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

RAYMOND KAN and GUOFU ZHOU∗

This version: June, 2003

∗Kan is from the University of Toronto, Zhou is from Washington University in St. Louis. We are grateful for helpful discussions and comments of Martijn Cremers, Heber Farnsworth, William Goetzmann, Campbell Harvey, Yongmiao Hong, Ravi Jagannathan, Christopher Jones, Bruce Lehmann, Luboš Pástor, Jay Shanken, and seminar participants at Washington University in St. Louis. 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

In this paper, we propose to replace the widely used and firmly rejected multivariate normality assumption for the distribution of asset returns with a suitable multivariate $t$-distribution. Much of the asset pricing literature tries to explain expected returns on assets. In empirical studies, expected returns are often approximated by sample averages of realized returns. However, after replacing the normality assumption with a reasonable $t$-distribution, the most efficient estimator of expected return is drastically different from the sample average return which is heavily influenced by fat-tails in the return data. For example, the annual difference in expected return estimation under normal and $t$ is 2.964% for the Fama and French’s (1993, 1996) smallest size and book-to-market portfolio. In addition, there are also substantial differences in estimating alphas and testing asset pricing models when returns follow a multivariate $t$-distribution instead of a multivariate normal.
Ever since Fama (1965), Affleck-Graves and McDonald (1989), and Richardson and Smith (1993), 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.

The purpose of this paper is to advocate the use of a $t$-distribution instead of the normal for three reasons.\(^1\) First, it models financial data well in many circumstances. Theoretically, the $t$-distribution nests the normal as a special case, but it captures the observed fat tails of financial data. For example, the multivariate normality assumption of the joint distribution of Fama and French’s (1993) 25 assets returns and their 3 factors from January 1963 to December 2002 is unequivocally rejected by a kurtosis test with a $p$-value of less than 0.01%. On the other hand, such a test for a multivariate $t$-distribution has a $p$-value of 15.98%. Second, the $t$-distribution has become almost as tractable as the normal distribution. Traditionally, multivariate 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 multivariate normal distribution has been almost the only choice in analyzing a large number of assets due to its analytical formulae for parameter estimates. However, this is no longer a decisive advantage of the multivariate normal. Owing to the path breaking EM algorithm of Dempster, Laird and Rubin (1977), and especially Liu and Rubin (1995), there are explicit iterative formulae available to yield fast and monotonically convergent solutions to efficient estimation under $t$. This is important when comparing with alternative multivariate non-normal processes, such as the well-known GRACH models, that require numerical search which usually limits the number of assets to no more than ten (see, e.g., Bollerslev (2001)). In contrast, there are 28 assets and over 400 parameters in our later applications. The analytical iterations, however, take less than a minute to find the solutions, implying that our method is almost as easy to use as that under normality. It should also be pointed out that the generalized method of moments (GMM) estimators of Hansen (1982), one of the most widely used procedures in finance, allows for

\(^1\)Blattberg and Gonedes (1974) seems to be the first paper to use $t$-distribution to model stock returns in finance. Later applications of $t$ and generalized $t$ in the univariate case can be found in Theodossiou (1998) and references therein. Although MacKinlay and Richardson (1991), Zhou (1993) and Geczy (2001) use multivariate $t$-distributions, their analyses focus on how results under normality vary when under $t$ without providing the results estimated based on the $t$ assumption.
a much more general distributional assumption than the normal. However, the GMM estimates of important parameters, such as the expected asset returns, alphas and betas, are the same as the maximum likelihood estimates obtained under the normality assumption, except that the GMM enlarges the standard errors to account for non-normality. In contrast, the EM algorithm here provides the asymptotically most efficient estimates when the data is $t$ distributed. 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, Chamberlain (1983) and Owen and Rabinovitch (1983) show that the elliptical class, of which $t$ is a special case, is the largest class of distributions that validates the mean-variance framework, and hence the Capital Asset Pricing Model (CAPM) of Sharpe (1964) andLintner (1965) can still be valid under $t$. Therefore, there is little reason why the widely used normality assumption should not be replaced with the $t$-distribution.

Assuming that asset returns are $t$ distributed 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 0.210%/month when estimated under normality, but is only 0.127%/month when estimated under $t$ with 8 degrees of freedom, implying an annual difference of 0.996%. This difference is of economic significance, suggesting that the $t$-distribution plays an important role for the estimation of the cost of capital. Such differences are even larger for some of the 25 portfolios used by Fama and French (1993). For instance, the annual difference is 2.964% for the portfolio that is in the smallest size and book-to-market quintiles. In fact, the expected returns on all of the Fama and French’s (1993) 25 portfolios are lower when estimated 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 order to estimate expected return. In contrast, estimator under $t$-distribution assigns less weight to samples in the tails than the normal, so the estimate of expected return can be substantially different from the sample average in the presence of fat tails. Indeed, the realized returns have heavy fat tails on both sides and heavier on the positive side during our sample period. Assigning less weights to the fat tails results in a shift of the estimated mean leftward. In contrast, the standard deviations estimated under either normality or $t$ are fairly close. Therefore, it appears that the estimation of the mean is more sensitive to fat tails of the data.
Second, the estimated optimal portfolio weights of a mean-variance investor can be substantially different. When there are no short selling restrictions, the investor who uses sample mean and variance will come up with a very different portfolio from the one if he uses the maximum likelihood estimates of mean and variance under the $t$-distribution assumption. When short selling restrictions are imposed, there are still significant changes in the optimal portfolio weights.

Third, estimation of Jensen’s alpha relies critically upon the distributional assumption on the data. For some assets, alphas can increase 7 times or decrease by as much as 50% once the normality assumption is replaced by a $t$-distribution with 8 degrees of freedom. In contrast, the betas are almost unchanged except for a few cases. This appears to be consistent with the well-known fact that it is easier to estimate the second moments of asset returns while harder to estimate their first moments.

Fourth, 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 fact that estimation of Jensen’s alpha are more unreliable when the returns follow a multivariate $t$-distribution instead of a multivariate normal distribution. 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. This suggests that non-normality modeling by using the $t$ helps us not only in obtaining better estimates of asset expected returns, but also in providing more powerful tests of asset pricing restrictions.

The remainder of the paper is organized as follows. The next section provides first the empirical evidence on the necessity of modeling the data as $t$ distributed rather than the normal. Section II presents both the estimation technique under $t$ and a comparison of the results with those obtained under normality. Section III assesses asset pricing implications of the $t$-distribution. Section IV discusses some general issues and extensions. Section V concludes.
I. Why Multivariate $t$?

In this section, we provide a description of the data 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 $t$-distribution cannot be rejected.

A. 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. In addition, we also use the monthly return 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 1965 through December 2002.

B. Normality test

Our first question is whether the data can be adequately described by a normal distribution. To answer, let $x_t = (r_t', f_t')'$, where $r_t$ contains the excess returns of $N = 25$ test assets and $f_t$ contains 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,

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

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

where

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} x_t,$$  \hspace{1cm} (3)

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

---

$^2$We are grateful to Ken French for making this data available on his website.
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

$$
\delta_1 = E \left( \left( (x - \mu)' V^{-1} (y - \mu) \right)^3 \right), \quad \delta_2 = E \left( \left( (x - \mu)' V^{-1} (x - \mu) \right)^2 \right),
$$

(5)

where $\mu$ and $V$ are the population mean and covariance-matrix of $x$, and $y$ is a random variable that has the same probability density as $x$, but is independent of $x$. Under the normality assumption, $\delta_1$ is simply zero and $\delta_2$ is equal to $n(n + 2)$. The second property is that $D_1$ and $D_2$ are invariant to any linear transformations of the data. In other words, any non-singular repackaging of the assets will not alter the 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. With reliable accuracy, we use 10,000 draws in what follows.

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

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

(6)

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

Table I 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. Clearly multivariate normality implies univariate normality and not the reverse. Therefore, it is
not surprising that the \( p \)-value based on the multivariate kurtosis test is less than 0.01%. Hence, multivariate normality is unequivocally rejected by the data. On the other hand, if we assume that the data is from a \( t \)-distribution with degrees of freedom \( \nu = 10, 8 \) and 6, the \( p \)-value goes up from 0.39\% to 15.98\% and 94.11\%. Therefore, a \( t \)-distribution with \( \nu = 8 \) is not rejected by the data, neither is the one with \( \nu = 6 \).

Consider now both the univariate and multivariate skewness tests. The sample skewness statistics are provided in the second column. We find that the univariate values are in general very small so that there are many assets that pass the test. The multivariate skewness test, however, strongly rejects the normality assumption and even a \( t \) with \( \nu = 10 \) at the usual 5\% level. Nevertheless, the multivariate skewness test cannot reject a \( t \)-distribution with \( \nu = 8 \) or 6, a conclusion similar to what we obtain using the kurtosis test. The reason is that the finite sample variation of the skewness of a \( t \)-distribution is very large when \( \nu \) is small, so as to imply a large probability for observing a large sample skewness.\(^3\)

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

\[
\kappa \equiv \frac{D^2}{n(n+2)} - 1 = \frac{2}{\nu - 4},
\]

where the last equality follows for a \( t \)-distribution. Under normality, \( \kappa = 0 \), so \( \kappa \) measures the excess kurtosis relative to the normal. Equation (7) implies that the population kurtosis goes to infinity as \( \nu \) goes down to 4. Hence, no matter how large the sample kurtosis is, one can always find a \( t \)-distribution to describe it with a small enough \( \nu \). Although the \( p \)-value for the sample kurtosis test is greater with a smaller \( \nu \) (but greater than 4), the sample kurtosis can in fact falls out of a reasonable left tail of the distribution when \( \nu \) is too small. For example, the \( p \)-value of 94.11\% when \( \nu = 6 \) implies that the observed sample multivariate kurtosis falls into the 5.89\% mass of the distribution from the left, close to the usual 5\% rejection level if a left-sided test is used. So a value of \( \nu = 8 \) appears to be a better model than a model with \( \nu = 6 \). This will be confirmed later

\(^3\)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.
with an estimate of $\nu = 8.66$ based on the maximum likelihood method. In short, the normality is firmly rejected by the data, but a $t$-distribution with $\nu = 8$ is a reasonable model for it.

**II. Estimation under Multivariate $t$**

After rejecting normality and accepting $t$ as a good alternative distribution for the data in the previous section, we now proceed to present the EM algorithm that elegantly solves the parameter estimation problem under $t$. With these estimates, we can then address the impacts of the $t$ on expected returns, optimal portfolio weights as well as the utility gain of using $t$ instead of the normal.

