Financial Data and the Skewed Generalized T Distribution

Panayiotis Theodossiou
School of Business, Rutgers University, Camden, New Jersey 08102

This paper develops a skewed extension of the generalized t (GT) distribution, introduced by McDonald and Newey (1988). In particular, the paper derives the mathematical moments and other properties of the distribution and assesses its ability to fit the empirical distribution of several financial series characterized by skewness and excess kurtosis. In all cases the skewed GT provides an excellent fit to the empirical distribution of data.

(*Skewed Generalized t Distribution; Laplace Distribution; Power Exponential Distribution; Skewness; Leptokurtic; Value at Risk*)

1. Introduction

This paper develops a skewed version of the Generalized t (GT) distribution introduced by McDonald and Newey (1988); see also McDonald and Nelson (1989), Butler et al. (1990), Lye and Martin (1993), and McDonald and Xu (1995). The skewed GT is a flexible distribution accommodating the skewness and excess kurtosis often present in financial data. Several well-known distributions are nested within the skewed GT. These include the symmetric and skewed student's t, e.g., Hansen (1994), the power exponential, e.g., Subbotin (1923) and Box and Tiao (1962), the Laplace, the normal, and the uniform distributions, e.g., Johnson et al. (1995).

From the aforementioned set of probability distributions, the student's t is used by Bollerslev (1987) and Baillie and Bollerslev (1989) to model the distribution of foreign exchange rates; the power exponential distribution is used by Nelson (1991) to model the distribution U.S. stock market returns; Hsieh (1989), Theodossiou (1994) and Koutmos and Theodossiou (1994), to model the distribution of foreign exchange rates; and Akgiray et al. (1991), to model the distribution of precious metals prices. Applications of the symmetric GT to U.S. stock index returns are presented in Bollerslev et al. (1994, pp. 3017–3027). Note that the aforementioned distributions impose the restriction of symmetry, which is not always valid for financial data (e.g., Engle and Gonzales-Rivera 1991, and Markowitz and Usmen 1996).

Specifically, this paper derives the mathematical moments and other properties of the skewed GT and assesses its ability to fit the empirical (unconditional) distribution of several financial data series. The series are the S&P500 stock market index for the United States; the TSE300 stock market index for Canada; the Topix stock market index for Japan; the Canadian dollar and Japanese yen to the U.S. dollar exchange rates, and the price of gold. In all cases, the skewed GT provides a good fit to the empirical distribution of the data.

The paper is organized as follows. Section 2 provides many analytical results on the skewed GT distribution. Section 3 presents several applications of the skewed GT distribution to financial data. The last section presents the conclusions.

2. Mathematical Properties of the Skewed GT Distribution

DEFINITION 1. Let x be a random variable with the following probability density function (p.d.f.):

$$f(x|k,n,\lambda,\sigma^2)=\begin{cases} f_1 = C(1+(k/(n-2))\theta^{-k} \\ \quad\times(1-\lambda)^{-k}|x/\sigma|^k)^{-(n+1)/k} \\ \quad \text{for } x<0, \\ f_2 = C(1+(k/(n-2))\theta^{-k} \\ \quad\times(1+\lambda)^{-k}|x/\sigma|^k)^{-(n+1)/k} \\ \quad \text{for } x\geq 0, \end{cases} \tag{1}$$

where k, n, λ, and σ^2 are scaling parameters and C and

0025-1909/98/4412/1650$05.00

θ are normalizing constants ensuring that $f(\cdot)$ is a proper p.d.f. and σ^2 is the variance of the random variable x. Note that $k > 0$, $n > 2$, and $-1 < \lambda < 1$. The parameters k and n control the height and tails of the density. The skewness parameter λ controls the rate of descent of the density around $x = 0$. For example, in the case of positive skewness, $\lambda > 0$, the random variable x in f_1 is weighted by a value greater than unity and in f_2 is weighted by a value less than a unity, resulting in a density skewed to the right. The parameter n has the degrees of freedom interpretation in case $\lambda = 0$ and $k = 2$.

