# Modeling Non-normality Using Multivariate $t$: Implications for Asset Pricing

Raymond Kan and Guofu Zhou[1]

This version: February, 2006

[1]Kan is from the University of Toronto, Zhou is from Washington University in St. Louis. We are grateful for helpful discussions and comments of Giovanni Barone-Adesi, Martijn Cremers, Anna Dodonova, Heber Farnsworth, Wayne Ferson, Kenneth French, William Goetzmann, Campbell Harvey, Yongmiao Hong, Ravi Jagannathan, Christopher Jones, Lynda Khalaf, Bruce Lehmann, Andrew Lo, Ľuboš Pástor, Jay Shanken, seminar participants at Vanderbilt University and Washington University in St. Louis, and participants at the 2003 Northern Finance Meetings. Kan gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. We also thank the outstanding research assistance of Jun Tu.

# Modeling Non-normality Using Multivariate $t$: Implications for Asset Pricing


## ABSTRACT

Many important findings in empirical finance are based on the normality assumption, but this assumption is firmly rejected by the data due to fat tails of asset returns. In this paper, we propose the use of a multivariate $t$-distribution as a simple and powerful tool to examine the robustness of results that are based on the normality assumption. In particular, we find that, after replacing the normality assumption with a reasonable $t$-distribution, the most efficient estimator of the expected return of an asset is drastically different from the sample average return. For example, the annual difference in the estimated expected returns under normal and $t$ is 2.964% for the Fama and French's (1993, 1996) smallest size and book-to-market portfolio. In addition, there are also substantial differences in estimating Jensen's alphas, choosing optimal portfolios, and testing asset pricing models when returns follow a multivariate $t$-distribution instead of a multivariate normal.

Ever since Fama (1965), Affleck-Graves and McDonald (1989), Richardson and Smith (1993), and Dufour, Khalaf and Beaulieu (2003), among others, there is strong evidence that stock returns do not follow a normal distribution. Despite this, the normality assumption is still the working assumption of mainstream finance. The reason for the wide use of the normality assumption is not because it models financial data well, but due to its tractability that allows interesting economic questions to be asked and answered without substantial technical impediments. Thus, many important findings in empirical finance are based on the normality assumption. The question is that whether these findings are robust to alternative multivariate distributional assumptions.

The multivariate $t$ distribution is a well known alternative to the normal. In the econometrics literature, Chib, Tiwari and Jammalamadaka (1988), Van Praag and Wesselman (1989) and Osiewalski and Steel (1993), among others, provide attractive methodologies for analyzing elliptical models, of which $t$ is a special case. However, to our knowledge, the literature focuses on model errors, not the joint distribution of all the regression variables. In finance, it is exactly this joint modeling that is of interest. As it turns out, both parameter estimation and the associated asymptotic theory are substantially different from the usual case of modeling model errors.

In this paper, we advocate the use of a multivariate $t$-distribution to model jointly the stock returns, develop the associated asymptotic theory and examine asymptotic theory, and examine the robustness of some of the major empirical results that are based on the normality assumption. There are three major reasons for the use of a $t$-distribution.[1] First, it models financial data well in many circumstances. Theoretically, the $t$-distribution nests the normal as a special case, but it captures the observed fat tails of financial data. For example, the multivariate normality assumption of the joint distribution of Fama and French's (1993) 25 assets returns and their 3 factors from January 1963 to December 2002 is unequivocally rejected by a kurtosis test with a $p$-value of less than 0.01%. On the other hand, such a test for a multivariate $t$-distribution with 8 degrees of freedom has a $p$-value of 16.84%. Although the $t$-distribution is symmetric, its sample skewness is highly volatile that can generate the skewness of the data with high probability. Second, with the algorithms provided here, the $t$-distribution has become almost as tractable as the normal one. Traditionally, non-normal distributions, such as the $t$, do not yield easy parameter estimation,

[1]Blattberg and Gonedes (1974) seems the first to use $t$ to model stock returns in finance. Later applications of $t$ and generalized $t$ in the univariate case can be found in Theodossiou (1998) and references therein. Although MacKinlay and Richardson (1991), Zhou (1993), and Geczy (2001) use multivariate $t$, their analysis focus on how results under normality vary when under $t$ without providing the results estimated based on the $t$ assumption.

making their use limited to low dimensional problems. As a result, the normal distribution has been almost the only choice in analyzing a large number of assets due to its analytical formulas for parameter estimates. However, this is no longer a decisive advantage of the normal. Owing to the path breaking EM algorithm of Dempster, Laird and Rubin (1977), and especially Liu and Rubin (1995), explicit iterative formulas are available to obtain fast and monotonically convergent parameter estimates under the $t$. The third reason supporting the use of a $t$-distribution is that asset pricing theories that are valid under normality are usually also valid under $t$. For example, the well-known Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) is still valid under $t$ (see Chamberlain (1983) and Owen and Rabinovitch (1983)).

Comparing with the normal distribution, the $t$ adds only one more parameter. Nevertheless, this parsimonious extension allows us to capture a salient feature of the return data (i.e., the fat tails). Admittedly, the $t$-distribution does not describe all the features of the return data like time-varying volatility for which the well-known GRACH models are very useful.[2] However, there is little evidence of GARCH effects on the Fama and French data, or monthly data in general that are typically used for asset pricing and corporate studies. Moreover, the GRACH models require difficult numerical optimization to obtain the estimated parameters, which usually limits applications to no more than ten assets (see, e.g., Bollerslev, 2001). In contrast, there are 28 assets and over 400 parameters in our later applications. While it is impossible for us to solve the optimization problem in the GARCH framework, the EM analytical iterations under $t$ take less than a minute to find the solutions. Hence, the key advantage of the $t$ is its tractability, the same reason for the wide use of normality. It should also be pointed out that the widely used generalized method of moments (GMM) estimators of Hansen (1982) allows for a much more general distributional assumption than the normal. However, the GMM estimators of important parameters, such as the expected asset returns, alphas and betas, are the same as the those obtained under the normality assumption, except that the GMM standard errors are enlarged to account for non-normality. In contrast, the EM algorithm here provides the asymptotically most efficient estimates when the data is $t$ distributed.

Assuming that asset returns are $t$ rather than normally distributed, we find that our understanding of certain major issues in finance is drastically altered. First, there is a substantial and

[2] Moreover, GARCH models that are based on normal errors are still unable to explain the fat tails of the return data. Hence, recent work extends GARCH to allow $t$ errors (see, e.g., Alexander, 2002, and the references therein).

economically important difference in estimating expected returns of assets. For example, the expected excess return for Fama and French's (1993) SMB factor is 0.210%/month when estimated under normality, but is only 0.127%/month when estimated under $t$ with 8 degrees of freedom, implying an annual difference of 0.996%. This difference is of significant economic importance in estimating the cost of capital. Moreover, such differences are even larger for some of the 25 portfolios used by Fama and French (1993). For instance, the annual difference in estimated expected return is 2.964% for the portfolio that is in the smallest size and book-to-market quintiles. In fact, over our sample period, the estimated expected returns on all of the Fama and French's (1993) 25 portfolios are lower under $t$ than under normality. The intuition is that the normality assumption suggests using a sample average return which has equal weights on the observations in estimating the expected return. In contrast, the estimator under $t$-distribution assigns less weight to data points that are far away from the center, so the estimated expected return can be substantially different from the sample average in the presence of fat tails. Indeed, the returns over the months that are considered to be outliers during our sample period tend to have more positive returns than negative ones. Assigning less weights to these outliers results in a shift of the estimated mean leftward.[3] However, the standard deviations of the asset returns estimated under either normality or $t$ are fairly close. This suggests the estimation of the mean is more sensitive to fat tails of the data, consistent with the conventional wisdom that estimating asset standard deviation is easier than estimating its mean.

Second, estimation of Jensen's alpha relies critically upon the distributional assumption on the data. In finance, if both the asset returns and the factors are random and jointly $t$, the regression model residuals must be conditionally heteroskedastic, a case not studied in the econometrics and statistics literature on $t$ distributions. We develop both the estimation technique and the associated asymptotic theory, and apply them to examine both the Fama-French and a set of mutual fund data. We find that some alphas of the Fama-French portfolios can substantially change once the normality assumption is replaced by a suitable $t$. With the mutual fund data, the performance ranking of a mutual fund can change drastically under normal versus under $t$. In some cases, a loser fund with an estimated alpha of $-0.556$%/month under normality becomes a winner fund with estimated alpha of 0.378%/month under an optimally estimated $t$-distribution. On the other

[3]The shifting of the estimated mean to the left is specific to the sample period because it is possible that there might be more negative outliers than positive ones over another subperiod to result in the shifting rightward.

hand, a winner fund with an estimated alpha of 0.528%/month under normality turns into a loser with an estimated alpha of $-0.404\%$/month when estimated under the $t$.

Third, the $t$-distribution sheds new insights in testing asset pricing models. Due to strong rejection of the underlying normality assumption, one should be cautious in interpreting the results from the well-known Gibbons, Ross, and Shanken (1989, GRS) test that relies on the normality assumption. Indeed, MacKinlay and Richardson (1991) and Geczy (2001) both suggest that the GRS test statistic should be reduced to reflect the fat tails of the data. However, this reduction of the GRS test statistics comes at a cost. Namely, the test has lower power after the adjustment. We propose using a likelihood ratio test of the asset pricing restrictions that are based on the multivariate $t$-distribution. Interestingly, we find that there are indeed cases where the GRS or adjusted GRS test fail to reject, while our test based on the $t$-distribution does. In the case where all test results agree, it is still interesting to know the robustness of the conclusion because the data behaves more like $t$ than the normal. This suggests that non-normality modeling by using the $t$ helps us not only in obtaining better estimates of asset expected returns, but also in providing more powerful and reliable tests of asset pricing restrictions.

The remainder of the paper is organized as follows. The next section provides first the empirical evidence on the necessity of modeling the data as $t$ distributed rather than the normal. Section 2 presents both the estimation technique under $t$ and a comparison of the results with those obtained under normality. Section 3 discusses how performance evaluation of mutual funds differs under the normal and $t$ assumptions. Section 4 assesses asset pricing implications of the $t$-distribution. Section 5 discusses some general issues and extensions. Section 6 concludes.

## 1. Why Multivariate $t$?

In this section, we provide a description of the return data that we use, followed by a formal test of both univariate and multivariate normality, and a test of GARCH effects. The empirical results show that the multivariate normality assumption is unequivocally rejected by the data, but a suitable $t$-distribution cannot be rejected.

### 1.1. The data

In recent empirical studies, Fama and French's (1993) 25 portfolios, formed on size and book-to-market, are the standard test assets in empirical asset pricing studies. As a result, we will focus our analysis on these 25 portfolios plus their associated three factors to provide potentially highly valuable non-normality information on this widely used data set. The data are monthly returns available from French's website.[4] In addition, we also use the monthly returns on the one-month Treasury bill to construct the excess returns on the 25 size and book-to-market ranked portfolios. Altogether, there are $n = 28$ excess returns from July 1963 through December 2002.

### 1.2. Normality and GARCH tests

Our first question is whether the data can be adequately described by a normal distribution. To answer this, let $x_t = (r_t', \ f_t')'$, where $r_t$ represents the excess returns of $N = 25$ portfolios and $f_t$ represents the excess returns of $k = n - N = 3$ factors at time $t$. Following Mardia (1970) and many multivariate statistics books (e.g., Seber, 1984, p.142), we consider tests based on the following multivariate skewness and kurtosis,

$$D_1 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[(x_t-\hat{\mu})'\,\hat{V}^{-1}\,(x_s-\hat{\mu})\right]^3, \tag{1}$$

$$D_2 = \frac{1}{T}\sum_{t=1}^{T}\left[(x_t-\hat{\mu})'\,\hat{V}^{-1}\,(x_t-\hat{\mu})\right]^2, \tag{2}$$

where

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t, \tag{3}$$

$$\hat{V} = \frac{1}{T}\sum_{t=1}^{T}(x_t-\hat{\mu})(x_t-\hat{\mu})', \tag{4}$$

are the sample mean and covariance matrix of $x_t$, respectively. There are two desirable properties of $D_1$ and $D_2$. First, they converge, as sample size increases to infinity, to their population counterparts

$$\delta_1 = E\left(\left[(x-\mu)'V^{-1}(y-\mu)\right]^3\right), \qquad \delta_2 = E\left(\left[(x-\mu)'V^{-1}(x-\mu)\right]^2\right), \tag{5}$$

where $\mu$ and $V$ are the population mean and covariance-matrix of $x$, and $y$ is a random variable that has the same probability density as $x$, but is independent of $x$. Under the normality assumption, $\delta_1$ is simply zero, and $\delta_2$ is equal to $n(n+2)$. The second property is that $D_1$ and $D_2$ are invariant to any linear transformations of the data. In other words, any non-singular repackaging of the assets will

[4] We are grateful to Ken French for making this data available on his website. The Matlab programs for this paper will be available on our website.

not alter the multivariate skewness and kurtosis. Due to this invariance property, one can assume, without any loss of generality, that the true distribution has zero mean and unit covariance matrix for the purpose of computing the exact distribution of $D_1$ and $D_2$. As demonstrated by Zhou (1993), the exact distribution can be computed up to any desired accuracy by simulating samples from the standardized hypothetical true distribution of the data without specifying any unknown parameters. Tu and Zhou (2003) also use this idea to provide an exact test for normality. To achieve reliable accuracy, we use 10,000 draws in what follows.

This procedure can also be applied to test whether or not the data follow a suitable multivariate $t$-distribution with $\nu$ degrees of freedom. The $t$ density function is given by

$$f(x_t) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi\nu)^{\frac{n}{2}}\Gamma\left(\frac{\nu}{2}\right)|\Psi|^{\frac{1}{2}}}\left[1 + \frac{(x_t-\mu)'\Psi^{-1}(x_t-\mu)}{\nu}\right]^{-\frac{\nu+n}{2}}, \tag{6}$$

where $\Psi = (\nu - 2)V/\nu$ is a scale matrix whose use in place of $V$ is standard which simplifies formulas later. It is clear that this density approaches the normal as $\nu$ goes to infinity, and hence the usual normal distribution is a special limiting case of $t$. In order to apply the earlier procedure, one simulates data from a standard $t$-distribution and the empirical rejection rates can then be computed the same way as before.