**A. Comparison with Normal**

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

$$ \text{Avar}[\hat{\mu}] = V. $$

(8)

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

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

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

(9)

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 formulae to find the parameter estimate that maximizes the log-likelihood function under $t$, i.e., the solution to maximizing $L$. Starting from any initial estimate of $\mu$ and $\Psi$, say $\tilde{\mu}^{(1)} = \hat{\mu}$ and $\tilde{\Psi}^{(1)} = \hat{\Psi} = (\nu - 2)\tilde{V}/\nu$, we can obtain iterative estimates via

1. $u_t^{(i)} = \frac{\nu + n}{\nu + (x_t - \tilde{\mu}^{(i)})'[\tilde{\Psi}^{(i)}]^{-1}(x_t - \tilde{\mu}^{(i)})}$, \hspace{1cm} (10)

2. $\tilde{\mu}^{(i+1)} = \frac{\sum_{t=1}^{T} u_t^{(i)} x_t}{\sum_{t=1}^{T} u_t^{(i)}}$, \hspace{1cm} (11)

3. $\tilde{\Psi}^{(i+1)} = \frac{1}{T} \sum_{t=1}^{T} u_t^{(i)} (x_t - \tilde{\mu}^{(i+1)})(x_t - \tilde{\mu}^{(i+1)})'$, \hspace{1cm} (12)

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 (9), the log-likelihood function under $t$. Indeed, in our application to Fama-French 25 assets and three-factors, the algorithm converges with less than 100 iterations and it takes less than a minute to run on a PC.

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

$$f(\nu) = \phi \left( \frac{\nu + n}{2} \right) - \phi \left( \frac{\nu}{2} \right) + \log \left( \frac{\nu}{\nu + n} \right) + \frac{1}{T} \sum_{t=1}^{T} \left[ \log(u_t^{(i)}(\nu)) - u_t^{(i)}(\nu) \right] + 1 = 0,$$ \hspace{1cm} (13)

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

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

Hence, the earlier EM algorithm can be combined with this one so that it is still applicable when $\nu$ is treated as an unknown parameter. It should be noted that equation (13) does not admit an analytical solution, so the implementation is more complex than the earlier case of a known
\( \nu \). However, equation (13) involves only one variable and its solution is easy to find by using a line-search routine. Therefore, even with an unknown \( \nu \), practical implementation of the algorithm is still straightforward. Indeed, even if we treat \( \nu \) as unknown in implementing the EM algorithm, it still converges in less than a minute in our applications. Moreover, regardless of what starting value of \( \nu \) chosen, the algorithm has always quickly converged to an estimated value \( \tilde{\nu} = 8.66 \) for the data that studied in Section I.

Therefore, even if one is less willing to simply use several values of \( \nu \) to assess the sensitivity of \( \nu \) on the statistical inference, one can estimate \( \nu \) easily and then use this estimated value instead of the assumed ones to carry out both the statistical computations and economic evaluations. This approach clearly makes little qualitative differences 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 its accuracy. Based on Lange, Little and Taylor (1989), the asymptotic variance-covariance matrix of \( \tilde{\mu} \) is, for \( \nu > 2 \),

\[
\text{Avar}(\tilde{\mu}) = (1 - \rho)V, \quad \rho \equiv \frac{2n + 4}{\nu(\nu + n)}. \tag{15}
\]

When the data is \( t \) distributed rather than the normal, the sample mean \( \hat{\mu} \) is no longer the asymptotically most efficient estimate of \( \mu \), but the maximum likelihood estimator \( \tilde{\mu} \) is. The relative efficiency is measured by \( \rho \). The greater the \( \rho \), the better the maximum likelihood method. In our application with \( n = 28, \nu = 8 \), we have \( \rho = 0.2083 \), implying that the maximum likelihood estimator \( \tilde{\mu} \) is 20\% less volatile than the sample mean.\(^4\) It is interesting to observe that this improvement in estimation accuracy is independent of the parameter values of \( \mu \) and \( V \). In addition, the relative efficiency increases when \( n \) increases. Under normality, the sample average return of an asset is the most efficient estimator of its expected return, and the inclusion of other assets will not alter this estimate. In contrast, once the \( t \)-distribution is allowed, realized returns from one asset contain useful information for estimating the expected return of another asset. The greater the number of assets, the more efficient the \( \tilde{\mu} \). Moreover, the relative efficiency increases when \( \nu \) gets smaller. This makes intuitive sense: the smaller the \( \nu \), the greater the deviation of the data from normality, and hence a procedure incorporating non-normality gains more in estimation efficiency.

\(^4\)It should be noted that the asymptotic variance of \( \tilde{\mu} \) is the same whether \( \nu \) is known or unknown, so the efficiency gain of using \( \tilde{\mu} \) does not depend on whether we know \( \nu \) or not.
Similarly, one can ask what the efficiency gain is for estimating the variance of asset returns by
the maximum likelihood method under \( t \)-distribution. In the Appendix, we show that

\[
\text{Avar}[\hat{V}_{ii}] = (1 - \rho_v)\text{Avar}[\tilde{V}_{ii}],
\]

\[
\rho_v \equiv \frac{2[2n + 4 + \nu(n + 5)]}{\nu(\nu - 1)(\nu + n)},
\]

(16)

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

\[
\text{Avar}[\tilde{V}_{ii}^2] = (1 - \rho_v)\text{Avar}[\hat{V}_{ii}^2],
\]

(17)

so \( \rho_v \) is also the efficiency gain for estimating the standard deviation of asset returns by
the maximum likelihood method under \( t \)-distribution. When \( n = 28 \) and \( \nu = 8 \), we have \( \rho_v = 32.14\% \).

This says that compared with the sample variance or standard deviation, the maximum likelihood
procedure improves the estimation efficiency in estimating the variance or standard deviation by
about 32%.

Table II provides the estimation results for the expected returns and the standard deviations, the
fundamental parameters of the data. The second column reports the sample average returns, while
the next three columns are maximum likelihood estimates of expected return under a \( t \) distribution
with \( \nu = 10, 8 \) and 6, respectively. As discussed earlier, a value of \( \nu = 8 \) appears to be a good model
for the data, but the results on two other values are provided to assess the sensitivity of the results
to the specification of \( \nu \). It is striking that the expected returns estimated under \( t \) for all assets are
smaller than those estimated under normality. For example, the sample average excess returns for
the market (MKT) and size (SMB) factors are 0.410% and 0.210%, but their estimated expected
excess returns under \( t \) with \( \nu = 8 \) are only 0.358% and 0.127%, implying an annual difference of
0.624% and 0.996%. Such differences between the assets are even larger. For instance, the S1B1
asset has an annual difference of 2.964% between the two estimates of expected return.

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

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

(18)
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 exact maximum likelihood estimate is a solution of

\[
\sum_{t=1}^{T} w_t (r_t - \mu) = 0, \tag{19}
\]

or

\[
\tilde{\mu} = \sum_{t=1}^{T} w_t r_t, \tag{20}
\]

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 clearly that \( \delta_t \) measures how far the data is from its center. The larger the \( \delta_t \), the smaller the \( w_t \). Hence, outliers are weighted less than other data. In contrast, the sample mean weights every single data point equally with weight \( 1/T \). When the true distribution has indeed fatter tails than the normal, the sample mean become 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 its tail behavior. The upper part of Figure 1 provides a Q-Q plot of S1B1 against the normal, a standard diagnostic tool in statistics. If S1B1 follows a normal distribution, the points should all lie on the 45 degree line. In fact, S1B1 has much heavier positive and negative tails than the normal, and there is more mass on the positive tail. Hence, to use a \( t \) distribution to fit the data, the estimated mean must shift substantially leftward relative to the sample mean. In contrast, as shown by the lower part of Figure 1, the market does not have as heavy a positive tail as S1B1, so the mean of the market when estimated under \( t \) shifts less leftward than that of S1B1.

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 II, while there is a huge difference in the two estimates of expected returns, S1B1 has similar standard deviations using the sample one \( \hat{\sigma}_{11}^2 = 8.292\% \) (per month) and the maximum likelihood one \( \tilde{\sigma}_{11}^2 = 7.751\% \) under the \( t \)-distribution with 8 degrees of freedom. The same is also true for the market excess return whose standard deviations in the two cases are 4.509\% and 4.537\%, very close to each other. The small differences in estimating the standard deviations are consistent with the
well-known fact that it is easier to estimate the second-order moments while harder to estimate
the first-order ones. Indeed, the estimated standard error of $\hat{V}_{11}$ for S1B1 is only 0.356%, so the
estimate of 8.292% is very accurate. In contrast, the estimated standard error of $\hat{\mu}_1$ is 0.381%.
Recall that the sample mean is 0.135%, indicating that the estimate of the mean is very imprecise.
Therefore, the difference between this normal mean estimate and the $t$ one can be much greater
than the difference in estimating the standard deviations.

As means and variance-covariance matrix are the fundamental parameters of asset returns, the
differences in estimating the means, as we found above, are likely to cause differences in portfolio
selection and mean-variance utility maximization. Whether this is substantial or not is the subject
of the next two subsections.

B. Optimal Portfolio Choice

As mean-variance framework is the most widely used one in practice for which there are also many
insights available from asset pricing theories, we will use it here to assess the portfolio impact and
the economic significance of utilizing the $t$-distribution. We consider a mean-variance investor who
chooses portfolio weights $w$ so as to maximize the standard mean-variance objective function

$$U = E[r_{pt}] - \frac{\tau}{2} \text{Var}[r_{pt}], \quad (21)$$

where

$$r_{pt} = \sum_{i=1}^{N} w_i r_{it}, \quad (22)$$

and $r_t$, as we recall, is the excess returns on the $N$ test assets, and $\tau$ is interpreted as the coefficient
of relative risk aversion.\textsuperscript{5} When there are no short selling constraints, it is well-known that the
optimal portfolio weights are

$$w^* = \frac{1}{\tau} V^{-1} \mu. \quad (23)$$

Table III provides this unconstrained estimated optimal portfolio weights for a mean-variance-
optimizing investor with an initial wealth of $100 and a relative risk aversion $\tau = 1, 3$ and $9$, when the parameters are estimated under normality and $t$ with degrees of freedom $10, 8$ and $6$.

\textsuperscript{5}In our portfolio analysis, we do not include the three factors in the portfolio. This is because the three factors
are close to linear combinations of the excess returns on the test assets, and including the factors would make the
covariance matrix close to singular.
respectively. It is seen that there are huge short and long positions in the assets. For example, with $\tau = 1$, one would long $1624.0$ worth of S1B4 and short $1005.7$ of S4B2. This wide range of optimal weights compared with a naive diversification is a well-known practical problem. Green and Hollifield (1992) and Jagannathan and Ma (2002) analyze some of the associated theoretical reasons. However, what is of interest here is the difference made by using normal versus by using $t$.

