	13	$t_{(3)}$	0.329	0.330	0.325	0.320	0.493	0.446	0.447	0.470	0.465	0.629
250	13	$t_{(5)}$	0.500	0.499	0.553	0.550	0.582	0.636	0.635	0.677	0.675	0.702
250	13	Normal	0.678	0.677	0.767	0.769	0.652	0.815	0.815	0.859	0.858	0.798
250	13	$\chi^2_{(4)}$	0.682	0.681	0.702	0.693	0.674	0.811	0.784	0.816	0.810	0.786

Above are the results of a Monte Carlo simulation to assess the statistical performances of various asset pricing test statistics. In each simulation we construct a series of artificial returns by first taking the product $\hat{\beta}r_m$, where r_m is the series of excess returns on the CRSP value-weighted index as used in our empirical exercise, $\hat{\beta}$ is a $N \times 1$ vector of betas taken from our OLS estimates in our empirical exercise, and where N is determined in the second column. We then add an intercept (α) to the artificial returns. For the results in panel A, we set $\alpha = 0$; and for the results in panel B, we set $\alpha = 0.15$ if $N = 1$ and $\alpha = 0.075$ if $N = 13$. Last, we add a vector of residuals, $\breve{\varepsilon}$, taken from one of four distributions listed in the third column. On each set of simulated portfolio returns, we estimate the linear CAPM using OLS, GMM, and HLV estimators and then test mean-variance efficiency of the CAPM by constructing Wald test statistics as described in Section 1. We repeat this process 10,000 times and, given these results, construct measures of the size and power performance of the various mean-variance efficiency test statistics. Panel A reports the size performance, or performance under the null, of the various test statistics while panel B reports the power performance, or ability to reject the alternative, of the various test statistics. Individual cells represent p-values of the test statistics for nominal size levels of 5% and 10%.

set $N = 13$. The important region of these graphs is where **X** is between 0 and 0.1, the region of size values most often analyzed in hypothesis testing. For Figure 1a, all statistics are in close proximity to the 45-degree line. The J_{OLS} statistic is slightly oversized, consistent with Zhou (1993). Figure 1 (a and c) also shows that the $J_{\mathrm{C-OLS}}$ statistic tends to overcompensate for the oversized properties of the J_{OLS} statistic, leading to an undersized (or underrejecting) statistic. This overcompensation is consistent with the notion that the multiplicative adjustment factor to $J_{\mathrm{C-OLS}}$ found in Equation (7) is intended to correct for conditional heteroscedastic-induced kurtosis, not found in our simulations.

One other interesting aspect of the Figure 1 graphs is the size performance of the J_{GMM1} and J_{GMM2} statistics. For the cases where $N = 1$ (Figure 1a), both the GMM-based statistics are reasonably well sized. However, as either the kurtosis is increased or the number of portfolios in the system is increased, both J_{GMM1} and J_{GMM2} become oversized. For example, J_{GMM2} (J_{GMM1}) rejects the null 20% (more than 10%) of the time at a nominal level of 5% for the simulation with $T = 250$, $t_{(3)}$ residual, and $N = 13$, as seen in panel A of Table 5. The GMM test statistics are also oversized when the residual distribution is simulated as normal or $\chi^2_{(4)}$.

These GMM results appear surprising, given the minimal restrictions GMM places on the econometric model. However, others have found similar results in related studies. Altonji and Segal (1996) find that univariate GMM-based Wald tests are oversized in simulations where non-normal leptokurtotic distributions are used. Burnside and Eichenbaum (1996) find that GMM-based Wald tests are oversized and that the "size increases uniformly with the dimensionality of joint hypotheses" when residual distributions are simulated as normal distributions [see also Ferson and Foerster (1994)]. Burnside and Eichenbaum conclude that the difficulty associated with the estimation of the optimal weighting matrix W often results in underestimating the variance-covariance matrix of the parameters.[18]

Although both J_{GMM1} and J_{GMM2} appear to have size problems, the problem is more severe with the J_{GMM2} statistic. One possible explanation for the oversized property of J_{GMM2} can be found in Gallant and Jorgensen (1979). In their derivation of the J_{GMM2} statistic, the assumption of normality is required to obtain the asymptotic chi-square distribution of the test statistic. In fact, Gallant and Jorgensen (1979) suggest using transformations such as Box and Cox (1964) to normalize the data prior to construction of J_{GMM2}, a step rarely performed in practice.

