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

Raymond Kan and Guofu Zhou[1]

This version: September, 2016

[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, Canlin Li, 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 and the 2006 China International Conference in Finance. 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 finance are based on the normality assumption, but this assumption is firmly rejected by data due to fat tails. In this paper, we propose using a multivariate t-distribution, which fits well the data, as a simple alternative to examine the robustness of many existing results. We find that, under the multivariate t-distribution, the asymptotically most efficient estimator of the expected return of an asset can be drastically different from the sample average return. For example, the annual difference in the estimated expected returns under normal and t is 2.1% 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 distribution.

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 whether these findings are robust to alternative multivariate distributional assumptions.

The multivariate t-distribution is a well known alternative to the multivariate normal distribution. 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 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 July 1963 to December 2015 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 7 degrees of freedom has a p-value of 26.88%. Although the t-distribution is symmetric, its sample skewness is highly volatile that can generate the observed sample skewness in the data with high probability. Second, with the algorithms

[1] Blattberg and Gonedes (1974) seems the first to use t-distribution to model stock returns in finance. Later applications of t-distribution and generalized t-distribution 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 multivariate t without providing the results estimated based on the multivariate t-distribution assumption.

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, 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. However, the GARCH effect is generally much weaker in monthly return data as compared with daily return data, so it is not totally unreasonable to ignore the GARCH effect for a typical asset pricing study that involves monthly data. In addition, 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 difficult 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 economically important difference in estimating expected returns of assets. For example, the expected excess return for Fama and French's (1993) SMB factor is 0.221%/month when estimated under normality, but is only 0.102%/month when estimated under multivariate t with 7 degrees of freedom, implying an annual difference of 1.428%. 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 2.1% for the portfolio that is in the smallest size and book-to-market quintiles. In fact, over our sample period, most of the estimated expected returns 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.[2] 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 portfolios 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 -0.451%/month under normality becomes a winner fund

[2] 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.

with estimated alpha of 0.067%/month under an optimally estimated t-distribution. On the other hand, a winner fund with an estimated alpha of 1.328%/month under normality turns into a loser with an estimated alpha of -0.727%/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-distribution than the normal. This suggests that non-normality modeling by using the t-distribution 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 the empirical evidence that makes the case for modeling the data as t distributed rather than the normal. Section 2 presents both the estimation technique under t-distribution 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-distribution 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. The empirical results show that the multivariate normality assumption is unequivocally rejected by the data, but a suitable multivariate 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.[3] 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 2015.

1.2. Normality 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 μ 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, δ_1

[3] 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.

is simply zero, and δ_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 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 100,000 draws in what follows.

This procedure can also be applied to test whether or not the data follow a suitable multivariate t-distribution with ν degrees of freedom. The multivariate 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 multivariate normal as ν goes to infinity, and hence the usual multivariate normal distribution is a special limiting case of multivariate t. In order to apply the earlier procedure, one simulates data from a standard multivariate 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 multivariate t-distribution with degrees of freedom $\nu = 8$, 7 and 6, the p-value goes up from 4.18% to 26.88% and 80.44%. Therefore, a multivariate t-distribution with $\nu = 7$ 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 univariate skewness

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

γ_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 γ_1 or D_1. However, reporting γ_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 $\nu = 8$ at the usual 5% level. Nevertheless, the multivariate skewness test and many of the univariate skewness tests cannot reject a multivariate t-distribution with $\nu = 7$ 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 ν is small, so as to imply a large probability for observing a large sample skewness even though the true distribution, assumed multivariate 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.[4] 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 ν, 7 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 κ measures the excess kurtosis relative to the normal. Equation (8) implies that the population kurtosis goes to

[4] 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 (2010) for an excellent survey and new developments on skewness studies.

infinity as ν 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 ν. Although the p-value for the sample kurtosis test is greater with a smaller ν (but greater than 4), it does not imply a smaller ν actually fits the data better. It is because the sample kurtosis can in fact falls out of a reasonable left tail of the distribution when ν is too small. For example, the p-value of 80.04% when $\nu = 6$ implies that the observed sample multivariate kurtosis falls into the 19.96% mass of the distribution from the left, and it is no better than the p-value of 26.88% for the case of $\nu = 7$. Later on, we report an estimate of $\nu = 7.5385$ 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 ν, and we cannot be sure which one is closer to the true ν. For the sake of robustness, we report most of our results based on three different values of ν (6, 7 and 8), so we can determine whether our reported results are sensitive to the choice of degrees of freedom.

2. Impacts on Estimating Mean, Variance and Sharpe Ratios

After rejecting multivariate normality and accepting multivariate t-distribution 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 multivariate t-distribution. With these estimates, we can then address the impact of the multivariate 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 μ and V are their sample analogues, $\hat{\mu}$ and $\hat{V}$. The accuracy of the sample averages to estimating the population mean can be judged by its asymptotic variance-covariance matrix,[5]

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

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

[5] In this paper, we use $\mathrm{Avar}[\hat{\theta}]$ to stand for the limiting variance of $\sqrt{T}(\hat{\theta} - \theta)$ as $T \to \infty$, where $\hat{\theta}$ is a consistent estimator of θ.

will be different.

Indeed, under multivariate t-distribution, the asymptotically most efficient estimator of the parameters is the solution of maximizing the log-likelihood function based on the multivariate 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 multivariate t-distribution, i.e., the solution to maximizing $\mathcal{L}$. Starting from any initial estimate of μ and Ψ, say $\tilde{\mu}^{(1)} = \hat{\mu}$ and $\tilde{\Psi}^{(1)} = \hat{\Psi} = (\nu-2)\hat{V}/\nu$, we can obtain iterative estimates via

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

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

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

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 ν here is assumed known. This may be reasonable because the likely values for ν can be assessed by using the kurtosis test. When one is concerned about the fact that ν is unknown, one can treat ν as an additional parameter and estimates it directly from the data. Then the following extended algorithm due to Liu and Rubin

(1995) can be used. Starting with any initial estimate of ν, say $\tilde{\nu}^{(1)} = 7$, one can update a new estimate of ν in the $(i+1)$-th step by solving