THEOREM 1. *For the above probability density function* $f(\cdot)$,

$$C = 0.5kB(1/k, n/k)^{-3/2} \times B(3/k, (n-2)/k)^{1/2} S(\lambda)\sigma^{-1}, \tag{2}$$

$$\theta = (k/(n-2))^{1/k} B(1/k, n/k)^{1/2} \times B(3/k, (n-2)/k)^{-1/2} S(\lambda)^{-1}, \tag{3}$$

and

$$S(\lambda) = (1 + 3\lambda^2 - 4\lambda^2 B(2/k(n-1)/k)^2 \times B(1/k, n/k)^{-1} B(3/k, (n-2)/k)^{-1})^{1/2}, \tag{4}$$

where $B(\cdot)$ is the beta function. Furthermore, the r^{th} noncentered moment function of x for integer values of r is

$$M_r \equiv E(x^r) = 0.5((-1)^r(1-\lambda)^{(r+1)} + (1+\lambda)^{(r+1)})S(\lambda)^{-r} \times B((r+1)/k, (n-r)/k)B(1/k, n/k)^{-1-(r/2)} \times B(3/k, (n-2)/k)^{-r/2}\sigma^r \tag{5}$$

and the expected value of $|x|$ is

$$E(|x|) = (1+\lambda^2)S(\lambda)^{-1}B(2/k, (n-1)/k) \times B(1/k, n/k)^{-1/2}B(3/k, (n-2)/k)^{-1/2}\sigma. \tag{6}$$

PROOF. See the Appendix.

COROLLARY 1. *It follows from Theorem 1 that*

$$\mu \equiv E(x) = 2\lambda S(\lambda)^{-1}B(2/k, (n-1)/k) \times B(1/k, n/k)^{-1/2}B(3/k, (n-2)/k)^{-1/2}\sigma, \tag{7}$$

$$\mathrm{Var}(x) = \sigma^2, \tag{8}$$

$$E(x^3) = M_3 = 4\lambda(1+\lambda^3)S(\lambda)^{-3}B(4/k, (n-3)/k) \times B(1/k, n/k)^{1/2}B(3/k, (n-2)/k)^{-3/2}\sigma^3, \tag{9}$$

$$m_3 \equiv E(x-\mu)^3 = E(x^3) - 3\mu\sigma^2 - \mu^3, \tag{10}$$

$$E(x^4) = M_4 = (1 + 10\lambda^2 + 5\lambda^4)S(\lambda)^{-4}B(5/k, (n-4)/k) \times B(1/k, n/k)B(3/k, (n-2)/k)^{-2}\sigma^4, \tag{11}$$

$$m_4 = E(x-\mu)^4 = E(x^4) - 4\mu m_3 - 6\mu^2\sigma^2 - \mu^4. \tag{12}$$

THEOREM 2. *The probabilities*

$$P(x<0) = \int_{-\infty}^{0} f_1\, dx = (1-\lambda)/2, \tag{13}$$

$$P(x>0) = \int_{0}^{+\infty} f_2\, dx = (1+\lambda)/2. \tag{14}$$

PROOF. See the Appendix.

THEOREM 3. *The probability density function of the transformed variable $z \equiv x - \mu$ is*

$$f(z|\cdot) = \begin{cases} f_1 = C(1 + (k/(n-2))\theta^{-k} \times (1-\lambda)^{-k}|(z+\mu)/\sigma|^k)^{-(n+1)/k} & \text{for } z < -\mu, \\ f_2 = C(1 + (k/(n-2))\theta^{-k} \times (1+\lambda)^{-k}|(z+\mu)/\sigma|^k)^{-(n+1)/k} & \text{for } z \geq -\mu, \end{cases} \tag{15}$$

where the constants C, μ, and θ are defined by Equations (2), (3), *and* (7).

PROOF. See the Appendix.