If the $t$ parameter estimates of $\mu$ and $V$ (assuming $\nu = 8$) are used, the optimal portfolio weights change drastically. For example, when $\tau = 1$, the long position of $26.2$ in S1B3 changes to a short position of $417.0$ while the short position of $110.8$ in S3B4 changes to a long position of $156.7$. So far, the results are discussed for $\nu = 8$ and $\tau = 1$. It is clear that similar results are also found for $\nu = 10$ or 6, and $\tau = 3$ or 9.

As an alternative to the unconstrained optimal portfolio problem, we also consider one without short selling so that $w_i \geq 0$ for $i = 1, \ldots, n$. This is of interest as Geczy, Musto and Reed (2002) show that short selling is not an easy matter because the standard collateral for US equities is 102% of the shares’s value. However, in this case the optimal weights cannot be written out explicitly, but can be computed numerically in a straightforward fashion. Table IV provides the results. As seen from the table, even with such strong constraints, there are still substantial differences between the weights under normality and those under $t$. For example, based on the sample mean and variance, the estimated optimal portfolio has a zero position in S3B4, but based on the maximum likelihood estimates of mean and variance under the $t$-distribution, the estimated optimal portfolio has a positive investment of $72.4$ when $\nu = 8$ and $\tau = 1$. In addition, similar to those results of Pástor and Stambaugh (2000) which impose various margin requirements, we find that there are many zero positions in the assets after imposing the no short selling constraints.

Admittedly, the portfolio optimization problems considered here are far from complete because we assume an investor follows simple mean-variance strategies to choose his portfolio, but the analysis of an optimal portfolio rule under uncertainty goes beyond the scope of this paper. With
the significant differences in the estimated expected returns, it will be unusual to not see drastic
portfolio weight differences in any optimal portfolio choice problem.

C. Utility Gain

What is the cost of assuming the returns follow a normal distribution instead of the more reasonable
t-distribution? Suppose under normality, an investor makes his optimal portfolio choice based on
the sample estimates of $\mu$ and $V$. Namely, the investor chooses

$$\hat{w} = \frac{1}{\tau} \hat{V}^{-1} \hat{\mu}. \quad (24)$$

In contrast, the investor assumes the returns follow a multivariate t-distribution would compute
their optimal portfolio weights based on $\tilde{\mu}$ and $\tilde{V}$ and chooses

$$\tilde{w} = \frac{1}{\tau} \tilde{V}^{-1} \tilde{\mu}. \quad (25)$$

The portfolio weight differences are analyzed in the previous subsection, but this is not sufficient
to claim that there is value of using the t because, due to correlations among the payoffs of risky
positions, the performances of two different portfolios can be similar even though they are quite
different in position-by-position allocations.

Nevertheless, following Kandel and Stambaugh (1996), Pástor and Stambaugh (2000), and
Kirby and Ostdiek (2001), we can compute the expected utility gain from using $\tilde{w}$ instead of $\hat{w}$ for
making portfolio decisions. The expected utility gain is defined as

$$E[U(\tilde{w}) - U(\hat{w})] = E \left[ \tilde{w}' \mu - \frac{\tau}{2} \tilde{w}' \tilde{V}^{-1} \tilde{w} \right] - E \left[ \hat{w}' \mu - \frac{\tau}{2} \hat{w}' \hat{V}^{-1} \hat{w} \right], \quad (26)$$

where the expectation is taken with respect to the true underlying distribution. This measures the
utility gain in terms of certainty-equivalent return to an investor who switches from the optimal
portfolio selection based on normality to that based on $t$. Under this approach of comparing
portfolios formed by two different methods, we view each method’s portfolio decision as a function
of the sample and then compare the performances of the two methods across repeated random
samples. Note that although $\tilde{\mu}$ and $\tilde{V}$ are more precise estimators of $\mu$ and $V$ than $\hat{\mu}$ and $\hat{V}$,
the expected utility gain is not positive by construction. This is because the objective function
here is not estimation efficiency of the parameters, so a method that improves estimation of mean
and variance does not automatically imply improvement in portfolio performance. In addition, the sign and the magnitude of the expected utility gain can be specific to a particular application. Generally speaking, values over a couple of percentage points per year are deemed to be economically significant.

Table V provides the results on the utility gains which are computed by drawing 10,000 data sets from a multivariate $t$-distribution, with its parameters chosen based on the maximum likelihood estimates of mean and variance under $\nu = 8$, as reported in Table II. In the first half of Table V, we assume the investor knows about the degrees of freedom of the $t$-distribution. For each case, we also report two cases, one with sample size of $T = 474$ and the other with sample size of $T = 240$. It can be seen that when the risk aversion parameter, $\tau$, is equal to one, the utility gains, varying from a low of 72 basis points (bp) per month for $\nu = 10$ to a high of 160 bp per month for $\nu = 6$ when the sample size is $T = 474$. The expected utility gains are much more dramatic when the investor uses a shorter time series to estimate mean and variance. When $T = 240$, the expected utility gains for an investor with $\tau = 1$ can range from 185 bp per month to 406 bp per month. With higher $\tau$, the utility gains from using the maximum likelihood estimator is reduced, but yet they are still very significant. Once short selling is ruled out, however, the gains become much smaller, but yet for $T = 240$ and $\tau = 1$, we still can observe utility gains ranging from 7–15 bp per month. There are three patterns in the results. First, as $\nu$ goes down from 10 to 6, the gains increase, suggesting that the more deviations from normality, the greater the gains. Second, the gains increase as the risk aversion parameter decreases. This is expected because a less risk averse investor would invest more in risky assets, thereby increasing the impact of improved estimates of expected means. Third, the longer the time series, the less the utility gains. This is also expected because when $T$ is very large, both $\hat{\mu}$ and $\tilde{\mu}$ converges to the true mean, so the use of a more efficient estimator makes less of a difference. However, when only a short time series is available for estimating the optimal weights, the investor can benefit a lot if he uses a more efficient estimator of mean and variance.

In the second half of Table V, we consider a more realistic situation when the investor does not know the degrees of freedom of the $t$-distribution and he has to estimate it from the data. As can
be seen from Table V, the need to estimate the degrees to freedom adds some additional volatility to \( \hat{\mu} \) and \( \hat{V} \) in finite samples, and this reduces the utility gain of using the maximum likelihood estimator. Nevertheless, the reduction of utility gain is relatively small, and the pattern in the first half of Table V also carries forward to the case of unknown degrees of freedom.

Many studies in the classical framework, such as Ang and Bekaert (2002) and Guidolin and Timmermann (2002), usually examine the utility gains without imposing short-sale constraints. If so, then in the absence of short selling restrictions, the utility gains here are clearly of great economic importance. However, Tu and Zhou (2003) analyze the utility gains in a Bayesian framework and find that the gains are small. This is due to their use of \( \tau = 2.83 \) and their focus on the cases with margin requirements. Nevertheless, their results are completely consistent with those here when \( \tau = 3 \) and when no short selling is allowed.

III. Asset Pricing Tests

The test of multi-factor pricing model is usually cast in a regression framework. 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,
\[ r_t = \alpha + \beta f_t + \epsilon_t, \tag{27} \]
where \( \epsilon_t \) is an \( N \times 1 \) vector of residuals with zero mean and a non-singular covariance matrix. To relate \( \alpha \) and \( \beta \) to the earlier parameters \( \mu \) and \( V \), consider the corresponding partition
\[ \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{28} \]
Under the usual multivariate normal distribution, it is clear that the distribution of \( r_t \) conditional on \( f_t \) is also normal and
\[ E[r_t|f_t] = \mu_1 + V_{12}V_{22}^{-1}(f_t - \mu_2), \tag{29} \]
\[ \text{Var}[r_t|f_t] = V_{11} - V_{12}V_{22}^{-1}V_{21}. \tag{30} \]
Therefore, the parameters \( \alpha, \beta \) and the earlier parameters \( \mu, V \) obey the following relationship:
\[ \alpha = \mu_1 - \beta \mu_2, \quad \beta = V_{12}V_{22}^{-1}. \tag{31} \]
Denote $\Sigma$ as the usual notation for the covariance matrix of $\epsilon_t$,

$$\Sigma = \text{Var}[\epsilon_t] = V_{11} - V_{12}V_{22}^{-1}V_{21}. \quad (32)$$

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

$$\text{Var}[r_t|f_t] = \left[\frac{\nu - 2 + (f_t - \mu_2)V_{22}^{-1}(f_t - \mu_2)}{\nu + k - 2}\right]\Sigma. \quad (33)$$

This is a key feature of our $t$ model versus those in the econometric and statistical literature where $f_t$ is treated as fixed and $\epsilon_t$ is assumed to be $t$ distributed with a constant covariance matrix. In contrast, $f_t$ here is random and the conditional covariance matrix of $\epsilon_t$ is time-varying. As shown below, this will have important implications for both parameter estimation and asset pricing tests.