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

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

$$u_t^{(i+1)}(\nu) = \frac{\nu+n}{\nu + \left(x_t - \tilde{\mu}^{(i+1)}\right)'\left(\tilde{\Psi}^{(i+1)}\right)^{-1}\left(x_t - \tilde{\mu}^{(i+1)}\right)}. \quad (15)$$

Hence, the earlier EM algorithm can be combined with this one so that it is still applicable when ν 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 ν. 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 ν, practical implementation of the algorithm is still straightforward. Indeed, even if we treat ν 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 ν chosen, the algorithm has always quickly converged to an estimated value $\tilde{\nu} = 7.5385$ 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 ν to assess the sensitivity of ν on the statistical inference, one can estimate ν 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 $\nu > 2$,

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

This says that, when the data is multivariate t distributed rather than the multivariate normal, the sample mean $\hat{\mu}$ is no longer the asymptotically most efficient estimate of μ, but the maximum likelihood estimator $\tilde{\mu}$ is. The relative efficiency is measured by ρ. The greater the ρ, the better the maximum likelihood method. In our application with $n = 28$, $\nu = 7$, we have $\rho = 0.2449$, implying that the maximum likelihood estimator $\tilde{\mu}$ is 24% 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 μ 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, under the multivariate t-distribution, 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 ν gets smaller. This makes intuitive sense: the smaller the ν, 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 multivariate 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, ρ_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 ρ_v is also the efficiency gain for estimating the standard deviation of asset returns by the maximum likelihood method under multivariate t-distribution assumption. When $n = 28$ and $\nu = 8$, we have $\rho_v = 39.59\%$. This says that compared with the sample variance or standard deviation, the maximum likelihood procedure under t improves the estimation efficiency in estimating the variance or standard deviation by about 40%.

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 sample Sharpe ratios that are constructed based on the sample mean and variance. Given our improved estimates on the means and variances, it is naturally to ask the

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

question that how much the optimal ML estimates under multivariate t-distribution 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 sample 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 multivariate t is $\tilde{\theta}_p = \tilde{\mu}_p/\tilde{\sigma}_p = w'\tilde{\mu}/(w'\tilde{V}w)^{\frac{1}{2}}$. Based on (16), (17), and using the delta method, it can be shown that

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

and

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

In our applications here with $n = 28$ and $\nu = 7$, it is clear that $\mathrm{Avar}[\hat{\theta}_p] = 1 + \theta_p^2$ and $\mathrm{Avar}[\tilde{\theta}_p] = 0.755 + 0.604\theta_p^2$. Hence, regardless the exact combination of the portfolio, the new estimate that is based on the ML under multivariate t has at least a 24% reduction in asymptotic variance as compared with the traditional sample estimator.

2.2. Empirical results

After providing the estimation method and the associated asymptotic theory, we now present 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 $\nu = 8$, 7 and 6, respectively. As discussed earlier, a value of $\nu = 7$ 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 ν. It is striking that the expected returns estimated under t for most of the assets are smaller than those estimated under normality. For example, the sample average excess returns for the size (SMB) factor is 0.221%, but their estimated expected excess return under multivariate t with $\nu = 7$ are only 0.102%, implying an annual difference of 1.428%. Such differences in some of the portfolios are even larger. For instance, the S1B1 and S1B4 portfolios have an annual difference of 2.1% and 2.22% between the two estimates of expected return, respectively.

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 μ 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 δ_t measures how far the data is from its center. Since w_t is a decreasing function of δ_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 δ_t. In Figure 1, we provide plots of r_t against δ_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 δ_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 δ_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 $\hat{V}_{11}^{\frac{1}{2}} = 7.942\%$ (per month) and the maximum likelihood one $\tilde{V}_{11}^{\frac{1}{2}} = 7.582\%$ under the t-distribution with 7 degrees of freedom. The same is also true for the market excess return whose standard deviations in the two cases are 4.440% and 4.474%, very close to each other. The small differences in estimating the standard deviations are consistent with the general belief 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 0.316% when $\nu = 7$, so the estimate 7.942% is very accurate. While the estimated standard error of $\hat{\mu}_1$ is also 0.316%, the sample mean of S1B1 is only 0.221%, so a standard error of 0.316% suggests 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 one of the most widely reported measures despite its restrictive assumptions. We show in this section that a relaxation of the 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 multivariate t-distribution assumption, and then apply them to mutual fund performance analysis.

3.1. The multivariate 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 ϵ_t is an $N \times 1$ vector of residuals with zero mean and a non-singular covariance matrix. To relate α and β to the earlier parameters μ 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 α, β and the earlier parameters μ, V obey the following relationship:

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

Denote Σ as the covariance matrix of ϵ_t,

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

It should be noted that under the multivariate normality assumption, Σ is also the variance of ϵ_t conditional on f_t. However, once the normality assumption is removed, this will not necessarily be the case. Indeed, when the data follow a multivariate t-distribution with ν degrees of freedom, the mean of r_t conditional on f_t is still a linear function of f_t as above, but 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 multivariate t regression model versus those in the econometrics and statistics literature where f_t is treated as fixed and ϵ_t is assumed to be multivariate 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 ϵ_t is time-varying.[7] As shown below, this will have important implications for both parameter estimation and asset pricing tests.

[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.

To assess the estimation accuracy of the alphas and betas under multivariate t-distribution assumption, we need to derive the associated asymptotic standard errors of the estimates which are not available previously. The ML estimates of $\tilde{\alpha}$ and $\tilde{\beta}$ under multivariate t-distribution assumption are easily obtained from equation (28) by replacing μ 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 multivariate t-distribution, $\tilde{\alpha}$, 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 $\nu=7$, the percentage standard reduction of $\tilde{\alpha}$ ranges from 24.49% to 36.57%. 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 $\nu=7$, the percentage variance reduction of $\tilde{\beta}$ is 36.57%.

3.2. Empirical alpha estimates

Table 3 provides the estimates of the alphas and betas under normality and t (with $\nu=7$) for the Fama-French data. It is seen that there can be large differences in the alpha estimates. For example, the estimated alphas for S5B4 and S5B5 are -0.212% and -0.151% under multivariate normal distribution, but change to -0.138% and -0.224% under the multivariate t-distribution with $\nu = 7$. 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 multivariate t or multivariate normal. However, there are some substantial differences in the HML betas. The HML betas for S3B2 and S4B2 have reduced significantly from 0.173 and 0.190 to 0.054 and 0.080. 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 α 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). We consider only domestic equity funds. For funds with multiple classes, we only consider the class with the longest history. Out of this subset of funds, we select the ones with complete monthly return data in the last 5 or 10 years at the end of December 2015. In this way, we obtain 5266 and 1967 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 sample alphas, what is the percentage of funds with the sign of their alphas reversed when estimated based on a reasonable t-distribution? Moreover, what is the magnitude of the reversals? Similarly, we can ask the same questions for funds that have positive sample alphas.