THEOREM 4. *The random variable z has*

$$E(|z|) = E(|x|) - \lambda\mu + 2\int_{\mu}^{0} x f_1\, dx - 2\mu\int_{\mu}^{0} f_1\, dx \quad \text{for } \lambda < 0, \tag{16a}$$

$$E(|z|) = E(|x|) - \lambda\mu - 2\int_{0}^{\mu} x f_2\, dx - 2\mu\int_{0}^{\mu} f_2\, dx \quad \text{for } \lambda > 0, \tag{17b}$$

$$E(z) = 0, \tag{18}$$

$$\mathrm{Var}(z) = \sigma^2, \tag{19}$$

$$E(z^3) = m_3 = E(x^3) - 3\mu\sigma^2 - \mu^3, \tag{20}$$

$$E(z^4) = m_4 = E(x^4) - 4\mu m_3 - 6\mu^2\sigma^2 - \mu^4. \tag{21}$$

PROOF. See the Appendix.

REMARK 1. It can easily be shown that, other things being equal, $E(|z||\lambda = c) = E(|z||\lambda = -c)$, where c is a positive constant.

DEFINITION 2. The skewness and kurtosis measures are respectively

$$S_k = m_3/\sigma^3, \tag{22}$$

$$K_u = m_4/\sigma^4. \tag{23}$$

REMARK 2. For $\lambda = 0$, the skewed GT distribution presented by Equations (1) and (15) is symmetric and has all odd moments equal to zero.

REMARK 3. The skewed GT distribution generates for $\lambda = 0$ McDonald's and Newey's GT distribution; for $k = 2$, Hansen's skewed student's t distribution; for $\lambda = 0$ and $k = 2$, the student's t distribution; for $\lambda = 0$ and $n = \infty$, the Subbotin's power exponential distribution; for $\lambda = 0$, $k = 1$ and $n = \infty$, the Laplace distribution; for $\lambda = 0$, $k = 2$ and $n = 1$, the Cauchy distribution; for $\lambda = 0$, $k = 2$, and $n = \infty$, the normal distribution; and for $\lambda = 0$, $k = \infty$, and $n = \infty$, the uniform distribution.

Tables 1–4 present skewness and kurtosis values for various combinations of k and n of the skewed GT with $\lambda = 0$ (symmetric), 0.05, 0.1 and 0.2. The results of these tables provide additional insight into the relationship between the scaling parameters k, n, and λ and the shape of the skewed GT. It appears from the tables that smaller values of k and n result in larger values for the kurtosis (i.e., more leptokurtic p.d.f.s) and vice versa. Moreover, larger positive values of λ result in larger positive values for both skewness and kurtosis. Note that the respective negative values of λ will generate identical tables for skewness and kurtosis; however, all skewness values will be negative.

3. Applications of the Skewed Generalized t Distribution

3.1 Data Preliminary Statistics

The financial series used are the S&P500 stock market index for the United States; the TSE300 stock market index for Canada; the Topix stock market index for Japan; the Canadian dollar and Japanese yen exchange rates with respect to the U.S. dollar; and the U.S. dollar price of gold. The series are transformed into continuously compounded daily percentage returns (logarithmic changes) using the formula

$$y_t = 100^*(\ln(S_t) - \ln(S_{t-1})), \tag{24}$$

where ln is the natural logarithm and S_t is the level of each series at time t. Investigating the distributional properties of logarithmic changes is more appropriate. Unlike the levels of the series, logarithmic changes are stationary processes. Moreover, their distributional

Table 1 Skewed GT Distribution with $\lambda = 0$

	Kurtosis—K_u					
$\backslash n$ / k	5	6	8	16	30	∞
1	36	20	12.6000	8.0769	6.9402	6
1.25	20.4954	12.2806	8.3059	5.7550	5.0890	4.5272
1.5	14.1523	8.9039	6.3093	4.6032	4.1490	3.7620
1.75	10.9069	7.0968	5.1960	3.9325	3.5931	3.3026
2	9	6	4.5000	3.5000	3.2308	3
5	4.2853	3.1105	2.5544	2.2177	2.1359	2.0701
10	3.5501	2.6180	2.1941	1.9635	1.9168	1.8842
∞	3.2450	2.4037	2.0281	1.8401	1.8116	1.8014

Note. For $\lambda = 0$ the distribution is symmetric and skewness is equal to zero. The kurtosis for the Laplace distribution ($k = 1$ and $n = \infty$) is equal to six, for the normal distribution ($k = 2$ and $n = \infty$) is equal to three, and for the uniform distribution ($k = \infty$ and $n = \infty$) is equal to 1.8014. The kurtosis values for the student's t-distribution for various values of n (degrees of freedom) are given by the row corresponding to $k = 2$. Kurtosis measures are computed using equations (12) and (23).