To examine the power of a test statistic, or the test statistic's ability to reject a given alternative, we construct a graph entitled power-size plot. The power-size plot graphs the statistic's $\hat{F}^a(\mathbf{X})$ on the y-axis against the same statistic's $\hat{F}^n(\mathbf{X})$ on the x-axis. When this line is plotted for a specific test statistic, any size problems are removed by using the actual size, or $\hat{F}^n(\mathbf{X})$, along the x-axis and not the nominal size, $\mathbf{X}$. Consequently when we compare the power performance of two different test statistics, differences in power have been adjusted for any statistic's inclination to over- or underreject the hypothesis under the null. A statistic will have greater

[18] In our simulations, we use the White (1980) method to estimate W, given that no conditional heteroscedasticity was introduced in the residuals. However, in unreported simulations, we used methods such as Newey and West (1987) and Andrews (1991) and found that these methods led to an even greater size problem with J_{GMM1} and J_{GMM2}.

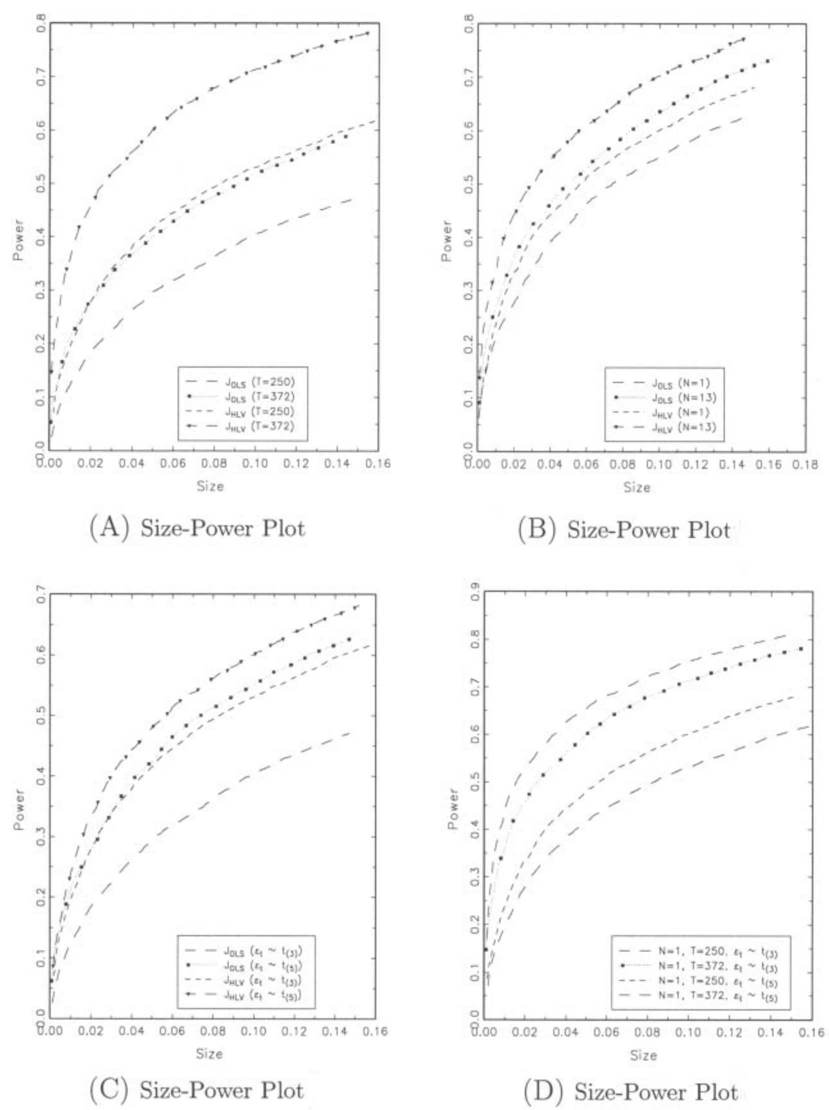

Figure 2
Various size-power plots from Monte Carlo exercise. (a) Reports simulations with $\varepsilon_t \sim t_{(3)}$ and $N=1$. (b) Reports simulations with $\varepsilon_t \sim t_{(5)}$ and $T=250$. (c) Reports simulations with $T=250$ and $N=1$. (d) Reports results from simulations of the J_{HLV} statistic.

power, or reject the alternative more often, if its power-size plot is located above the power-size plot of a competing statistic.