Table 4 reports the results using mutual funds in the past five years from January 2011 to December 2015. 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 estimated alphas under normality and under t whose degrees of freedom is treated as unknown and estimated from the data. For example, iPath ETN Global Carbon Class A shares's alpha changes from a negative value of -0.451% per month to a positive value of 0.067% 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 7.6 (not reported in the table), which implies a distribution that is quite different from the normal. For comparison, we also provide those alphas under fixed degrees of freedom for the t-distribution, and find similar results. Among all the 5266 funds there are 2.28% of them that have reversals from negative alphas to positive ones. To assess the magnitude of the reversals, the average difference of the reversed funds, 0.081%/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 in the first panel. For example, Direxon Daily Small Cap Bear 3X shares has a huge alpha value of 1.328% per month under the normality assumption, but this value is significantly reduced to a negative alpha of -0.727% per month. In addition, the average magnitude of the reversals is now 0.151%, quite a bit more than the 0.081% in the previous panel.

While the percentage of funds that see reversal of the signs of their estimated alphas is not too high, it does not imply the t-distribution has small impact on the estimated alphas. There can be funds that see substantial changes in its estimated alphas under but yet the signs of the estimated alphas remain the same under both distributional assumptions. To further assess the difference in estimated alphas 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 11.62% 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 1.86% of the funds whose alpha estimates differ by 3% or more per year.

Table 5 provides corresponding results for 1967 funds whose return data is available for 10 years from January 2006 to December 2015. With the increase in the length of sample period to 10 years, we now see even more reversals of estimated alphas in Table 5 than in the case of Table 4, mostly due to the fact that there is a higher probability of observing more extreme observations with a longer sample period. Overall, the fund returns data suggests economically significant differences in 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 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 multivariate normality assumption, this can be tested by the well-known Gibbons, Ross, and Shanken (1989) test,

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

where $\hat{\alpha}$ and $\hat{\Sigma}$ are obtained from either linear regressions or from the relations between them and $\hat{\mu}$ and $\hat{V}$. Under the multivariate normality assumption, it is well-known that the GRS test is simply a transformation of the likelihood ratio test. However, once the multivariate normality assumption on $(r_t',\ f_t')'$ is replaced by a multivariate t-distribution assumption, one should be cautious in using the GRS test as it is no longer valid under the multivariate t-distribution assumption on $(r_t',\ f_t')'$.

Now, if the returns follow a multivariate t-distribution with ν 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. Under the null, $(\mu,\ \Psi)$ can be mapped into $(\beta,\ \Psi_\epsilon,\ \mu_2,\ \Psi_{22})$ by the following relation:

$$\mu = \begin{bmatrix} \beta\mu_2 \\ \mu_2 \end{bmatrix}, \tag{38}$$

$$\Psi = \begin{bmatrix} \beta\Psi_{22}\beta' + \Psi_\epsilon & \beta\Psi_{22} \\ \Psi_{22}\beta' & \Psi_{22} \end{bmatrix}, \tag{39}$$

where $\Psi_\epsilon = (\nu-2)\Sigma/\nu$. With an initial estimate of $(\beta^{(1)}, \Psi_\epsilon^{(1)}, \mu_2^{(1)}, \Psi_{22}^{(1)})$, the EM algorithm for the estimation of $(\beta,\ \Psi_\epsilon,\ \mu_2,\ \Psi_{22})$ can be obtained by the following iterative procedure:

$$u_t^{(i)} = \frac{\nu+n}{\nu + \left(r_t - \tilde{\beta}^{(i)} f_t\right)' \left(\Psi_\epsilon^{(i)}\right)^{-1} \left(r_t - \tilde{\beta}^{(i)} f_t\right) + \left(f_t - \tilde{\mu}_{2r}^{(i)}\right)' \left(\Psi_{22r}^{(i)}\right)^{-1} \left(f_t - \tilde{\mu}_{2r}^{(i)}\right)}, \tag{40}$$

$$\tilde{Y}^{(i)} = \left[r_1\sqrt{u_1^{(i)}},\ r_2\sqrt{u_2^{(i)}},\ \ldots,\ r_T\sqrt{u_T^{(i)}}\right]', \tag{41}$$

$$\tilde{X}^{(i)} = \left[f_1\sqrt{u_1^{(i)}},\ f_2\sqrt{u_2^{(i)}},\ \ldots,\ f_T\sqrt{u_T^{(i)}}\right]', \tag{42}$$

$$\tilde{\beta}^{(i+1)} = \left(\tilde{Y}^{(i)\prime}\tilde{X}^{(i)}\right)\left(\tilde{X}^{(i)\prime}\tilde{X}^{(i)}\right)^{-1}, \tag{43}$$

$$\tilde{\Psi}_\epsilon^{(i+1)} = \frac{1}{T}\left(\tilde{Y}^{(i)} - \tilde{X}^{(i)}\tilde{\beta}^{(i+1)\prime}\right)'\left(\tilde{Y}^{(i)} - \tilde{X}^{(i)}\tilde{\beta}^{(i+1)\prime}\right), \tag{44}$$

$$\tilde{\mu}_{2r}^{(i+1)} = \frac{\sum_{t=1}^T u_t^{(i)} f_t}{\sum_{t=1}^T u_t^{(i)}}, \tag{45}$$

$$\tilde{\Psi}_{22r}^{(i+1)} = \frac{1}{T}\sum_{t=1}^T u_t^{(i)}\left(f_t - \tilde{\mu}_{2r}^{(i+1)}\right)\left(f_t - \tilde{\mu}_{2r}^{(i+1)}\right)', \tag{46}$$

