Evaluation of Robust Regression Estimation Methods and Intercept Bias: A Capital Asset Pricing Model Application

James McDonald
Brigham Young University

Richard A. Michelfelder
Rutgers University

Panayiotis Theodossiou*
Rutgers University

July 2007

*Contact: Panayiotis Theodossiou, School of Business, Rutgers University, 227 Penn Street, Camden, NJ 08102, fax: +856 874 0014, tel: +856 225-6633 e-mail: theodoss@camden.rutgers.edu.

We would like to thank the participants at the American Statistical Association and Multinational Finance Society Annual Meetings for helpful comments on earlier versions of this paper. We would also like to thank the Whitcomb Center for Research in Financial Services for providing research support through the use of the WRDS system.

Evaluation of Robust Regression Estimation Methods and Intercept Bias: A Capital Asset Pricing Model Application

Abstract

Robust estimation techniques, such as LAD, M and L_k and quasi-maximum likelihood techniques based on symmetric probability density functions, such as student's T, Laplace, GED and GT, are often used instead of OLS to obtain more efficient regression parameters in thick-tail data. The empirical and theoretical results presented in this paper show that when skewness is present in the data, symmetric robust estimation techniques produce biased regression intercepts. The simulation results favor the T and GT regression estimators in leptokurtic symmetric data and the skewed T and skewed GT regression estimators in skewed data. In normal data, the OLS estimators are preferred because of their simplicity (JEL: G12, C13, C14, C15).

Keywords: CAPM; quasi maximum likelihood estimators; robust estimators; skewed generalized t.

Managerial Relevance Statement

The CAPM is an important tool for pricing systematic risk and measuring the performance of investment funds. The CAPM's slope (beta) is used in the computation of security risk premium and the intercept (Jensen's alpha) is used to measure of portfolio performance. This paper evaluates the ability of several robust methods in estimating the CAPM's parameters. Also, it presents theoretical and empirical findings showing that in the presence of skewness in the data (most of US stocks are positively skewed) standard robust estimation techniques produce biased estimates of alphas. The paper provides a suggestion as to which robust estimation techniques are the best for the estimation of CAPM's parameter. The findings are significant and important to portfolio managers, investors, regulators, litigators and in general all those interested in obtaining unbiased and efficient estimators of the CAPM's alpha and beta.

I. Introduction

It is a generally accepted finding that stock returns are usually non-normally distributed, and their empirical distributions often exhibit both kurtosis and skewness; e.g., Mandelbrot (1963), Fama (1965), Fama, Fisher, Jensen and Roll (1969), Fielitz and Smith (1972), Francis (1975) and McDonald and Nelson (1989). A major issue investigated in this paper is the effect of skewness in stock returns on the estimation of the capital asset pricing model (CAPM), a two-parameter linear model that involves regressing excess stock returns on excess stock market returns.

Estimating CAPM regressions with ordinary least squares (OLS) produces unbiased and efficient parameter estimates when regression errors are normally distributed. When the error distribution is thick-tailed and/or skewed, OLS produces inefficient estimates. The latter necessitates the use of robust, quasi-maximum likelihood, or partially adaptive estimation techniques.

Some robust estimation techniques commonly used include the least absolute deviation (LAD) estimator which minimizes the sum of the absolute value of the regression errors or, more generally, L_k estimators which minimize the sum of absolute values of the regression errors raised to the power k for fixed, but unrestricted value of k. Still more general robust estimators include the L_k estimators where the data endogenize the selection of the value of the parameter k or M-estimators which minimize a general function of the errors over the parameter values. LAD and L_k (with k predetermined or determined by the data) are both special cases of M-estimation. See Hampel (1974), Huber (1981), Koenker (1982), Koenker and Basset (1978), and Koenker and Basset (1982) for a more thorough discussion of these estimation techniques.

Another type of robust estimator involves the choice of alternatives to the normal density for regression estimation such as generalized T (GT), generalized error (GED), student's T, and Laplace, which is the LAD. There are many robust estimation methods and we do not attempt to develop an exhaustive discussion of all of them in this paper. They can be placed into two categories, one that is based on outlier-resistant methods in choosing regression parameter estimates, and the other the choice of probability density function (pdf) and parameters for specifying the likelihood function. These types may overlap. Boyer, McDonald, and Newey (2003) differentiate robust or outlier-resistant estimators into re-weighted least squares or least median squares, and partially adaptive estimators. Partially adaptive estimation procedures can be viewed as being quasi-maximum likelihood estimators because they maximize a log-likelihood function corresponding to an approximating error distribution over both regression and distributional parameters. Hinich and

Talwar (1975), Chan and Lakonishok (1992), Yohai and Zamar (1997), Martin and Simin (2003) also have developed outlier resistant methods for efficiency improvement.

The above robust techniques address efficiency due to kurtosis, but do not account for skewness. If skewness is present, the estimated CAPM intercept will be biased downwards in positively skewed data and upwards in negatively skewed data. The size of the bias is directly related to the extent of skewness and kurtosis. The CAPM intercept, also known as Jensen's alpha, developed in Jensen (1968), is frequently used as a measure of stock portfolio performance. A biased alpha can lead to erroneous decisions on stock valuation, portfolio selection, and mutual fund investment evaluation. Moreover, stocks with biased alphas can lead to biased and inefficient portfolios; see Frankfurter, Phillips, and Seagle (1974).

The above problem is important to address if the regression intercept has a meaningful interpretation or if the regression model is used in forecasting. A solution to the problem of both parameter inefficiency and intercept bias in regression is to use "flexible" probability distribution functions (pdf's), that is, those that accommodate both kurtosis and skewness. Such flexible pdf's include the skewed generalized T (SGT) of Theodossiou (1998), the skewed generalized error (SGED) of Theodossiou (2001), the inverse hyperbolic sine (IHS) of Johnson (1949) and the exponential generalized beta of the second kind (EGB2) of McDonald and Xu (1995). Hansen, McDonald, and Theodossiou (2001) include some additional discussion of these distributions and applications.

This paper provides theoretical, empirical, and simulation verifications of the intercept bias and shows how to address the bias problem. It also evaluates many well-known robust estimation methods using nine error distributions nested within the SGT that have various restrictions in accommodating thick-tails and skewness. The SGT nesting provides us with an integrated framework for making comparisons of the techniques such as LAD, trimmed regression quantile, L_k, and M estimators and reaching conclusions on the relative efficiency of the regression parameters. The paper employees the entire universe of publicly traded stocks in the U.S. which had at least four years of usable data over the period 1995 to 2004.

The next section of the paper discusses the CAPM estimation, the properties of the SGT, and the bias in the intercept. Section III reviews the empirical results. Section IV discusses the simulation results involving CAPM regressions where normal, thick-tailed, and skewed error pdf's are used to simulate regression errors. Section V concludes the paper.

II. SGT-CAPM Estimation

The estimation of CAPM's parameters is accomplished by fitting the following regression equation to each stock's returns data:

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \varepsilon_{i,t} \ , \quad \text{for } i = 1, 2, ..., N \text{ and } t = 1, 2, ..., T_i, \tag{1}$$

where $r_{i,t} = R_{i,t} – R_{f,t}$ and $r_{M,t} = R_{M,t} – R_{f,t}$ are excess returns from the risk free rate $R_{f,t}$ for stock i and the market, α_i and β_i are the alpha and beta for the stock, $\varepsilon_{i,t}$ is a regression error term for each individual stock's return generating process having zero mean and constant variance, T_i is the sample size for the stock, and N is the number of stocks. Note in the above CAPM specification the value for each stock's alpha implied by the theory is zero since $R_{f,t}$ is subtracted from both sides of the equations and otherwise $R_{f,t}$ is the theoretical intercept. As such, the above equation is often used to test the validity of the CAPM model for stocks and other assets as well as to assess the performance of stocks and mutual funds. A positive alpha would indicate that the stock or mutual fund had superior returns relative to risk and vice versa.