Table 2 Skewed GT Distribution with $\lambda = 0.05$

			Skewness—S_k			
$\backslash n$ k	5	6	8	16	30	∞
1	.5490	.4411	.3523	.2671	.2386	.2115
1.25	.3981	.3209	.2569	.1952	.1744	.1546
1.5	.3125	.2512	.2006	.1517	.1353	.1197
1.75	.2578	.2061	.1636	.1228	.1092	.0963
2	.2201	.1747	.1375	.1023	.0907	.0797
5	.0922	.0657	.0456	.0290	.0242	.0200
10	.0637	.0405	.0238	.0115	.0085	.0063
∞	.0505	.0285	.0131	.0027	.0008	.0001
			Kurtosis—K_u			
$\backslash n$ k	5	6	8	16	30	∞
1	36.4693	20.2226	12.7150	8.1315	6.9809	6.0299
1.25	20.7459	12.4008	8.3684	5.7847	5.1110	4.5432
1.5	14.3133	8.9805	6.3488	4.6216	4.1625	3.7717
1.75	11.0227	7.1510	5.2234	3.9449	3.6022	3.3090
2	9.0897	6.0411	4.5203	3.5090	3.2372	3.0045
5	4.3151	3.1215	2.5586	2.2190	2.1368	2.0706
10	3.5720	2.6252	2.1963	1.9640	1.9170	1.8843
∞	3.2640	2.4095	2.0296	1.8403	1.8117	1.8014

Note. Skewness measures are computed using equations (10) and (22). Kurtosis measures are computed using equations (12) and (23).

properties are more relevant for financial equilibrium models. A few values for each data series falling outside the range of plus-minus six standard deviations from the mean are dropped from the sample.

Table 5 presents several preliminary statistics on the data. The first row gives the Dickey-Fuller F-values for testing the null hypothesis that each $\ln(S_t)$ series follows a random walk process with a drift, e.g., Dickey and Fuller (1979) and (1981). All F-values are lower than their critical value at the one-percent level of 8.27, providing support to the null hypothesis. The second row gives the Dickey-Fuller F-values for the series of logarithmic changes. Unlike the levels, these F-values are lower than their critical value at the one-percent level of 8.27, providing support to the hypothesis that the series of logarithmic changes are stationary processes.

The third and fourth rows give the statistics for skewness (b_1) and excess kurtosis (b_2). These are calculated using the formulae $b_1 = m_3 / m_2^{3/2}$ and $b_2 = (m_4 / m_2^2) - 3$, where m_j is the estimate for the j^{th} moment around the mean. Under the null hypothesis of normality, the two statistics are normally distributed with standard errors $\text{se}(b_1) = \sqrt{6/T}$ and $\text{se}(b_2) = \sqrt{24/T}$, where T is the sample size. The results for kurtosis imply that all the y_t series exhibit excess kurtosis beyond that of the normal distribution. Moreover, the results for skewness imply that the y_t series for the Topix and the Japanese yen are negatively skewed and the series for the Canadian dollar are positively skewed.

The fifth row gives the Bera-Jarque (BJ) statistic for detecting departures of the data from normality. The BJ statistic is calculated using the formula $\text{BJ} = T(b_1^2/6 + b_2^2/24)$, where b_1, b_2, and T are defined above. Under the null hypothesis of normality, the BJ statistic is distributed as $\chi^2(2)$ with 2 degrees of freedom. All BJ values are greater than their critical value at the one-percent level of 9.21, indicating that the y_t series are nonnormal.