Consider first the estimation accuracy of the alphas and betas under $t$. The estimates under $t$, $\hat{\alpha}$ and $\hat{\beta}$, are obtained from equation (31) by replacing $\mu$ and $V$ with their maximum likelihood estimates. The key issue is how accurate $\hat{\alpha}$ and $\hat{\beta}$ are as 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:

$$\text{Avar} \left[\begin{array}{c} \hat{\alpha} \\ \text{vec}(\hat{\beta}) \end{array}\right] = \left(\frac{\nu + n + 2}{\nu + n}\right) \left[\begin{array}{cc} \left(\frac{\nu - 2}{\nu}\right) + \mu_2'V_{22}^{-1}\mu_2 & -\mu_2'V_{22}^{-1} \\ -V_{22}^{-1}\mu_2 & V_{22}^{-1} \end{array}\right] \otimes \Sigma. \quad (34)$$

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

$$\text{Avar} \left[\begin{array}{c} \hat{\alpha} \\ \text{vec}(\hat{\beta}) \end{array}\right] = \left[1 + \left(\frac{\nu - 2}{\nu}\right)\mu_2'V_{22}^{-1}\mu_2 - \left(\frac{\nu - 2}{\nu}\right)\mu_2'V_{22}^{-1}\right] \otimes \Sigma. \quad (35)$$

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

$$1 - \left(\frac{\nu + n + 2}{\nu + n}\right) \left[\begin{array}{c} \left(\frac{\nu - 2}{\nu}\right) + \mu_2'V_{22}^{-1}\mu_2 \\ 1 + \left(\frac{\nu - 2}{\nu}\right)\mu_2'V_{22}^{-1}\mu_2 \end{array}\right] = \frac{2}{\nu + n} \left[\frac{n + 4}{\nu - 4} + \frac{n + 4}{\nu - 4}\right] \mu_2'V_{22}^{-1}\mu_2. \quad (36)$$

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 $\mu_2'V_{22}^{-1}\mu_2 \to \infty$. 17
When \( n = 28 \) and \( \nu = 8 \), the percentage variance reduction of \( \tilde{\alpha} \) ranges from 20.83% to 29.63%. Similarly, the percentage improvement of \( \tilde{\beta} \) is:

\[
1 - \frac{\nu + n + 2}{\nu + n} = \frac{2(n + 4)}{(\nu - 2)(\nu + n)}.
\]

(37)

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

Table VI provides the estimates of the alphas and betas under normality and \( t \) (with \( \nu = 8 \)) for the Fama-French data. It is seen that there are large differences in the alpha estimates. For example, the alphas for S1B2 and S3B4 are \(-0.004\) and \(0.011\) under normality, but change to \(-0.082\) and \(0.071\), about 20 times and 7 times larger in absolute value, respectively. In contrast, the differences in the beta estimates are much smaller in percentage terms. For both the MKT and SMB factors, the betas are virtually the same under either \( t \) or normal. However, there are some substantial differences in the HML betas. The betas for S3B2 and S4B2 have reduced significantly from 0.221 and 0.263 to 0.095 and 0.141. Overall, it appears that the usual OLS betas are fairly accurately estimated, but the alphas are not. As a result, there is a great value of obtaining more accurate estimate of \( \alpha \) for both asset pricing tests and performance evaluation.

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

\[
H_0 : \quad \alpha = 0_N
\]

(38)
in the multivariate regressions of

\[
\mathbf{r}_t = \mathbf{\alpha} + \mathbf{\beta}_t \mathbf{f}_t + \mathbf{\epsilon}_t, \quad t = 1, \ldots, T.
\]

(39)

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

\[
\text{GRS} = \left( \frac{T - N - k}{N} \right) \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1 + \hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}} \sim F_{N,T-N-k},
\]

(40)

where \( \hat{\alpha} \) and \( \hat{\Sigma} \) are obtained from either linear regressions or from the relations between them and \( \hat{\mu} \) and \( \hat{V} \).\(^6\) If the normality is violated, and if the regression residuals are homoskedastic (having

\(^6\)Unlike earlier in the paper, the sample estimate \( \hat{V} \) here is, with minor scaling change, \( \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{\mu})(x_t - \hat{\mu})' \).
constant covariance matrix), the null hypothesis can be tested by the standard likelihood ratio test,

$$LRT_n = \left( T - \frac{N}{2} - k - 1 \right) \log \left( 1 + \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + (1 + \kappa) \hat{\mu}_2' \hat{V}_{22}^{-1} \hat{\mu}_2} \right) \overset{\sim}{\sim} \chi^2_N. \quad (41)$$

Note that 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.\(^7\) It is easy to see that the GRS test is simply a suitable transformation of this likelihood ratio test. However, once the normality assumption is replaced by a joint $t$ return assumption, one should be cautious in using the above tests as they are not strictly valid under $t$. To overcome this problem, Geczy (2001) provides an interesting correction to the asymptotic distribution under the $t$-distribution assumption rather than the normal,

$$LRT_n = \left( T - \frac{N}{2} - k - 1 \right) \log \left( 1 + \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + (1 + \kappa) \hat{\mu}_2' \hat{V}_{22}^{-1} \hat{\mu}_2} \right) \overset{\sim}{\sim} \chi^2_N, \quad (42)$$

where $\kappa$ is the standardized kurtosis as defined by (7). In addition, Geczy (2001) also suggests an approximate $F$-test

$$GRS_a = \left( \frac{T - N - k}{N} \right) \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + (1 + \kappa) \hat{\mu}_2' \hat{V}_{22}^{-1} \hat{\mu}_2} \sim F_{N,T-N-k}. \quad (43)$$

Notice that $LRT_n$ and $GRS_a$ are computed based on the parameter estimates obtained under the normality assumption. As shown earlier, these estimates are inefficient when the returns are $t$ rather than normal, so the associated tests might be less powerful than those based on the $t$ estimates.

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

$$u_t^{(i)} = \frac{\nu + n}{\nu + (r_t - \hat{\beta}^{(i)} f_t)' [\hat{\Psi}_r^{(i)}]^{-1} (r_t - \hat{\beta}^{(i)} f_t) + (f_t - \hat{\mu}_2)' [\hat{\Psi}_{22}^{(i)}]^{-1} (f_t - \hat{\mu}_2')}, \quad (44)$$

$$\tilde{Y}^{(i)} = [r_1 \sqrt{u_1^{(i)}}, r_2 \sqrt{u_2^{(i)}}, \ldots, r_T \sqrt{u_T^{(i)}}], \quad (45)$$

$$\tilde{X}^{(i)} = [f_1 \sqrt{u_1^{(i)}}, f_2 \sqrt{u_2^{(i)}}, \ldots, f_T \sqrt{u_T^{(i)}}]. \quad (46)$$

\(^7\)See Muirhead (1982, Theorem 10.5.5) for a derivation of this Bartlett correction.
where the iteration can start from, say, the estimates under normal as before. With the restricted parameter estimates denoted by \( \tilde{\mu}_r \) and \( \tilde{\Psi}_r \), we can compute the likelihood ratio test under \( t \):

\[
\text{LRT}_t \equiv 2 \left( \frac{T - \left( \frac{N}{2} \right) - k - 1}{T} \right) \left[ \log L(\tilde{\mu}, \tilde{\Psi}) - \log L(\tilde{\mu}_r, \tilde{\Psi}_r) \right] \sim \chi^2_N,
\]

where \( \log L(\cdot, \cdot) \) is the log-likelihood function under \( t \) given by (9), and the adjustment factor \( [T - (N/2) - k - 1]/T \) is introduced to improve the small sample properties of the test statistic.

As the CAPM of Sharpe (1964) and Lintner (1965) is of fundamental importance in finance, it is of interest to test it first. This amounts to testing (38) with the single and theory-motivated market factor in (39). Table VII reports the testing results based on the GRS test and the likelihood ratio tests. Over the entire sample period of July 1963 to December 2002, all of the tests reject the CAPM strongly with virtually zero \( p \)-values, whether we assume the underlying distribution is normal or \( t \). However, given the strong rejection of the normality assumption, one cannot draw firm conclusions about the rejections from test statistics that assume normality. With the Geczy’s tests and those new tests developed here also reaffirming the rejections reached by the GRS test under normality, one can claim the rejection is indeed caused by the failure of the model rather than the violation of the restrictive normality assumption.

However, it is not always the case that these tests give the same conclusion. For example, in the first subperiod of July 1963 to June 1983, we find that both the GRS test and the likelihood ratio test that are based on the OLS estimators of alpha do not reject the CAPM at the usual 5% level, but the likelihood ratio test computed under \( t \) with \( \nu = 10, 8 \) or 6, suggests rejection of the CAPM at the 5% level. This is clearly a case where the \( t \) based test makes a difference by suggesting the
CAPM fails to hold in the first subperiod while the other tests are unable to do so.\footnote{Although not reported here, we apply the same analysis to 10 size-sorted portfolios of the NYSE and find cases where the two likelihood ratio tests give conflicting conclusions on the validity of an asset pricing model.} For the second subperiod, all the tests reject the CAPM strongly as they do for the entire sample period.

Consider now testing the Fama-French three factors model where \( f_t \) in (39) are the market (MKT), book-to-market (HML) and size (SML) factors. Table VIII reports the testing results of this three factors model. For the entire sample period July 1963 to December 2002, the GRS statistic has a value of 3.021 under normality, implying a \( p \)-value of less than 0.01\% from its exact \( F \)-distribution. The likelihood ratio test, \( \text{LRT}_n \), has a value of 71.58, implying also a \( p \)-value less than 0.01\% from its asymptotic \( \chi^2_{25} \) distribution. Hence, imposing normality, both tests strongly reject the Fama-French three factors model.

When the data is modeled by a \( t \) with \( \nu = 8 \), the likelihood ratio test computed under such a \( t \), \( \text{LRT}_t \), is 86.59, implying also a \( p \)-value less than 0.01\% from its asymptotic \( \chi^2_{25} \) distribution. Geczy’s adjusted GRS and \( \text{LRT}_n \) also have \( p \)-values less than 0.01\%. So, all three tests reach the same conclusion of forcefully rejecting the three factors model for the period of July 1967 to December 2002.

By dividing the data into two subperiods, we find that all of the tests fail to reject the model in the first subperiod, but they reject the model strongly in the second subperiod. The first subperiod result is consistent with Fama and French (1993) who reported that the GRS test had a \( p \)-value of 3.9\% based on the data from July 1967 to December 1991. Their result can be replicated here if we exclude the data of the last 11 years. Therefore, the strong rejection for the entire sample period is driven by the addition of 11 years of data, which increases both the power of the test as well as the magnitude of the statistic because the deviations of the alphas from zero are greater than those for the shorter sample period used by Fama and French (1993).

### IV. Extensions and Future Research

Like many empirical asset pricing and corporate studies, we assume that the asset returns are jointly iid over time. Moreover, they have a multivariate \( t \)-distribution at any time. Although the
multivariate $t$-distribution is still restrictive, it is more general and realistic than the normality assumption and yet it contains the normal as a special limiting case.

While the iid 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 recent 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 (2003), among others, model the dynamics of conditional variables based on normality assumption. The trade-off between the two approaches is precision and generality. Clearly, the $t$-distribution advocated here can be used to offer some more generality than the normal as well as improving the estimation accuracy of parameters.

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

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

Finally, it is worth noting that the more the asset return deviates from normality, the greater difference it tends to make in estimating the asset’s expected return and alpha by using the max-
imum 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.\textsuperscript{9} Another important issue is on mutual fund performance where it is well-known that the estimation of alpha is very inaccurate due to outliers. In all these applications, it is sensible to estimate alpha based on the $t$-distribution as it may provide a more reliable estimate than the usual OLS estimator of $\alpha$.