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

$$\mathrm{LRT} \equiv 2\left(\frac{T-(N/2)-k-1}{T}\right)\left[\log\mathcal{L}(\tilde{\mu},\tilde{\Psi}) - \log\mathcal{L}(\tilde{\mu}_r,\tilde{\Psi}_r)\right] \overset{A}{\sim} \chi_N^2, \tag{47}$$

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]

In the above procedure, we assume the degrees of freedom ν is known. When ν is unknown, we need to modify the procedure. Under the alternative, we can easily estimate ν by introducing an extra step in the iteration to update ν as outlined in (14). Under the null of $\alpha = 0_N$, we can also update ν in the $(i+1)$-th step by solving (14) except that $u_t^{(i+1)}(\nu)$ is defined as

$$u_t^{(i+1)}(\nu) = \frac{\nu+n}{\nu + \left(x_t - \tilde{\mu}_r^{(i+1)}\right)'\left(\tilde{\Psi}_r^{(i+1)}\right)^{-1}\left(x_t - \tilde{\mu}_r^{(i+1)}\right)}. \tag{48}$$

Therefore, we can compute the likelihood ratio test under the multivariate t-distribution assumption

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

even when the degrees of freedom are unknown:

$$\mathrm{LRT} \equiv 2\left(\frac{T-(N/2)-k-1}{T}\right)\left[\log \mathcal{L}(\tilde{\mu}, \tilde{\Psi}, \tilde{\nu}) - \log \mathcal{L}(\tilde{\mu}_r, \tilde{\Psi}_r, \tilde{\nu}_r)\right] \overset{A}{\sim} \chi^2_N, \tag{49}$$

where $\tilde{\nu}$ and $\tilde{\nu}_r$ are the ML estimates of ν under the alternative and null, respectively.

As the CAPM of Sharpe (1964) and Lintner (1965) is of fundamental importance in finance, it is of interest to use it as a first example to illustrate our test. This amounts to testing (35) with the single and theory-motivated market factor in (36). Table 6 reports the testing results based on the GRS test and the likelihood ratio test when the test assets are Fama and French's (1993) 25 size and book-to-market ranked portfolios. Over the entire sample period of July 1963 to December 2015 as well as two subperiods, both tests reject the CAPM strongly with virtually zero p-values, whether we assume the underlying distribution is multivariate normal or multivariate t. However, given the strong rejection of the multivariate normality assumption, one cannot draw firm conclusions about the rejections from test statistics that assume multivariate normality. With the test developed here also reaffirming the rejections reached by the GRS test under multivariate normality, one can claim that the rejection is indeed caused by the failure of the model rather than the violation of the restrictive multivariate normality assumption. Hence, even in cases both the GRS and the multivariate 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, there 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. In Table 7, we report the multivariate tests of the Fama and French (1993) 3-factor model using the same test assets as in Table 6. While both the GRS test and the likelihood ratio test under the multivariate t-distrubution reject the Fama-French 3-factor model using the full sample period, there are difference in the results of these two tests in the subperiods. For example, in the first subperiod of July 1963 to September 1989, we find that the GRS test does not reject the Fama-French at even the 10% level, but the likelihood ratio test computed under multivariate t with $\nu = 8$, 7, 6, or unknown degrees of freedom all suggest rejection of the Fama-French 3-factor model at the 5% level. This is clearly a case where the multivariate t based test makes a difference by suggesting the Fama-French 3-factor model fails to hold in the first subperiod while the GRS test is unable to do so.[9]

[9] Although not reported here, we apply the same analysis to 10 size-sorted portfolios of the NYSE and find cases

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

$$\mathrm{GRS}_g = \left(\frac{T-N-k}{N}\right) \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1+(1+\kappa)\hat{\mu}_2'\hat{V}_{22}^{-1}\hat{\mu}_2} \ \sim\ F_{N,T-N-k}, \tag{50}$$

where κ is the standardized kurtosis as defined by (8). Although not reported in the 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 17.07%, GRS_g has p-values of 20.34%, 21.46% and 23.74% under t with 8, 7 and 6 degrees of freedom, respectively. An intuition for GRS_g having less power than LRT is that the former is based on inefficient parameter estimates when the returns are multivariate t rather than multivariate 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 restrictive, it is more general and more realistic than the widely used multivariate 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.

An often asked question is why we single out the multivariate t-distribution from the class of elliptical distributions, of which the multivariate t-distribution is only a special case and there are countless others. The major reason is that the multivariate t-distribution appears to be the simplest

where the two likelihood ratio tests give conflicting conclusions on the validity of an asset pricing model.

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 multivariate 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 multivariate normality assumption. The trade-off between the two approaches is precision and generality. Clearly, the multivariate 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 departure from the normality assumption. 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 re-

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

jected by the data, suitable multivariate t-distributions pass standard skewness and kurtosis tests. In addition, parameter estimation and tests under multivariate t-distribution can now be implemented almost as easily as under the multivariate 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 multivariate 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 July 1963 to December 2015, 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 multivariate normal versus under the multivariate t. In addition, the results on multivariate tests of asset pricing models can also be sensitive to the multivariate normality assumption too.

In both statistics and econometrics, multivariate t-distribution is widely used for robust analysis of data with fat tails. As asset returns do have large fat tails, the multivariate t-distribution should play a similar role in finance. Hence, our 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 multivariate t, equation (10), is that the terms do not combine to yield tractable solutions. However, it is well-known that a multivariate 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.

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,

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

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

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

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

Based on the following identities

$$[D_n' (\Psi^{-1} \otimes \Psi^{-1}) D_n]^{-1} = D_n^+ (\Psi \otimes \Psi) D_n^{+\prime}, \tag{A.8}$$

$$D_n^{+\prime} D_n' \mathrm{vec}(\Psi^{-1}) = D_n D_n^+ \mathrm{vec}(\Psi^{-1}) = D_n \mathrm{vech}(\Psi^{-1}) = \mathrm{vec}(\Psi^{-1}), \tag{A.9}$$

$$\mathrm{vec}(\Psi^{-1})' \mathrm{vec}(\Psi) = \mathrm{tr}(\Psi^{-1}\Psi) = n, \tag{A.10}$$

$$(\Psi^{-1} \otimes \Psi^{-1}) \mathrm{vec}(\Psi) = \mathrm{vec}(\Psi^{-1} \Psi \Psi^{-1}) = \mathrm{vec}(\Psi^{-1}), \tag{A.11}$$

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

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

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

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

In particular, we have

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

A combination of (A.6) and (A.14) yields (17). This complete the proof.