Table 3 Skewed GT Distribution with $\lambda = 0.1$

$\backslash n$ k	5	6	8	16	30	∞
			Skewness—S_k			
1	1.0851	.8724	.6972	.5292	.4729	.4194
1.25	.7886	.6359	.5094	.3873	.3461	.3070
1.5	.6199	.4986	.3982	.3015	.2689	.2380
1.75	.5120	.4095	.3251	.2443	.2173	.1917
2	.4375	.3473	.2736	.2037	.1805	.1586
5	.1841	.1311	.0911	.0579	.0483	.0400
10	.1273	.0809	.0476	.0229	.0170	.0126
∞	.1012	.0570	.0261	.0054	.0016	.0002
			Kurtosis—K_u			
$\backslash n$ k	5	6	8	16	30	∞
1	37.8391	20.8732	13.0518	8.2919	7.1005	6.1176
1.25	21.4807	12.7540	8.5523	5.8720	5.1759	4.5905
1.5	14.7874	9.2066	6.4653	4.6760	4.2026	3.8006
1.75	11.3646	7.3112	5.3045	3.9819	3.6290	3.3281
2	9.3552	6.1629	4.5806	3.5357	3.2564	3.0180
5	4.4041	3.1545	2.5712	2.2230	2.1393	2.0722
10	3.6377	2.6469	2.2031	1.9655	1.9178	1.8848
∞	3.3208	2.4268	2.0342	1.8408	1.8118	1.8014

Note. Skewness measures are computed using equations (10) and (22). Kurtosis measures are computed using equations (12) and (23).

The sixth row gives the Kolmogorov-Smirnov (KS) statistics for testing the null hypothesis of normality. This is calculated using the formula KS $= \max|F_E(y_t) - F_G(y_t)|$, for $t = 1, 2, \ldots, T$, where $F_E(y_t) = t/(T-1)$ and $F_G(y_t)$ are the empirical and normal cumulative distributions for y_t, respectively. All KS values are greater than their respective critical values at the one-percent level of significance, rejecting the null hypothesis of normality for the data. The critical values for KS are calculated using the formula $1.63/\sqrt{T}$; e.g., L. H. Miller (1956).

These preliminary results indicate that all series exhibit excess kurtosis and, in some cases, skewness. As such, the series cannot be normally distributed. The results, however, provide no information as to the true parametric probability distribution of the series.

3.2 Maximum Likelihood Estimation

The Maximum Likelihood (ML) estimators for the parameters of the skewed GT distribution can be obtained from the maximization of the sample log-likelihood function

$$L(\beta) = \sum_{t=1}^{T} L_t(\beta) = \sum_{t=1}^{T} \log[C(1 + (k/(n-2))\theta^{-k} \times (1 + \text{sign}(x_t)\lambda)^{-k}|x_t/\sigma|^k)^{-(n+1)/k}], \quad (25)$$

with respect to column vector $\beta = (k, n, \lambda, \sigma^2)'$, where $-1 < \lambda < 1$, $\sigma^2 > 0$, C is given by Equation (2), θ is given by equation (3), T is the sample size, $z_t = y_t - \bar{y}$, $\bar{y}$ is the sample mode, $x_t = z_t + \mu$, μ is given by Equation (7), $\text{sign}(x_t) = -1$ for $x_t < 0$ and $\text{sign}(x_t) = +1$ for $x_t \geq 0$, and $L_t(\beta)$ is the likelihood function for the t^{th} observation.