Figures 1(b and d) and 2(a–d) are representative power-size plots. Figure 1(b and d) illustrates the power-size plots for J_{OLS}, $J_{\mathrm{C-OLS}}$, J_{GMM1}, J_{GMM2}, and J_{HLV} statistics, while Figure 2(a–d) shows power-size plots for only the J_{OLS} and J_{HLV} statistics. Five percent and 10% levels of the power curves are found in panel B of Table 5 as well. In general, these graphs indicate that J_{HLV} has the greatest power to reject the alternative in almost all cases. J_{HLV}'s improvement over the competing statistics appears to be a function of sample size (T), the number of portfolios (N), and the degree of excess kurtosis.[19]

In Figure 2a, a larger sample size is seen to improve the relative performance of J_{HLV}. This improvement in power is likely due to the nonparametric component of the HLV estimation method. Figure 2b illustrates that increasing N, or the number of portfolios, also increases the relative performance of J_{HLV}. However, this improvement in power appears to be smaller in magnitude than increasing T. As seen in Figure 2c, increasing kurtosis leads to greater relative power of J_{HLV}. Figure 2d shows how the combination of residual distribution and sample size influence J_{HLV}. From this graph we see that sample size again appears to play the dominant role in the performance of J_{HLV}.

Results in panel B of Table 5 also demonstrate the superior power properties of J_{HLV}. For example, in the case of $T = 250$, a $t_{(3)}$ residual, and $N = 1$, J_{HLV} rejects the alternative hypothesis ($\alpha = .15$) more than 16% more often than J_{OLS} and J_{C-OLS} at a size level of 5%. Increasing T to 372 (keeping a $t_{(3)}$ residual and $N=1$), J_{HLV} rejects the alternative hypothesis more than 20% more often than J_{OLS} and J_{C-OLS}. J_{HLV}'s improvement increases to 18%, for the simulation with $T = 250$, a $t_{(3)}$ residual, and $N = 13$, again at a size level of 5%. The simulations using a $t_{(5)}$ residual distribution show parallel but smaller improvements.

We find that the relative power of J_{HLV} remains remarkably strong even when the residuals are drawn from either a normal distribution or a skewed $\chi^2_{(4)}$ distribution, as evidenced in panel B of Table 5. While in both cases J_{HLV} exhibits the lowest power of all competing statistics, the loss in power is quite small relative to OLS/GMM in these cases.

In some cases the GMM-based J_{GMM1} and J_{GMM2} statistics are seen to have greater power than the J_{OLS} and J_{C-OLS} statistics in the presence of leptokurtosis, as seen in Figure 1(b and d). In fact, Table 5 shows that, under normality, the J_{GMM1} and J_{GMM2} statistics have the greatest power; however, these improvements seem to be of small benefit given their frequent problems with overrejecting.

Clearly the simulations indicate that J_{HLV} is well sized for all distributions tested and has superior power properties to J_{OLS}, J_{GMM1}, and J_{GMM2} in the presence of symmetric kurtosis. Simulations with richer frameworks (i.e., conditional heteroscedasticity in returns and/or

[19] The fact that in all the graphs both J_{OLS} and J_{C-OLS} have nearly identical power curves should not be troubling. In fact, because this graphing exercise corrects for any size problems, these statistics, by definition, should have identical power curves.

residuals, different forms of skewness, etc.) in the future should add more to the picture of the HLV method's performance relative to alternative estimation and testing methods.