Table 1 reports the results. Consider first both the univariate and multivariate sample kurtosis of the data which are in the seventh column of the table. It is seen that the univariate values are all greater than 3, the population value under normality. Indeed, the $p$-values of the univariate kurtosis test, reported in the next column in percent, all reject normality for each of the assets. Given the strong rejection of the univariate kurtosis test, it is not surprising that the $p$-value based on the multivariate kurtosis test is less than 0.01%. Hence, multivariate normality is unequivocally rejected by the data. On the other hand, if we assume that the data is from a $t$-distribution with degrees of freedom $\nu = 10, 8$ and 6, the $p$-value goes up from 0.34% to 16.84% and 94.11%. Therefore, a $t$-distribution with $\nu = 8$ is not rejected by the data, neither is the one with $\nu = 6$.

Consider now both the univariate and multivariate skewness tests. The sample skewness statistics are provided in the second column. For the measure of univariate skewness, what we actually report in Table 1 is the more common measure of skewness

$$\gamma_1 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t - \hat{\mu}}{\hat{V}^{\frac{1}{2}}}\right)^3. \tag{7}$$

$\gamma_1$ is related to the Mardia's measure of skewness by the relation $D_1 = \gamma_1^2$, so for a two-tailed test of zero skewness, it does not really matter whether we use $\gamma_1$ or $D_1$. However, reporting $\gamma_1$ allows us to find out if the returns are positively skewed or negatively skewed. From Table 1, we find that the skewness of individual portfolios are mostly negative but in general very small so that there are many portfolios that pass the test even under the normality assumption. The multivariate skewness test, however, strongly rejects the normality assumption and even a $t$ with $\nu = 10$ at the usual 5% level. Nevertheless, the multivariate skewness test and many of the univariate skewness tests cannot reject a $t$-distribution with $\nu = 8$ or 6, a conclusion similar to what we obtain using the kurtosis test. The reason is that the finite sample variation of the sample skewness of a $t$-distribution is very large when $\nu$ is small, so as to imply a large probability for observing a large sample skewness even though the true distribution, assumed $t$ here, is actually symmetric.

Although in the entire period we find many portfolios have negative sample skewness, further examination of the data shows that the sign of the sample skewness is not stable across subperiods. This suggests that the negative sample skewness is only a sample specific phenomenon, not necessarily a feature of the data that we have to model here.[5] In contrast to the behavior of the sample skewness, the sample kurtoses are all very large and significantly different from normality across subperiods. This indicates that the fat tails are indeed a salient feature of the data that we have to account for and we do so here by advocating the use of the $t$-distribution.

A question arises as to which value of $\nu$, 8 or 6, is a better model for the data. To understand the impact of the degrees of freedom on the $p$-values of the kurtosis test, consider as an alternative a popular kurtosis measure, the standardized one:

$$\kappa \equiv \frac{D_2}{n(n+2)} - 1 = \frac{2}{\nu - 4}, \tag{8}$$

where the last equality follows for a $t$-distribution. Under normality, $\kappa = 0$, so $\kappa$ measures the excess kurtosis relative to the normal. Equation (8) implies that the population kurtosis goes to infinity as $\nu$ goes down to 4. Hence, no matter how large the sample kurtosis is, one can always find a $t$-distribution to describe it with a small enough $\nu$. Although the $p$-value for the sample kurtosis test is greater with a smaller $\nu$ (but greater than 4), the sample kurtosis can in fact falls out of a reasonable left tail of the distribution when $\nu$ is too small. For example, the $p$-value of 94.11%

[5] A skewed $t$-distribution of Branco and Dey (2001) may be useful in applications where the skewness is too large to use the standard $t$, but not here as the skewness estimate is insignificant. See Harvey, Liechty, Liechty, and Müller (2004) for an excellent survey and new developments on skewness studies.

when $\nu = 6$ implies that the observed sample multivariate kurtosis falls into the 5.89% mass of the distribution from the left, close to the usual 5% rejection level if a one-sided test is used. So based on the sample kurtosis, a value of $\nu = 8$ appears to be a slightly better than $\nu = 6$. Later on, we report an estimate of $\nu = 8.66$ based on the maximum likelihood method, which is determined by the entire distribution of the data rather than by a particular moment. However, these are all point estimates of $\nu$, and we cannot be sure which one is closer to the true $\nu$. For the sake of robustness, we report most of our results based on three different values of $\nu$ (6, 8 and 10), so we can determine whether our reported results are sensitive to the choice of degrees of freedom.

Another question is that why $t$ is a better choice than the well-known GARCH models for the Fama-French data which are in the monthly frequency typically used in asset pricing and corporate studies. As pointed out in the introduction, the reasons are that the monthly data have little GARCH effects and the $t$ is more tractable in high-dimensional problems. Additionally, the $t$ has far fewer parameters than the GARCH model, and so its finite sample performance should be much better. To verify that the Fama-French data have little GARCH effects, the last column of Table 1 provides the P-values of the standard GARCH(1,1) test (see, e.g., Bollerslev, Engle and Nelson, 1994). The test runs an AR(1) regression on the fitted residuals and examines whether the $TR^2$ of the regression is large enough according to an asymptotic $\chi_1^2$ distribution. It is seen that all the P-values are greater than the 5% level except for the second asset. So, there is little evidence of GARCH effects. However, there might be a concern on the small sample performance of the asymptotic test. But this concern turn out not an issue at all. Although the results are not reported here, simulations show that the asymptotic test in fact over-rejects the null when the data is either normal or $t$ distributed. So, there is little evidence of GARCH effects in the monthly Fama-French data.

## 2. Impacts on Estimating Mean, Variance and Sharpe Ratios

After rejecting normality and accepting $t$ as a good alternative distribution for the data in the previous section, we now proceed to present the EM algorithm that elegantly solves the parameter estimation problem under $t$. With these estimates, we can then address the impact of the $t$-distribution assumption on estimated expected returns, variances and Sharpe ratios.

### 2.1. EM Algorithms and asymptotic theory

Under normality, the asymptotically most efficient estimate of $\mu$ and $V$ are their sample analogues, $\hat{\mu}$ and $\hat{V}$. The accuracy of the sample averages to estimating the expected mean can be judged by its asymptotic variance-covariance matrix,

$$\mathrm{Avar}[\hat{\mu}] = V. \tag{9}$$

This expression is in fact exact for all jointly independent and identically distributed (i.i.d.) returns. The sample average $\hat{\mu}$ is the asymptotically most efficient estimator under normality because it is in this case also the maximum likelihood estimator. However, as shown below, it will no longer be the most efficient estimator once the normality assumption is removed since the likelihood function will be different.

Indeed, under multivariate $t$, the asymptotically most efficient estimator of the parameters is the solution of maximizing the log-likelihood function based on the $t$ density,

$$\log \mathcal{L} = \text{constant} - \frac{T}{2}\log|\Psi| - \frac{\nu+n}{2}\sum_{t=1}^{T}\log\left[1 + \frac{(x_t-\mu)'\Psi^{-1}(x_t-\mu)}{\nu}\right]. \tag{10}$$

Unlike the log-likelihood function in the normal case, this one does not allow the combination of terms to yield a simple explicit solution to its maximum. Moreover, a direct numerical optimization is extremely difficult as the number of parameters is $434 = n + n(n+1)/2$, where $n = 28$ in our application to Fama and French's (1993) 25 assets plus 3 factors.

Fortunately, with the path breaking EM algorithm of Dempster, Laird and Rubin (1977), and especially Liu and Rubin (1995), we can use the following explicit iterative formulas to find the parameter estimate that maximizes the log-likelihood function under $t$, i.e., the solution to maximizing $\mathcal{L}$. Starting from any initial estimate of $\mu$ and $\Psi$, say $\tilde{\mu}^{(1)} = \hat{\mu}$ and $\tilde{\Psi}^{(1)} = \hat{\Psi} = (\nu-2)\hat{V}/\nu$, we can obtain iterative estimates via

$$u_t^{(i)} = \frac{\nu+n}{\nu + (x_t-\tilde{\mu}^{(i)})'[\tilde{\Psi}^{(i)}]^{-1}(x_t-\tilde{\mu}^{(i)})}, \tag{11}$$

$$\tilde{\mu}^{(i+1)} = \frac{\sum_{t=1}^{T} u_t^{(i)} x_t}{\sum_{t=1}^{T} u_t^{(i)}}, \tag{12}$$

$$\tilde{\Psi}^{(i+1)} = \frac{1}{T}\sum_{t=1}^{T} u_t^{(i)}(x_t-\tilde{\mu}^{(i+1)})(x_t-\tilde{\mu}^{(i+1)})', \tag{13}$$

where $u_t^{(i)}$ is an auxiliary variable whose meaning as well as why the algorithm works are discussed in the Appendix. Clearly, the above EM algorithm is simple to program and easy to implement.

Mathematically, the solutions monotonically converge to $\tilde{\mu}$ and $\tilde{\Psi}$ that maximize equation (10), the log-likelihood function under $t$. Indeed, in our application to Fama-French 25 assets and three-factors, the algorithm converges with less than 100 iterations and it takes less than a minute to run on a PC.

However, we should remark that the degrees of freedom $\nu$ here is assumed known. This may be reasonable because the likely values for $\nu$ can be assessed by using the kurtosis test. When one is concerned about the fact that $\nu$ is truly unknown, one can treat $\nu$ as an additional parameter and estimates it directly from the data. Then the following extended algorithm can be used. Starting with any initial estimate of $\nu$, say $\tilde{\nu}^{(1)} = 8$, one can update a new estimate of $\nu$ by solving

$$f(\nu) = \phi\left(\frac{\nu+n}{2}\right) - \phi\left(\frac{\nu}{2}\right) + \log\left(\frac{\nu}{\nu+n}\right) + \frac{1}{T}\sum_{t=1}^{T}\left[\log(u_t^{(i)}(\nu)) - u_t^{(i)}(\nu)\right] + 1 = 0, \tag{14}$$

where $\phi(\nu) = \mathrm{d}\log\Gamma(\nu)/\mathrm{d}\nu$ is the digamma function and

$$u_t^{(i)}(\nu) = \frac{\nu+n}{\nu + (x_t - \tilde{\mu}^{(i)})'[\tilde{\Psi}^{(i)}]^{-1}(x_t - \tilde{\mu}^{(i)})}. \tag{15}$$

Hence, the earlier EM algorithm can be combined with this one so that it is still applicable when $\nu$ is treated as an unknown parameter. It should be noted that equation (14) does not admit an analytical solution, so the implementation is more complex than the earlier case of a known $\nu$. However, equation (14) involves only one variable and its solution is easy to find by using a line-search routine. Therefore, even with an unknown $\nu$, practical implementation of the algorithm is still straightforward. Indeed, even if we treat $\nu$ as unknown in implementing the EM algorithm, it still converges in less than a minute in our applications. Moreover, regardless of what starting value of $\nu$ chosen, the algorithm has always quickly converged to an estimated value $\tilde{\nu} = 8.66$ for the Fama-French data set that we studied earlier in Section 1.

Therefore, even if one is less willing to simply use several values of $\nu$ to assess the sensitivity of $\nu$ on the statistical inference, one can estimate $\nu$ easily and then use this estimated value instead in carrying out both the statistical computations and economic evaluations. This approach clearly makes little qualitative difference in our applications here.

While the EM algorithm provides an elegant solution to the maximum likelihood estimation problem, it is only valuable if there is an efficiency gain over the sample averages. Like the normality case, a simple analytical expression is available to assess the accuracy of the $t$ estimates. Based on

Lange, Little and Taylor (1989), the asymptotic variance-covariance matrix of $\tilde{\mu}$ is, for $\nu > 2$,

$$\mathrm{Avar}(\tilde{\mu}) = (1-\rho)V, \qquad \rho \equiv \frac{2n+4}{\nu(\nu+n)}. \tag{16}$$

This says that, when the data is $t$ distributed rather than the normal, the sample mean $\hat{\mu}$ is no longer the asymptotically most efficient estimate of $\mu$, but the maximum likelihood estimator $\tilde{\mu}$ is. The relative efficiency is measured by $\rho$. The greater the $\rho$, the better the maximum likelihood method. In our application with $n = 28$, $\nu = 8$, we have $\rho = 0.2083$, implying that the maximum likelihood estimator $\tilde{\mu}$ is 20% less volatile than the sample mean.[6] It is interesting to observe that this improvement in estimation accuracy is independent of the parameter values of $\mu$ and $V$. In addition, the relative efficiency increases when $n$ increases. Under normality, the sample average return of an asset is the asymptotically most efficient estimator of its expected return, and the inclusion of other assets will not alter this estimate. In contrast, once the $t$-distribution is allowed, realized returns from one asset contain useful information on estimating the expected return of another asset, as shown later by empirical results and a simple analytical example. The greater the number of assets, the more efficient the $\tilde{\mu}$. Moreover, the relative efficiency increases when $\nu$ gets smaller. This makes intuitive sense: the smaller the $\nu$, the greater the deviation of the data from normality, and hence the greater the gain from using a procedure that incorporates non-normality into estimation.