The maximization of the above log-likelihood function is troublesome, however. In many occasions, the iterative algorithm "overshoots" and assigns values to σ^2, (which are negative) and to λ (which lie outside the unit interval), resulting in a breakdown of the algo-

Table 4 **Skewed GT Distribution with $\lambda = 0.2$**

			Skewness—S_k			
$\backslash n$ k	5	6	8	16	30	∞
1	2.0739	1.6711	1.3391	1.0200	.9127	.8107
1.25	1.5188	1.2272	.9854	.7514	.6724	.5971
1.5	1.2010	.9677	.7744	.5878	.5249	.4651
1.75	.9964	.7981	.6347	.4780	.4255	.3758
2	.8543	.6792	.5359	.3997	.3544	.3118
5	.3657	.2604	.1810	.1150	.0959	.0794
10	.2544	.1617	.0950	.0458	.0339	.0251
∞	.2029	.1144	.0524	.0107	.0033	.0004
			Kurtosis—K_u			
$\backslash n$ k	5	6	8	16	30	∞
1	42.7957	23.2397	14.2843	8.8833	7.5433	6.4438
1.25	24.1900	14.0619	9.2371	6.1996	5.4200	4.7690
1.5	16.5592	10.0547	6.9045	4.8823	4.3548	3.9106
1.75	12.6557	7.9185	5.6131	4.1232	3.7322	3.4016
2	10.3656	6.6281	4.8117	3.6385	3.3305	3.0700
5	4.7544	3.2844	2.6209	2.2387	2.1493	2.0784
10	3.8987	2.7329	2.2301	1.9714	1.9211	1.8866
∞	3.5476	2.4962	2.0525	1.8427	1.8122	1.8014

Note. Skewness measures are computed using equations (10) and (22). Kurtosis measures are computed using equations (12) and (23).

rithm. This problem can be avoided by substituting the following monotonic functions for λ and σ^2

$$\lambda = 2(1 + \exp(-\Lambda))^{-1} - 1 \tag{26}$$

and

$$\sigma^2 = \exp(V), \tag{27}$$

into Equation (25), and maximizing the log-likelihood function with respect to the parameters k, n, Λ, and V. Let the transformed log-likelihood function be

$$l(\varphi) = L(k, n, \lambda(\Lambda), \sigma^2(V)) = \sum_{t=1}^{T} l_t(\varphi), \tag{28}$$

where $\varphi = (k, n, \Lambda, V)'$.

The maximization of $l(\varphi)$ is performed using the Berndt et al. (1974) algorithm. That is, the ML estimators for φ are obtained using the recursive equation

$$\varphi^{s+1} = \varphi^s + M(\varphi^s)^{-1}S(\varphi^s), \tag{29}$$

where φ^s denotes the parameter vector generated by the s^{th} iteration of the algorithm, $S(\varphi^s) = \Sigma_{t=1}^{T} \partial l_t / \partial\varphi$ is a vector of the first (partial) derivatives of $l(\varphi)$ with respect to φ, and $M(\varphi^s) = \Sigma_{t=1}^{T} (\partial l_t / \partial\varphi)(\partial l_t / \partial\varphi')$ is a matrix of cross products of the first derivatives. Both $S(\varphi^s)$ and $M(\varphi^s)$ are extremely complex functions of φ, and, as such, they are difficult to derive analytically. Therefore, their evaluation is accomplished using a relatively accurate numerical method suggested by Press et al. (1992, pp. 186–188).

The recursive procedure described by the above system of equations is repeated until the first-order conditions for maximum are met, i.e., $S(\varphi^s) \approx 0$. The starting (initial) values for the algorithm are arbitrarily set to $\varphi^0 = (1.5, 15, 0, \ln(s^2))'$, where s^2 is the OLS estimator for the variance of x_t. Convergence of the algorithm is assumed when the absolute values of all first derivatives, $|S(\varphi^s)|$, fall below 10^{-3}. In most cases, the algorithm gives a global maximum in less