A.3. Proof of the second algorithm, (40)–(46):

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

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

Note that the first part of the likelihood function has parameters β and Ψ_ϵ and the second part has parameters μ_2 and Ψ_{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 β and Ψ_ϵ 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. This completes the proof.

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 $K_{mn}\mathrm{vec}(A) = \mathrm{vec}(A')$ 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

$$\mathrm{Avar}[\tilde{\mu}] = \left(\frac{\nu + n + 2}{\nu + n}\right)\Psi, \tag{A.23}$$

$$\mathrm{Avar}[\mathrm{vech}(\tilde{\Psi})] = \frac{2(\nu + n + 2)}{\nu + n}\left[D_n^+(\Psi \otimes \Psi)D_n^{+\prime} + \frac{\mathrm{vech}(\Psi)\mathrm{vech}(\Psi)'}{\nu}\right]. \tag{A.24}$$

We first prove that

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

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

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

It follows that

$$\frac{\partial \mathrm{vec}(\beta)}{\partial \mathrm{vec}(\Psi_{12})'} = \Psi_{22}^{-1} \otimes I_N, \tag{A.27}$$

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

Also, note that

$$\begin{aligned}\mathrm{vec}(\Psi_{12}) &= \mathrm{vec}([I_N,\ 0_{N\times k}]\Psi[0_{k\times N},\ I_k]') \\ &= ([0_{k\times N},\ I_k] \otimes [I_N,\ 0_{N\times k}])D_n\mathrm{vech}(\Psi), \end{aligned} \tag{A.29}$$

$$\begin{aligned}\mathrm{vech}(\Psi_{22}) &= D_k^+\mathrm{vec}([0_{k\times N},\ I_k]\Psi[0_{k\times N},\ I_k]') \\ &= D_k^+([0_{k\times N},\ I_k] \otimes [0_{k\times N},\ I_k])D_n\mathrm{vech}(\Psi). \end{aligned} \tag{A.30}$$

[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

$$
\begin{aligned}
&\text{Avar}[\text{vec}(\tilde{\beta})] \\
&= \text{Avar}\left[\left[\Psi_{22}^{-1} \otimes I_N,\ (\Psi_{22}^{-1} \otimes -\beta) D_k\right] \begin{bmatrix} [0_{k\times N},\ I_k] \otimes [I_N,\ 0_{N\times k}] \\ D_k^{+}\left([0_{k\times N},\ I_k] \otimes [0_{k\times N},\ I_k]\right) \end{bmatrix} D_n \text{vech}(\tilde{\Psi})\right] \\
&= \text{Avar}\left[\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) D_n \text{vech}(\tilde{\Psi})\right] \\
&= \frac{2(\nu+n+2)}{\nu+n}\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) D_n \left[D_n^{+}(\Psi\otimes\Psi)D_n^{+\prime} + \frac{\text{vech}(\Psi)\text{vech}(\Psi)'}{\nu}\right] D_n' \\
&\quad \times \left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
&= \left(\frac{\nu+n+2}{\nu+n}\right)\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)\left[(I_{n^2}+K_n)(\Psi\otimes\Psi) + \frac{2\text{vec}(\Psi)\text{vec}(\Psi)'}{\nu}\right] \\
&\quad \times \left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)', \qquad \text{(A.31)}
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
&\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right) K_n(\Psi\otimes\Psi)\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
&= K_{Nk}\left([I_N,\ -\beta] \otimes [0_{k\times N},\ \Psi_{22}^{-1}]\right)(\Psi\otimes\Psi)\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
&= K_{Nk}\left([\Psi_\epsilon,\ 0_{N\times k}] \otimes [\beta',\ I_k]\right)\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
&= 0_{Nk\times Nk}, \qquad \text{(A.34)}
\end{aligned}
$$

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

we have

$$
\begin{aligned}
\text{Avar}[\text{vec}(\tilde{\beta})] &= \left(\frac{\nu+n+2}{\nu+n}\right)\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)(\Psi\otimes\Psi)\left([0_{k\times N},\ \Psi_{22}^{-1}] \otimes [I_N,\ -\beta]\right)' \\
&= \left(\frac{\nu+n+2}{\nu+n}\right)(\Psi_{22}^{-1}\otimes\Psi_\epsilon). \qquad \text{(A.36)}
\end{aligned}
$$

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

$$
\frac{\partial\alpha}{\partial\mu'} = [I_N,\ -\beta], \qquad \text{(A.37)}
$$

$$
\frac{\partial\alpha}{\partial\text{vec}(\beta)'} = -\mu_2'\otimes I_N. \qquad \text{(A.38)}
$$

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 Ψ_{22} and Ψ_ϵ but not in terms of the variance of f_t and ϵ_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 ν is unknown because the information matrix is block diagonal.

For the asymptotic variance of the OLS estimators $\hat{\alpha}$ 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.

References

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

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., 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.

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 mulitfactor pricing models, volatility bounds and portfolio performance, Chapter 12 in G. M. Constantinides, M. Harris and R. M. Stultz (ed.): Handbook of the Economics of Finance, Elsevier, North Holland, 743–800.

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, Quantitative Finance 10, 469–485.

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.

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.

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 25 size and book-to-market ranked portfolios and three factors based on monthly returns from July 1963 through December 2015. 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 8, 7, and 6, respectively.