Similarly, one can ask what the efficiency gain is for estimating the variance of asset returns by the maximum likelihood method under $t$-distribution. In the Appendix, we show that

$$\mathrm{Avar}[\tilde{V}_{ii}] = (1-\rho_v)\mathrm{Avar}[\hat{V}_{ii}], \qquad \rho_v \equiv \frac{2[2n+4+\nu(n+5)]}{\nu(\nu-1)(\nu+n)}. \tag{17}$$

Again, the improvement in estimation accuracy, $\rho_v$, is independent of the true parameters. In addition, as $\mathrm{Avar}[\tilde{V}_{ii}^{\frac{1}{2}}] = \mathrm{Avar}[\tilde{V}_{ii}]/(4V_{ii})$ and $\mathrm{Avar}[\hat{V}_{ii}^{\frac{1}{2}}] = \mathrm{Avar}[\hat{V}_{ii}]/(4V_{ii})$, we have

$$\mathrm{Avar}[\tilde{V}_{ii}^{\frac{1}{2}}] = (1-\rho_v)\mathrm{Avar}[\hat{V}_{ii}^{\frac{1}{2}}], \tag{18}$$

so $\rho_v$ is also the efficiency gain for estimating the standard deviation of asset returns by the maximum likelihood method under $t$. When $n = 28$ and $\nu = 8$, we have $\rho_v = 32.14\%$. This says that compared with the sample variance or standard deviation, the maximum likelihood procedure

[6]It should be noted that the asymptotic variance of $\tilde{\mu}$ is the same whether $\nu$ is known or unknown, so the efficiency gain of using $\tilde{\mu}$ does not depend on whether we know $\nu$ or not.

under $t$ improves the estimation efficiency in estimating the variance or standard deviation by about 32%.

Besides the fundamental parameters of asset means and variances, Sharpe ratios of a given portfolio is of great interest in practice. Because of this, Lo (2002) devotes an entire article to the derivation of the asymptotic theory for estimating them. He answers the question that how accurately the usual Sharpe ratios are estimated by using the sample mean and variance. Given our improved estimates on the means and variances, it is naturally to ask the question that how much the optimal ML estimates under $t$ can improve upon the usual estimate of the Sharpe ratio.

Let $w$ be an $n \times 1$ vector of portfolio weights, and $R_{pt} = w'R_t$ be the return of the portfolio at time $t$. Then the theoretical and unobservable Sharpe ratio of the portfolio is $\theta_p = \mu_p/\sigma_p = w'\mu/(w'Vw)^{\frac{1}{2}}$. The usual estimate is $\hat{\theta}_p = \hat{\mu}_p/\hat{\sigma}_p = w'\hat{\mu}/(w'\hat{V}w)^{\frac{1}{2}}$, and the estimate that is based on the ML under $t$ is $\tilde{\theta}_p = \tilde{\mu}_p/\tilde{\sigma}_p = w'\tilde{\mu}/(w'\tilde{V}w)^{\frac{1}{2}}$. Based on (17) and using the delta method, it can be shown that

$$\mathrm{Var}[\hat{\theta}_p] = 1 + \frac{\theta_p^2(\nu-1)}{2(\nu-4)} \tag{19}$$

and

$$\mathrm{Var}[\tilde{\theta}_p] = \frac{\nu+n+2}{\nu(\nu+n)}\left[\nu - 2 + \frac{\theta_p^2(\nu+1)}{2}\right]. \tag{20}$$

In our applications here with $n = 25$ and $\nu = 8$, it is clear that $\mathrm{Var}[\hat{\theta}_p] = 1 + 0.875\theta_p^2$ and $\mathrm{Var}[\tilde{\theta}_p] = 0.795 + 0.596\theta_p^2$. Hence, regardless the exact combination of the portfolio, the new estimate that is based on the ML under $t$ reduces the estimation error by about 20%.

### 2.2. Empirical results

After providing the estimation method and the associated asymptotic theory, we are ready now for the empirical results on the expected returns and the standard deviations, the fundamental parameters of the Fama-French data. The second column of Table 2 reports the sample average returns, while the next three columns are the maximum likelihood estimates of the expected returns under a $t$-distribution with $\nu = 10, 8$ and 6, respectively. As discussed earlier, a value of $\nu = 8$ appears to be a good model for the data, but the results on two other values are provided to assess the sensitivity of the results to the specification of $\nu$. It is striking that the expected returns estimated under $t$ for all assets are *smaller* than those estimated under normality. For example, the sample average excess returns for the market (MKT) and size (SMB) factors are 0.410% and

0.210%, but their estimated expected excess returns under $t$ with $\nu = 8$ are only 0.358% and 0.127%, implying an annual difference of 0.624% and 0.996%. Such differences in some of the portfolios are even larger. For instance, the S1B1 portfolio has an annual difference of 2.964% between the two estimates of expected return.

To understand further the intuition why the difference is so large for S1B1, consider, for simplicity, that we try to fit its returns using a univariate $t$-distribution whose log-likelihood function is

$$\log \mathcal{L} = \text{constant} - \frac{T}{2}\log(\psi) - \frac{\nu+1}{2}\sum_{t=1}^{T}\log\left(1 + \frac{(r_t-\mu)^2}{\nu\psi}\right), \tag{21}$$

where $r_t$ is the return on S1B1 at time $t$ and $\mu = E[r_t]$. It is easy to see, from the score function, that the maximum likelihood estimator of $\mu$ is a solution of

$$\sum_{t=1}^{T} w_t(r_t - \mu) = 0, \tag{22}$$

or

$$\tilde{\mu} = \sum_{t=1}^{T} w_t r_t, \tag{23}$$

where $w_t = c/(\nu + \delta_t)$ with $\delta_t = (r_t - \mu)^2/\psi$ and $c$ is a constant such that $\sum_{t=1}^{T} w_t = 1$. It is clear that $\delta_t$ measures how far the data is from its center. Since $w_t$ is a decreasing function of $\delta_t$, outliers are weighted less than other data points in the computation of $\tilde{\mu}$. In contrast, the sample mean weights all data points equally with weight $1/T$. When the true distribution has fatter tails than the normal, the sample mean becomes less efficient when compared with the maximum likelihood estimator, and the estimated mean under $t$ can shift leftward or rightward depending on the tail behavior of the actual data. In a multivariate setting, similar results follow. Now, to see why the expected return of S1B1 estimated under $t$ is much smaller than its sample mean, we need to examine the relationship between $r_t$ and $\delta_t$. In Figure 1, we provide plots of $r_t$ against $\delta_t$ for S1B1 and MKT. The upper part of Figure 1 provides the plot for S1B1. As we can see, for the months that are considered to be outliers (i.e., large $\delta_t$) by the multivariate $t$-distribution, the portfolio S1B1 have mostly large positive returns. By down-weighting these large positive monthly returns, the resulting maximum likelihood estimate of the mean of S1B1 is therefore substantially lower than the sample mean. In contrast, as shown by the lower part of Figure 1, while the market returns are mostly positive during the months that have large $\delta_t$, they are not unusually large. Therefore, the mean of the market when estimated under $t$ is not all that far away from the sample mean.

Using a similar data set, Knez and Ready (1997) also find that the size effect is significantly reduced if one drops a small percentage of influential observations from the entire sample (see their Tables III and VI). However, dropping influential observations completely cannot be easily justified statistically, so it is unclear whether one should rely more on the sample mean from the trimmed sample or on that from the original data. Instead of dropping outliers, our approach simply puts less weights on the outliers. Such a strategy can be justified statistically because it is based on the likelihood principle to improve estimation efficiency. As a result, we can have more faith here that the size effect is indeed smaller than what is shown by the sample mean of SMB.

In contrast to the sharp differences in the estimated means, the standard deviations are not much different when estimated under either normal or $t$. For example, as shown in Table 2, while there is a huge difference in the two estimates of expected returns, S1B1 has similar standard deviations using the sample one $\hat{V}_{11}^{\frac{1}{2}} = 8.283\%$ (per month) and the maximum likelihood one $\tilde{V}_{11}^{\frac{1}{2}} = 7.751\%$ under the $t$-distribution with 8 degrees of freedom. The same is also true for the market excess return whose standard deviations in the two cases are 4.505% and 4.537%, very close to each other. The small differences in estimating the standard deviations are consistent with the well-known fact that it is easier to estimate the second moments than the first moments of returns. Indeed, the estimated standard error of $\hat{V}_{11}^{\frac{1}{2}}$ for S1B1 is only 0.356%, so the estimate 8.283% is very accurate. In contrast, the estimated standard error of $\hat{\mu}_1$ is 0.381%. Recall that the sample mean is only 0.135%, indicating that the estimate of the mean is very imprecise. Therefore, the difference between this normal mean estimate and the $t$ one can be much greater than the difference in estimating the standard deviations.

## 3. Jensen's Alpha

In evaluation of mutual fund performance, Jensen's alpha is the most widely reported measure despite its restrictive assumptions. We show in this section that a relaxation of the unrealistic normality assumption to a more reasonable $t$ can generate drastically different rankings for mutual funds. To see this, we develop first the theory associated with estimating the alphas and betas under $t$, and then apply them to mutual fund performance analysis.

### 3.1. The $t$-regression model

Recall that $x_t = (r_t',\ f_t')'$, where $r_t$ contains the excess returns of $N$ test assets and $f_t$ contains the excess returns of $k$ $(= n - N)$ factors. Then we have the usual multivariate regression,

$$r_t = \alpha + \beta f_t + \epsilon_t, \tag{24}$$

where $\epsilon_t$ is an $N \times 1$ vector of residuals with zero mean and a non-singular covariance matrix. To relate $\alpha$ and $\beta$ to the earlier parameters $\mu$ and $V$, consider the corresponding partition

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \qquad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{25}$$

Under the usual multivariate normal distribution, it is clear that the distribution of $r_t$ conditional on $f_t$ is also normal and

$$E[r_t|f_t] = \mu_1 + V_{12}V_{22}^{-1}(f_t - \mu_2), \tag{26}$$

$$\mathrm{Var}[r_t|f_t] = V_{11} - V_{12}V_{22}^{-1}V_{21}. \tag{27}$$

Therefore, the parameters $\alpha$, $\beta$ and the earlier parameters $\mu$, $V$ obey the following relationship:

$$\alpha = \mu_1 - \beta\mu_2, \qquad \beta = V_{12}V_{22}^{-1}. \tag{28}$$

Denote $\Sigma$ as the usual notation for the covariance matrix of $\epsilon_t$,

$$\Sigma = \mathrm{Var}[\epsilon_t] = V_{11} - V_{12}V_{22}^{-1}V_{21}. \tag{29}$$

It should be noted that under the normality assumption, $\Sigma$ is also the variance of $\epsilon_t$ conditional on $f_t$. However, once the normality assumption is removed, this will not necessarily be the case. Indeed, if the data follow a $t$-distribution with $\nu$ degrees of freedom, although the mean of $r_t$ conditional on $f_t$ is still a linear function of $f_t$ as above, the conditional covariance matrix is no longer a constant, but rather a quadratic function of $f_t$:

$$\mathrm{Var}[r_t|f_t] = \left[\frac{\nu - 2 + (f_t - \mu_2)'V_{22}^{-1}(f_t - \mu_2)}{\nu + k - 2}\right]\Sigma. \tag{30}$$

This says that the conditional variance of the $t$-regression residuals vary with time, and hence is conditionally heteroskedastic.

The conditionally heteroskedasticity is a key feature of our $t$ model versus those in the econometrics and statistics literature where $f_t$ is treated as fixed and $\epsilon_t$ is assumed to be $t$ distributed with a *constant* covariance matrix (see, e.g., Chib, Tiwari and Jammalamadaka, 1988, Van Praag and Wesselman, 1989, Osiewalski and Steel, 1993, and references therein). In contrast, $f_t$ here is jointly random with the asset returns, and the conditional covariance matrix of $\epsilon_t$ is time-varying.[7] As shown below, this will have important implications for both parameter estimation and asset pricing tests.

To assess the estimation accuracy of the alphas and betas under $t$, we need to derive the associated asymptotic standard errors of the estimates which are not available previously. The estimates under $t$, $\tilde{\alpha}$ and $\tilde{\beta}$, are easily obtained from equation (28) by replacing $\mu$ and $V$ with their maximum likelihood estimates. The key issue is how accurate $\tilde{\alpha}$ and $\tilde{\beta}$ are when compared with the OLS estimates. It can be shown (see the Appendix) that the $N(k+1)$ parameter formed by them has an asymptotic variance-covariance matrix:

$$\text{Avar}\begin{bmatrix} \tilde{\alpha} \\ \text{vec}(\tilde{\beta}) \end{bmatrix} = \left(\frac{\nu+n+2}{\nu+n}\right)\begin{bmatrix} \left(\frac{\nu-2}{\nu}\right) + \mu_2' V_{22}^{-1}\mu_2 & -\mu_2' V_{22}^{-1} \\ -V_{22}^{-1}\mu_2 & V_{22}^{-1} \end{bmatrix} \otimes \Sigma. \tag{31}$$

In contrast, the usual OLS estimators $\hat{\alpha}$ and $\hat{\beta}$ have an asymptotic variance-covariance matrix of

$$\text{Avar}\begin{bmatrix} \hat{\alpha} \\ \text{vec}(\hat{\beta}) \end{bmatrix} = \begin{bmatrix} 1 + \left(\frac{\nu-2}{\nu-4}\right)\mu_2' V_{22}^{-1}\mu_2 & -\left(\frac{\nu-2}{\nu-4}\right)\mu_2' V_{22}^{-1} \\ -\left(\frac{\nu-2}{\nu-4}\right)V_{22}^{-1}\mu_2 & \left(\frac{\nu-2}{\nu-4}\right)V_{22}^{-1} \end{bmatrix} \otimes \Sigma. \tag{32}$$

It follows that the percentage improvement of the maximum likelihood estimator under $t$, $\tilde{\alpha}$, over $\hat{\alpha}$ is

$$1 - \left(\frac{\nu+n+2}{\nu+n}\right)\left[\frac{\left(\frac{\nu-2}{\nu}\right) + \mu_2' V_{22}^{-1}\mu_2}{1 + \left(\frac{\nu-2}{\nu-4}\right)\mu_2' V_{22}^{-1}\mu_2}\right] = \frac{2}{\nu+n}\left[\frac{\frac{n+2}{\nu} + \left(\frac{n+4}{\nu-4}\right)\mu_2' V_{22}^{-1}\mu_2}{1 + \left(\frac{\nu-2}{\nu-4}\right)\mu_2' V_{22}^{-1}\mu_2}\right]. \tag{33}$$

The lower bound of the percentage improvement is $2(n+2)/(\nu(\nu+n))$, which is reached when $\mu_2' V_{22}^{-1}\mu_2 \to 0$. The upper bound is $2(n+4)/((\nu-2)(\nu+n))$, which is reached when $\mu_2' V_{22}^{-1}\mu_2 \to \infty$. When $n = 28$ and $\nu = 8$, the percentage standard reduction of $\tilde{\alpha}$ ranges from 20.83% to 29.63%. Similarly, the percentage improvement of $\tilde{\beta}$ is:

$$1 - \frac{\frac{\nu+n+2}{\nu+n}}{\frac{\nu-2}{\nu-4}} = \frac{2(n+4)}{(\nu-2)(\nu+n)}. \tag{34}$$

When $n = 28$ and $\nu = 8$, the percentage variance reduction of $\tilde{\beta}$ is 29.63%.