V. Conclusions

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

Applying multivariate $t$-distributions to Fama and French’s (1993) 25 portfolio returns and their 3 factors from January 1963 to December 2002, we find that there are drastic differences in estimating the expected asset returns, alphas and betas, in constructing the optimal portfolio and in testing asset pricing models. It appears that the proposed approach is useful in a number of areas to ask how sensitive the results are to the usual normality assumption. Leading examples in this regard are cost of capital estimation, performance evaluations, IPO abnormal returns and risk management.

\begin{footnote}{Ritter (1991) raises some of the interesting issues and Lyon, Barber, and Tsai (1999) and references therein provide some of the latest methodologies.}
\end{footnote}
Appendix

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 formulae 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 statistics literature, so we provide a derivation for them here.

Proof of the first algorithm, (10)–(12):

As noted earlier, the key difficulty associated with maximizing the log-likelihood function under $t$, equation (9), is that the terms do not combine to yield tractable solutions. However, it is well-known that a $t$ distribution is a mixture of normal. That is, there exists $u_t \sim \chi^2_\nu/\nu$ such that, conditional on $u_t$, $x_t$ is normal:

$$x_t \sim N(\mu, \Psi/u_t). \quad (A1)$$

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

$$L(x_t|u_t) = \frac{n^2}{2} \left[ \sum_{t=1}^{T} \log(u_t) - T \log(2\pi) \right] - \frac{T}{2} \log(|\Psi|) - \frac{1}{2} \sum_{t=1}^{T} u_t(x_t - \mu)'\Psi^{-1}(x_t - \mu), \quad (A2)$$

which can be obviously maximized with

$$\tilde{\mu} = \frac{\sum_{t=1}^{T} u_t x_t}{\sum_{t=1}^{T} u_t}, \quad (A3)$$

$$\tilde{\Psi} = \frac{1}{T} \sum_{t=1}^{T} u_t(x_t - \tilde{\mu})(x_t - \tilde{\mu}). \quad (A4)$$

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

$$E[u_t|x_t; \mu, \Psi] = \frac{\nu + n}{\nu + (x_t - \mu)'\Psi^{-1}(x_t - \mu)}. \quad (A5)$$

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 \( u_t \). Continuing iterations may converge to the solution that maximizes the unconditional log-likelihood function, equation (9). Fortunately, for our problems here and many other models, the EM algorithm indeed converges and it even converges monotonically. \( Q.E.D. \)

**Proof of the asymptotic variance-covariance matrix for \( \tilde{V} \), (16):**

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

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

(A6)

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

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

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

(A7)

Based on the following identities

\[
[D_n' (\Psi^{-1} \otimes \Psi^{-1}) D_n]^{-1} = D_n^+ (\Psi \otimes \Psi) D_n^+, \quad (A8)
\]

\[
D_n^+ D_n' \text{vec}(\Psi^{-1}) = D_n D_n^+ \text{vec}(\Psi^{-1}) = D_n \text{vech}(\Psi^{-1}) = \text{vec}(\Psi^{-1}), \quad (A9)
\]

\[
\text{vec}(\Psi^{-1}) \text{vec}(\Psi)' = \text{tr}(\Psi^{-1} \Psi) = n, \quad (A10)
\]

\[
(\Psi^{-1} \otimes \Psi^{-1}) \text{vec}(\Psi) = \text{vec}(\Psi^{-1} \Psi \Psi^{-1}) = \text{vec}(\Psi^{-1}), \quad (A11)
\]

we can analytically invert \( J_{\psi\psi} \) as

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

(A12)

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

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

(A13)
In particular, we have
\[
\text{Acov} [\tilde{V}_{ij}, \tilde{V}_{kl}] = \left( \frac{2(\nu + n + 2)}{\nu(\nu + n)} \right) V_{ij} V_{kl} + \left( \frac{\nu + n + 2}{\nu + n} \right) (V_{ik} V_{jl} + V_{il} V_{jk}). \tag{A14}
\]
A combination of (A6) and (A14) yields (16). \(Q.E.D.\)

Proof of the second algorithm, (44)–(50):

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
\[
x_t \sim N(\mu, \Psi/u_t). \tag{A15}
\]
Conditional on \(f_t\) and \(u_t\) and under the assumption that \(\alpha = 0_N\), we have
\[
\tilde{r}_t | f_t, u_t \sim N(\beta f_t, \Psi_{\epsilon}/u_t). \tag{A16}
\]
Therefore, conditional on \(u_t\), the log-likelihood function of \((r'_t, f'_t)'\) is
\[
\mathcal{L}(r_t, f_t | u_t) = \mathcal{L}(r_t | f_t, u_t) + \mathcal{L}(f_t | u_t)
\]
\[
= \frac{N}{2} \left[ \sum_{t=1}^{T} \log(u_t) - T \log(2\pi) \right] - \frac{T}{2} \log(|\Psi_{\epsilon}|) - \frac{1}{2} \sum_{t=1}^{T} (r_t - \beta f_t)'(\Psi_{\epsilon}/u_t)^{-1}(r_t - \beta f_t)
\]
\[
+ \frac{k}{2} \left[ \sum_{t=1}^{T} \log(u_t) - T \log(2\pi) \right] - \frac{T}{2} \log(|\Psi_{22}|) - \frac{1}{2} \sum_{t=1}^{T} [u_t(f_t - \mu_2)'\Psi_{22}^{-1}(f_t - \mu_2)]. \tag{A17}
\]
Note that the first part of the likelihood function has parameters \(\beta\) and \(\Psi_{\epsilon}\) and the second part has parameters \(\mu_2\) and \(\Psi_{22}\). So we can maximize them separately. For the second part, it is clear that
\[
\tilde{\mu}_2 = \frac{\sum_{t=1}^{T} u_t f_t}{\sum_{t=1}^{T} u_t}, \tag{A18}
\]
\[
\tilde{\Psi}_{22} = \frac{1}{T} \sum_{t=1}^{T} u_t (f_t - \tilde{\mu}_2)'(f_t - \tilde{\mu}_2)'. \tag{A19}
\]
Therefore, we can focus our attention no the first part of the conditional likelihood function. Denote \(\bar{Y} = [r_1\sqrt{u_1}, r_2\sqrt{u_2}, \ldots, r_T\sqrt{u_T}]\) and \(\bar{X} = [f_1\sqrt{u_1}, f_2\sqrt{u_2}, \ldots, f_T\sqrt{u_T}]\), we can write the first part as
\[
\mathcal{L}(r_t | f_t, u_t) = \frac{N}{2} \left[ \sum_{t=1}^{T} \log(u_t) - T \log(2\pi) \right] - \frac{T}{2} \log(|\Psi_{\epsilon}|) - \frac{1}{2} \sum_{t=1}^{T} (\tilde{Y}_t - \beta \tilde{X}_t)'\Psi_{\epsilon}^{-1}(\tilde{Y}_t - \beta \tilde{X}_t), \tag{A20}
\]
which has the standard form of the multivariate normality case and hence, conditional on \( u_t \), the maximum likelihood estimator of \( \beta \) and \( \Psi_\epsilon \) under the null are

\[
\tilde{\beta} = (\tilde{Y}'\tilde{X})(\tilde{X}'\tilde{X})^{-1}, \\
\tilde{\Psi}_\epsilon = \frac{1}{T}(\tilde{Y} - \tilde{X}\tilde{\beta}')(\tilde{Y} - \tilde{X}\tilde{\beta}').
\]

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

**Proof of the asymptotic variance-covariance matrix for \( \tilde{\alpha} \) and \( \tilde{\beta} \), (34):**

In our derivation, we make use of the commutation matrix and the duplication matrix.\(^{10} \) Duplication matrix was defined earlier in the Appendix. For the commutation matrix \( K_{mn} \), it is defined as the unique \( mn \times mn \) consisting of 0’s and 1’s such that \( K_{mn}\text{vec}(A) = \text{vec}(A') \) for an \( m \times n \) matrix \( A \). If \( m = n \), \( K_{nn} \) is simply denoted as \( K_n \). Commutation matrix allows us to commute two matrices in a Kronecker product. 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 we have

\[
\text{Avar}[\tilde{\mu}] = \left( \frac{\nu + n + 2}{\nu + n} \right) \Psi, \\
\text{Avar}[\text{vech}(\tilde{\Psi})] = \frac{2(\nu + n + 2)}{\nu + n} \left[ D_\nu^+(\Psi \otimes \Psi)D_\nu^+ + \frac{\text{vech}(\Psi)\text{vech}(\Psi)'}{\nu} \right].
\]

We first prove that

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

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

\[
\text{vec}(\beta) = (\Psi_{22}^{-1} \otimes I_N)\text{vec}(\Psi_{12}) = (I_k \otimes \Psi_{12})\text{vec}(\Psi_{22}^{-1}).
\]

It follows that

\[
\frac{\partial \text{vec}(\beta)}{\partial \text{vec}(\Psi_{12})'} = \Psi_{22}^{-1} \otimes I_N, \\
\frac{\partial \text{vec}(\beta)}{\partial \text{vech}(\Psi_{22})'} = (I_k \otimes \Psi_{12})\frac{\partial \text{vec}(\Psi_{22}^{-1})}{\partial \text{vech}(\Psi_{22})'}
\]

\[
= -(I_k \otimes \Psi_{12})(\Psi_{22}^{-1} \otimes \Psi_{22}^{-1})D_k \\
= (\Psi_{22}^{-1} \otimes -\beta)D_k.
\]

\(^{10}\text{See Harville (1997, Chapter 16) for a review of the properties of the commutation and the duplication matrices.}\)
Also, note that
\[
\begin{align*}
\text{vec}(\Psi_{12}) &= \text{vec}([I_N, O_{N \times k}]\Psi[O_{k \times N}, I_k]^t) \\
&= ([O_{k \times N}, I_k] \otimes [I_N, O_{N \times k}])D_n\text{vec}(\Psi), \quad \text{(A29)} \\
\text{vech}(\Psi_{22}) &= D_k^+\text{vec}([O_{k \times N}, I_k]\Psi[O_{k \times N}, I_k]^t) \\
&= D_k^+([O_{k \times N}, I_k] \otimes [O_{k \times N}, I_k])D_n\text{vec}(\Psi). \quad \text{(A30)}
\end{align*}
\]