Estimates for the alpha and beta of each stock are obtained from the maximization of the sample log-likelihood function

$$\max_{\theta_i} l(\theta_i) = \sum_{t=1}^{T_i} \ln f\left(r_{i,t} \middle| r_{M,t}, \theta_i\right), \tag{2}$$

where f is the hypothesized probability density function for $r_{i,t}$ and $\theta_i = [\alpha_i, \beta_i, ...]$ is a parameter vector for the alpha, beta and other distributional parameters. As in other studies, we use the non-centered SGT log-likelihood specification rather than the centered specification because it does not require the existence of the first and second moments and it is easier to estimate. See Theodossiou (1998) for additional the estimation details on the description and moments of the SGT.

The non-centered SGT specification for stock's i returns is

$$f\left(r_{i,t} \mid r_{M,t}\right) = .5 k_i \left(\frac{n_i + 1}{k_i}\right)^{-\frac{1}{k_i}} B\left(\frac{n_i}{k_i}, \frac{1}{k_i}\right)^{-1} \varphi_i^{-1} \left(1 + \frac{\left|u_{i,t}\right|^{k_i}}{\left(\left(n_i + 1\right)/k_i\right)\left(1 + sign(u_{i,t})\lambda_i\right)^{k_i} \varphi_i^{k_i}}\right)^{-\frac{n_i+1}{k_i}} \tag{3}$$

and

$$u_{i,t} = r_{i,t} - \left(m_i + \beta_i r_{M,t}\right) \tag{4}$$

where $u_{i,t}$ is a deviation of $r_{i,t}$ from its conditional mode $m_i + \beta_i\, r_{M,t}$, φ_i is a scaling constant related to the standard deviation, when it exists, $B(\cdot)$ is the beta function, sign is the sign function taking the value of –1 for negative values of $u_{i,t}$ and 1for positive values of $u_{i,t}$, λ_i is a skewness parameter obeying the constraint $-1< \lambda_i < 1$, and k_i and n_i are positive kurtosis parameters.

The parameter k_i controls mainly the shape of the conditional density around the mode of $r_{i,t}$. Specifically, values of k_i below two ($k_i < 2$) result in density functions that are leptokurtic relative to the normal distribution (i.e., peaked around the mode) and values of k_i greater than two ($k_i > 2$) result in density functions that are platykurtic relative to the normal distribution. As k_i grows larger, the SGT density function approaches that of the uniform distribution. The parameter n_i controls mainly the tails of the density. As n_i gets smaller the tails of the SGT become fatter and as n_i gets infinitely large, the SGT approaches the SGED and for $k= 2$ the normal distribution. Note that standardized values for skewness and kurtosis (see equations (9) and (10) for the specifications of standardized skewness and kurtosis) in the ranges $(-\infty, \infty)$ and $(1.8, \infty)$ can be modeled with the SGT.

The non-centered SGT is defined for any value of $n_i > 0$ and can be used in the estimation of the parameters m_i and β_i regardless of the existence of the first and second moments of the distribution. Note that the moments of the SGT exist up to the value n_i; see McDonald and Newey (1988) and Theodossiou (1998).

When $n_i > 1$, the conditional expected value of $r_{i,t}$ is equal to

$$\begin{aligned} E\left(r_{i,t} \mid r_{M,t}\right) &= \alpha_i + \beta_i r_{M,t} + E\left(\varepsilon_{i,t}\right) = \alpha_i + \beta_i r_{M,t} \\ &= m_i + \beta_i r_{M,t} + E\left(u_{i,t}\right) = m_i + \beta_i r_{M,t} + \rho_i \varphi_i \end{aligned} \tag{5}$$

where

$$\rho_i = 2\lambda_i B\left(\frac{n_i}{k_i}, \frac{1}{k_i}\right)^{-1} \left(\frac{n_i+1}{k_i}\right)^{\frac{1}{k_i}} B\left(\frac{n_i-1}{k_i}, \frac{2}{k_i}\right). \tag{6}$$

Thus, the regression intercept of equation (1) is equal to

$$\alpha_i = m_i + \rho_i \varphi_i \,. \tag{7}$$

When $n_i > 2$, the conditional variance of returns exists and is equal to

$$\sigma_i^2 = \operatorname{var}\left(r_{i,t} \mid r_{M,t}\right) = \left(\gamma_i - \rho_i^2\right)\varphi_i^2 \,, \tag{8}$$

where

$$\gamma_i = \left(1+3\lambda_i^2\right) B\left(\frac{n_i}{k_i},\frac{1}{k_i}\right)^{-1}\left(\frac{n_i+1}{k_i}\right)^{\frac{2}{k_i}} B\left(\frac{n_i-2}{k_i},\frac{3}{k_i}\right).$$

Equations (7) and (8) are necessary to compute the intercept and variance of a regression model when a non-centered density likelihood specification is used. Equation (7) can be rewritten as

$$\alpha_i - m_i = \rho_i \varphi_i \ .$$

This equation provides the adjustment factor for the intercept when the non-centered SGT log-likelihood specification is used. Note that in the case of: a) negatively skewed SGT, $\lambda_i < 0$ and $\rho_i < 0$, b) symmetric SGT, $\lambda_i = 0$ and $\rho_i = 0$ and c) positively skewed SGT, $\lambda_i > 0$ and $\rho_i > 0$. The latter adjustment factor will be negative for negatively skewed returns and positive for positively skewed returns.

The skewness of $r_{i,t}$, for $n_i > 3$, is

$$SK_i = \frac{m_{3,i}}{\sigma_i^3} = \frac{A_{3,i} - 3\gamma_i\rho_i + 2\rho_i^3}{\left(\gamma_i - \rho_i^2\right)^{3/2}} \tag{9}$$

where $A_{3,i} = 4\lambda_i\left(1+\lambda_i^2\right) B\left(\frac{n_i}{k_i},\frac{1}{k_i}\right)^{-1}\left(\frac{n_i+1}{k_i}\right)^{\frac{3}{k_i}} B\left(\frac{n_i-3}{k_i},\frac{4}{k_i}\right)$.

The kurtosis of $r_{i,t}$, for $n_i > 4$, is

$$KU_i = \frac{m_{4,i}}{\sigma_i^4} = \frac{A_{4,i} - 4A_{3,i}\rho_i + 6\gamma_i\rho_i^2 - 3\rho_i^4}{\left(\gamma_i - \rho_i^2\right)^2} \tag{10}$$

where $A_{4,i} = \left(1+10\lambda_i^2+5\lambda_i^4\right) B\left(\frac{n_i}{k_i},\frac{1}{k_i}\right)^{-1}\left(\frac{n_i+1}{k_i}\right)^{\frac{4}{k_i}} B\left(\frac{n_i-4}{k_i},\frac{5}{k_i}\right)$; see the appendix A for the derivations of the moments, skewness and kurtosis equations.