### 3.2. Empirical alpha estimates

Table 3 provides the estimates of the alphas and betas under normality and $t$ (with $\nu = 8$) for the Fama-French data. It is seen that there are large differences in the alpha estimates. For

[7] Relying on this property, Laplante (2003) provides an interesting model of the joint distribution of returns and information signals in his study of market timing with conditional heteroskedasticity.

example, the estimated alphas for S1B2 and S3B4 are $-0.004$ and $0.011$ under normality, but change to $-0.082$ and $0.071$ under the $t$-distribution with $\nu = 8$, about 20 times and 6 times larger in absolute value, respectively. In contrast, the differences in the beta estimates are much smaller in percentage terms. For both the MKT and SMB factors, the betas are virtually the same under either $t$ or normal. However, there are some substantial differences in the HML betas. The HML betas for S3B2 and S4B2 have reduced significantly from 0.221 and 0.263 to 0.095 and 0.141. Overall, it appears that the usual OLS betas are fairly accurately estimated, but the alphas are not. As a result, there is a great value of obtaining more accurate estimate of $\alpha$ for both asset pricing tests and performance evaluation.

We now examine Jensen's alpha for mutual funds which are more relevant in practice. The mutual fund returns data are available from the Center for Research in Securities Prices (CRSP). There are about 10,000 mutual funds at the end of December 2002. We consider only those funds with complete data in the last 5 or 10 years. In this way, we obtain 8996 and 3152 funds, respectively.

Consistent with many studies and reports, we compute a mutual fund's alpha based on the traditional CAPM. There are two questions of interest. First, among funds with negative alphas (the losers), what is the percentage of funds whose rankings are reversed based on a reasonable $t$-distribution? Moreover, what is the magnitude of the reversals? Similarly, we can ask the same questions for winners which have positive Jensen's alphas.

Table 4 reports the results using mutual funds in the past five years from January 1998 to December 2002. The first panel provides the results on reversals from under-performance to over-performance. The panel reports five funds that have the greatest reversals as measured by the difference of Jensen's alphas under normality and under $t$ whose degrees of freedom is treated as unknown and estimated from the data. For example, One Group Diversified's alpha changes from a negative value of $-0.556\%$ (per month) to a positive value of $0.378\%$ (per month). The reason for the huge shift of the alpha value is because the estimated degrees of freedom of the $t$-distribution has a value of 3.5 (not reported in the table), which implies a distribution that is far different from the normal. For comparison, we also provide those alphas under fixed degrees of freedom for the $t$-distribution, and find similar results. However, among all the 8996 funds there are only 0.8% reversals from negative alphas to positive ones. This is not too high, but is expected as many loser funds have large negative alphas which can be shifted leftward under $t$ substantially, but still remain negative. To assess the magnitude of the reversals, the average difference of the reversed funds, 0.136%/month, is reported at the last row of the first panel.

The next panel provides the corresponding results for reversals from over-performance to under-performance. One apparent feature is that the magnitude of the reversals seems much larger than the losers to winners case. For example, PBHG Select Equity Fund has a huge alpha value of 1.532% under the normality assumption, but this value is reduced enormously to a negative alpha of $-0.374\%$. In addition, the magnitude of the reversals is now 0.376%, almost three times the loser case. Moreover, the percentage of the reversals also increases to 1.4%.

To further assess the difference in estimating alpha under the normal and $t$, the last panel of the table reports the percentage of funds for which the difference in estimated alphas is greater than 1% to 5% per year. As reported in the table, there are 13.9% of the funds whose alpha estimates under the two distributional assumptions differ by 1% or more per year. This is clearly a high percentage, which indicates many mutual funds have outliers in their monthly returns. In fact, there are still 2.6% of the funds whose alpha estimates differ by 5% or more per year.

Table 5 provides corresponding results for 3152 funds whose data is available for 10 years from January 1993 to December 2002. Despite the increase in horizon, the results are little changed from the 5 years' case. The fund returns data suggests economically significant differences in Jensen's alpha estimates under normality versus under the $t$. Clearly, it is important to examine the economic reasons why fund return data is so fat-tailed and what institutional or compensation design may exasperate it. However, such a study goes beyond the scope of the current paper as our focus here is to provide empirical evidence that the estimated Jensen's alpha is sensitive to the underlying normality assumption. Another question is that what alternative measures can one use that are less sensitive to the normality assumption? Refining Jensen's alpha, Cohen, Coval and Pástor (2003) provide an interesting and more accurate measure that pools information across funds. It is of interest to examine whether this new measure captures some of the kurtosis of the data.

## 4. Asset Pricing Tests

The popular method for testing the factor pricing model is a multivariate test of the following standard parametric restrictions:

$$H_0 : \quad \alpha = 0_N \tag{35}$$

in the multivariate regressions of

$$r_t = \alpha + \beta f_t + \epsilon_t, \quad t = 1, \ldots, T. \tag{36}$$

Under normality, this can be tested by the well-known Gibbons, Ross, and Shanken (1989) test,

$$\mathrm{GRS} = \left(\frac{T-N-k}{N}\right) \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1+\hat{\mu}_2'\hat{V}_{22}^{-1}\hat{\mu}_2} \;\sim\; F_{N,T-N-k}, \tag{37}$$

where $\hat{\alpha}$ and $\hat{\Sigma}$ are obtained from either linear regressions or from the relations between them and $\hat{\mu}$ and $\hat{V}$. It is well-known that the GRS test is simply a suitable transformation of the standard likelihood ratio test. However, once the normality assumption is replaced by a joint $t$ return assumption, one should be cautious in using the GRS test as it is not strictly valid under $t$.

Now, if the returns follow a multivariate $t$-distribution with $\nu$ degrees of freedom, we can also easily estimate the parameters under the null to obtain a likelihood ratio test based on the likelihood function under $t$. Denote the restricted parameter estimates by $\tilde{\mu}_r$ and $\tilde{\Psi}_r$. Under the null, $(\mu,\ \Psi)$ can be mapped one-to-one into $(\beta,\ \Psi_\epsilon,\ \mu_2,\ \Psi_{22})$, where $\Psi_\epsilon = (\nu-2)\Sigma/\nu$, and hence the EM algorithm for their estimation can be conveniently written in terms of the latter parameters:

$$u_t^{(i)} = \frac{\nu+n}{\nu+(r_t-\tilde{\beta}^{(i)}f_t)'[\Psi_\epsilon^{(i)}]^{-1}(r_t-\tilde{\beta}^{(i)}f_t)+(f_t-\tilde{\mu}_{2r}^{(i)})'[\Psi_{22r}^{(i)}]^{-1}(f_t-\tilde{\mu}_{2r}^{(i)})}, \tag{38}$$

$$\tilde{Y}^{(i)} = [r_1\sqrt{u_1^{(i)}},\ r_2\sqrt{u_2^{(i)}},\ \ldots,\ r_T\sqrt{u_T^{(i)}}]', \tag{39}$$

$$\tilde{X}^{(i)} = [f_1\sqrt{u_1^{(i)}},\ f_2\sqrt{u_2^{(i)}},\ \ldots,\ f_T\sqrt{u_T^{(i)}}]', \tag{40}$$

$$\tilde{\beta}^{(i+1)} = (\tilde{X}^{(i)\prime}\tilde{X}^{(i)})^{-1}(\tilde{X}^{(i)\prime}\tilde{Y}^{(i)}), \tag{41}$$

$$\tilde{\Psi}_\epsilon^{(i+1)} = \frac{1}{T}(\tilde{Y}^{(i)}-\tilde{X}^{(i)}\tilde{\beta}^{(i+1)})'(\tilde{Y}^{(i)}-\tilde{X}^{(i)}\tilde{\beta}^{(i+1)}), \tag{42}$$

$$\tilde{\mu}_{2r}^{(i+1)} = \frac{\sum_{t=1}^T u_t^{(i)} f_t}{\sum_{t=1}^T u_t^{(i)}}, \tag{43}$$

$$\tilde{\Psi}_{22r}^{(i+1)} = \frac{1}{T}\sum_{t=1}^{T} u_t^{(i)}(f_t-\tilde{\mu}_{2r}^{(i+1)})(f_t-\tilde{\mu}_{2r}^{(i+1)})', \tag{44}$$

where the iteration can start from, say, the estimates under normal as before. With the restricted parameter estimates denoted by $\tilde{\mu}_r$ and $\tilde{\Psi}_r$, we can compute the likelihood ratio test under $t$:

$$\mathrm{LRT}_t \equiv 2\left(\frac{T-(N/2)-k-1}{T}\right)\left[\log\mathcal{L}(\tilde{\mu},\tilde{\Psi})-\log\mathcal{L}(\tilde{\mu}_r,\tilde{\Psi}_r)\right] \overset{A}{\sim} \chi_N^2, \tag{45}$$

where $\log \mathcal{L}(\cdot, \cdot)$ is the log-likelihood function under $t$ given by (10). Note that, analogous to the normality case, we use the Bartlett correction factor $T-(N/2)-k-1$ instead of $T$ in the likelihood ratio test statistic because it can substantially improve the small sample properties of the likelihood ratio test statistic.[8]

As the CAPM of Sharpe (1964) and Lintner (1965) is of fundamental importance in finance, it is of interest to use it as an example to illustrate our test. This amounts to testing (35) with the single and theory-motivated market factor in (36). Table 10 reports the testing results based on the GRS test and the likelihood ratio test. Over the entire sample period of July 1963 to December 2002, both tests reject the CAPM strongly with virtually zero $p$-values, whether we assume the underlying distribution is normal or $t$. However, given the strong rejection of the normality assumption, one cannot draw firm conclusions about the rejections from test statistics that assume normality. With the test developed here also reaffirming the rejections reached by the GRS test under normality, one can claim that the rejection is indeed caused by the failure of the model rather than the violation of the restrictive normality assumption. Hence, even in cases both the GRS and the $t$-distribution-based tests have the same conclusions, the latter is still of interest because it says that the GRS conclusion is robust in those case. Without the latter, it is of no way of knowing whether the GRS test is reliable at all due to its false assumptions.

More interestingly, though, it is not always the case that the two tests give rise to the same conclusion. For example, in the first subperiod of July 1963 to June 1983, we find that the GRS test does not reject the CAPM at the usual 5% level, but the likelihood ratio test computed under $t$ with $\nu = 10$, 8 or 6, suggests rejection of the CAPM at the 5% level. This is clearly a case where the $t$ based test makes a difference by suggesting the CAPM fails to hold in the first subperiod while the GRS test is unable to do so.[9] For the second subperiod, both the tests reject the CAPM strongly as they do for the entire sample period.

There are two questions on the rejection by $\mathrm{LRT}_t$. First, does the rejection of $\mathrm{LRT}_t$ due to small sample problem? Without doing a full blown simulation experiment on the finite sample distribution of $\mathrm{LRT}_t$, we cannot give a definite answer. But this is unlikely because $\mathrm{LRT}_t$ for the normal case gives a $p$-value that is very close to the one from the GRS test, which indicates

[8]See Muirhead (1982, Theorem 10.5.5) for a derivation of this Bartlett correction.

[9]Although not reported here, we apply the same analysis to 10 size-sorted portfolios of the NYSE and find cases where the two likelihood ratio tests give conflicting conclusions on the validity of an asset pricing model.

that with the Bartlett adjustment and the sample size that we have, the asymptotic distribution provides a reasonably good approximation of the finite sample distribution of $\mathrm{LRT}_t$. Second, can a suitably adjusted GRS test that is valid under $t$ reject the null? Geczy (2001) provides an adjusted GRS test that has an approximate $F$-test under $t$ distributed returns,

$$\mathrm{GRS}_g = \left(\frac{T-N-k}{N}\right)\frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1+(1+\kappa)\hat{\mu}_2'\hat{V}_{22}^{-1}\hat{\mu}_2} \;\sim\; F_{N,T-N-k}, \tag{46}$$

where $\kappa$ is the standardized kurtosis as defined by (8). Although not reported table, the $p$-values of this test are not much different from those of the standard GRS test. For example, in the first subperiod where the GRS test does not reject and has a $p$-value of 9.03%, $\mathrm{GRS}_g$ has $p$-values of 9.33%, 9.18% and 9.13% under $t$ with 6, 8 and 10 degrees of freedom, respectively. An intuition for $\mathrm{GRS}_g$ having less power than $\mathrm{LRT}_t$ is that the former is based on inefficient parameter estimates when the returns are $t$ rather than normal, while the latter utilizes the optimal estimates from the maximum likelihood method.

## 5. Extensions and Future Research

Like many empirical asset pricing and corporate finance studies, we assume that the asset returns are i.i.d. over time. Moreover, they have a multivariate $t$-distribution at any time. Although the multivariate $t$-distribution is still restrictive, it is more general and realistic than the widely used normality assumption, and contains the normal as a special limiting case. In statistics, multivariate $t$-distributions are used extensively for robustness analysis of data that exhibit fat tails (see, e.g., Lange, Little and Taylor (1989), Vasconcellos and Cordeiro, (2000), and references therein). Because stock returns do have fat tails, the multivariate $t$-distribution is particularly relevant in finance.