Using the delta method, we have
\[
\begin{align*}
\text{Avar}[\text{vec}(\hat{\beta})] &= \text{Avar} \left[ [\Psi_{22}^{-1} \otimes I_N, (\Psi_{22}^{-1} \otimes -\beta)D_k] \left[ [O_{k \times N}, I_k] \otimes [I_N, O_{N \times k}] \right] \right] D_n\text{vec}(\hat{\Psi}) \\
&= \text{Avar} \left( ([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])D_n\text{vec}(\hat{\Psi}) \right) \\
&= \frac{2(\nu + n + 2)}{\nu + n} \left( ([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])D_n \left[ D_n^+(\Psi \otimes \Psi)D_n^+ + \frac{\text{vech}(\Psi)(\text{vech}(\Psi))^t}{\nu} \right] \right. \\
&\quad \times \left. \left( (I_n^2 + K_n)(\Psi \otimes \Psi) + \frac{2\text{vec}(\Psi)\text{vec}(\Psi)^t}{\nu} \right) \right) \\
&\quad \times \left( ([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])' \right), \quad \text{(A31)}
\end{align*}
\]

where the third equality follows from the identity
\[
D_k^+D_k^+(A \otimes A)D_n = (A \otimes A)D_n \quad \text{(A32)}
\]
for a \(k \times n\) matrix \(A\), and the fourth equality follows from the identity
\[
2D_nD_n^+(\Psi \otimes \Psi)D_n^+D_n' = \frac{1}{2}(I_n^2 + K_n)(\Psi \otimes \Psi)(I_n^2 + K_n) = (I_n^2 + K_n)(\Psi \otimes \Psi) \quad \text{(A33)}
\]
because \(2D_nD_n^+ = I_n^2 + K_n\). Using (A31) and the following identities
\[
\begin{align*}
([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])K_n(\Psi \otimes \Psi) ([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])' &= K_{nk} ([I_N, -\beta] \otimes [O_{k \times N}, \Psi_{22}^{-1}])(\Psi \otimes \Psi) ([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])' \\
&= K_{nk} ([\Psi e \otimes O_{N \times k} \otimes [\beta', I_k]) ([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])' \\
&= O_{nk \times nk}, \quad \text{(A34)}
\end{align*}
\]
\[
([O_{k \times N}, \Psi_{22}^{-1}] \otimes [I_N, -\beta])\text{vec}(\Psi) = \text{vec} ([I_N, -\beta]\Psi[O_{k \times N}, \Psi_{22}^{-1}]') = 0_{nk}, \quad \text{(A35)}
\]
we have
\[
\text{Avar}[\text{vec}(\bar{\beta})] = \left( \frac{\nu + n + 2}{\nu + n} \right) \left( [I_N, -\beta] \otimes (\Psi^{-1} \otimes \Psi) ([I_N, -\beta]^\prime \otimes [I_N, -\beta]) \right)
\]
\[
= \left( \frac{\nu + n + 2}{\nu + n} \right) (\Psi^{-1} \otimes \Psi). \tag{A36}
\]

As \( \alpha = \mu_1 - \beta \mu_2 \), we have
\[
\frac{\partial \alpha}{\partial \mu^l} = [I_N, -\beta], \tag{A37}
\]
\[
\frac{\partial \alpha}{\partial \text{vec}(\bar{\beta})^\prime} = -\mu_2^l \otimes I_N. \tag{A38}
\]

Using the delta method and the fact that \( \bar{\mu} \) and \( \bar{\beta} \) are asymptotically independent, we have
\[
\text{Avar} \left[ \frac{\bar{\alpha}}{\text{vec}(\bar{\beta})} \right] = (-\mu_2^l \otimes I_N)(\Psi^{-1} \otimes \Psi) = -\mu_2^l \Psi^{-1} \otimes \Psi. \tag{A39}
\]

Similarly, the asymptotic covariance of \( \bar{\alpha} \) and \( \text{vec}(\bar{\beta}) \) is given by
\[
(-\mu_2^l \otimes I_N)\text{Avar}[\text{vec}(\bar{\beta})] = (-\mu_2^l \otimes I_N)(\Psi^{-1} \otimes \Psi) = -\mu_2^l \Psi^{-1} \otimes \Psi. \tag{A40}
\]

Note that the expressions so far are written in terms of \( \Psi_{22} \) and \( \Psi_\epsilon \) but not in terms of the variance of \( f_t \) and \( \epsilon_t \). For comparison with the asymptotic variance of \( \hat{\alpha} \) and \( \text{vec}(\hat{\beta}) \), 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 \( \bar{\alpha} \) and \( \text{vec}(\bar{\beta}) \) as
\[
\text{Avar} \left[ \frac{\bar{\alpha}}{\text{vec}(\bar{\beta})} \right] = \frac{\nu + n + 2}{\nu + n} \left[ \begin{array}{cc}
\left( \frac{\nu - 2}{\nu - 4} \right) \mu_2^l V_{22}^{-1} \mu_2 & -\mu_2^l V_{22}^{-1} \\
-\mu_2^l V_{22}^{-1} \mu_2 & V_{22}^{-1}
\end{array} \right] \otimes \Sigma, \tag{A41}
\]
which is the expression in (34). Although not provided here, it can be shown that the asymptotic variance of \( \bar{\alpha} \) and \( \bar{\beta} \) remains the same even when the degrees of freedom \( \nu \) is unknown.

For the asymptotic variance of the OLS estimator \( \hat{\alpha} \) and \( \text{vec}(\hat{\beta}) \) under the multivariate \( t \)-distribution, we have from Geczy (2001) that
\[
\text{Avar} \left[ \frac{\hat{\alpha}}{\text{vec}(\hat{\beta})} \right] = \left[ \begin{array}{cc}
1 + \left( \frac{\nu - 2}{\nu - 4} \right) & \mu_2^l V_{22}^{-1} \mu_2 \\
\mu_2^l V_{22}^{-1} \mu_2 & \left( \frac{\nu - 2}{\nu - 4} \right)\mu_2^l V_{22}^{-1}
\end{array} \right] \otimes \Sigma. \tag{A42}
\]

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


Figure 1
Q-Q Pots of S1B1 and MKT

The figure presents the Q-Q plots of the monthly excess returns of S1B1 and MKT over the period of July 1963 to December 2002. S1B1 is the portfolio that has the smallest size and book-to-market out of the 25 Fama and French (1993) portfolios, and MKT is the value-weighted combined NYSE-AMEX-NASDAQ market portfolio.
Table I

Normality Tests

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

<table>
<thead>
<tr>
<th></th>
<th>Student-( t ) with df</th>
<th>Student-( t ) with df</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skewness</td>
<td>Normal</td>
</tr>
<tr>
<td>Univariate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1B1</td>
<td>0.00</td>
<td>95.09</td>
</tr>
<tr>
<td>S1B2</td>
<td>0.00</td>
<td>84.22</td>
</tr>
<tr>
<td>S1B3</td>
<td>0.01</td>
<td>40.14</td>
</tr>
<tr>
<td>S1B4</td>
<td>0.02</td>
<td>20.75</td>
</tr>
<tr>
<td>S1B5</td>
<td>0.01</td>
<td>31.11</td>
</tr>
<tr>
<td>S2B1</td>
<td>0.08</td>
<td>1.07</td>
</tr>
<tr>
<td>S2B2</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td>S2B3</td>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>S2B4</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td>S2B5</td>
<td>0.09</td>
<td>0.85</td>
</tr>
<tr>
<td>S3B1</td>
<td>0.09</td>
<td>0.57</td>
</tr>
<tr>
<td>S3B2</td>
<td>0.38</td>
<td>0.00</td>
</tr>
<tr>
<td>S3B3</td>
<td>0.38</td>
<td>0.00</td>
</tr>
<tr>
<td>S3B4</td>
<td>0.13</td>
<td>0.22</td>
</tr>
<tr>
<td>S3B5</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>S4B1</td>
<td>0.03</td>
<td>12.83</td>
</tr>
<tr>
<td>S4B2</td>
<td>0.31</td>
<td>0.00</td>
</tr>
<tr>
<td>S4B3</td>
<td>0.18</td>
<td>0.01</td>
</tr>
<tr>
<td>S4B4</td>
<td>0.00</td>
<td>58.43</td>
</tr>
<tr>
<td>S4B5</td>
<td>0.06</td>
<td>3.34</td>
</tr>
<tr>
<td>S5B1</td>
<td>0.03</td>
<td>12.62</td>
</tr>
<tr>
<td>S5B2</td>
<td>0.12</td>
<td>0.26</td>
</tr>
<tr>
<td>S5B3</td>
<td>0.07</td>
<td>1.66</td>
</tr>
<tr>
<td>S5B4</td>
<td>0.00</td>
<td>78.38</td>
</tr>
<tr>
<td>S5B5</td>
<td>0.03</td>
<td>11.29</td>
</tr>
<tr>
<td>SMB</td>
<td>0.26</td>
<td>0.00</td>
</tr>
<tr>
<td>HML</td>
<td>0.01</td>
<td>38.40</td>
</tr>
<tr>
<td>MKT</td>
<td>0.23</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Multivariate