Table 5 Preliminary Statistics

Parameters	S&P500 (U.S.)	TSE300 (Canada)	Topix (Japan)	Can-$/US-$	Jap-yen/US-$	Gold in US-$
Dickey-Fuller F-value for $\ln(S_t)$	5.342*	2.755*	3.091*	.8714*	1.86*	5.297*
Dickey-Fuller F-value for y_t	956.9	363.6	715.8	888.2	902.0	1,494
Skewness	−.0160	.0392	−.2474	.2049	−.3566	.0680
	(.0400)*	(.0554)*	(.0474)	(.0403)	(.0402)	(.0331)*
Excess Kurtosis	2.6509	2.8655	3.7600	2.4375	2.9307	5.4101
	(.0801)	(.1107)	(.0949)	(.0805)	(.0805)	(.0661)
Bera-Jarque	1,095.52	670.38	1,597.01	942.65	1,404.43	6,697.23
KS	.0495	.0676	.0777	.0535	.0612	.0869
Critical value at 1%	.0267	.0368	.0316	.0268	.0268	.0220
# of observations	3,746	1,970	2,675	3,705	3,705	5,498
Period of data	1/3/80–	1/5/87–	1/5/84–	1/3/80–	1/3/80–	1/3/73–
	10/25/94	10/25/94	10/26/94	10/3/94	10/3/94	2/22/95

Notes. Series are expressed as continuously compounded daily returns (logarithmic changes). Standard errors for the estimates are included in parentheses. The Dickey-Fuller F-value is for testing the null hypothesis that a series is a random walk process with a trend. Its critical value at the one-percent level of significance is 8.27. The statistics for skewness is $b_1 = m_3/m_2^{3/2}$ and for kurtosis is $b_2 = (m_4/m_2^2) - 3$, where m_j is the estimate for the j^{th} moment around the mean. The Bera-Jarque statistic for testing normality is $BJ = T(b_1^2/6 + b_2^2/24)$. This is asymptotically distributed as $\chi^2(2)$ with 2 degrees of freedom. Its critical value at the one-percent level is 9.21. KS is the Kolmogorov-Smirnov statistic for testing the null hypothesis of normality. Unless otherwise noted, all statistics are statistically significant or they reject the null hypothesis at the one-percent level of significance.

* implies statistically insignificant estimators or acceptance of the null hypothesis at the one-percent level.

than thirty iterations. Let $\varphi^* = (k^*, n^*, \Lambda^*, V^*)$ be the ML estimators for φ. Then, the ML estimators for λ and σ^2 are $\lambda^* = (1 + \exp(-\Lambda^*))^{-1} - 1$ and $\sigma^{2*} = \ln(V^*)$.

Define

$$A(\beta^*) = \sum_{t=1}^{T} \partial^2 L_t / \partial\beta\partial\beta', \tag{30}$$

$$B(\beta^*) = \sum_{t=1}^{T} (\partial L_t / \partial\beta)(\partial L_t / \partial\beta') \tag{31}$$

to be respectively the Hessian matrix and cross product of first derivatives of $L(\beta)$ with respect to the vector β, evaluated numerically at $\beta^* = (k^*, n^*, \lambda^*, \sigma^{2*})'$. It is well known from the theory of ML estimation that the variance matrix of β^* of a well-specified model can be obtained using one of the following two equations:

$$\text{var}(\beta^*) = -A(\beta^*)^{-1}, \tag{32}$$

$$\text{var}(\beta^*) = B(\beta^*)^{-1}. \tag{33}$$

This paper employs Equation (33) to calculate the variance of β^* because the errors associated with the numerical computation of the first derivatives used in Equation (33) are relatively smaller than those of the second derivatives used in Equation (32), e.g., Press et al. (1992, pp. 186–188).

The variance matrix of β^* for a mis-specified model is given by the quasi-ML

$$\text{Var}(\beta^*) = A(\beta^*)^{-1}B(\beta^*)A(\beta^*)^{-1}, \tag{34}$$

e.g., Bollerslev and Wooldridge (1988). However, Engle and Gonzalez (1991), using Monte Carlo simulations, find that the quasi-ML estimators of the standard errors are inefficient compared to those under the correct specification when the density is highly nonnormal and skewed.