Since Engle's (1982) path-breaking paper, the ARCH and the related GRACH models are shown to successfully model conditional heteroskedasticity of the data in finance. Recently, Vorkink (2003) applies semiparametric models to asset pricing tests. As mentioned earlier, a common weakness of these models is that they are, with too many parameters requiring numerical search to maximize the objective function, very difficult (if not impossible) to apply to large dimensional problems such as those analyzed in this paper. For example, Vorkink (2003) still uses univariate estimates to conduct the multivariate tests. In addition, the GARCH effect is primarily pronounced in high frequency return data, and is much weaker in monthly return data. Hence, ignoring the time-varying volatility in monthly return data is not grossly unrealistic. Although the $t$-distribution advocated in this paper is by no means the best model for the real world data, it is an attractive alternative to the common normal distribution due to its ease of use and its ability to capture the salient feature of the data by adding only one more parameter relative to the normal. As many interesting problems are multivariate in nature, the use of the $t$-distribution goes far beyond the scope of the paper.

An often asked question is why we single out the $t$ from the class of elliptical distributions, of which the $t$-distribution is only a special case and there are countless others. The major reason is that the $t$ appears to be the simplest distribution that nests the normal and is almost as tractable as the normal. Unless it is rejected by the data and a better alternative is found, the $t$-distribution should serve as a more reasonable model than the normal. Although it seems possible to extend the EM algorithm to some other elliptical distributions, the value of such extension is unknown, and is yet to be established by future research.

While the i.i.d. assumption is popular in testing the unconditional version of asset pricing models, it usually rules out the use of conditional information. Ferson (2003) provides an excellent review of testing conditional asset pricing models based on the generalized method of moments (GMM) of Hansen (1982), while Cremers (2002), Pástor and Stambaugh (2002) and Avramov (2004), among others, model the dynamics of conditional variables based on normality assumption. The trade-off between the two approaches is precision and generality. Clearly, the $t$-distribution advocated here can also be used in a conditional set-up to offer some more generality than the normal, and at the same time, to provide improvements in estimation accuracy of parameters.

Finally, it is worth noting that the more the asset return deviates from normality, the greater the difference it tends to make in estimating the asset's expected return and alpha by using the maximum likelihood method under $t$. This seems to have implications in measuring the abnormal returns of corporate events, of which long-term performance of IPOs is a leading example. As part of future research, it is of interest to examine how much of the abnormal performance may simply be due to estimation errors in estimating the benchmark from an asset pricing model.[10] In fact, for any hypotheses or studies that rely on the first moments of the asset returns, the methodology of the current paper may be applied to study the robustness of the results to normality assumption.

[10] Ritter (1991) raises some of the interesting issues and Lyon, Barber, and Tsai (1999) and references therein provide some of the latest methodologies.

Another important issue is why asset returns have fat tails. An understanding of the underlying economic reasons associated with a rational decision model is of fundamental importance, serving as yet another direction for future research.

## 6. Conclusion

In this paper, we attempt to provide convincing arguments for the wide use of multivariate $t$-distributions in finance. In contrast with the multivariate normal distribution which is firmly rejected by the data, suitable $t$-distributions pass standard skewness and kurtosis tests. In addition, parameter estimation and tests under $t$ can now be implemented almost as easily as under the normality case. So, it appears that multivariate $t$-distributions are promising in modeling financial data and answering interesting economic questions. Of course, we are not claiming that multivariate $t$-distributions are the best models. In fact, they should by construction be less realistic than other parameters rich models such as the well-known GARCH family. But the monthly data that are typically used for asset pricing tests and corporate studies have little GARCH effects. A key issue is that, for large dimensional problems, the multivariate $t$-distribution is tractable while GARCH models and the like are not.

Applying multivariate $t$-distributions to Fama and French's (1993) 25 portfolio returns and their 3 factors from January 1963 to December 2002, we find that there are drastic differences in estimating the expected asset returns. There are also large reversals in ranking mutual fund performance based on Jensen's alphas under the normal versus under the $t$. In addition, testing asset pricing models is sensitive to the normality assumption too.

In both statistics and econometrics, multivariate $t$-distributions are widely used for robust analysis of data with fat tails. As asset returns do have large fat tails, the $t$-distributions should play a similar role in finance. Hence, the proposed approach seems useful in a number of areas to ask how sensitive the results are to the usual normality assumption. Leading examples in this regard are the estimation of the cost of capital, performance evaluation, event studies, risk management, and any hypotheses that rely heavily on the first moments of the asset returns.

**Appendix A**

In this appendix, the intuition and proofs of the EM algorithms are provided for easy understanding and completeness of the paper, though they follow directly from Dempster, Laird and Rubin (1977) and Liu and Rubin (1995). However, explicit formulas for the asymptotic variance-covariance matrix of $\tilde{V}$ and that of the alphas and betas under the multivariate $t$-distribution are, to our knowledge, not available in the statistics literature, and hence their derivations are provided here.

*A.1. Proof of the first algorithm, (11)–(13):*

As noted earlier, the key difficulty associated with maximizing the log-likelihood function under $t$, equation (10), is that the terms do not combine to yield tractable solutions. However, it is well-known that a $t$ distribution is an infinite mixture of the normals. That is, there exists $u_t \sim \chi^2_\nu/\nu$ such that, conditional on $u_t$, $x_t$ is normal:

$$x_t \sim N(\mu, \Psi/u_t). \tag{A.1}$$

Suppose we had observations on all the $u_t$'s, then the conditional log-likelihood function:

$$\mathcal{L}(x_t|u_t) = \frac{n}{2}\left[\sum_{t=1}^{T}\log(u_t) - T\log(2\pi)\right] - \frac{T}{2}\log(|\Psi|) - \frac{1}{2}\sum_{t=1}^{T} u_t(x_t-\mu)'\Psi^{-1}(x_t-\mu), \tag{A.2}$$

which can be obviously maximized with

$$\tilde{\mu} = \frac{\sum_{t=1}^{T} u_t x_t}{\sum_{t=1}^{T} u_t}, \tag{A.3}$$

$$\tilde{\Psi} = \frac{1}{T}\sum_{t=1}^{T} u_t(x_t-\tilde{\mu})(x_t-\tilde{\mu})'. \tag{A.4}$$

However, the $u_t$'s are in fact unobserved. The idea of Dempster, Laird and Rubin (1977) and Liu and Rubin (1995) is that, we can estimate them by using their expected values conditional on the parameters and the data. This is the E-step of the algorithm, and the expectation is easily obtained as

$$E[u_t|x_t;\mu,\Psi] = \frac{\nu+n}{\nu+(x_t-\mu)'\Psi^{-1}(x_t-\mu)}. \tag{A.5}$$

Although we do not know the true parameters, the above provides an estimate of $u_t$ with any initial estimates of the parameters. Then we can maximize the conditional log-likelihood function easily. This is the M-step. Intuitively, the maximization should update our knowledge on the parameter estimates which can be used in turn to update a new estimate for $u_t$. Intuitively, continuing iterations may converge to the solution that maximizes the unconditional log-likelihood function, equation (10). Fortunately, for our problems here and many other models, the EM algorithm indeed converges and it even converges monotonically. *Q.E.D.*

*A.2. Proof of the asymptotic variance-covariance matrix for $\tilde{V}$, (17):*

First, the asymptotic covariance between the sample estimates $\hat{V}_{ij}$ and $\hat{V}_{kl}$ is known,

$$\text{Acov}[\hat{V}_{ij}, \hat{V}_{kl}] = \left(\frac{2}{\nu - 4}\right) V_{ij} V_{kl} + \left(\frac{\nu - 2}{\nu - 4}\right) (V_{ik} V_{jl} + V_{il} V_{jk}), \tag{A.6}$$

which follows from Muirhead (1982, p.42 and p.49). So the key is to obtain $\text{Acov}[\tilde{V}_{ij}, \tilde{V}_{kl}]$.

Define $D_n$ as an $n^2 \times n(n+1)/2$ duplication matrix such that $D_n \text{vech}(\tilde{V}) = \text{vec}(\tilde{V})$, where $\text{vec}(V)$ is an $n^2 \times 1$ column vector by stacking up the columns of $V$, and $\text{vech}(V)$ is an $n(n+1)/2 \times 1$ column vector by stacking up the columns of $V$, but with its supradiagonal elements deleted. Let $D_n^+ = (D_n' D_n)^{-1} D_n'$, we have $D_n^+ \text{vec}(\tilde{V}) = \text{vech}(\tilde{V})$. Lange, Little, and Taylor (1989) provide the formula for the individual elements of the information matrix of $\psi = \text{vech}(\Psi)$. With some simplification, we can write the information matrix of $\psi$ as

$$J_{\psi\psi} = \frac{1}{2(\nu + n + 2)} \left[(\nu + n) D_n' (\Psi^{-1} \otimes \Psi^{-1}) D_n - D_n' \text{vec}(\Psi^{-1}) \text{vec}(\Psi^{-1})' D_n\right]. \tag{A.7}$$

Based on the following identities

$$\begin{aligned}
[D_n' (\Psi^{-1} \otimes \Psi^{-1}) D_n]^{-1} &= D_n^+ (\Psi \otimes \Psi) D_n^{+\prime}, && \text{(A.8)}\\
D_n^{+\prime} D_n' \text{vec}(\Psi^{-1}) &= D_n D_n^+ \text{vec}(\Psi^{-1}) = D_n \text{vech}(\Psi^{-1}) = \text{vec}(\Psi^{-1}), && \text{(A.9)}\\
\text{vec}(\Psi^{-1})' \text{vec}(\Psi) &= \text{tr}(\Psi^{-1}\Psi) = n, && \text{(A.10)}\\
(\Psi^{-1} \otimes \Psi^{-1}) \text{vec}(\Psi) &= \text{vec}(\Psi^{-1}\Psi\Psi^{-1}) = \text{vec}(\Psi^{-1}), && \text{(A.11)}
\end{aligned}$$

we can analytically invert $J_{\psi\psi}$ as

$$J_{\psi\psi}^{-1} = \frac{2(\nu + n + 2)}{\nu + n} \left[D_n^+ (\Psi \otimes \Psi) D_n^{+\prime} + \frac{\text{vech}(\Psi)\text{vech}(\Psi)'}{\nu}\right]. \tag{A.12}$$

This implies that the asymptotic variance of $\text{vech}(\tilde{V})$ is

$$\text{Avar}[\text{vech}(\tilde{V})] = \frac{2(\nu + n + 2)}{\nu + n} \left[D_n^+ (V \otimes V) D_n^{+\prime} + \frac{\text{vech}(V)\text{vech}(V)'}{\nu}\right]. \tag{A.13}$$

In particular, we have

$$\text{Acov}[\tilde{V}_{ij}, \tilde{V}_{kl}] = \left(\frac{2(\nu + n + 2)}{\nu(\nu + n)}\right) V_{ij} V_{kl} + \left(\frac{\nu + n + 2}{\nu + n}\right) (V_{ik} V_{jl} + V_{il} V_{jk}). \tag{A.14}$$

A combination of (A.6) and (A.14) yields (17). *Q.E.D.*

*A.3. Proof of the second algorithm, (38)–(44):*

Similar to the first case, suppose we observe $u_t$ where $u_t \sim \chi^2_\nu/\nu$. Then conditional on $u_t$, we have

$$x_t \sim N(\mu, \Psi/u_t). \tag{A.15}$$

Conditional on $f_t$ and $u_t$ and under the assumption that $\alpha = 0_N$, we have

$$r_t|f_t, u_t \sim N(\beta f_t, \Psi_\epsilon/u_t). \tag{A.16}$$

Therefore, conditional on $u_t$, the log-likelihood function of $(r_t', \ f_t')'$ is

$$\begin{aligned}
& \mathcal{L}(r_t, f_t|u_t) \\
= \; & \mathcal{L}(r_t|f_t, u_t) + \mathcal{L}(f_t|u_t) \\
= \; & \frac{N}{2}\left[\sum_{t=1}^{T}\log(u_t) - T\log(2\pi)\right] - \frac{T}{2}\log|\Psi_\epsilon| - \frac{1}{2}\sum_{t=1}^{T}(r_t - \beta f_t)'(\Psi_\epsilon/u_t)^{-1}(r_t - \beta f_t) \\
& + \frac{k}{2}\left[\sum_{t=1}^{T}\log(u_t) - T\log(2\pi)\right] - \frac{T}{2}\log|\Psi_{22}| - \frac{1}{2}\sum_{t=1}^{T}\left[u_t(f_t - \mu_2)'\Psi_{22}^{-1}(f_t - \mu_2)\right].
\end{aligned} \tag{A.17}$$

Note that the first part of the likelihood function has parameters $\beta$ and $\Psi_\epsilon$ and the second part has parameters $\mu_2$ and $\Psi_{22}$. So we can maximize them separately. For the second part, it is clear that

$$\tilde{\mu}_2 = \frac{\sum_{t=1}^{T} u_t f_t}{\sum_{t=1}^{T} u_t}, \tag{A.18}$$

$$\tilde{\Psi}_{22} = \frac{1}{T}\sum_{t=1}^{T} u_t(f_t - \tilde{\mu}_2)(f_t - \tilde{\mu}_2)'. \tag{A.19}$$

Therefore, we can focus our attention to the first part of the conditional likelihood function. Denote $\tilde{Y} = [r_1\sqrt{u_1}, \ r_2\sqrt{u_2}, \ \ldots, \ r_T\sqrt{u_T}]'$ and $\tilde{X} = [f_1\sqrt{u_1}, \ f_2\sqrt{u_2}, \ \ldots, \ f_T\sqrt{u_T}]'$, we can write the first part as

$$\mathcal{L}(r_t|f_t, u_t) = \frac{N}{2}\left[\sum_{t=1}^{T}\log(u_t) - T\log(2\pi)\right] - \frac{T}{2}\log|\Psi_\epsilon| - \frac{1}{2}\sum_{t=1}^{T}(\tilde{Y}_t - \beta\tilde{X}_t)'\Psi_\epsilon^{-1}(\tilde{Y}_t - \beta\tilde{X}_t), \tag{A.20}$$

which has the standard form of the multivariate normality case and hence, conditional on $u_t$, the maximum likelihood estimator of $\beta$ and $\Psi_\epsilon$ under the null are

$$\tilde{\beta} = (\tilde{Y}'\tilde{X})(\tilde{X}'\tilde{X})^{-1}, \tag{A.21}$$

$$\tilde{\Psi}_\epsilon = \frac{1}{T}(\tilde{Y} - \tilde{X}\tilde{\beta}')'(\tilde{Y} - \tilde{X}\tilde{\beta}'). \tag{A.22}$$