5. Conclusion

The distribution of most financial asset returns are known to exhibit kurtosis in excess of the normal distribution. Yet estimation and testing of the CAPM and other return models has largely relied on the assumption of multivariate normality. Using the method developed in Hodgson, Linton, and Vorkink (2002), based on the general assumption of elliptical return distributions, we investigate the performance of numerous asset pricing models on two recognized monthly return datasets: size-sorted and momentum-sorted portfolio stock returns. We start by finding support for the elliptical assumption as a description of stock returns based on Beran's (1979) test statistic.

In the course of our investigation, we find that, contrary to the OLS and GMM estimators, the Hodgson, Linton, and Vorkink (2002) estimator fails to reject the linear CAPM on the group of size-sorted portfolios. We find that the OLS/GMM rejection of the CAPM is driven by sensitivity to outliers in the size-sorted data. Outliers in the momentum-sorted dataset, primarily related to January returns, also cause the OLS-estimated mispricing associated with the linear CAPM to be understated. Consequently the puzzle associated with momentum may be larger than previously believed.

Whereas the Hodgson, Linton, and Vorkink (2002) estimation and testing methodology will likely be supplanted by even more optimal econometric methods, particularly ones that can efficiently incorporate both skewness and kurtosis, this research does illustrate some valuable insights. We observe how sensitive asset pricing estimations and tests can be to the data and chosen econometric methodology. In particular, we observe how outliers can, and do, play a dominant role in the conclusions of these tests. For both of the datasets analyzed in this research, and most likely all equity return datasets, this better-specified econometric method appears to add meaningfully to asset pricing investigations.

Appendix A

In this appendix we discuss how to construct the adaptive HLV estimator ($\tilde{\theta}$). We refer the reader to HLV for a comprehensive discussion of the construction and asymptotic properties of the estimator.[20]

[20] Gauss code for implementing the HLV estimator can be found at the following web address: http://marriottschool.byu.edu/emp/tsp/adcapm.html.

Our general approach to estimating $\tilde{\theta}$ will involve using a Newton–Raphson step as follows:

$$\tilde{\theta} = \theta - T^{-\frac{1}{2}}\,\hat{\mathcal{I}}_T^{-1}\hat{\Delta}_T, \tag{14}$$

where $\hat{\theta}$ is a preliminary estimator and $\hat{\Delta}_T$ and $\hat{\mathcal{I}}_T$ are estimates of the first and second standardized derivatives of the likelihood function $L_T(\theta) = \sum_{t=1}^{T} \ln p(r_{i,t} - \alpha_i + \boldsymbol{\beta_i}\mathbf{F_i})$, respectively. Given the linear model, we can rewrite $\hat{\Delta}_T$ as follows:

$$\hat{\Delta}_T(\hat{\theta}) = T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{Z}_t^{\mathrm{T}}\hat{\varphi}_t(\hat{u}_{i,t}),$$

where $\hat{\varphi}_t(\hat{u}_{i,t})$ is a consistent estimator of the score vector $\varphi(u_t)$ and $\mathbf{Z}_t = \{1, \mathbf{F}_t\}$. We provide an algorithm for computing nonparametric estimates of these quantities ($\hat{\Delta}_T$, $\hat{\mathcal{I}}_T^{-1}$) while imposing the restriction that the errors $\{u_{i,t}\}$ have an elliptical distribution. Recall from Section 1 that we can state the elliptical density $p(u)$ to follow the function $g(u^{\mathrm{T}}\Sigma^{-1}u) = g(\varepsilon^{\mathrm{T}}\varepsilon)$, where $\varepsilon = \Sigma^{-1/2}u$ is a spherically symmetric random variable with density $f(\varepsilon) = g(\varepsilon^{\mathrm{T}}\varepsilon) = g(v)$ with $v = \varepsilon^{\mathrm{T}}\varepsilon$. We will actually construct an estimate of $p(u)$ indirectly, from directly estimating the density of the random variable v. From Muirhead (1982), the density of v, which we shall denote $h(v)$, is

$$h(v) = c_N v^{1/2-1} g(v),$$

where $c_N = \pi^{1/2}/\Gamma(1/2)$.