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>172.41</td>
<td>0.00</td>
<td>0.38</td>
<td>5.91</td>
<td>61.40</td>
<td>1159.66</td>
<td>0.00</td>
<td>0.39</td>
<td>15.98</td>
<td>94.11</td>
</tr>
</tbody>
</table>
Table II
Estimation of Mean and Standard Deviation under Normality versus under $t$
The table reports the maximum likelihood estimates of means and standard deviations of Fama-French benchmark portfolios and factors based on monthly returns from July 1963 to December 2002, assuming the returns are generated from multivariate normality or multivariate $t$-distribution with degrees of freedom 10, 8, and 6, respectively.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$-distribution</td>
</tr>
<tr>
<td></td>
<td>Normal df=10 df=8 df=6</td>
</tr>
<tr>
<td>S1B1</td>
<td>0.135 -0.112 -0.117 -0.121</td>
</tr>
<tr>
<td>S1B2</td>
<td>0.703 0.469 0.464 0.459</td>
</tr>
<tr>
<td>S1B3</td>
<td>0.784 0.536 0.530 0.523</td>
</tr>
<tr>
<td>S1B4</td>
<td>0.995 0.771 0.766 0.761</td>
</tr>
<tr>
<td>S1B5</td>
<td>1.070 0.876 0.871 0.865</td>
</tr>
<tr>
<td>S2B1</td>
<td>0.305 0.125 0.121 0.116</td>
</tr>
<tr>
<td>S2B2</td>
<td>0.567 0.418 0.416 0.413</td>
</tr>
<tr>
<td>S2B3</td>
<td>0.830 0.715 0.713 0.711</td>
</tr>
<tr>
<td>S2B4</td>
<td>0.892 0.768 0.765 0.762</td>
</tr>
<tr>
<td>S2B5</td>
<td>0.926 0.753 0.748 0.743</td>
</tr>
<tr>
<td>S3B1</td>
<td>0.334 0.237 0.236 0.235</td>
</tr>
<tr>
<td>S3B2</td>
<td>0.646 0.546 0.544 0.541</td>
</tr>
<tr>
<td>S3B3</td>
<td>0.659 0.588 0.587 0.585</td>
</tr>
<tr>
<td>S3B4</td>
<td>0.792 0.705 0.702 0.699</td>
</tr>
<tr>
<td>S3B5</td>
<td>0.937 0.782 0.776 0.770</td>
</tr>
<tr>
<td>S4B1</td>
<td>0.459 0.306 0.300 0.294</td>
</tr>
<tr>
<td>S4B2</td>
<td>0.435 0.317 0.313 0.309</td>
</tr>
<tr>
<td>S4B3</td>
<td>0.635 0.517 0.512 0.506</td>
</tr>
<tr>
<td>S4B4</td>
<td>0.758 0.621 0.614 0.608</td>
</tr>
<tr>
<td>S4B5</td>
<td>0.850 0.694 0.687 0.679</td>
</tr>
<tr>
<td>S5B1</td>
<td>0.386 0.334 0.329 0.323</td>
</tr>
<tr>
<td>S5B2</td>
<td>0.415 0.383 0.380 0.377</td>
</tr>
<tr>
<td>S5B3</td>
<td>0.455 0.421 0.419 0.416</td>
</tr>
<tr>
<td>S5B4</td>
<td>0.559 0.486 0.481 0.476</td>
</tr>
<tr>
<td>S5B5</td>
<td>0.523 0.429 0.421 0.412</td>
</tr>
<tr>
<td>SMB</td>
<td>0.210 0.126 0.127 0.129</td>
</tr>
<tr>
<td>HML</td>
<td>0.432 0.401 0.399 0.397</td>
</tr>
<tr>
<td>MKT</td>
<td>0.410 0.361 0.358 0.354</td>
</tr>
</tbody>
</table>
Table III
Optimal Portfolio Weights when Short Selling is Allowed

The table reports the unconstrained optimal investments on 25 size and book-to-market ranked portfolios for an investor with a $100 portfolio and a quadratic utility function with relative risk aversion $\tau = 1, 3, \text{and} 9$, when the parameters are estimated under normality and $t$ with degrees of freedom 10, 8, and 6, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 1$</th>
<th></th>
<th>$\tau = 3$</th>
<th></th>
<th>$\tau = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$-distribution</td>
<td></td>
<td>$t$-distribution</td>
<td></td>
<td>$t$-distribution</td>
</tr>
<tr>
<td></td>
<td>Normal df=10 df=8 df=6</td>
<td></td>
<td>Normal df=10 df=8 df=6</td>
<td></td>
<td>Normal df=10 df=8 df=6</td>
</tr>
<tr>
<td></td>
<td>Unconstrained</td>
<td></td>
<td>Unconstrained</td>
<td></td>
<td>Unconstrained</td>
</tr>
<tr>
<td>S1B1</td>
<td>1545.2 -1600.9 -1518.0 -1366.2</td>
<td>515.1 -533.6 -506.0 -455.4</td>
<td>-171.7 -177.9 -168.7 -151.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1B2</td>
<td>444.6 604.4 576.3 521.5</td>
<td>148.2 201.5 192.1 173.8</td>
<td>49.4 67.2 64.0 57.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1B3</td>
<td>26.2 -422.3 -417.0 -391.5</td>
<td>8.7 -140.8 -139.0 -130.5</td>
<td>2.9 -46.9 -46.3 -43.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1B4</td>
<td>1624.0 1420.3 1346.1 1212.6</td>
<td>541.3 473.4 448.7 404.2</td>
<td>180.4 157.8 149.6 134.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1B5</td>
<td>363.7 609.1 581.7 526.1</td>
<td>121.2 203.0 193.9 175.4</td>
<td>40.4 67.7 64.6 58.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2B1</td>
<td>-74.4 -103.4 -104.4 -100.9</td>
<td>-24.8 -34.5 -34.8 -33.6</td>
<td>-8.3 -11.5 -11.6 -11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2B2</td>
<td>-440.6 -397.4 -373.3 -332.5</td>
<td>-146.9 -132.5 -124.4 -110.8</td>
<td>-49.0 -44.2 -41.5 -36.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2B3</td>
<td>668.2 801.9 769.3 702.7</td>
<td>222.7 267.3 256.4 234.2</td>
<td>74.2 89.1 85.5 78.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2B4</td>
<td>105.4 157.7 154.1 143.4</td>
<td>35.1 52.6 51.4 47.8</td>
<td>11.7 17.5 17.1 15.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2B5</td>
<td>-372.1 -403.2 -384.9 -348.8</td>
<td>-124.0 -134.4 -128.3 -116.3</td>
<td>-41.3 -44.8 -42.8 -38.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3B1</td>
<td>-406.7 -152.9 -136.1 -113.6</td>
<td>-135.6 -51.0 -45.4 -37.9</td>
<td>-45.2 -17.0 -15.1 -12.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3B2</td>
<td>41.7 190.3 188.3 177.6</td>
<td>13.9 63.4 62.8 59.2</td>
<td>4.6 21.1 20.9 19.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3B3</td>
<td>-347.1 -95.0 -79.5 -60.8</td>
<td>-115.7 -31.7 -26.5 -20.3</td>
<td>-38.6 -10.6 -8.8 -6.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3B4</td>
<td>-110.8 153.3 156.7 152.3</td>
<td>-36.9 51.1 52.2 50.8</td>
<td>-12.3 17.0 17.4 16.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3B5</td>
<td>368.3 228.6 213.1 188.3</td>
<td>122.8 76.2 71.0 62.8</td>
<td>40.9 25.4 23.7 20.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4B1</td>
<td>1218.1 883.7 826.0 731.6</td>
<td>406.0 294.6 275.3 243.9</td>
<td>135.3 98.2 91.8 81.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4B2</td>
<td>-1005.7 -1138.9 -1083.3 -978.9</td>
<td>-335.2 -379.6 -361.1 -326.3</td>
<td>-111.7 -126.5 -120.4 -108.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4B3</td>
<td>-250.1 -340.6 -333.5 -311.4</td>
<td>-83.4 -113.5 -112.1 -103.8</td>
<td>-27.8 -37.8 -37.1 -34.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4B4</td>
<td>307.7 269.9 254.7 228.2</td>
<td>102.6 90.0 84.9 76.1</td>
<td>34.2 30.0 28.3 25.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4B5</td>
<td>-110.0 -177.6 -173.0 -160.4</td>
<td>-36.7 -59.2 -57.7 -53.5</td>
<td>-12.2 -19.7 -19.2 -17.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5B1</td>
<td>253.6 178.5 161.5 136.6</td>
<td>84.5 59.5 53.8 45.5</td>
<td>28.2 19.8 17.9 15.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5B2</td>
<td>-137.1 58.7 65.7 69.8</td>
<td>-45.7 19.6 21.9 23.3</td>
<td>-15.2 6.5 7.3 7.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5B3</td>
<td>410.3 252.0 236.4 210.8</td>
<td>136.8 84.0 78.8 70.3</td>
<td>45.6 28.0 26.3 23.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5B4</td>
<td>-75.0 -28.0 -26.0 -22.9</td>
<td>-25.0 -9.3 -8.7 -7.6</td>
<td>-8.3 -3.1 -2.9 -2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5B5</td>
<td>-133.4 -148.6 -143.1 -131.1</td>
<td>-44.5 -49.5 -47.7 -43.7</td>
<td>-14.8 -16.5 -15.9 -14.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cash</td>
<td>-723.6 -699.5 -658.0 -582.4</td>
<td>-687.1 -166.5 -152.7 -127.5</td>
<td>8.5 11.2 15.8 24.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table IV
Optimal Portfolio Weights under the Constraints of No Short Selling

The table reports the optimal investments on 25 size and book-to-market ranked portfolios under the constraints of no short selling for an investor with a $100 portfolio and a quadratic utility function with relative risk aversion $\tau = 1, 3, \text{ and } 9$, when the parameters are estimated under normality and $t$ with degrees of freedom $10, 8, \text{ and } 6$, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$t = 1$</th>
<th>$t = 3$</th>
<th>$t = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal df=10 df=8 df=6</td>
<td>Normal df=10 df=8 df=6</td>
<td>Normal df=10 df=8 df=6</td>
</tr>
<tr>
<td>S1B1</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S1B2</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S1B3</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S1B4</td>
<td>15.7 0.0 0.0 0.0 0.0</td>
<td>5.2 0.0 0.0 0.0 0.0</td>
<td>1.7 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S1B5</td>
<td>178.5 115.2 106.3 92.6</td>
<td>59.5 38.4 35.4 30.9</td>
<td>19.8 12.8 11.8 10.3</td>
</tr>
<tr>
<td>S2B1</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S2B2</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S2B3</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S2B4</td>
<td>0.0 42.0 48.4 52.8 0.0</td>
<td>0.0 14.0 16.1 17.6 0.0</td>
<td>0.0 4.7 5.4 5.9 0.0</td>
</tr>
<tr>
<td>S2B5</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S3B1</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S3B2</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S3B3</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S3B4</td>
<td>0.0 78.3 72.4 63.2 0.0</td>
<td>0.0 26.1 24.1 21.1 0.0</td>
<td>0.0 8.7 8.0 7.0 0.0</td>
</tr>
<tr>
<td>S3B5</td>
<td>96.6 75.8 67.5 56.1 0.0</td>
<td>32.2 25.3 22.5 18.7 0.0</td>
<td>10.7 8.4 7.5 6.2 0.0</td>
</tr>
<tr>
<td>S4B1</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S4B2</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S4B3</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S4B4</td>
<td>50.2 0.0 0.0 0.0 0.0</td>
<td>16.7 0.0 0.0 0.0 0.0</td>
<td>5.6 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S4B5</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S5B1</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S5B2</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S5B3</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S5B4</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>S5B5</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
<td>0.0 0.0 0.0 0.0 0.0</td>
</tr>
<tr>
<td>Cash</td>
<td>-241.1 -211.2 -194.5 -164.7 -13.7 -3.7 1.8 11.8</td>
<td>62.1 65.4 67.3 70.6</td>
<td></td>
</tr>
</tbody>
</table>

Cash
Table V
Expected Utility Gains

The table reports the expected utility gains (in percent) for a mean-variance-optimizing investor with relative risk aversion $\tau = 1, 3$ and $9$, when the investor switches from an optimal portfolio using mean and variance estimated under the normality assumption to an optimal portfolio using mean and variance estimated under the $t$-distribution assumption. The gains are based on 10,000 draws of data sets from a $t$-distribution with various degrees of freedom. Both the unconstrained and no short selling conditions are considered. In the first half of the table, the parameters under the $t$-distribution are estimated assuming the degrees of freedom are known whereas in the second half of the tables, the parameters are estimated assuming the degrees of freedom are unknown. For each case, we consider two different lengths of time series ($T = 474$ and $240$) that are used to estimate the optimal portfolio weights.