This accomplishes the M-step. The E-step is clearly the same as the earlier case. *Q.E.D.*

*A.4. Proof of the asymptotic variance-covariance matrix for $\tilde{\alpha}$ and $\tilde{\beta}$, (31):*

In the derivation below, we use the commutation matrix[11] in addition to he duplication matrix defined earlier in Appendix A.2. Commutation matrix allows us to commute two matrices in a Kronecker product, and is defined as the unique $mn \times mn$ matrix $K_{mn}$ consisting of 0's and 1's such that $K_{mn}\text{vec}(A) = \text{vec}(A')$ for any $m \times n$ matrix $A$. If $m = n$, $K_{nn}$ is simply denoted as $K_n$. Let $A$ be $m \times n$, $B$ be $p \times q$. We have $K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$. From Lange, Little, and Taylor (1989) and our earlier results, we know that $\tilde{\mu}$ and $\tilde{\Psi}$ are asymptotically independent and

$$\begin{aligned} \text{Avar}[\tilde{\mu}] &= \left(\frac{\nu + n + 2}{\nu + n}\right)\Psi, & \text{(A.23)} \\ \text{Avar}[\text{vech}(\tilde{\Psi})] &= \frac{2(\nu + n + 2)}{\nu + n}\left[D_n^+(\Psi \otimes \Psi)D_n^{+\prime} + \frac{\text{vech}(\Psi)\text{vech}(\Psi)'}{\nu}\right]. & \text{(A.24)} \end{aligned}$$

We first prove that

$$\text{Avar}[\text{vec}(\tilde{\beta})] = \left(\frac{\nu + n + 2}{\nu + n}\right)\Psi_{22}^{-1} \otimes \Psi_\epsilon. \tag{A.25}$$

Since $\beta = \Psi_{12}\Psi_{22}^{-1}$, we have

$$\text{vec}(\beta) = (\Psi_{22}^{-1} \otimes I_N)\text{vec}(\Psi_{12}) = (I_k \otimes \Psi_{12})\text{vec}(\Psi_{22}^{-1}). \tag{A.26}$$

It follows that

$$\begin{aligned} \frac{\partial \text{vec}(\beta)}{\partial \text{vec}(\Psi_{12})'} &= \Psi_{22}^{-1} \otimes I_N, & \text{(A.27)} \\ \frac{\partial \text{vec}(\beta)}{\partial \text{vech}(\Psi_{22})'} &= (I_k \otimes \Psi_{12})\frac{\partial \text{vec}(\Psi_{22}^{-1})}{\partial \text{vech}(\Psi_{22})'} \\ &= -(I_k \otimes \Psi_{12})(\Psi_{22}^{-1} \otimes \Psi_{22}^{-1})D_k \\ &= (\Psi_{22}^{-1} \otimes -\beta)D_k. & \text{(A.28)} \end{aligned}$$

Also, note that

$$\begin{aligned} \text{vec}(\Psi_{12}) &= \text{vec}([I_N,\ \text{O}_{N\times k}]\Psi[\text{O}_{k\times N},\ I_k]') \\ &= ([\text{O}_{k\times N},\ I_k] \otimes [I_N,\ \text{O}_{N\times k}])D_n\text{vech}(\Psi), & \text{(A.29)} \\ \text{vech}(\Psi_{22}) &= D_k^+\text{vec}([\text{O}_{k\times N},\ I_k]\Psi[\text{O}_{k\times N},\ I_k]') \\ &= D_k^+([\text{O}_{k\times N},\ I_k] \otimes [\text{O}_{k\times N},\ I_k])D_n\text{vech}(\Psi). & \text{(A.30)} \end{aligned}$$

[11]See Harville (1997, Chapter 16) for a review of the properties of the commutation and the duplication matrices.

Using the delta method, we have

$$
\begin{aligned}
& \mathrm{Avar}[\mathrm{vec}(\tilde{\beta})] \\
= \;& \mathrm{Avar}\left[\left[\Psi_{22}^{-1} \otimes I_N,\ (\Psi_{22}^{-1} \otimes -\beta) D_k\right] \begin{bmatrix} [\mathrm{O}_{k\times N},\ I_k] \otimes [I_N,\ \mathrm{O}_{N\times k}] \\ D_k^{+}\left([\mathrm{O}_{k\times N},\ I_k] \otimes [\mathrm{O}_{k\times N},\ I_k]\right) \end{bmatrix} D_n \mathrm{vech}(\tilde{\Psi})\right] \\
= \;& \mathrm{Avar}\left[\left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) D_n \mathrm{vech}(\tilde{\Psi})\right] \\
= \;& \frac{2(\nu+n+2)}{\nu+n} \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) D_n \left[D_n^{+}(\Psi \otimes \Psi) D_n^{+\prime} + \frac{\mathrm{vech}(\Psi)\mathrm{vech}(\Psi)'}{\nu}\right] \\
& D_n' \times \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
= \;& \left(\frac{\nu+n+2}{\nu+n}\right) \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) \left[(I_{n^2} + K_n)(\Psi \otimes \Psi) + \frac{2\mathrm{vec}(\Psi)\mathrm{vec}(\Psi)'}{\nu}\right] \\
& \times \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)', \qquad \text{(A.31)}
\end{aligned}
$$

where the third equality follows from the identity

$$D_k D_k^{+}(A \otimes A) D_n = (A \otimes A) D_n \qquad \text{(A.32)}$$

for a $k \times n$ matrix $A$, and the fourth equality follows from the identity

$$2 D_n D_n^{+}(\Psi \otimes \Psi) D_n^{+\prime} D_n' = \frac{1}{2}(I_{n^2} + K_n)(\Psi \otimes \Psi)(I_{n^2} + K_n) = (I_{n^2} + K_n)(\Psi \otimes \Psi) \qquad \text{(A.33)}$$

because $2 D_n D_n^{+} = I_{n^2} + K_n$. Using (A.31) and the following identities

$$
\begin{aligned}
& \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) K_n (\Psi \otimes \Psi) \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
= \;& K_{Nk} \left([I_N,\ -\beta] \otimes [\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}]\right) (\Psi \otimes \Psi) \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
= \;& K_{Nk} \left([\Psi_{\epsilon},\ \mathrm{O}_{N\times k}] \otimes [\beta',\ I_k]\right) \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
= \;& \mathrm{O}_{Nk\times Nk}, \qquad \text{(A.34)}
\end{aligned}
$$

$$\left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) \mathrm{vec}(\Psi) = \mathrm{vec}\left([I_N,\ -\beta] \Psi [\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}]'\right) = 0_{Nk}, \qquad \text{(A.35)}$$

we have

$$
\begin{aligned}
\mathrm{Avar}[\mathrm{vec}(\tilde{\beta})] &= \left(\frac{\nu+n+2}{\nu+n}\right) \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) (\Psi \otimes \Psi) \left([\mathrm{O}_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
&= \left(\frac{\nu+n+2}{\nu+n}\right) (\Psi_{22}^{-1} \otimes \Psi_{\epsilon}). \qquad \text{(A.36)}
\end{aligned}
$$

As $\alpha = \mu_1 - \beta\mu_2$, we have

$$
\begin{aligned}
\frac{\partial \alpha}{\partial \mu'} &= [I_N,\ -\beta], \qquad \text{(A.37)} \\
\frac{\partial \alpha}{\partial \mathrm{vec}(\beta)'} &= -\mu_2' \otimes I_N. \qquad \text{(A.38)}
\end{aligned}
$$

Using the delta method and the fact that $\tilde{\mu}$ and $\tilde{\beta}$ are asymptotically independent, we have

$$\begin{aligned}\text{Avar}[\tilde{\alpha}] &= \left(\frac{\nu+n+2}{\nu+n}\right)\left([I_N,\ -\beta]\Psi[I_N,\ -\beta]' + (-\mu_2'\otimes I_N)(\Psi_{22}^{-1}\otimes\Psi_\epsilon)(-\mu_2\otimes I_N)\right)\\ &= \left(\frac{\nu+n+2}{\nu+n}\right)(1+\mu_2'\Psi_{22}^{-1}\mu_2)\Psi_\epsilon. \end{aligned} \tag{A.39}$$

Similarly, the asymptotic covariance between $\tilde{\alpha}$ and $\text{vec}(\tilde{\beta})$ is given by

$$\begin{aligned}\text{Acov}[\tilde{\alpha},\text{vec}(\tilde{\beta})] &= (-\mu_2'\otimes I_N)\text{Avar}[\text{vec}(\tilde{\beta})]\\ &= (-\mu_2'\otimes I_N)\left(\frac{\nu+n+2}{\nu+n}\right)(\Psi_{22}^{-1}\otimes\Psi_\epsilon)\\ &= \left(\frac{\nu+n+2}{\nu+n}\right)(-\mu_2'\Psi_{22}^{-1}\otimes\Psi_\epsilon). \end{aligned} \tag{A.40}$$

Note that the expressions so far are written in terms of $\Psi_{22}$ and $\Psi_\epsilon$ but not in terms of the variance of $f_t$ and $\epsilon_t$. For comparison with the asymptotic variance of $\hat{\alpha}$ and $\hat{\beta}$, we use the fact that $\Sigma = \nu\Psi_\epsilon/(\nu-2)$ and $V_{22} = \nu\Psi_{22}/(\nu-2)$ and write the asymptotic variance of $\tilde{\alpha}$ and $\text{vec}(\tilde{\beta})$ as

$$\text{Avar}\begin{bmatrix}\tilde{\alpha}\\ \text{vec}(\tilde{\beta})\end{bmatrix} = \frac{\nu+n+2}{\nu+n}\begin{bmatrix}\left(\frac{\nu-2}{\nu}\right)+\mu_2'V_{22}^{-1}\mu_2 & -\mu_2'V_{22}^{-1}\\ -V_{22}^{-1}\mu_2 & V_{22}^{-1}\end{bmatrix}\otimes\Sigma, \tag{A.41}$$

which is the expression in (31). Although not provided here, it can be shown that the asymptotic variance of $\tilde{\alpha}$ and $\tilde{\beta}$ remains the same even when the degrees of freedom $\nu$ is unknown.

For the asymptotic variance of the OLS estimator $\hat{\alpha}$ and $\text{vec}(\hat{\beta})$ under the multivariate $t$-distribution, we have from Geczy (2001) that

$$\text{Avar}\begin{bmatrix}\hat{\alpha}\\ \text{vec}(\hat{\beta})\end{bmatrix} = \begin{bmatrix}1+\left(\frac{\nu-2}{\nu-4}\right)\mu_2'V_{22}^{-1}\mu_2 & -\left(\frac{\nu-2}{\nu-4}\right)\mu_2'V_{22}^{-1}\\ -\left(\frac{\nu-2}{\nu-4}\right)V_{22}^{-1}\mu_2 & \left(\frac{\nu-2}{\nu-4}\right)V_{22}^{-1}\end{bmatrix}\otimes\Sigma. \tag{A.42}$$

This completes the proof. *Q.E.D.*

## References


Affleck-Graves, J., and B. McDonald, 1989. Nonnormalities and tests of asset pricing theories, Journal of Finance 44, 889–908.

Alexander, C., 2002, Market Models: A Guide to Financial Data Analysis, Wiley, New York.

Avramov, D., 2004, Stock return predictability and asset pricing models, Review of Financial Studies 17, 699–738.

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

Bollerslev, T., R. F. Engle, and D. B. Nelson, 1994, ARCH models, in R. F. Engle and D. McFadden (ed.), Handbook of Econometrics IV, North-Holland, Amsterdam.

Bollerslev, T., 2001, Financial econometrics: past developments and future challenges, Journal of Econometrics 100, 41–51.

Branco, M. D., and D.K. Dey, 2001, A general class of multivariate skew-elliptical distributions, Journal of Multivariate Analysis 79, 99–113.

Chamberlain, G., 1983, A characterization of the distributions that imply mean-variance utility functions, Journal of Economic Theory 29, 185–201.

Chib, S., R. C. Tiwari, and R. Jammalamadaka, 1988, Bayes prediction in regressions with elliptical errors, Journal of Econometrics 38, 349–360.

Cohen, R., J. Coval, and Ľ. Pástor, 2005, Judging fund managers by the company they keep, Journal of Finance 60, 1057–1096.

Cremers, M., 2002, Stock return predictability: A Bayesian model selection perspective, Review of Financial Studies 15, 1223–1249.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977, Maximum likelihood from in-complete data via the EM algorithm, Journal of the Royal statistical Society Series B 39, 1–38.

Dufour, J., L. Khalaf, and M. Beaulieu, 2003, Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models, Oxford Bulletin of Economics and Statistics 65, 891–906.

Engle, R. F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation, Econometrica 50, 987–1007.

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

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

Fama E. F., and K. R. French, 1996, Multifactor explanations of asset pricing anomalies, Journal of Finance 51, 55–84.

Ferson, W. E., 2003, Tests of multifactor pricing models, volatility bounds and portfolio performance, NBER working paper 9441.

Geczy, C. C., 2001, Some generalized tests of mean-variance efficiency and multifactor model performance, Working Paper, University of Pennsylvania.

Geczy, C. C., D. K. Musto, and A. V. Reed, 2002, Stocks are special too: An analysis of the equity lending market, Journal of Financial Economics 66, 241–269.

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

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

Harvey, C. R., J. Liechty, M. Liechty, and P. Müller, 2003, Portfolio selection with higher moments, Working Paper, Duke University.

Harville, D. A., 1997, Matrix Algebra from a Statistician's Perspective, Springer-Verlag, New York.

Knez, P. J., and M. J. Ready, 1997, On the robustness of size and book-to-market in cross-sectional regressions, Journal of Finance 52, 1355–1382.