For computational reasons we directly estimate the density of the random variable $y = \tau(v)$, where $\tau(v) = (v^{\zeta} - 1)/\zeta$ and $\zeta = 1/2$.[21] We use direct kernel estimates of the density of y, denoted $\gamma(y)$, to indirectly obtain consistent estimates of the score and information of $p(u)$. Our algorithm for estimating φ and $\mathcal{I}$ proceeds according to the following steps.

First, obtain $\hat{\theta}$ (by OLS, for example) and construct the associated residuals $\{\hat{u}_{i,t}\}_{t=1}^{T}$ and standardized residuals $\{\hat{\varepsilon}_t\}_{t=1}^{T}$, where $\hat{\varepsilon}_t = \hat{\Sigma}^{-1/2}\hat{u}_{i,t}$, $\hat{\Sigma}_t = \hat{c}^{-1}\hat{\Sigma}_{u_i}$, $\hat{\Sigma}_{u_i} = (T-k)^{-1}\Sigma_{t=1}^{T}\hat{u}_{i,t}\hat{u}_{i,t}$, and $\hat{c} = \det\hat{\Sigma}_{u_i}$. Then compute the transformed sequence $\{\hat{y}_t\}_{t=1}^{T}$, where $\hat{y}_t = \tau(\hat{v}_t)$ with $\hat{v} = \hat{\varepsilon}_t^{T}\hat{\varepsilon}_t$.

Following the construction of the series $\{\hat{y}_t\}_{t=1}^{T}$, we then estimate $\gamma_t(y)$ and $\gamma_t'(y)$ using nonparametric methods in the following manner:

$$\hat{\gamma}_t(y) = \frac{1}{T-1}\sum_{\substack{s=1\\ s\neq t}}^{T} K_{bw_T}(y - \hat{y}_s); \quad \hat{\gamma}_t'(y) = \frac{1}{T-1}\sum_{\substack{s=1\\ s\neq t}}^{T} K'_{bw_T}(y - \hat{y}_s),$$

where $K_{bw_T}(\cdot)$ represents a kernel function with bandwidth parameter bw_T.

The next step involves constructing the indirect estimates of the score and information of $p(u)$ as follows:

$$\hat{\varphi}_t(\hat{u}_{i,t}) = \hat{\Sigma}^{-1/2}\hat{\varepsilon}_t\left[s(\hat{v}_t) + \tau'(\hat{v}_t)\frac{\hat{\gamma}_t'}{\hat{\gamma}_t}(\hat{y}_t)\right],$$

where $s(v) = (1/2)v^{-1} - (J_\tau'/J_\tau)\{\tau(v)\}\tau'(v)$ and $J_\tau(z) = |\partial\tau^{-1}(z)/\partial y|$.[22] We then construct the score and information estimators for θ as

$$\hat{\Delta}_T(\hat{\theta}) = -T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{Z}_t^{\mathrm{T}}\hat{\varphi}_t(\hat{u}_{i,t}), \tag{15}$$

$$\hat{\mathcal{I}}_T(\hat{\theta}) = T^{-\frac{1}{2}}\sum_{t=1}^{T}\mathbf{Z}_t^{\mathrm{T}}\hat{\Omega}_p\mathbf{Z}_t T^{-\frac{1}{2}} \tag{16}$$

[21] HLV discusses the transformation $y = \tau(v) = (v^{\zeta} - 1)/\zeta$ and also provides a motivation for our choice of $\zeta = 1/2$.

[22] HLV discuss trimming conditions that the theory requires for estimation. In practice, little or no trimming is required for the estimation method to be well behaved.

where $\hat{\Omega}_p = (1/T)\sum_{t=1}^{T}\hat{\varphi}_t(\hat{u}_{i,t})\hat{\varphi}_t(\hat{u}_{i,t})^{\mathrm{T}}$. With the estimates $\hat{\Delta}_T(\hat{\theta})$ and $\hat{\mathcal{I}}_T(\hat{\theta})$ in hand, compute the adaptive estimator $\tilde{\theta}$ given in Equation (14).