		p-value (%)					p-value (%)			
			Student-t with df					Student-t with df		
	Skew.	Normal	8	7	6	Kurt.	Normal	8	7	6
					Univariate					
S1B1	−0.036	71.08	85.59	87.20	89.34	5.232	0.00	12.05	21.08	36.88
S1B2	0.040	67.47	83.75	85.47	87.98	6.328	0.00	4.36	8.42	17.53
S1B3	−0.225	2.04	27.34	33.34	41.55	5.370	0.00	10.39	18.48	33.37
S1B4	−0.147	12.85	46.28	51.64	58.65	6.016	0.00	5.64	10.62	21.40
S1B5	−0.217	2.47	28.83	34.82	43.03	6.252	0.00	4.65	8.90	18.37
S2B1	−0.344	0.06	11.43	16.25	23.91	4.510	0.00	29.61	43.97	63.43
S2B2	−0.478	0.00	4.47	7.42	12.99	5.438	0.00	9.66	17.38	31.72
S2B3	−0.486	0.00	4.25	7.06	12.55	5.863	0.00	6.39	12.00	23.61
S2B4	−0.444	0.00	5.60	8.92	14.98	5.863	0.00	6.39	11.99	23.60
S2B5	−0.406	0.01	7.35	11.14	17.87	5.884	0.00	6.27	11.80	23.29
S3B1	−0.389	0.01	8.22	12.27	19.32	4.467	0.00	31.34	46.00	65.42
S3B2	−0.533	0.00	3.18	5.42	10.29	5.731	0.00	7.23	13.45	25.73
S3B3	−0.483	0.00	4.33	7.19	12.72	5.145	0.00	13.34	22.95	39.47
S3B4	−0.333	0.09	12.41	17.38	25.16	5.346	0.00	10.62	18.91	33.93
S3B5	−0.380	0.01	8.82	13.02	20.18	5.954	0.00	5.94	11.14	22.24
S4B1	−0.254	0.91	22.15	28.07	36.40	4.827	0.00	19.69	31.53	50.10
S4B2	−0.576	0.00	2.46	4.38	8.60	5.820	0.00	6.64	12.44	24.30
S4B3	−0.520	0.00	3.44	5.85	10.87	6.158	0.00	5.03	9.54	19.53
S4B4	−0.252	0.94	22.40	28.31	36.62	5.139	0.00	13.46	23.10	39.65
S4B5	−0.367	0.02	9.67	14.01	21.36	5.348	0.00	10.61	18.89	33.88
S5B1	−0.233	1.64	25.85	31.82	40.06	4.663	0.00	24.30	37.43	56.65
S5B2	−0.383	0.01	8.62	12.74	19.89	4.820	0.00	19.86	31.76	50.35
S5B3	−0.288	0.32	17.39	22.91	31.12	5.085	0.00	14.34	24.29	41.26
S5B4	−0.555	0.00	2.80	4.86	9.37	6.481	0.00	3.88	7.59	16.06
S5B5	−0.217	2.52	29.03	35.01	43.23	4.317	0.00	38.34	53.62	72.22
SMB	0.513	0.00	3.58	6.08	11.19	8.780	0.00	1.08	2.35	5.68
HML	0.033	73.00	86.60	88.10	90.05	5.219	0.00	12.23	21.34	37.27
MKT	−0.523	0.00	3.37	5.74	10.74	4.925	0.00	17.39	28.51	46.58
					Multivariate					
	159.04	0.00	3.60	14.51	49.42	1246.81	0.00	4.18	26.88	80.04

Table 2
Estimation of Mean and Standard Deviation under Multivariate Normality versus under Multivariate 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 2015, assuming that the returns are generated from a multivariate normal or t-distribution with degrees of freedom 8, 7, and 6, respectively.

	Mean				Standard Deviation			
		t-distribution				t-distribution		
	Normal	df=8	df=7	df=6	Normal	df=8	df=7	df=6
S1B1	0.232	0.058	0.057	0.056	7.942	7.432	7.582	7.812
S1B2	0.777	0.607	0.605	0.604	6.906	6.483	6.615	6.817
S1B3	0.763	0.600	0.598	0.596	5.990	5.817	5.936	6.117
S1B4	0.988	0.805	0.803	0.800	5.706	5.453	5.564	5.732
S1B5	1.079	0.953	0.950	0.948	6.014	5.806	5.923	6.102
S2B1	0.460	0.364	0.364	0.364	7.177	6.997	7.141	7.360
S2B2	0.738	0.642	0.642	0.642	5.985	5.923	6.045	6.232
S2B3	0.871	0.809	0.808	0.807	5.424	5.392	5.503	5.672
S2B4	0.915	0.857	0.856	0.855	5.218	5.136	5.241	5.402
S2B5	0.981	0.862	0.859	0.857	6.043	5.879	5.999	6.182
S3B1	0.492	0.477	0.478	0.479	6.610	6.403	6.535	6.736
S3B2	0.780	0.735	0.735	0.735	5.477	5.431	5.542	5.712
S3B3	0.736	0.717	0.717	0.717	5.003	5.000	5.103	5.260
S3B4	0.861	0.807	0.806	0.805	4.895	4.895	4.996	5.150
S3B5	1.017	0.894	0.892	0.889	5.620	5.503	5.615	5.786
S4B1	0.589	0.526	0.525	0.524	5.913	5.677	5.794	5.971
S4B2	0.582	0.539	0.540	0.540	5.127	5.097	5.203	5.363
S4B3	0.679	0.655	0.654	0.654	4.979	4.914	5.016	5.171
S4B4	0.851	0.799	0.797	0.796	4.795	4.768	4.868	5.020
S4B5	0.801	0.779	0.779	0.778	5.667	5.537	5.651	5.824
S5B1	0.467	0.493	0.492	0.492	4.633	4.625	4.722	4.868
S5B2	0.510	0.541	0.541	0.541	4.444	4.433	4.526	4.666
S5B3	0.520	0.575	0.577	0.578	4.299	4.244	4.332	4.466
S5B4	0.481	0.559	0.560	0.561	4.644	4.460	4.551	4.690
S5B5	0.658	0.628	0.627	0.625	5.332	5.240	5.348	5.513
SMB	0.221	0.103	0.102	0.101	3.094	2.834	2.892	2.980
HML	0.348	0.325	0.324	0.323	2.809	2.564	2.616	2.695
MKT	0.499	0.531	0.532	0.532	4.440	4.383	4.474	4.612

Table 3
Alpha and Beta Estimation under Multivariate Normality versus under Multivariate 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 2015. 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 and factors are multivariate t-distributed with 7 degrees of freedom.