The SGT nests several well known pdf's as special cases, such as the generalized t (GT), skewed t (ST), student's t, Cauchy, skewed generalized error distribution (SGED), generalized error distribution (GED) and Laplace; Moreover, SGT (with restrictions) log-likelihood specification yield the L_k estimator, MAD (or LAD) estimator, and trimmed regression quantile estimator as special cases; see the Appendix B for more details.

III. Sampling and Estimation

We considered the population of all 11,001 common stocks in the University of Chicago's Center for Research in Security Prices (CRSP) database that were publicly traded between January 1, 1995 and December 31, 2004 on the NYSE, the AMEX, and the NASDAQ. Any stock was removed that did not have at least four years of data (1,000 trading-day returns) during the 10 year period. The time series observations of rates of return for individual stocks range between 1,000 and 2,519 observations. We avoided survival bias by not removing stocks that were de-listed or stopped trading due to liquidation from bankruptcy, dropped from trading on the exchange, merged, or exchanged for other stock. The resulting universe of 6,502 stocks were used in the analysis.

Stock returns ($R_{i,t}$) are daily holding period rates of return for each stock obtained from the CRSP database. The risk-free rate of return ($R_{f,t}$) is the daily return on the one-month US Treasury Bill that compounds to the monthly return for a specific month. The stock market return ($R_{m,t}$) from the CRSP database is the value-weighted daily return on all of the stocks in the CRSP database. The daily excess stock market return {$r_{m,t}$ or ($R_{m,t} – R_{f,t}$)}, the independent variable in the regressions, is from the Fama and French files of the CRSP database.

All 6,502 stocks were used to compute regression estimates of alphas and betas using OLS, LAD, SLAD, GED, SGED, student's T, ST, GT, and SGT. We used the OLS estimates to initially analyze the pdf characteristics of the residuals as OLS is commonly used by practitioners and many researchers of the CAPM.

A bivariate relative frequency table for the OLS regressions residuals skewness is presented on table 1. The bottom row shows that 88% (5,728 stocks) of all regression residuals have standardized kurtosis greater than 4. Specifically, about 32% (2,078 stocks) have kurtosis between 4 and 8, 17% (1,128 stocks) between 8 and 12 and 39% (2,522 stocks) greater than 12. These results strongly support the hypothesis that the distribution of CAPM residuals are leptokurtic relative to the normal distribution. The last column of the table presents the results for the standardized skewness. According to the table, 88% (5,694 stocks) of the CAPM's skewness values are outside the –0.2 to 0.2 range, which roughly constitutes the confidence interval at the 5% level of significance (see table 1 notes for standard errors). Of these, about 9% (600 stocks) are less than –0.2 and 79% (5,094 stocks) are greater than 0.2, implying that the overwhelming majority of CAPM residuals are significantly positively skewed.

Interestingly the bivariate results of table 1 show that only about 3% (190 stocks) are approximately normally distributed as they exhibit skewness in the range of –0.2 to 0.2 and kurtosis in the range of 0 to 4. Thus, about 97% (6,312 stocks) of CAPM residuals exhibit skewness and / or kurtosis. The results depict a positive relation between absolute skewness and kurtosis. In conclusion, the results of table 1 establish that robust estimation methods are required to obtain more efficient estimates of the CAPM's parameters as the majority of the CAPM regression residuals have both skewed and thick-tailed distributions.

We developed a frequency table of the Jarque-Bera (JB) statistic for the regressions residuals. The JB statistic tests the null hypothesis that the pdf of the regression residuals are normal. It performs a joint test for skewness and excess kurtosis and is χ^2 distributed with two degrees of freedom. The results reject the null hypothesis of normality for all of the stocks' regression residuals. The JB table is available upon request.

Table 2 presents the frequency distribution of the betas estimated with the SGT. It shows that about 80% of the betas range from 0 to 1.5. This is a reasonable range for the majority of beta levels. The associated estimated intercepts, although not shown but discussed below, are adjusted for the bias due to the structure of the SGT.

The first column of table 3 presents the frequency $N(\lambda_i)$ and relative frequency $P(\lambda_i)$ of the skewness parameter λ_i, for all 6,502 CAPM regressions, estimated using the SGT likelihood specification. Observe that about 21% (1,391 stocks) of the λ_i's are negative and about 79% (5,111 stocks) are positive. The last two columns give the number and fraction of λ's in each class interval that are statistically significant at the 5% and 1% levels, respectively. The t-values for the estimated λ_i are based on robust standard errors. Notice that as we move away from zero, the fraction of statistically significant λ's in each class interval increases. The bottom row shows that about 50% (3,242 stocks) of the CAPM residuals exhibit significant positive or significant negative skewness. Of these, 21% (680 stocks) of the regression residuals exhibit negative skewness and 79% (2,562 stocks) exhibit positive skewness. For the overall sample, the percentage of stocks with significant negative and significant positive skewness in CAPM residuals are respectively 10% (= 680/6,502) and 39% (= 2,562/6,502).

We developed frequency tables of the SGT kurtosis parameters, n_i and k_i (not presented for brevity and available upon request). The n_i parameter determines the thickness of the pdf's tails. Lower (higher) values of n_i reflects thicker (thinner) tails for the SGT. About 51% (3,316 stocks) of the n_i's range between 2 and 10 and are substantially lower than the normal pdf value of $n_i = 30$. Less than 1% (52 stocks) of the estimated n_i's are less than one. Therefore the majority of the residuals have thick tailed pdf's, thereby driving the need for robust-efficient estimators.

The k_i parameter determines the degree of leptokurtosis (platykurtosis). The value of k_i for OLS and the normal pdf is 2. Values of k_i less than (greater than) 2 reflects leptokurtic (platykurtic) pdf's. We find that k_i is less than 1.75 for over 85% (5,527 stocks) of all residuals. The mean of k_i is 1.12 for all 6,502 SGT regressions. This is further evidence of mainly thick-tailed and peaked residual pdf's.

Note that to obtain the correct α_i value for the regression intercept, the quantity $\rho_i\varphi_i$ has to be added to the estimated mode intercept m_i, i.e.,$\alpha_i = m_i + \rho_i\varphi_i$; see equation 7. Note, however, that the latter equation is only defined for values of $n_i > 1$. The first column of table 4 presents the frequency $N(\rho_i\varphi_i)$ and relative frequency $P(\rho_i\varphi_i)$ of the intercept adjustment factor $\rho_i\varphi_i$, due to skewness, for 6,450 SGT- estimated CAPM regressions; i.e., 52 regressions yielded estimated values for $n_i < 1$. The last two columns give the number and fraction of $\rho_i\varphi_i$'s in each class interval that are statistically significant at the 5% and 1% levels, respectively. Notice that the results are quite analogous to those of table 3. This is a byproduct of the fact that the adjustment factor is driven mainly by the skewness parameter λ_i. Interestingly, the adjustment factor is in many instances greater than .5 or 50 basis points.

The estimated CAPM intercept bias due to skewness $b(\alpha_i)$ is computed as the difference between intercepts estimated using the SGT and GT (symmetric) likelihood specification, i.e., $b(\alpha_i) = \alpha_{SGT,i} - \alpha_{GT,i}$. Similarly, table 5 presents the relative frequency and significance results for the intercept bias $b(\alpha_i)$. The results show a significant difference in the intercepts of the skewed and symmetric likelihood specifications. In 67% (4,334 stocks) of the cases, the bias is statistically significant at the 5% level and 51% (3,324 stocks) are significant at the 1% level. Recall that in table 3, 50% of the regressions' residuals distributions had significant skewness. The fraction of stocks that have significant bias with negative bias is 23% (999 stocks) and significant positive bias is 77% (3,335 stocks).