We use Silverman's (1986) rule-of-thumb bandwidth to calculate the optimal bandwidth for the density γ estimation and find the optimal value to be approximately $4.25T^{-1/5}$ when using the bi-quartic kernel,

$$K(u) = \frac{15}{16}(1 - u^2)^2\mathbb{I}\{|\mathrm{u}| \leq 1\},$$

the kernel used in our nonparametric estimations. We find that the analogous optimal bandwidth for γ' is $4.0T^{-1/7}$. These bandwidths are optimal in terms of minimizing the mean integrated square error of the density estimate. We found our results to be quite robust to the particular value of the bandwidth choice, although all of the estimations and simulations reported use the proposed bandwidth values.

Appendix B

In this appendix we describe a test for ellipticity developed by Beran (1979) and discuss its implementation. We use $\|A\| = [\mathrm{tr}(A^{\mathrm{T}}A)]^{1/2}$ to denote the euclidean norm of a vector or matrix A. Suppose we have a series of the standardized regression residuals $\hat{\varepsilon}_t = \hat{\Sigma}^{-1/2}\hat{u}_t$ for $t = 1, \ldots, T$. These residuals will be used in the construction of a test of the null hypothesis that the true $\{u_t\}$ are draws from an elliptical density which imply that the true $\{\varepsilon_t\}$ are *i.i.d.* draws from a spherically symmetric density. The test utilizes a couple of unique features of spherically symmetric random variables. The first is that the standardized random variable $\varepsilon_t/\|\varepsilon_t\|$. is uniformly distributed on the $N-1$-dimensional unit hypersphere. The second is that this standardized random vector is independent of the vector's "length," viz. $\|\varepsilon_t\|$.

In the following we describe the construction of the test, provide some intuition as to how it incorporates the aforementioned characteristics, and then state the test's asymptotic distribution under the null.

The first step involves ranking the distances $\{||\hat{\varepsilon}_t||\}_{t=1}^{T}$ and then dividing these ranks by $T+1$. Let $\{R_t\}_{t=1}^{T}$ denote these ranks. The directional vector $\varepsilon/\|\varepsilon\|$ can be represented in terms of its polar coordinates $\Xi = (\xi_1, \ldots, \xi_{N-1})$ as follows:

$$\frac{\varepsilon}{||\varepsilon||} = (\cos(\xi_1), \sin(\xi_1)\cos(\xi_2), \ldots, \sin(\xi_1)\sin(\xi_2), \ldots, \sin(\xi_{N-1})).$$

We then define the coordinates of $\hat{\varepsilon}_t/||\hat{\varepsilon}_t||$ by Ξ_t.

Let $\{a_l : l \geq 1\}$ be the family of functions orthonormal with respect to the Lebesgue measure on [0, 1] and orthogonal to the constant function on [0, 1]. Furthermore, let $\{b_\ell : m \geq 1\}$ denote another family of orthonormal functions with respect to the uniform measure on $[0, \pi]^{p-2} \times [0, 2\pi)$ and orthogonal to the constant function on this domain (we use Legendre's polynomials as described in HLV). Beran (1979) proposed a statistic of the form

$$S_T = \sum_{l=1}^{L_T}\sum_{m=1}^{M_T}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}a_l(R_t)b_m(\Xi_t)\right]^2. \tag{17}$$

Using the fact that, under the null, R_t and Ξ_t both have uniform distributions, it follows from our assumptions on $\{a_l\}$ and $\{b_m\}$ that $E[a_l] = E[b_m] = 0$ for all l, m. Given that R_t and Ξ_t are independent under the null implies that $E[a_l(R_t)b_m(\Xi_t)] = 0$ for all l, m. If $\{\varepsilon_t\}$ follows a spherically symmetric distribution, then S_T should be close to zero; otherwise it should be far from zero. The following proposition gives the asymptotic distribution of S_T.

Proposition 1. [Beran (1979)] *Suppose that the functions* $\{a_k, b_\ell : k, \ell \geq 1\}$ *satisfy certain conditions contained in Beran (1979). Then the null limiting distribution of* $(2L_TM_T)^{-\frac{1}{2}}[S_T - L_TM_T]$ *as* $T \to \infty$ *is* $N(0, 1)$.