	α (%)		β_{MKT}		β_{HML}		β_{SMB}	
Portfolio	Normal	t_7	Normal	t_7	Normal	t_7	Normal	t_7
S1B1	−0.521	−0.580	1.100	1.086	−0.282	−0.236	1.364	1.341
S1B2	−0.017	−0.075	0.975	1.001	0.050	0.074	1.311	1.217
S1B3	−0.045	−0.095	0.935	0.939	0.296	0.259	1.078	1.084
S1B4	0.160	0.096	0.879	0.885	0.439	0.406	1.065	1.027
S1B5	0.122	0.124	0.962	0.946	0.684	0.650	1.078	1.101
S2B1	−0.190	−0.187	1.132	1.117	−0.381	−0.454	0.985	1.028
S2B2	−0.007	−0.006	1.015	1.021	0.125	0.036	0.881	0.919
S2B3	0.087	0.118	0.964	0.962	0.389	0.291	0.762	0.829
S2B4	0.087	0.114	0.960	0.952	0.557	0.493	0.698	0.740
S2B5	−0.031	−0.046	1.078	1.080	0.801	0.747	0.884	0.877
S3B1	−0.073	−0.032	1.104	1.099	−0.428	−0.461	0.738	0.738
S3B2	0.081	0.113	1.041	1.019	0.173	0.054	0.541	0.621
S3B3	−0.001	0.036	0.991	0.984	0.423	0.323	0.427	0.520
S3B4	0.078	0.069	0.974	0.974	0.595	0.519	0.409	0.498
S3B5	0.081	0.036	1.078	1.057	0.796	0.707	0.548	0.635
S4B1	0.108	0.053	1.078	1.072	−0.413	−0.415	0.394	0.364
S4B2	−0.061	−0.074	1.064	1.048	0.190	0.080	0.208	0.295
S4B3	−0.038	−0.034	1.050	1.035	0.439	0.338	0.178	0.274
S4B4	0.110	0.074	1.005	1.013	0.560	0.499	0.200	0.226
S4B5	−0.110	−0.094	1.162	1.138	0.787	0.724	0.262	0.323
S5B1	0.169	0.144	0.965	0.965	−0.375	−0.429	−0.241	−0.250
S5B2	0.027	0.026	0.997	0.995	0.092	0.016	−0.210	−0.187
S5B3	−0.006	0.008	0.952	0.956	0.311	0.255	−0.256	−0.221
S5B4	−0.212	−0.138	1.036	1.021	0.644	0.541	−0.222	−0.194
S5B5	−0.151	−0.224	1.110	1.116	0.801	0.821	−0.107	−0.081

Table 4
Estimated Alphas of Mutual Funds under Normality versus under t (2011/1–2015/12)

Based on monthly data from January 2011 to December 2015, the first panel of the table reports the estimated monthly Jensen's alpha (in percentage) of five mutual funds estimated under 5 different distributional assumptions: the multivariate normal, the multivariate t-distribution with unknown, 8, 7, and 6 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 absolute difference in the two estimated alphas. The second panel provides the corresponding results for funds with estimated reversed from positive to negative. The third panel reports the percentage of funds that have an annualized absolute difference over 1% to 5%, respectively, in their estimated alphas under the normal and the t-distribution assumptions with unknown degrees of freedom.

		degrees of freedom of t-distribution			
Fund	Normal	unknown	8	7	6
iPath ETN Global Carbon/A	−0.451	0.067	0.047	0.100	0.166
Ultra Series Fund: Large Cap Value Fund/II	−0.229	0.091	0.053	0.059	0.067
MML Managed Volatility Fund	−0.128	0.167	0.042	0.059	0.078
MML Large Cap Value Fund	−0.149	0.146	0.022	0.038	0.057
Franklin Focused Core Equity Fund/A	−0.230	0.008	−0.051	−0.037	−0.021
Percentage of reversals	2.28				
Average absolute difference in alpha	0.081				
Direxion Daily Small Cap Bear 3X Shares	1.328	−0.727	−0.607	−0.621	−0.638
Direxion Financial Bear 3X Shares	0.833	−1.156	−1.057	−1.068	−1.082
Columbia Abs. Ret. Currency & Inc. Fund/W	0.246	−0.157	−0.156	−0.161	−0.165
Highland L/S Healthcare Fund/Z	0.241	−0.040	−0.005	−0.029	−0.038
MassMutual Select Fund. Growth Fund/A	0.064	−0.005	−0.006	−0.011	−0.018
Percentage of reversals	0.62				
Average absolute difference in alpha	0.151				

	Absolute difference in annual alphas				
	1%	2%	3%	4%	5%
Percentage of Funds	11.62	4.08	1.86	1.18	0.70

Table 5
Estimated Alphas of Mutual Funds under Normality versus under t (2006/1–2015/12)

Based on monthly data from January 2006 to December 2015, the the first panel of the table reports the estimated monthly Jensen's alpha (in percentage) of five mutual funds estimated under 5 different distributional assumptions: the multivariate normal, the multivariate t-distribution with unknown, 8, 7, and 6 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 absolute difference in the two estimated alphas. The second panel provides the corresponding results for funds with estimated reversed from positive to negative. The third panel reports the percentage of funds that have an annualized absolute difference over 1% to 5%, respectively, in their estimated alphas under the normal and the t-distribution assumptions with unknown degrees of freedom.

		degrees of freedom of t-distribution			
Fund	Normal	unknown	8	7	6
Forward Real Estate L/S Fund/A	−0.372	0.106	0.013	0.030	0.049
Davis Real Estate Fund/A	−0.247	0.101	0.011	0.029	0.050
SSgA Clarion Real Estate Fund/N	−0.056	0.279	0.207	0.226	0.247
American Century Capital R.E. Fund/I	−0.099	0.215	0.120	0.137	0.158
PowerShares S&P 500 High Quality Portfolio	−0.107	0.201	0.082	0.098	0.117
Percentage of reversals	4.88				
Average absolute difference in alpha	0.108				
Reynolds Blue Chip Growth Fund	0.415	−0.123	0.002	−0.019	−0.042
Fidelity Select Portfolios: Computers Portfolio	0.031	−0.227	−0.151	−0.167	−0.185
Burnham Financial Industries Fund/C	0.191	−0.042	0.013	−0.003	−0.024
Centaur Total Return Fund	0.173	−0.038	0.024	0.011	−0.004
Brown Capital Mgmt. Mid Company Fund/I	0.064	−0.126	−0.106	−0.117	−0.130
Percentage of reversals	4.63				
Average absolute difference in alpha	0.076				

	Absolute difference in annual alphas				
	1%	2%	3%	4%	5%
Percentage of Funds	18.10	6.25	2.29	0.76	0.41

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 multivariate 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:

$$r_{it} = \alpha_i + \beta_i MKT_t + \epsilon_{it}, \quad t = 1, \ldots, T$$

where the data are monthly returns from July 1963 through December 2015. The tests are carried out for the entire sample period as well as for two subperiods.