These results provide strong empirical support that when skewness is present in the data, the use of symmetric log-likelihood specifications or "symmetric type" robust estimation techniques will result in biased regression intercepts. This issue along with the issue of efficiency of various estimators is further investigated in the next section using simulations.

IV. Simulations and Estimator Performance

We use Monte Carlo simulations to assess the effects of tail thickness and/or skewness of the error distribution on the relative efficiency of the various regression estimators of CAPM's intercept. In addition to OLS, LAD or Laplace, GED, student's T and GT -based estimators, used in prior studies, we consider the skewed specifications of Laplace (SLAD), GED, student T and GT estimators. These results extend those reported in Manski (1984) and McDonald and White (1993).

Specifically for the simulations, we use the CAPM model

$$r_t = \alpha + \beta r_{M,t} + \varepsilon_t \ ,$$

where $r_{M,t}$ is the excess market return for the entire sampling period (i.e., 2,519 observations), and ε_t and r_t are a randomly generated error terms and stock returns. A value of zero for the alpha and one for the beta (i.e., α = 0 and β = 1), are used in all random samples.

Following McDonald and White (1993), the regression errors are generated using the (1) normal, (2) mixed-normal (thick-tailed variance-contaminated), and (3) skewed log-normal distributions. Specifically, the normal error term is generated by $\varepsilon_1 = \sigma z$, where $z \sim N(0,1)$. The thick-tailed variance contaminated error distribution is generated by $\varepsilon_2 = \sigma\,[w\,z_1 + (1-w)\,z_2]$, where $z_1 \sim N(0,1/9)$, $z_2 \sim N(0,9)$, and w= 1 with probability 0.9 and w = 0 with a probability of 0.1. This distribution is symmetric and has a standardized kurtosis of 24.33. The log-normal distribution is generated by $\varepsilon_3 = \sigma\,(e^{0.5\,z} - e^{0.125})/(e^{0.5} - e^{0.25})^{0.5}$, where $z \sim N(0,1)$. This distribution has standardized skewness 1.75 and standardized kurtosis of 8.898. The standard deviation σ is computed using the equation $\sigma = [(1/R^2)-1]^{0.5}\ |\beta|\ \sigma_M$, for R^2= 0.0879 (average CAPM-R-square for all stocks in the sample), β =1 and σ_M = 1.1195 (standard deviation of $r_{M,t}$ in the sample).

One thousand and fifty replications of samples are generated with the same nine error distributions to estimate the alpha and beta parameters. Table 6 presents the means of the regression alphas for the normal, normal mixture, and log-normal samples and associated t-statistics in parentheses. Note that for the case of the normal and normal mixture random data the hypothesis of unbiased estimates of the regression intercept cannot be rejected at traditional levels of statistical significance for any of the estimators. In the case of log-normal (skewed) random data all of the symmetric models except OLS, i.e., the LAD, GED, student's T and GT provide biased estimates of the regression intercept. On the other hand, the skewed models provide unbiased estimates of the regression intercepts. In the case of the slope, however, all models, symmetric and non-symmetric, provide unbiased estimates of the regression slopes.

Table 7 provides root mean squared errors (RMSE) of the nine estimation procedures in each of the three simulation samples for the intercept and slope estimators. The RMSE, computed as the square root of the sum of the sample variance of each estimator and the square of its sample bias, measures how close the estimator is to the true parameter.

Panel A of table 7, presents the results for the intercept estimator. In the case of the normal random sample, all models except LAD exhibit similar RMSE performance, thus there appears to be little efficiency loss for the intercept estimators relative to the OLS estimator. For the mixed normal random sample, the student's T and GT estimators are the best. The latter estimators are slightly better than the LAD and GED estimators. The remaining estimators are clearly inferior. For the log-normal random sample, OLS and all skewed estimators, including the ST and SGT, exhibit similar performance.

Panel B of table 7, presents the results for the slope estimator. In the case of the normal random sample, all slope estimators, except for those of LAD and SLAD, exhibit similar RMSE performance. For the mixed-normal sample, the student's T, ST, GT and SGT slope estimators are the best. Their RMSE values are about 86% of those of GED and SGED, 71% of those of LAD and SLAD and 40% of that of OLS. Finally, in the case of the log-normal sample, the ST and SGT appear to be the best slope estimators, followed closely by the SGED estimator. The RMSE values of ST and GT are about 80% of those of SLAD and GT, 75% of that of student's T and about 63% of those of OLS, LAD and GED.

In sum, the simulation results for the intercept and slope estimators show that a) all models, except the LAD and SLAD, exhibit similar performances in the normal random sample, b) the student's T and GT are best estimators in the mixed-normal sample and c) the ST and SGT are the best estimators in the log-normal sample. Overall, the results favor the T and GT estimators in leptokurtic symmetric data and ST and SGT estimators in skewed data. In normal samples, the OLS estimators are preferred because of their simplicity.

V. Summary and Concluding Remarks

This paper introduces a general class of quasi-maximum likelihood regression estimators based on the skewed generalized t (SGT) distribution. This class of estimators includes the OLS, the least absolute deviation (LAD), the L_k, the trimmed regression quantile estimators, M-estimators, and the quasi-maximum likelihood estimators of symmetric and skewed student's T, Laplace, GED and GT probability distributions. As such, the SGT distribution provides a unified framework to investigate the impact of skewness on the estimated regression parameters of the various estimators and compare their relative efficiency in diverse types of data.

The importance and relevance of the various robust estimation techniques and impact of skewness on the estimated regression parameters is demonstrated using the capital asset pricing model (CAPM), which involves regressing individual stock excess returns on market excess returns, and the entire universe of all publicly traded U.S. stocks with at least four years of data. A preliminary analysis of CAPM's regression residuals depict that about 97% of the stocks exhibit significant skewness and/or excess kurtosis, 79% of them exhibit significant positive skewness and 9% of them exhibit significant negative skewness. These results provide overwhelming support for the use of robust type estimation techniques for CAPM's estimation.

Further empirical and theoretical analysis shows that when skewness is present in the data, quasi-maximum likelihood estimation techniques based on symmetric probability distributions produce biased estimates for the regression intercepts. The latter bias is is negative in negatively skewed data and positive in positively skewed data. The latter has significant implications for finance, since the CAPM intercept, known as the Jensen's alpha is frequently used in portfolio selection and stock and mutual fund valuation.

Simulation results using a normal, mixed-normal (fat tails) and log-normal (positively skewed) random samples show that a) all models, except the LAD and skewed LAD, exhibit similar performances in the normal random sample, b) the student's T and GT are the best estimators in the mixed-normal sample and c) the skewed student's T (ST) and SGT are the best estimators in the log-normal sample. Overall, the results favor the student's T and GT estimators in leptokurtic symmetric data and the ST and SGT estimators in skewed data. In normal samples, the OLS estimators is preferred because of their simplicity.

The above findings are relevant and important to researcher in other fields interested in unbiased and efficient regression estimators. That is, estimated regression models with biased intercepts will produce biased forecasts or predictions. Examples of skewed and/or leptokurtic data from other fields include a) building electricity usage data, e.g., Parti and Parti (1980), Hartman (1983), Bartels and Fiebig (1990); b) economic housing price data, e.g., Hansen, McDonald, and Turley (2006); c) meteorological solar radiation predictions and wind shear analysis data, e.g., Younes and Muneer (2006), Kanji (1985), Jones and McLachan (1990); and d) aeronautical flight navigation risk analysis, e.g., Hsu (1979).