Lange, K. L., R. A. Little, and J. G. Taylor, 1989, Robust statistical modeling using the $t$ distribution, Journal of the American Statistical Association 408, 881–896.

Laplante, M. J., 2003, Conditional market timing with heteroskedasticity, PhD dissertation, University of Washington.

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.

Liu, C., and D. B. Rubin, 1995, ML estimation of the $t$ distribution using EM and its Extensions, ECM and ECME, Statistica Sinica 5, 19–39.

Lo, A., 2002, The statistics of Sharpe ratios, Financial Analysts Journal 58, 36–53.

Lyon, J. D., B. M. Barber, and C. Tsai, 1999, Improved methods for tests of long-run abnormal stock returns, Journal of Finance 54, 165–201.

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

Mardia, K. V., 1970, Measures of Multivariate skewness and kurtosis with applications, Biometrika 57, 519–530.

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

Osiewalski, J., and M.F. Steel, 1993, Bayesian marginal equivalence of elliptical regression models, Journal of Econometrics 59, 391–403.

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.

Pástor, Ľ., and R. F. Stambaugh, 2002, Mutual fund performance and seemingly unrelated assets, Journal of Financial Economics 63, 315–349.

Richardson, M., and T. Smith, 1993, A test for multivariate normality in stock returns, Journal of Business 66, 295–321.

Ritter, J. R., 1991, The long-run performance of initial public offerings, Journal of Finance 46, 3–27.

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

Seber, G.F., 1984, Multivariate Observations, Wiley, New York.

Shao, J., and D. Tu, 1995, The Jackknife and Bootstrap, Springer-Verlag, New York.

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

Theodossiou, P., 1998, Financial data and the skewed generalized $t$ distribution, Management Science 44, 1650–1661.

Tu, J., and G. Zhou, 2004, Data-generating process uncertainty: what difference does it make in portfolio decisions? Journal of Financial Economics 72, 385–421.

Vasconcellos, K.P., and G.M. Cordeiro, 2000, Bias corrected estimates in multivariate student $t$ regression models, Communications in Statistics 29, 797–822.

Van Praag, B.M.S., and B. M. Wesselman, 1989, Elliptical multivariate analysis, Journal of Econometrics 41 189–203 .

Vorkink, K.P., 2003, Return distributions and improved tests of asset pricing models Review of Financial Studies, 16, 845–874.

Zhou, G., 1993, Asset pricing tests under alternative distributions, Journal of Finance 48, 1927–1942.

Table 1
Normality Test of the Fama-French Portfolios

The table reports the univariate and multivariate sample skewness and kurtosis measures of the Fama-French benchmark portfolios and factors based on monthly returns from July 1963 through December 2002. In addition, it also reports the $p$-values of the skewness and kurtosis tests if the data is assumed to be drawn form a univariate or multivariate normal distribution, or a univariate or multivariate $t$-distribution with degrees of freedom 10, 8, and 6, respectively.

| | | $p$-value (%) | | | | | $p$-value (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Student-$t$ with df | | | | | Student-$t$ with df | | | |
| | Skew. | Normal | 10 | 8 | 6 | Kurt. | Normal | 10 | 8 | 6 | GARCH |
| | | | | | Univariate | | | | | | |
| S1B1 | 0.01 | 94.83 | 96.96 | 97.39 | 98.01 | 5.23 | 0.00 | 4.90 | 12.47 | 33.31 | 61.45 |
| S1B2 | 0.02 | 83.43 | 90.55 | 91.74 | 93.76 | 6.10 | 0.00 | 1.83 | 5.79 | 18.51 | 0.88 |
| S1B3 | −0.10 | 40.52 | 62.87 | 67.42 | 75.89 | 6.07 | 0.00 | 1.89 | 5.87 | 18.80 | 85.47 |
| S1B4 | −0.14 | 20.42 | 46.46 | 53.02 | 63.99 | 6.47 | 0.00 | 1.40 | 4.30 | 14.71 | 77.14 |
| S1B5 | −0.11 | 31.24 | 55.74 | 61.21 | 70.95 | 6.87 | 0.00 | 1.02 | 3.38 | 11.79 | 26.63 |
| S2B1 | −0.29 | 1.12 | 15.33 | 21.56 | 35.77 | 4.43 | 0.00 | 15.75 | 30.76 | 60.72 | 88.20 |
| S2B2 | −0.48 | 0.00 | 2.90 | 6.40 | 14.67 | 5.57 | 0.00 | 3.23 | 9.12 | 25.99 | 63.88 |
| S2B3 | −0.52 | 0.00 | 2.24 | 5.25 | 12.79 | 6.55 | 0.00 | 1.29 | 4.13 | 14.02 | 94.46 |
| S2B4 | −0.41 | 0.01 | 5.22 | 10.37 | 20.65 | 6.67 | 0.00 | 1.18 | 3.83 | 13.13 | 21.08 |
| S2B5 | −0.29 | 0.91 | 14.39 | 20.54 | 34.72 | 6.96 | 0.00 | 0.96 | 3.18 | 11.30 | 33.75 |
| S3B1 | −0.31 | 0.50 | 12.54 | 18.73 | 32.31 | 4.21 | 0.00 | 22.40 | 40.24 | 70.63 | 76.96 |
| S3B2 | −0.61 | 0.00 | 1.14 | 2.88 | 8.72 | 5.89 | 0.00 | 2.34 | 6.72 | 20.95 | 33.83 |
| S3B3 | −0.62 | 0.00 | 1.14 | 2.88 | 8.72 | 5.81 | 0.00 | 2.57 | 7.30 | 21.90 | 44.12 |
| S3B4 | −0.36 | 0.19 | 8.47 | 14.13 | 26.33 | 5.84 | 0.00 | 2.44 | 7.06 | 21.47 | 89.49 |
| S3B5 | −0.45 | 0.00 | 3.92 | 8.07 | 17.55 | 6.91 | 0.00 | 0.98 | 3.25 | 11.47 | 80.39 |
| S4B1 | −0.17 | 12.77 | 38.07 | 44.91 | 57.14 | 4.47 | 0.00 | 14.91 | 29.31 | 59.07 | 62.45 |
| S4B2 | −0.55 | 0.00 | 1.71 | 4.28 | 11.04 | 6.03 | 0.00 | 1.95 | 6.03 | 19.21 | 5.70 |
| S4B3 | −0.42 | 0.01 | 4.54 | 9.21 | 19.13 | 6.06 | 0.00 | 1.91 | 5.95 | 18.94 | 10.44 |
| S4B4 | −0.06 | 58.63 | 75.20 | 78.04 | 83.76 | 5.15 | 0.00 | 5.42 | 13.56 | 35.30 | 51.98 |
| S4B5 | −0.24 | 3.12 | 21.91 | 28.54 | 42.50 | 5.60 | 0.00 | 3.07 | 8.90 | 25.33 | 63.18 |
| S5B1 | −0.17 | 12.53 | 37.69 | 44.59 | 56.87 | 4.47 | 0.00 | 14.88 | 29.30 | 59.04 | 18.4 |
| S5B2 | −0.34 | 0.24 | 9.27 | 15.07 | 27.58 | 4.74 | 0.00 | 9.36 | 21.03 | 47.99 | 40.61 |
| S5B3 | −0.27 | 1.69 | 17.44 | 23.85 | 38.19 | 5.35 | 0.00 | 4.15 | 11.06 | 30.46 | 7.05 |
| S5B4 | 0.03 | 78.26 | 87.20 | 88.84 | 91.64 | 4.43 | 0.00 | 15.74 | 30.76 | 60.69 | 31.91 |
| S5B5 | −0.18 | 11.00 | 35.89 | 42.61 | 54.93 | 3.81 | 0.21 | 43.73 | 64.03 | 87.50 | 91.85 |
| SMB | 0.51 | 0.00 | 2.30 | 5.42 | 12.96 | 7.97 | 0.00 | 0.47 | 1.80 | 7.09 | |
| HML | 0.10 | 38.74 | 61.60 | 66.14 | 75.02 | 5.17 | 0.00 | 5.33 | 13.35 | 34.83 | |
| MKT | −0.48 | 0.00 | 2.94 | 6.44 | 14.82 | 4.91 | 0.00 | 7.46 | 17.31 | 42.61 | |
| | | | | | Multivariate | | | | | | |
| | 172.41 | 0.00 | 0.37 | 6.50 | 61.29 | 1159.66 | 0.00 | 0.34 | 16.84 | 94.11 | |

Table 2
Estimation of Mean and Standard Deviation under Normality versus under $t$

The table reports the maximum likelihood estimates of means and standard deviations (in percentage per month) of Fama-French benchmark portfolios and factors based on monthly returns from July 1963 to December 2002, assuming that the returns are generated from a multivariate normal or $t$-distribution with degrees of freedom 10, 8, and 6, respectively.

| | Mean | | | | Standard Deviation | | | |
|---|---|---|---|---|---|---|---|---|
| | | $t$-distribution | | | | $t$-distribution | | |
| | Normal | df=10 | df=8 | df=6 | Normal | df=10 | df=8 | df=6 |
| S1B1 | 0.135 | −0.112 | −0.117 | −0.121 | 8.283 | 7.567 | 7.751 | 8.146 |
| S1B2 | 0.703 | 0.469 | 0.464 | 0.459 | 7.095 | 6.497 | 6.656 | 6.996 |
| S1B3 | 0.784 | 0.536 | 0.530 | 0.523 | 6.120 | 5.744 | 5.886 | 6.189 |
| S1B4 | 0.995 | 0.771 | 0.766 | 0.761 | 5.689 | 5.345 | 5.476 | 5.755 |
| S1B5 | 1.070 | 0.876 | 0.871 | 0.865 | 5.961 | 5.654 | 5.793 | 6.090 |
| S2B1 | 0.305 | 0.125 | 0.121 | 0.116 | 7.567 | 7.222 | 7.405 | 7.790 |
| S2B2 | 0.567 | 0.418 | 0.416 | 0.413 | 6.118 | 5.948 | 6.102 | 6.423 |
| S2B3 | 0.830 | 0.715 | 0.713 | 0.711 | 5.415 | 5.258 | 5.391 | 5.672 |
| S2B4 | 0.892 | 0.768 | 0.765 | 0.762 | 5.178 | 4.994 | 5.120 | 5.385 |
| S2B5 | 0.926 | 0.753 | 0.748 | 0.743 | 5.765 | 5.541 | 5.681 | 5.977 |
| S3B1 | 0.334 | 0.237 | 0.236 | 0.235 | 6.909 | 6.625 | 6.796 | 7.154 |
| S3B2 | 0.646 | 0.546 | 0.544 | 0.541 | 5.510 | 5.411 | 5.550 | 5.841 |
| S3B3 | 0.659 | 0.588 | 0.587 | 0.585 | 4.984 | 4.916 | 5.044 | 5.310 |
| S3B4 | 0.792 | 0.705 | 0.702 | 0.699 | 4.734 | 4.646 | 4.767 | 5.018 |
| S3B5 | 0.937 | 0.782 | 0.776 | 0.770 | 5.380 | 5.211 | 5.343 | 5.622 |
| S4B1 | 0.459 | 0.306 | 0.300 | 0.294 | 6.163 | 5.883 | 6.034 | 6.351 |
| S4B2 | 0.435 | 0.317 | 0.313 | 0.309 | 5.205 | 5.136 | 5.269 | 5.547 |
| S4B3 | 0.635 | 0.517 | 0.512 | 0.506 | 4.899 | 4.815 | 4.941 | 5.204 |
| S4B4 | 0.758 | 0.621 | 0.614 | 0.608 | 4.688 | 4.633 | 4.756 | 5.010 |
| S4B5 | 0.850 | 0.694 | 0.687 | 0.679 | 5.417 | 5.307 | 5.446 | 5.735 |
| S5B1 | 0.386 | 0.334 | 0.329 | 0.323 | 4.885 | 4.842 | 4.970 | 5.236 |
| S5B2 | 0.415 | 0.383 | 0.380 | 0.377 | 4.602 | 4.601 | 4.726 | 4.981 |
| S5B3 | 0.455 | 0.421 | 0.419 | 0.416 | 4.381 | 4.283 | 4.396 | 4.629 |
| S5B4 | 0.559 | 0.486 | 0.481 | 0.476 | 4.310 | 4.247 | 4.362 | 4.597 |
| S5B5 | 0.523 | 0.429 | 0.421 | 0.412 | 4.798 | 4.751 | 4.877 | 5.137 |
| SMB | 0.210 | 0.126 | 0.127 | 0.129 | 3.280 | 2.897 | 2.967 | 3.118 |
| HML | 0.432 | 0.401 | 0.399 | 0.397 | 3.003 | 2.701 | 2.766 | 2.906 |
| MKT | 0.410 | 0.361 | 0.358 | 0.354 | 4.505 | 4.422 | 4.537 | 4.777 |

Table 3
Alpha and Beta Estimation under Normality versus under $t$

The table reports the maximum likelihood estimates of alphas (in percent) and betas in the Fama-French three-factor model for 25 size and book-to-market ranked portfolios based on monthly returns from July 1963 through December 2002. Two sets of maximum likelihood estimates are reported. The first set assumes the returns and factors are multivariate normally distributed, and the second set assumes the returns are multivariate $t$-distributed with 8 degrees of freedom.