<table>
<thead>
<tr>
<th></th>
<th>$T = 474$</th>
<th>$T = 240$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 1$</td>
<td>$\tau = 3$</td>
</tr>
<tr>
<td>d.f.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.720</td>
<td>0.240</td>
</tr>
<tr>
<td>8</td>
<td>1.020</td>
<td>0.340</td>
</tr>
<tr>
<td>6</td>
<td>1.604</td>
<td>0.535</td>
</tr>
<tr>
<td>Unconstrained (Known Degrees of Freedom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Short Selling (Known Degrees of Freedom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.038</td>
<td>0.013</td>
</tr>
<tr>
<td>8</td>
<td>0.049</td>
<td>0.016</td>
</tr>
<tr>
<td>6</td>
<td>0.076</td>
<td>0.025</td>
</tr>
<tr>
<td>Unconstrained (Unknown Degrees of Freedom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Short Selling (Unknown Degrees of Freedom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.694</td>
<td>0.231</td>
</tr>
<tr>
<td>8</td>
<td>0.995</td>
<td>0.332</td>
</tr>
<tr>
<td>6</td>
<td>1.528</td>
<td>0.509</td>
</tr>
<tr>
<td>No Short Selling (Unknown Degrees of Freedom)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.034</td>
<td>0.011</td>
</tr>
<tr>
<td>8</td>
<td>0.049</td>
<td>0.016</td>
</tr>
<tr>
<td>6</td>
<td>0.074</td>
<td>0.025</td>
</tr>
</tbody>
</table>
Table VI

Alpha and Beta Estimation under Normality versus under $t$

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

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>$\alpha$ (%)</th>
<th>$\beta_{MKT}$</th>
<th>$\beta_{HML}$</th>
<th>$\beta_{SMB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal $t_8$</td>
<td>Normal $t_8$</td>
<td>Normal $t_8$</td>
<td>Normal $t_8$</td>
</tr>
<tr>
<td>S1B1</td>
<td>−0.450</td>
<td>1.061</td>
<td>−0.325</td>
<td>1.384</td>
</tr>
<tr>
<td></td>
<td>−0.004</td>
<td>0.969</td>
<td>0.074</td>
<td>1.322</td>
</tr>
<tr>
<td></td>
<td>0.043</td>
<td>0.926</td>
<td>0.292</td>
<td>1.118</td>
</tr>
<tr>
<td></td>
<td>0.209</td>
<td>0.901</td>
<td>0.460</td>
<td>1.037</td>
</tr>
<tr>
<td></td>
<td>0.150</td>
<td>0.975</td>
<td>0.675</td>
<td>1.089</td>
</tr>
<tr>
<td>S2B1</td>
<td>−0.190</td>
<td>1.116</td>
<td>−0.401</td>
<td>1.003</td>
</tr>
<tr>
<td>S2B2</td>
<td>−0.119</td>
<td>1.031</td>
<td>0.177</td>
<td>0.890</td>
</tr>
<tr>
<td></td>
<td>0.084</td>
<td>0.987</td>
<td>0.420</td>
<td>0.760</td>
</tr>
<tr>
<td></td>
<td>0.085</td>
<td>0.981</td>
<td>0.590</td>
<td>0.715</td>
</tr>
<tr>
<td></td>
<td>−0.031</td>
<td>1.086</td>
<td>0.774</td>
<td>0.840</td>
</tr>
<tr>
<td>S3B1</td>
<td>−0.071</td>
<td>1.084</td>
<td>−0.445</td>
<td>0.725</td>
</tr>
<tr>
<td>S3B2</td>
<td>0.007</td>
<td>1.060</td>
<td>0.221</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>−0.072</td>
<td>1.022</td>
<td>0.507</td>
<td>0.441</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>1.001</td>
<td>0.669</td>
<td>0.387</td>
</tr>
<tr>
<td>S3B5</td>
<td>0.015</td>
<td>1.109</td>
<td>0.826</td>
<td>0.528</td>
</tr>
<tr>
<td>S4B1</td>
<td>0.141</td>
<td>1.052</td>
<td>−0.448</td>
<td>0.379</td>
</tr>
<tr>
<td>S4B2</td>
<td>−0.173</td>
<td>1.100</td>
<td>0.263</td>
<td>0.207</td>
</tr>
<tr>
<td>S4B3</td>
<td>−0.065</td>
<td>1.082</td>
<td>0.513</td>
<td>0.164</td>
</tr>
<tr>
<td>S4B4</td>
<td>0.025</td>
<td>1.039</td>
<td>0.616</td>
<td>0.198</td>
</tr>
<tr>
<td>S4B5</td>
<td>−0.049</td>
<td>1.039</td>
<td>0.839</td>
<td>0.260</td>
</tr>
<tr>
<td>S5B1</td>
<td>0.210</td>
<td>0.961</td>
<td>−0.380</td>
<td>−0.256</td>
</tr>
<tr>
<td>S5B2</td>
<td>−0.024</td>
<td>1.039</td>
<td>0.145</td>
<td>−0.237</td>
</tr>
<tr>
<td>S5B3</td>
<td>−0.026</td>
<td>0.988</td>
<td>0.288</td>
<td>−0.232</td>
</tr>
<tr>
<td>S5B4</td>
<td>−0.088</td>
<td>1.007</td>
<td>0.642</td>
<td>−0.204</td>
</tr>
<tr>
<td>S5B5</td>
<td>−0.237</td>
<td>1.049</td>
<td>0.817</td>
<td>−0.111</td>
</tr>
</tbody>
</table>
Table VII
Multivariate Tests of the CAPM

The table reports both the Gibbons, Ross, and Shanken (1989) test and the likelihood ratio tests for the CAPM restrictions

\[ H_0 : \alpha = 0_N \]

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:

\[ r_{it} = \alpha_i + \beta_iMKT_t + \epsilon_t, \]

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

<table>
<thead>
<tr>
<th>Model</th>
<th>GRS</th>
<th>p-value (%)</th>
<th>LRT_n p-value (%)</th>
<th>LRT_t p-value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>July 1963 — December 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>3.985</td>
<td>0.00</td>
<td>92.27</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=10)</td>
<td>3.974</td>
<td>0.00</td>
<td>92.04</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=8)</td>
<td>3.969</td>
<td>0.00</td>
<td>91.93</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=6)</td>
<td>3.953</td>
<td>0.00</td>
<td>91.59</td>
<td>0.00</td>
</tr>
<tr>
<td>July 1963 — June 1983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1.434</td>
<td>9.03</td>
<td>34.93</td>
<td>8.96</td>
</tr>
<tr>
<td>t (df=10)</td>
<td>1.432</td>
<td>9.13</td>
<td>34.87</td>
<td>9.05</td>
</tr>
<tr>
<td>t (df=8)</td>
<td>1.430</td>
<td>9.18</td>
<td>34.85</td>
<td>9.10</td>
</tr>
<tr>
<td>t (df=6)</td>
<td>1.427</td>
<td>9.33</td>
<td>34.77</td>
<td>9.25</td>
</tr>
<tr>
<td>July 1983 — December 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>5.276</td>
<td>0.00</td>
<td>107.80</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=10)</td>
<td>5.254</td>
<td>0.00</td>
<td>107.45</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=8)</td>
<td>5.243</td>
<td>0.00</td>
<td>107.27</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=6)</td>
<td>5.211</td>
<td>0.00</td>
<td>106.75</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table VIII

Multivariate Tests of Fama-French Three-Factor Models

The table reports both the Gibbons, Ross, and Shanken (1989) test and the likelihood ratio tests for the standard factor pricing restrictions

\[ H_0 : \alpha = 0_N \]

in regressions of the returns of Fama-French 25 size and book-to-market ranked portfolios on their 3 factors:

\[ r_{it} = \alpha_i + \beta_{i1}MKT_t + \beta_{i2}HML_t + \beta_{i3}SMB_t + \epsilon_t, \]

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

<table>
<thead>
<tr>
<th>Model</th>
<th>GRS</th>
<th>p-value (%)</th>
<th>LRTn</th>
<th>p-value (%)</th>
<th>LRTt</th>
<th>p-value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>July 1963 — December 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>3.021</td>
<td>0.00</td>
<td>71.58</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t (df=10)</td>
<td>2.970</td>
<td>0.00</td>
<td>70.44</td>
<td>0.00</td>
<td>85.65</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=8)</td>
<td>2.945</td>
<td>0.00</td>
<td>69.89</td>
<td>0.00</td>
<td>86.59</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=6)</td>
<td>2.872</td>
<td>0.00</td>
<td>68.29</td>
<td>0.00</td>
<td>87.62</td>
<td>0.00</td>
</tr>
<tr>
<td>July 1963 — June 1983</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.913</td>
<td>58.67</td>
<td>22.86</td>
<td>58.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t (df=10)</td>
<td>0.896</td>
<td>61.04</td>
<td>22.46</td>
<td>60.92</td>
<td>26.17</td>
<td>39.87</td>
</tr>
<tr>
<td>t (df=8)</td>
<td>0.888</td>
<td>62.19</td>
<td>22.26</td>
<td>62.07</td>
<td>26.33</td>
<td>39.02</td>
</tr>
<tr>
<td>t (df=6)</td>
<td>0.864</td>
<td>65.50</td>
<td>21.69</td>
<td>65.38</td>
<td>26.51</td>
<td>38.08</td>
</tr>
<tr>
<td>July 1983 — December 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>4.638</td>
<td>0.00</td>
<td>97.12</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t (df=10)</td>
<td>4.545</td>
<td>0.00</td>
<td>95.54</td>
<td>0.00</td>
<td>109.89</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=8)</td>
<td>4.500</td>
<td>0.00</td>
<td>94.77</td>
<td>0.00</td>
<td>111.05</td>
<td>0.00</td>
</tr>
<tr>
<td>t (df=6)</td>
<td>4.370</td>
<td>0.00</td>
<td>92.54</td>
<td>0.00</td>
<td>112.40</td>
<td>0.00</td>
</tr>
</tbody>
</table>