Appendix A- Adjustment Factor and SGT Moments

The *s*th non-centered moment of $u = r - (m + \beta r_M)$ for integer values of $s < n$ is

$$M_s = C(-1)^s \int_0^\infty u^s \left(1 + \frac{|u|^k}{((n+1)/k)(1-\lambda)^k \varphi^k}\right)^{-\frac{n+1}{k}} du + C\int_0^\infty u^s \left(1 + \frac{|u|^k}{((n+1)/k)(1+\lambda)^k \varphi^k}\right)^{-\frac{n+1}{k}} du \ .$$

Gradshteyn and Ryzhik (1994, p. 341) show that

$$\int_0^\infty u^s \left(1 + qu^k\right)^{-d} du = k^{-1} q^{-(s+1)/k} B\left(\frac{dk-(s+1)}{k}, \frac{s+1}{k}\right)$$

where $0 < (s+1)/k < d$, $q \neq 0$ and $n \neq 0$.

Letting $q = ((n+1)/k)(1-\lambda)^{-k} \varphi^{-k}$ or $q = ((n+1)/k)(1+\lambda)^{-k} \varphi^{-k}$ and $d = (n+1)/k$ and substituting into the M_s equation gives

$$M_s = \left[(-1)^s (1-\lambda)^{s+1} + (1+\lambda)^{s+1}\right] C k^{-1} \left(\frac{n+1}{k}\right)^{\frac{s+1}{k}} B\left(\frac{n-s}{k}, \frac{s+1}{k}\right) \varphi^{s+1} .$$

For *f*(.) to be proper probability density function,

$$M_0 = 2Ck^{-1} \left(\frac{n+1}{k}\right)^{\frac{1}{k}} B\left(\frac{n}{k}, \frac{1}{k}\right) \varphi = 1 ,$$

thus

$$C = .5k \left(\frac{n+1}{k}\right)^{-\frac{1}{k}} B\left(\frac{n}{k}, \frac{1}{k}\right)^{-1} \varphi^{-1} .$$

Substitution of C into the M_s equation gives,

$$M_s = 0.5\left[(-1)^s (1-\lambda)^{s+1} + (1+\lambda)^{s+1}\right] B\left(\frac{n}{k}, \frac{1}{k}\right)^{-1} \left(\frac{n+1}{k}\right)^{\frac{s}{k}} B\left(\frac{n-s}{k}, \frac{s+1}{k}\right) \varphi^s .$$

The expected value of *u*, provided that $n > 1$, is

$$E(u) = M_1 = 2\lambda B\left(\frac{n}{k}, \frac{1}{k}\right)^{-1} \left(\frac{n+1}{k}\right)^{\frac{1}{k}} B\left(\frac{n-1}{k}, \frac{2}{k}\right) \varphi = \rho\varphi ,$$

where $\rho = 2\lambda B\left(\frac{n}{k}, \frac{1}{k}\right)^{-1} \left(\frac{n+1}{k}\right)^{\frac{1}{k}} B\left(\frac{n-1}{k}, \frac{2}{k}\right)$.

The second non-centered moment of *u*, provided that $n > 2$, is

$$E\left(u^2\right) = M_2 = 0.5\left[(-1)^2(1-\lambda)^3 + (1+\lambda)^3\right] B\left(\frac{n}{k},\frac{1}{k}\right)^{-1}\left(\frac{n+1}{k}\right)^{\frac{2}{k}} B\left(\frac{n-2}{k},\frac{3}{k}\right)\varphi^2$$

$$= \left(1+3\lambda^2\right) B\left(\frac{n}{k},\frac{1}{k}\right)^{-1}\left(\frac{n+1}{k}\right)^{\frac{2}{k}} B\left(\frac{n-2}{k},\frac{3}{k}\right)\varphi^2 = \gamma\varphi^2 .$$

In this case the variance of *u* is

$$\sigma^2 = E\left(u^2\right) - E\left(u\right)^2 = \left(\gamma - \rho^2\right)\varphi^2 .$$

where $\gamma - \tau^2 > 0$ (see below). In the above expression the variance, expressed in terms of φ, exists for as long as $n > 2$, although the value of φ exists for any value of $n > 0$. Note that

$$\gamma - \tau^2 = \left(1+3\lambda^2\right) B\left(\frac{n}{k},\frac{1}{k}\right)^{-1} B\left(\frac{n-2}{k},\frac{3}{k}\right) - 4\lambda^2 B\left(\frac{n}{k},\frac{1}{k}\right)^{-2} B\left(\frac{n-1}{k},\frac{2}{k}\right)^2$$

$$= \left[1+3\lambda^2 - 4\lambda^2 B\left(\frac{n}{k},\frac{1}{k}\right)^{-1} B\left(\frac{n-2}{k},\frac{3}{k}\right)^{-1} B\left(\frac{n-1}{k},\frac{2}{k}\right)^2\right] B\left(\frac{n}{k},\frac{1}{k}\right)^{-1} B\left(\frac{n-2}{k},\frac{3}{k}\right)$$

$$= S\left(\lambda\right) B\left(\frac{n}{k},\frac{1}{k}\right)^{-1} B\left(\frac{n-2}{k},\frac{3}{k}\right) > 0 ,$$

because $S(\lambda) > 0$ (the latter can be easily proven using the Stirling's approximation of the gamma function).

The third non-centered moment of *u*, provided that $n > 3$, is

$$E\left(u^3\right) = M_3 = 0.5\left[(-1)^3(1-\lambda)^4 + (1+\lambda)^4\right] B\left(\frac{n}{k},\frac{1}{k}\right)^{-1}\left(\frac{n+1}{k}\right)^{\frac{3}{k}} B\left(\frac{n-3}{k},\frac{4}{k}\right)\varphi^3$$

$$= 4\lambda\left(1+\lambda^2\right) B\left(\frac{n}{k},\frac{1}{k}\right)^{-1}\left(\frac{n+1}{k}\right)^{\frac{3}{k}} B\left(\frac{n-3}{k},\frac{4}{k}\right)\varphi^3 = A_3\varphi^3$$

The third center moment is,

$$m_3 = E\left(u - M_1\right)^3 = Eu^3 - 3M_1Eu^2 + 3M_1^2Eu - M_1^3$$

$$= A_3\varphi^3 - 3\gamma\rho\varphi^3 + 2\rho^3\varphi^3 = \left(A_3 - 3\gamma\rho + 2\rho^3\right)\varphi^3 = A_3\varphi^3$$

The fourth non-centered moment of *u*, provided that $n > 4$, is

$$E\left(u^4\right) = M_3 = 0.5\left[\left(-1\right)^4\left(1-\lambda\right)^5 + \left(1+\lambda\right)^5\right] B\left(\frac{n}{k},\frac{1}{k}\right)^{-1}\left(\frac{n+1}{k}\right)^{\frac{4}{k}} B\left(\frac{n-4}{k},\frac{5}{k}\right)\phi^4$$

$$= \left(1+10\lambda^2+5\lambda^4\right) B\left(\frac{n}{k},\frac{1}{k}\right)^{-1}\left(\frac{n+1}{k}\right)^{\frac{4}{k}} B\left(\frac{n-4}{k},\frac{5}{k}\right)\varphi^4 = A_4\varphi^4 .$$