| | $\alpha$ (%) | | $\beta_{MKT}$ | | $\beta_{HML}$ | | $\beta_{SMB}$ | |
|---|---|---|---|---|---|---|---|---|
| Portfolio | Normal | $t_8$ | Normal | $t_8$ | Normal | $t_8$ | Normal | $t_8$ |
| S1B1 | −0.450 | −0.569 | 1.061 | 1.065 | −0.325 | −0.256 | 1.384 | 1.367 |
| S1B2 | −0.004 | −0.082 | 0.969 | 0.994 | 0.074 | 0.085 | 1.322 | 1.236 |
| S1B3 | 0.043 | −0.053 | 0.926 | 0.939 | 0.292 | 0.272 | 1.118 | 1.085 |
| S1B4 | 0.209 | 0.155 | 0.901 | 0.892 | 0.460 | 0.406 | 1.037 | 1.024 |
| S1B5 | 0.150 | 0.137 | 0.975 | 0.958 | 0.675 | 0.626 | 1.089 | 1.111 |
| S2B1 | −0.190 | −0.229 | 1.116 | 1.117 | −0.401 | −0.453 | 1.003 | 1.031 |
| S2B2 | −0.119 | −0.100 | 1.031 | 1.033 | 0.177 | 0.074 | 0.890 | 0.918 |
| S2B3 | 0.084 | 0.136 | 0.987 | 0.974 | 0.420 | 0.320 | 0.760 | 0.798 |
| S2B4 | 0.085 | 0.119 | 0.981 | 0.968 | 0.590 | 0.517 | 0.715 | 0.735 |
| S2B5 | −0.031 | −0.032 | 1.086 | 1.079 | 0.774 | 0.718 | 0.840 | 0.846 |
| S3B1 | −0.071 | −0.060 | 1.084 | 1.095 | −0.445 | −0.472 | 0.725 | 0.729 |
| S3B2 | 0.007 | 0.062 | 1.060 | 1.023 | 0.221 | 0.095 | 0.518 | 0.614 |
| S3B3 | −0.072 | 0.011 | 1.022 | 0.999 | 0.507 | 0.379 | 0.441 | 0.529 |
| S3B4 | 0.011 | 0.070 | 1.001 | 0.981 | 0.669 | 0.558 | 0.387 | 0.462 |
| S3B5 | 0.015 | 0.021 | 1.109 | 1.077 | 0.826 | 0.737 | 0.528 | 0.595 |
| S4B1 | 0.141 | 0.058 | 1.052 | 1.058 | −0.448 | −0.453 | 0.379 | 0.354 |
| S4B2 | −0.173 | −0.163 | 1.100 | 1.069 | 0.263 | 0.141 | 0.207 | 0.290 |
| S4B3 | −0.065 | −0.059 | 1.082 | 1.059 | 0.513 | 0.402 | 0.164 | 0.246 |
| S4B4 | 0.025 | −0.006 | 1.039 | 1.041 | 0.616 | 0.555 | 0.198 | 0.209 |
| S4B5 | −0.049 | −0.067 | 1.177 | 1.155 | 0.839 | 0.758 | 0.260 | 0.308 |
| S5B1 | 0.210 | 0.202 | 0.961 | 0.951 | −0.380 | −0.452 | −0.256 | −0.264 |
| S5B2 | −0.024 | 0.016 | 1.039 | 1.033 | 0.145 | 0.057 | −0.237 | −0.220 |
| S5B3 | −0.026 | −0.005 | 0.988 | 0.976 | 0.288 | 0.254 | −0.232 | −0.209 |
| S5B4 | −0.088 | −0.080 | 1.007 | 1.010 | 0.642 | 0.558 | −0.204 | −0.177 |
| S5B5 | −0.237 | −0.262 | 1.049 | 1.068 | 0.817 | 0.787 | −0.111 | −0.101 |

Table 4
Estimated Alphas of Mutual Funds under Normality versus under $t$ (1998/1–2002/12)

Based on monthly data from January 1998 to December 2002, the first panel of the table reports the Jensen's alpha of five mutual funds estimated under 5 different distributional assumptions: the normal, the $t$-distribution with unknown degree of freedom, and $t$ with 6, 8, and 10 degrees of freedom. The panel also reports the percentage of funds that reverse from a negative alpha when estimated under the normality assumption to a positive alpha when estimated under the $t$-distribution assumption with unknown degrees of freedom, together with their average difference in the two estimated alphas. The second panel provides the corresponding results for funds whose rankings are reversed from a positive alpha to a negative alpha. The third panel reports the percentage of funds that have an annualized difference over 1% to 5%, respectively, in their estimated alphas under the normal and the $t$ assumptions with unknown degrees of freedom.

| | | $t$-distribution | | | |
|---|---|---|---|---|---|
| Fund | Normal | unknown | df=6 | df=8 | df=10 |
| One Group Diversified Mid Cap/B | −0.556 | 0.378 | 0.415 | 0.426 | 0.430 |
| Smith Barney Peachtree Growth Fund/Y | −0.413 | 0.023 | 0.003 | −0.022 | −0.043 |
| iShares:MSCI Germany Index Fund | −0.114 | 0.271 | 0.316 | 0.245 | 0.195 |
| Excelsior Latin America Fund | −0.100 | 0.249 | 0.570 | 0.476 | 0.406 |
| Merrill Lynch World Income Fund/C | −0.070 | 0.252 | 0.241 | 0.219 | 0.202 |
| Percentage of reversals | 0.8 | | | | |
| Average difference in alpha | 0.136 | | | | |
| PBHG Select Equity Fund | 1.352 | −0.374 | -0.050 | 0.130 | 0.272 |
| Thurlow Growth Fund | 1.301 | −0.343 | −0.441 | −0.174 | 0.020 |
| iShares:MSCI Malaysia Index Fund | 1.601 | −0.017 | 0.026 | 0.144 | 0.244 |
| US Global Investors Fds:World Prec Minerals | 0.687 | −0.212 | −0.403 | −0.248 | −0.128 |
| Emerald Select Technology Fund/A | 0.528 | −0.270 | −0.404 | −0.237 | −0.121 |
| Percentage of reversals | 1.4 | | | | |
| Average difference in alpha | 0.376 | | | | |

| | Difference in alphas | | | | |
|---|---|---|---|---|---|
| | 1% | 2% | 3% | 4% | 5% |
| Percentage of Funds | 13.9 | 8.4 | 6.0 | 4.0 | 2.6 |

Table 5
Estimated Alphas of Mutual Funds under Normality versus under $t$ (1993/1–2002/12)

Based on monthly data from January 1993 to December 2002, the first panel of the table reports the Jensen's alpha of five mutual funds estimated under different distributional assumptions: the normal, the $t$-distribution with unknown degree of freedom, and $t$-distributions with 6, 8, and 10 degrees of freedom. The panel also reports the percentage of funds that reverse from a negative alpha when estimated under the normality assumption to a positive alpha when estimated under the $t$-distribution assumption with unknown degrees of freedom, together with their average difference in the two estimated alphas. The second panel provides the corresponding results for funds whose rankings are reversed from a positive alpha to a negative alpha. The third panel reports the percentage of funds that have an annualized difference over 1% to 5%, respectively, in their estimated alphas under the normal and the $t$ assumptions with unknown degrees of freedom.

| | | $t$-distribution | | | |
|---|---|---|---|---|---|
| Fund | Normal | unknown | df=6 | df=8 | df=10 |
| Neuberger Berman Guardian/Trust | −0.438 | 0.223 | 0.199 | 0.187 | 0.177 |
| Wilmington Short Intermediate Bond/Instl | −0.157 | 0.461 | 0.460 | 0.461 | 0.461 |
| Excelsior Latin America Fund | −0.256 | 0.250 | 0.353 | 0.273 | 0.213 |
| Morgan Stanley High Yield Securities/D | −0.257 | 0.226 | 0.122 | 0.066 | 0.026 |
| AIM High Yield Fund/A | −0.002 | 0.375 | 0.270 | 0.223 | 0.189 |
| Percentage of reversals | 0.5 | | | | |
| Average difference in alpha | 0.263 | | | | |
| Green Century Balanced Fund | 0.153 | −0.454 | −0.352 | −0.287 | −0.237 |
| Monterey:OCM Gold Fund | 0.311 | −0.269 | −0.318 | −0.262 | −0.211 |
| First American Small Cap Growth/A | 0.204 | −0.306 | −0.209 | −0.156 | −0.118 |
| Brown Advisory Small Cap Growth Fund/A | 0.026 | −0.448 | −0.401 | −0.347 | −0.309 |
| New Alternatives Fund | 0.058 | −0.392 | −0.336 | −0.305 | −0.278 |
| Percentage of reversals | 1.9 | | | | |
| Average difference in alpha | 0.227 | | | | |

| | Difference in alphas | | | | |
|---|---|---|---|---|---|
| | 1% | 2% | 3% | 4% | 5% |
| Percentage of Funds | 15.3 | 6.9 | 3.2 | 1.1 | 0.6 |

Table 6
Multivariate Tests of the CAPM

The table reports both the Gibbons, Ross, and Shanken (1989) test and the likelihood ratio test under the $t$ distribution for the CAPM restrictions

$$H_0: \quad \alpha = 0_N$$

in regressions of the excess returns of Fama-French 25 size and book-to-market ranked portfolios on the excess return on the market portfolio:

$$r_{it} = \alpha_i + \beta_i MKT_t + \epsilon_t,$$

where the data are monthly returns from July 1963 through December 2002. The tests are carried out for the entire sample period as well as for two subperiods.

| Model | GRS | $p$-value (%) | $\text{LRT}_t$ | $p$-value (%) |
|---|---|---|---|---|
| | | July 1963 — December 2002 | | |
| Normal | 3.985 | 0.00 | | |
| $t$ (df=10) | | | 106.67 | 0.00 |
| $t$ (df=8) | | | 107.73 | 0.00 |
| $t$ (df=6) | | | 108.93 | 0.00 |
| | | July 1963 — June 1983 | | |
| Normal | 1.434 | 9.03 | | |
| $t$ (df=10) | | | 41.21 | 2.18 |
| $t$ (df=8) | | | 41.65 | 1.96 |
| $t$ (df=6) | | | 42.17 | 1.72 |
| | | July 1983 — December 2002 | | |
| Normal | 5.276 | 0.00 | | |
| $t$ (df=10) | | | 123.89 | 0.00 |
| $t$ (df=8) | | | 125.17 | 0.00 |
| $t$ (df=6) | | | 126.68 | 0.00 |

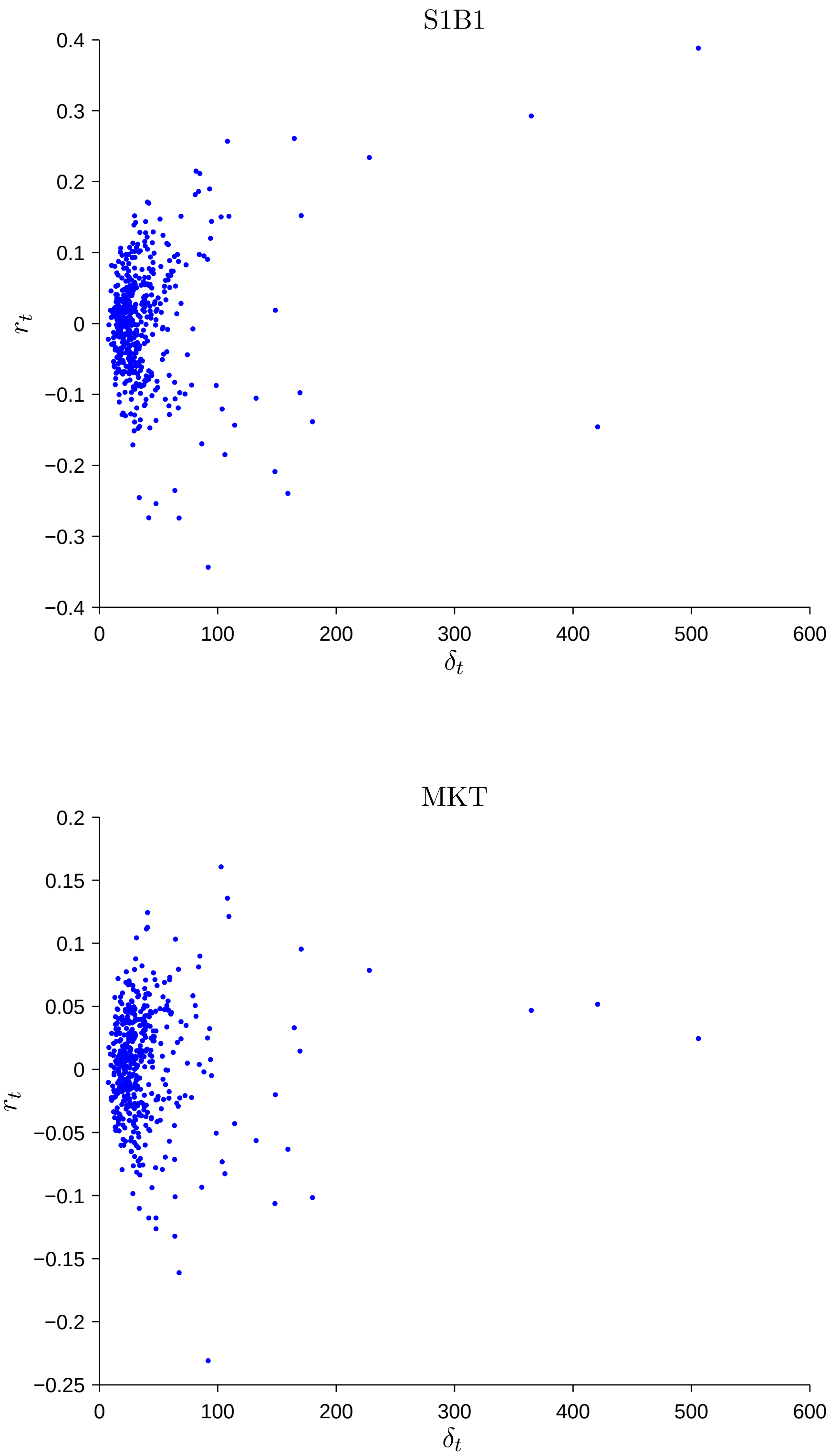


**Figure 1**
**Plots of returns vs. distance measure from the center for S1B1 and MKT**

The figure presents the plots of the monthly excess returns ($r_t$) of S1B1 and MKT vs. a measure of its distance from the center ($\delta_t$) of the mutlivariate $t$-distribution over the period of July 1963 to December 2002. S1B1 is the portfolio that has the smallest size and book-to-market out of the 25 Fama and French (1993) portfolios, and MKT is the value-weighted combined NYSE-AMEX-NASDAQ market portfolio.