In our empirical tests we set $L_T = M_T = 7$. The statistic discussed in Section 2 and listed in Table 3 (ES_T) is constructed as

$$ES_T = (2L_T M_T)^{-\frac{1}{2}}[S_T - L_T M_T], \tag{18}$$

and from the proposition above follows an asymptotic standard normal distribution.

References

Affleck-Graves, J., and B. McDonald, 1989, "Nonnormalities and Tests of Asset-Pricing Theories," *Journal of Finance*, 44, 889–908.

Altonji, J., and L. Segal, 1996, "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353–366.

Andrews, D., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Beran, R., 1979, "Testing for Ellipsoidal Symmetry of a Multivariate Density," *Annals of Statistics*, 7, 150–162.

Berk, J, 1997, "Necessary Conditions for the CAPM," *Journal of Economic Theory*, 73, 245–257.

Bollerslev, T., R. Chou, and K. Kroner, 1992, "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

Box, G., and D. Cox., 1964, "An Analysis of Transformations," *Journal of the Royal Statistical Society Series B* 211–264.

Burnside, C., and M. Eichenbaum, 1996, "Small-Sample Properties of GMM-Based Wald Tests," *Journal of Business & Economic Statistics*, 14, 294–308.

Campbell, J., W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Chamberlain, G., 1983, "A Characterization of the Distributions that Imply Mean-Variance Utility Functions," *Journal of Economic Theory*, 29, 185–201.

Davidson, R., and J., MacKinnon, 1994, "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *The Manchester School*, 66, 1–26.

Dittmar, R., 2002, "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross-Section of Equity Returns," *Journal of Finance*, 57, 369–403.

Fama, E., 1965, "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34–105.

Fama, E., and K. French, 1992, "The Cross-Section of Expected Returns," *Journal of Finance*, 47, 427–465.

Fama, E., and K. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Ferson, W., and S. Foerster, 1994, "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29–55.

Gallant, A. R., and D. W. Jorgenson, 1979, "Statistical Inference for a System of Simultaneous, Non-Linear, Implicit Equations in the Context of Instrumental Variable Estimation," *Journal of Econometrics*, 11, 275–302.

Géczy, C., 1998, "Some Generalized Tests of Mean-Variance Efficiency and Performance," working paper, University of Pennsylvania.

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

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Harvey, C., 2002, "The Specification of Conditional Expectations," forthcoming in *Journal of Empirical Finance*.

Harvey, C., and A. Siddique, 2000, "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 40, 1263–1295.

Harvey, C., and G. Zhou, 1993, "International Asset Pricing with Alternative Distributional Specifications," *Journal of Empirical Finance*, 1, 107–131.

Hodgson, D., O. Linton, and K. Vorkink, 2002, "Testing the Capital Asset Pricing Model Efficiently under Elliptical Symmetry: A Semiparametric Approach," forthcoming in *Journal of Applied Econometrics*.

Hodgson, D., and K. Vorkink, 2002, "Efficient Estimation of Conditional Asset Pricing Models," forthcoming in *Journal of Business and Economic Statistics*.

Keim, D., 1983, "Size Related Anomalies and Stock Return Seasonality: Further Empirical Evidence," *Journal of Financial Economics*, 12, 13–32.

Knez, P., and M. Ready, 1997, "On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions," *Journal of Finance*, 52, 1355–1382.

Lintner, J., 1965, "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

MacKinlay, A. C., and M. Richardson, 1991, "Using Generalized Method of Moments to Test Mean-Variance Efficiency," *Journal of Finance*, 46, 511–527.

Mandlebrot, B., 1963, "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

Mardia, K., 1970, "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika*, 57, 519–530.

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

Newey, W., and K. West, 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Owen, J., and R. Rabinovitch, 1983, "On the Class of Elliptical Distributions and Their Applications to the Theory of Portfolio Choice," *Journal of Finance*, 38, 745–752.

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

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

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

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

White, H., 1994, *Estimation, Inference, and Specification Analysis*, Cambridge University Press, New York.

Zhou, G., 1993, "Asset Pricing Tests Under Alternative Distributions," *Journal of Finance*, 48, 1927–1942.