The fourth centered moment of u is

$$m_4 = E\left(u-M_1\right)^4 = Eu^4 - 4M_1Eu^3 + 6M_1^2Eu^2 - 4M_1^3Eu + M_1^4$$

$$= \left(A_4 - 4A_3\rho + 6\gamma\rho^2 - 3\rho^4\right)\varphi^4 .$$

The skewness and kurtosis measures are

$$SK = \frac{m_3}{\sigma^3} = \frac{A_3 - 3\gamma\rho + 2\rho^3}{\left(\gamma-\rho^2\right)^{3/2}}$$

and

$$KU = \frac{m_3}{\sigma^4} = \frac{A_4 - 4A_3\rho + 6\gamma\rho^2 - 3\rho^4}{\left(\gamma-\rho^2\right)^2}$$

Appendix B - Popular Distributions Nested by the SGT

The skewed generalized t (SGT) distribution, developed by Theodossiou (1998),

$$f = .5k\left(\frac{n+1}{k}\right)^{-\frac{1}{k}} B\left(\frac{n}{k},\frac{1}{k}\right)^{-1} \varphi^{-1}\left(1+\frac{|u|^k}{\left((n+1)/k\right)\left(1+sign(u)\lambda\right)^k \varphi^k}\right)^{-\frac{n+1}{k}}$$

“nests” several well known distributions.

For $\lambda = 0$ it gives the generalized t (GT) of McDonald and Newey (1988)

$$f = .5k\left(\frac{n+1}{k}\right)^{-\frac{1}{k}} B\left(\frac{n}{k},\frac{1}{k}\right)^{-1} \varphi^{-1}\left(1+\frac{|u|^k}{\left((n+1)/k\right)\varphi^k}\right)^{-\frac{n+1}{k}}.$$

McDonald (1989), McDonald and Newey (1988) and McDonald and Nelson (1989) used the GT to develop partially adaptive estimation of regression models. Butler et al. (1990) discussed the robust estimation of CAPM using the GT.

For $k = 2$ it gives the Hansen’s (1994) skewed t (ST)

$$f = \left(\frac{n+1}{2}\right)^{-\frac{1}{2}} B\left(\frac{n}{2},\frac{1}{2}\right)^{-1} \phi^{-1}\left(1+\frac{|u|^2}{\left((n+1)/2\right)\left(1+sign(u)\lambda\right)^2 \varphi^2}\right)^{-\frac{n+1}{2}},$$

used in autoregressive conditional density estimation.

For $k = 2$ and $\lambda = 0$ it gives the student’s t distribution,

$$f = \left(\frac{n+1}{2}\right)^{-\frac{1}{2}} B\left(\frac{n}{2},\frac{1}{2}\right)^{-1} \phi^{-1}\left(1+\frac{|u|^2}{\left((n+1)/2\right)\varphi^2}\right)^{-\frac{n+1}{2}},$$

often used in log-likelihood specifications of data characterized by excess kurtosis, e.g., Bollerslev (1987).

For $n = 1$ and $\lambda = 0$ it gives the Cauchy distribution

$$f = (\pi\phi)^{-1}\left(1+\frac{|u|^2}{\varphi^2}\right)^{-1}.$$

For $n = \infty$ it gives the skewed generalized error distribution (SGED) of Theodossiou (2001)

$$f = 0.5k\Gamma\left(\frac{1}{k}\right)^{-1}\varphi^{-1}\exp\left(-\frac{|u|^k}{(1+sign(u)\lambda)^k\varphi^k}\right),$$

and for $\lambda = 0$ the generalized error distribution (GED)

$$f = 0.5k\Gamma\left(\frac{1}{k}\right)^{-1}\varphi^{-1}\exp\left(-\frac{|u|^k}{\varphi^k}\right).$$

The GED, introduced by Subbotin (1923), was used by Box and Tiao (1962) to model prior densities in Bayesian estimation, by Zeckhauser and Thompson (1970), Nelson (1991) and many others to model the distribution of financial return data.

For $k = 2$ and $\varphi = \sqrt{2}\sigma$ and $\lambda = 0$ the SGED gives the normal distribution.

$$f = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{|u|^2}{2\sigma^2}\right)$$

For $k = 1$, it gives the skewed Laplace distribution

$$f = 0.5\varphi^{-1}\exp\left(-\frac{|u|}{(1+sign(u)\lambda)\varphi}\right)$$

and for $k = 1$ and $\lambda = 0$ the Laplace distribution

$$f = 0.5\varphi^{-1}\exp\left(-\frac{|u|}{\varphi}\right).$$

The Laplace distribution has found some very interesting applications. For example, Hsu (1979) used the Laplace to model the distribution of position errors in navigation, Kanji (1985) and Jones and McLachan (1990) to model the distribution of wind shear data and Bagchi, Hayya and Ord (1983) to model demand during lead and slow times.

Interestingly, maximum likelihood estimation using the GED, Laplace and skewed Laplace specifications yield some very well known estimators often used in regression estimation. Specifically, the GED log-likelihood specification, for a fix value of k, yields the L_k estimator

$$(\alpha,\beta)_{GED} = \arg\left(\min_{\alpha,\beta}\sum_{t=1}^{T}|u_t|^k\right);$$

Note that for $k = 2$, the above gives the OLS estimator.

The Laplace log-likelihood specification yields the Lad or MAD estimator

$$(\alpha,\beta)_{Laplace} = \arg\left(\min_{\alpha,\beta}\sum_{t=1}^{T}|u_t|\right),$$

the skewed Laplace (SL) log-likelihood specification yields the (SLAD) the trimmed regression quantile (TRQ) estimator of Koenker and Bassett (1978); see also Chan and Lakonishok (1992),

$$(\alpha,\beta)_{SLAD} = \arg\left(\min_{\alpha,\beta}\sum_{t=1}^{T}\rho(u_t)\right)$$

where $\rho(u_t) = u_t/(1-\lambda)$ for $u_t \leq 0$ and $\rho(u_t) = u_t/(1+\lambda)$ for $u_t > 0$, where $-1 < \lambda < 1$; note in the TRQ literature $1/(1-\lambda) = \theta$ and $1/(1+\lambda) = 1-\theta$, with $0 < \theta < 1$.

References

Bagchi, U., J.C. Hayya, and J.K. Ord, 1983, "The Hermite Distribution as a Model of Demand During Lead-Time for Slow Moving Items," *Decision Sciences*, 14, 447-466.

Bartels, R., and D.J. Fiebig, 1990, "Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Load," *The Energy Journal*, 11, 79-97.

Bollerslev, T., 1987, "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.

Box, G. E. P., and G.C. Tiao, 1962, "A Further Look at Robustness Via Bayes Theorem." *Biometrika,* 49, 419-432.

Boyer, B. H.., J. B. McDonald, and W. K. Newey, 2003, "A Comparison of Partially Adaptive and Reweighted Least Squares Estimation," *Econometric Reviews*, 22, 115-134.

Butler, R. J., J. B. McDonald, R. D. Nelson, and S. White, 1990, "Robust and Partially Adaptive Estimation of Regression Models." *Review of Economics and Statistics*, 72, 321-327.

Chan, L. K. C., and J. Lakonishok, 1992, "Robust Measurement of Beta." *Journal of Financial and Quantitative Analysis*, 27, 265-282.

Fama, E., 1965, "The Behavior of Stock Market Movements." *Journal of Business*, 38, 1749-1778.

Fama, E., L. Fisher, M. Jensen, and R. Roll, 1969, "The Adjustment of Stock Prices to New Information." *International Economic Review*, 10, 1-21.