Model	GRS	p-value (%)	LRT	p-value (%)
		July 1963 — December 2015		
Normal	4.620	0.00	107.69	0.00
t (df=8)			121.95	0.00
t (df=7)			122.37	0.00
t (df=6)			122.82	0.00
t (unknown)			122.15	0.00
		July 1963 — September 1989		
Normal	2.097	0.21	50.08	0.21
t (df=8)			64.18	0.00
t (df=7)			64.52	0.00
t (df=6)			64.90	0.00
t (unknown)			62.75	0.00
		October 1989 — December 2015		
Normal	4.306	0.00	95.15	0.00
t (df=8)			90.55	0.00
t (df=7)			90.70	0.00
t (df=6)			90.88	0.00
t (unknown)			90.87	0.00

Table 7
Multivariate Tests of the Fama-French 3-Factor Model

The table reports both the Gibbons, Ross, and Shanken (1989) test and the likelihood ratio test under the multivariate t-distribution for the 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 Fama-French 3-factor model:

$$r_{it} = \alpha_i + \beta_{1,i} MKT_t + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \epsilon_{it}, \quad t = 1, \ldots, T,$$

where the data are monthly returns from July 1963 through December 2015. The tests are carried out for the entire sample period as well as for two subperiods.

Model	GRS	p-value (%)	LRT	p-value (%)
		July 1963 — December 2015		
Normal	3.809	0.00	90.09	0.00
t (df=8)			99.93	0.00
t (df=7)			100.20	0.00
t (df=6)			100.48	0.00
t (unknown)			100.05	0.00
		July 1963 — September 1989		
Normal	1.281	17.13	31.57	17.07
t (df=8)			43.81	1.14
t (df=7)			44.06	1.07
t (df=6)			44.31	1.00
t (unknown)			42.58	1.56
		October 1989 — December 2015		
Normal	4.292	0.00	94.82	0.00
t (df=8)			89.30	0.00
t (df=7)			89.43	0.00
t (df=6)			89.59	0.00
t (unknown)			89.60	0.00

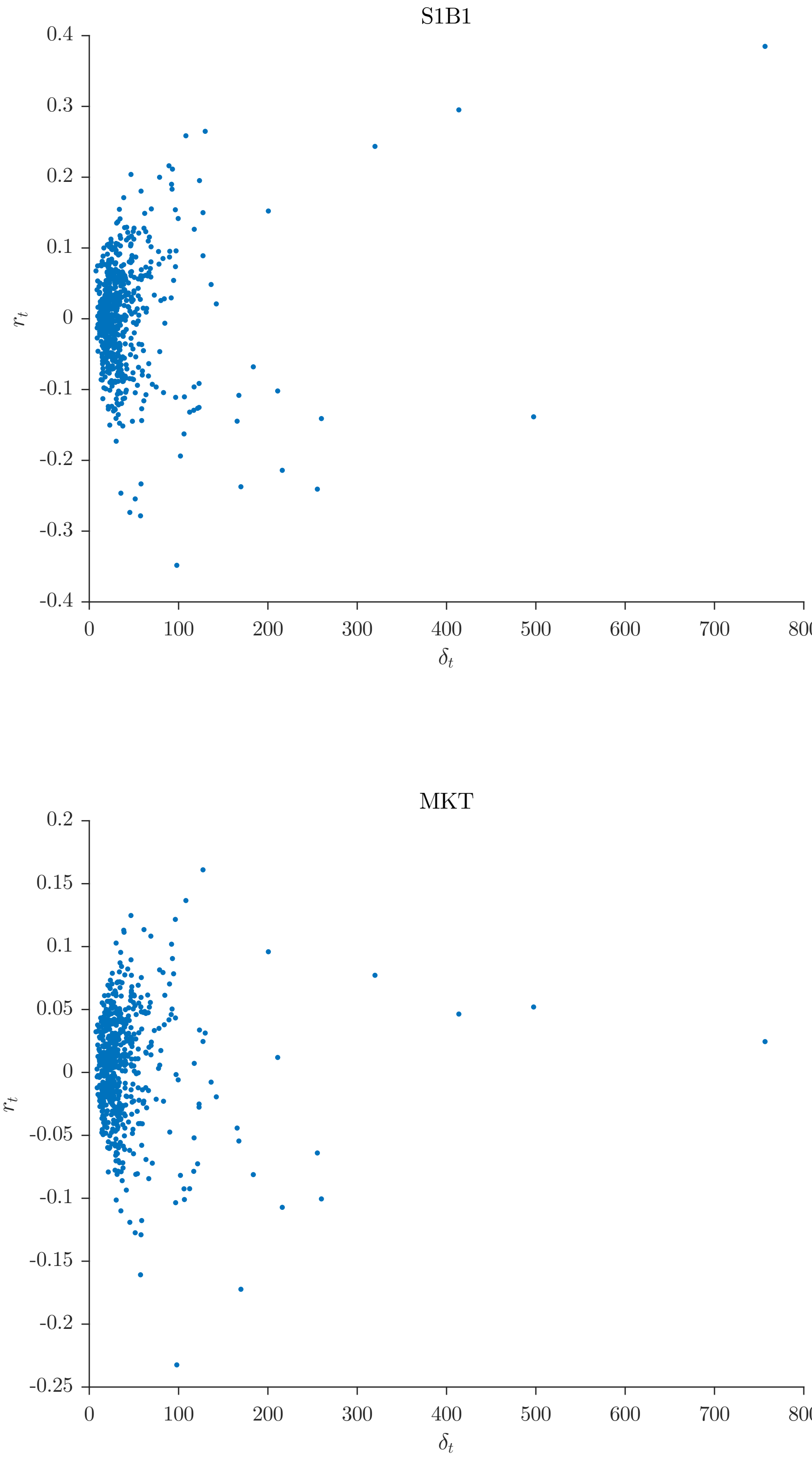

Figure 1
Plots of returns vs. distance measure from the center for S1B1 and MKT

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 (δ_t) of the mutlivariate t-distribution over the period of July 1963 to December 2015. 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.