Fielitz, B.D., and E. W. Smith, 1972, "Asymmetric Distributions of Stock Price Changes." *Journal of the American Statistical Association,* 67, 813-814.

Francis, J.C., 1975, "Skewness and Investor's Decisions." *Journal of Financial And Quantitative Analysis,* 10, 163-172.

Frankfurter, G.M., H.E. Phillips, J.P. Seagle, 1974, "Bias in Estimating Portfolio Alphas and Beta Scores." *Review of Economics and Statistics*, 56, 412-414.

Gradshteyn, I.S., and I.M. Ryzhik, 1994, *Tables of Integrals, Series, and Products,* 5^{th} ed.,, edited by Alan Jeffrey, New York: Academic Press.

Hampel, F.R., 1974, "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Society*, 69, 383-393.

Hansen, B. E., 1994, "Autoregressive Conditional Density Estimation." *International Economic Review*, 35, 705-730.

Hansen, J. V., J.B. McDonald, and R.S. Turley, 2006, "Partially Adaptive Robust Estimation of Regression Models and Applications," *European Journal of Operations Research*, 170, 132-143

Hansen, C. B., J.B. McDonald, and P. Theodossiou, 2001, "Some Flexible Parametric Models for Skewed and Leptokurtic Data," Brigham Young University.

Hartman, R.S., 1983, "The Estimation of Short-Run Household Electricity Demand Using Pooled Aggregate Data," *Journal of Business and Economic Statistics*, 1, 127-135.

Hinich, Melvin T., and Prem P. Talwar., 1975, "A Simple Method for Robust Regression." *Journal of the American Statistical Association*, 70, 113-119.

Hsu, D.A., 1979, "Long-Tailed Distributions for Position Errors in Navigation," *Applied Statistics*, 28, 62-72.

Huber, P.J., 1981, *Robust Statistics,* New York: Wiley.

Jensen, Michael C., 1968, "The Performance of Mutual Funds in the Period 1945-1964." *Journal of Finance*, 23, 389-416.

Johnson, N. L., 1949, "A Systems of Frequency Curves Generated by Methods of Translation," *Biometrica*, 36, 149-176.

Jones, P.N., and McLachlan, G.L., 1990, "Laplace-Normal Mixtures Fitted to Wind Shear Data," *Journal of Applied Statistics*, 17, 283-291.

Kanji, G.K., 1985, "A Mixture Model for Wind Shear Data," *Journal of Applied Statistics*, 12, 49-58.

Koenker, R., 1982, "Robust Methods in Econometrics," *Econometric Reviews*, 1, 213-255.

Koenker, R., and G. Bassett, 1978, "Asymptotic Theory of Least Absolute Regression Error." *Journal of the*

American Statistical Association, 73, 618-622.
Koenker, R., and G. Bassett, 1982, "Robust Methods in Econometrics." (with discussion), *Econometric Reviews,* 1, 213-255.
Mandelbrot, B., 1963, "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.
Manski, C. F., 1984, "Adaptive Estimation of Non-Linear Regression Models." *Econometric Reviews,* 3, 145-194.
Martin, R. G., and T. T. Simin, 2003, "Outlier-Resistant Estimates of Beta ," *Financial Analysts Journal*, 59, 56-69.
McDonald, J.B., 1989, "Partially Adaptive Estimation of ARMA Models," *International Journal of Forecasting*, 5, 217-230.
McDonald, J. B. and R. T. Nelson, 1989, "Alternative Beta Estimation for the Market Model using Partially Adaptive Estimation." *Communications in Statistics: Theory and Methods*, 18, 4039-4058.
McDonald, J. B., and W. K. Newey, 1988, "Partially Adaptive Estimation of Regression Models Via the Generalized T Distribution." *Econometric Theory* , 4, 428-457.
McDonald, J. B. and S. B. White,1993, "A Comparison of Some Robust, Adaptive, and Partially Adaptive Estimators of Regression Models." *Econometric Reviews*, 12, 103-124.
McDonald, J. B. and Y. J. Xu, 1995, "A Generalization of the Beta Distribution with Applications." *Journal of Econometrics,* 66, 133-152. Errata 69 (1995), 427-428.
Nelson, D., 1991, "A Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347-370.
Parti, M. and C. Parti, 1980, "The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector," *The Bell Journal of Economics*, 11, 309-321.
Subbotin, M.T.H., 1923, " On the Law of Frequency of Error." *Mathematicheskii Sbornik,* 31, 296-301.
Theodossiou, P., 1998, "Financial Data and the Skewed Generalized t Distribution."*Management Science*, 44, 1650-1661.
Theodossiou, P., 2001, "Skewness and Kurtosis in Financial Data and the Pricing of Options." Rutgers University.
Yohai, V.J. and R.H. Zamar, 1997. "Optimal Locally Robust M-estimates of Regression," *Journal of Statistical Planning and Inference*, 64, 309-324.
Younes, S., and T. Muneer, 2006, "Improvements in Solar Radiation Models Based on Cloud Data," *Building Services and Engineering Research Technology Journal*, 27, 41-54
Zeckhauser, R. and M. Thompson, 1970. "Linear Regression with Nonnormal Error Terms," *Review of Economics and Statistics*, 52, 280-286.

Table 1. Bivariate Relative Frequency of Skewness (*SK*) and Kurtosis (*KU*)

$SK \setminus KU$	$0 \leq KU < 4$	$4 \leq KU < 8$	$8 \leq KU < 12$	$12 \leq KU$	$N(SK)$
$-8.3 \leq SK < -0.6$	0 0.00	2 0.0003	16 0.0025	288 0.0443	306 0.0471
$-0.6 \leq SK < -0.2$	2 0.0003	78 0.0120	88 0.0135	126 0.0194	294 0.0452
$-0.2 \leq SK < 0.2$	190 0.0292	363 0.0558	144 0.0221	111 0.0171	808 0.124
$0.2 \leq SK < 0.6$	470 0.0723	724 0.1114	170 0.0261	115 0.0177	1,479 0.2275
$0.6 \leq SK < 40$	112 0.0172	911 0.1401	710 0.1092	1,882 0.2894	3,615 0.5560
$N(KU)$	774 0.1190	2,078 0.3196	1,128 0.1735	2,522 0.3879	6,502 1

Notes: Sample sizes used in the estimation of stock betas range between 1,001 and 2,519 observations. Standard errors for skewness, computed as $(6/T)^{0.5}$, where T is the number of observations, range between 0.002 and 0.077. Standard errors for kurtosis, computed as $(24/T)^{0.5}$, range between 0.095 and 0.155.

Table 2. Beta Frequency Distribution: SGT Model

Range		Number of Stocks	Fraction of Total Sample
–0.25	0.00	972	0.1495
0.00	0.25	1,927	0.2964
0.25	0.50	785	0.1207
0.50	0.75	1,070	0.1646
0.75	1.00	774	0.1190
1.00	1.25	397	0.0611
1.25	1.50	230	0.0354
1.50	1.75	135	0.0208
1.75	2.00	98	0.0151
2.00	2.25	72	0.0111
Above 2.25		42	0.0065
Total		6,502	1

Notes: Average beta of all stocks: 0.4738.

Table 3. Relative Frequency of the Skewness Parameter λ

Ranges	$N(\lambda)$/ $P(\lambda)$	Significant at the 5% level	1% level
$\lambda < -0.05$	170	126	122
	0.0261	0.7412	0.7176
$-0.050 \leq \lambda < -0.025$	377	234	221
	0.0580	0.6207	0.5862
$-0.025 \leq \lambda < 0.000$	844	320	285
	0.1298	0.3791	0.3377
$0.000 \leq \lambda < 0.025$	1,263	229	212
	0.1942	0.1813	0.1679
$0.025 \leq \lambda < 0.050$	1,247	186	111
	0.1918	0.1492	0.0890
$0.050 \leq \lambda < 0.075$	986	603	217
	0.1516	0.6116	0.2201
$0.075 \leq \lambda < 0.100$	708	647	502
	0.1089	0.9138	0.7090
$0.100 \leq \lambda < 0.125$	408	400	381
	0.0627	0.9804	0.9338
$0.125 \leq \lambda < 0.150$	261	261	259
	0.0401	1	0.9923
$0.150 \leq \lambda$	238	236	236
	0.0366	0.9916	0.9916
Total	6,502	3,242	2,546
Fraction of Total	1	0.4986	0.3916

Notes: The t-values used for significance tests are based on robust standard errors.

Table 4. Relative Significance of Non-Centered SGT Adjustment Factor $\rho_i \varphi_i$

$\rho_i \varphi_i$	$N(\rho_i \varphi_i)$/ $P(\rho_i \varphi_i)$	Significant at the 5% level	1% level
$\rho_i \varphi_i < -0.5$	118 0.0181	87 0.7373	86 0.7288
$-0.50 \leq \rho_i \varphi_i < -0.25$	275 0.0423	176 0.6400	167 0.6073
$-0.25 \leq \rho_i \varphi_i < 0.00$	973 0.1496	404 0.4152	365 0.3751
$0.00 \leq \rho_i \varphi_i < 0.25$	2,987 0.4631	813 0.2722	497 0.1664
$0.25 \leq \rho_i \varphi_i < 0.50$	910 0.1400	630 0.6923	425 0.4670
$0.50 \leq \rho_i \varphi_i < 0.75$	482 0.0741	430 0.8921	350 0.7261
$0.75 \leq \rho_i \varphi_i < 1.00$	301 0.0463	290 0.9635	256 0.8505
$1.00 \leq \rho_i \varphi_i < 1.25$	187 0.0288	179 0.9572	172 0.9198
$1.25 \leq \rho_i \varphi_i < 1.50$	103 0.0158	103 1.0000	102 0.9903
$1.50 \leq \rho_i \varphi_i$	114 0.0175	114 1	114 1
Total Fraction of Total	6,450 1	3,226 0.5002	2,534 0.3929

Notes: For the computation of the significance frequency rates of the intercept adjustment factor we used the robust t-values of λ's. The above table is based on a sample of 6,450 stocks, because 52 stocks had estimated values for $n_i < 1$, thus $\rho_i \varphi_i$ was not defined.

Table 5. Relative Frequency of Intercept Bias Due to Skewness: SGT vs. GT

$b(\alpha_i) = \alpha_{SGT,i} - \alpha_{GT,i}$	$N(b(\alpha_i))$/ $P(b(\alpha_i))$	Significant at the 5% level	1% level
$b(\alpha_i) < -0.75$	37	29	27
	0.0057	0.7838	0.7297
$-0.75 \leq b(\alpha_i) < -0.50$	81	69	53
	0.0125	0.8519	0.6543
$-0.50 \leq b(\alpha_i) < -0.25$	272	211	174
	0.0418	0.7757	0.6397
$-0.25 \leq b(\alpha_i) < 0.00$	1,035	690	591
	0.1592	0.6667	0.5710
$0.00 \leq b(\alpha_i) < 0.25$	4,243	2,514	1,720
	0.6526	0.5925	0.4054
$0.25 \leq b(\alpha_i) < 0.50$	659	646	585
	0.1014	0.9803	0.8877
$0.50 \leq b(\alpha_i) < 0.75$	141	141	140
	0.0217	1	0.9929
$0.75 \leq b(\alpha_i)$	34	34	34
	0.0052	1	1
Total	6,502	4,334	3,324
Fraction of Total	1	0.6666	0.5112

Note: The test statistics for the bias significance are based on the log-likelihood ratio test statistic of the SGT (unrestricted) and GT (restricted) models. The latter ratio follows chi-square distribution with one degree of freedom.

Table 6. Mean of Intercept and Slope Based on 1,050 Simulations of 2,519 Obervations

	OLS	LAD	SLAD	GED	SGED	T	ST	GT	SGT
A. Intercept									
Normal	0.00068	0.00059	0.00093	0.00012	0.00049	0.00118	0.00087	0.00101	0.00094
	(0.009)	(0.006)	(0.013)	(0.002)	(0.007)	(0.016)	(0.012)	(0.014)	(0.013)
Mix-Normal	–0.00047	–0.00021	–0.00029	–0.00015	–0.00063	–0.00056	–0.0009	–0.00037	–0.00137
	(–0.006)	(–0.005)	(–0.003)	(–0.003)	(–0.019)	(–0.018)	(–0.014)	(–0.012)	(–0.017)
Log-Normal	–0.00013	–0.75824	–0.00561	0.19764	–0.01107	–0.62898	–0.02649	–0.42671	–0.02572
	(–0.002)	(–9.708)*	(0.078)	(2.165)*	(–0.135)	(–8.228)*	(–0.368)	(–4.189)*	(–0.358)
B. Slope									
Normal	1.00268	1.00311	1.00365	1.0027	1.00269	1.00256	1.00263	1.00262	1.00266
	(0.041)	(0.037)	(0.043)	(0.041)	(0.041)	(0.039)	(0.040)	(0.040)	(0.040)
Mix-Normal	1.00089	0.99931	0.99929	0.99864	0.99861	0.99877	0.99873	0.99919	0.99917
	(0.014)	(–0.020)	(–0.020)	(–0.047)	(–0.048)	(–0.048)	(–0.049)	(–0.032)	(–0.033)
Log-Normal	0.99962	0.99742	0.99697	0.99929	0.99956	0.99861	0.99893	0.99894	0.99879
	(–0.006)	(–0.039)	(–0.063)	(–0.017)	(–0.006)	(–0.026)	(–0.027)	(–0.021)	(–0.030)

Notes: Intercept t-statistics (in parentheses) test the null hypothesis of no difference from zero. Slope t-statistics test the null hypothesis of no difference from one. * refers to statistical significance at the 5% level for a two-tailed test.

Table 7. Root Mean Squared Error of Intercept and Slope Based on 1,050 Simulations of 2,519 Observations

	OLS	LAD	SLAD	GED	SGED	T	ST	GT	SGT
A. Intercept									
Normal	0.0730	0.0925	0.0734	0.0732	0.0733	0.0740	0.0730	0.0731	0.0730
Mix-Normal	0.0731	0.0404	0.0849	0.0341	0.0582	0.0306	0.0642	0.0296	0.0804
Log-Normal	0.0721	0.7623	0.0723	0.2177	0.0829	0.6336	0.0768	0.4387	0.0764
B. Slope									
Normal	0.0656	0.0834	0.0847	0.0660	0.0659	0.0662	0.0665	0.0659	0.0661
Mix-Normal	0.0631	0.0346	0.0346	0.0287	0.0291	0.0258	0.0258	0.0251	0.0250
Log-Normal	0.0634	0.0657	0.0479	0.0634	0.0427	0.0534	0.0400	0.0493	0.0402