3.3 Estimated Distribution of Financial Data

Table 6 presents the estimates for the parameters of the distributions of the six series. Standard errors for the estimates, calculated using Equation (33), are presented in parentheses. Each combination of the location parameters k and n indicates that the distribution of each series exhibits kurtosis beyond that permitted by the normal distribution. Specifically, the kurtosis values for the series, calculated using Formula (23), range between

Table 6 Skewed GT Estimates

Parameters	S&P500 (U.S.)	TSE300 (Canada)	Topix (Japan)	Can-$ to US-$	Jap-yen to US-$	Gold US-$
k	1.4497	1.6162	1.4273	1.6771	1.4170	1.2504
	(.1192)	(.2189)	(.1587)	(.1552)	(.1237)	(.0906)
n	9.6622	4.6639	4.6706	6.0192	7.8277	4.7325
	(3.4863)	(1.5183)	(1.3329)	(1.5528)	(2.687)	(.9360)
λ	.0152	.0172	−.0234	.0648	−.0753	.0094
	(.0170)*	(.0227)*	(.0169)*	(.0183)	(.0161)	(.0099)*
σ^2	.8000	.6316	1.2267	.0709	.4367	2.1720
	(.0285)	(.0614)	(.111)	(.0032)	(.0185)	(.1402)
μ	.0198	.0191	−.0351	.0250	−.0712	.0181
S_k	.0576	.1097	−.1743	.2807	−.3295	.0818
K_u	5.8081	16.9706	21.4394	7.5868	7.0738	26.5992
Log L	−4,724.20	−2,146.90	−3,726.11	−154.19	−3,487.27	−9,034.64
LR	332.38	273.68	520.21	342.41	450.84	1,498.66
KS	.0075*	.0127*	.0129*	.0144*	.0134*	.0109*
Critical value at 1%	.0267	.0368	.0316	.0268	.0268	.0220

Notes. Series are expressed as continuously compounded daily returns (logarithmic changes). Standard errors for the estimators are included in parentheses. μ is computed using equation (7). S_k and K_u are the skewness and kurtosis measures calculated using equations (22) and (23). LR is the log-likelihood ratio for testing the null hypothesis that the series are distributed as normal against the alternative hypothesis that the series are distributed as skewed GT. The LR follows $\chi^2(3)$. Its critical value at the one-percent level of significance is 11.34. KS is the Kolmogorov-Smirnov statistic for testing the null hypothesis that the data follow the skewed GT distribution with parameters k, n, and λ given by the table. Its critical values at the one-percent level are given by the last row. Unless otherwise noted, all statistics are statistically significant or they reject the null hypothesis at the one-percent level of significance.

* implies statistically insignificant estimators or acceptance of the null hypothesis at the one-percent level.

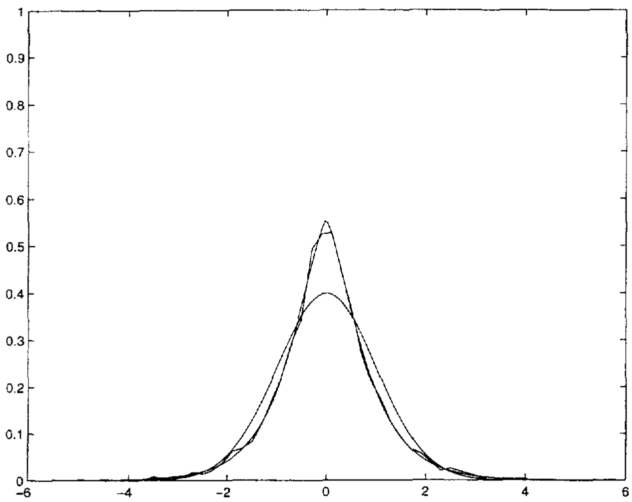
5.8081 (S&P500) and 26.5992 (gold). The skewness parameter λ is statistically insignificant at the one-percent level for the S&P500, TSE300, Topix, and gold, indicating that the distributions of these series are symmetric. On the other hand, the skewness parameter for the Canadian dollar and the Japanese yen are both statistically significant at the one-percent level. For the Canadian dollar it is positive, and for the Japanese yen it is negative, implying that the distribution of percentage changes for the Canadian dollar is positively skewed and for Japanese yen is negatively skewed.

Because the normal distribution is nested within the skewed GT distribution, a log-likelihood ratio (LR) statistic could be used to test the null hypothesis that the data are normally distributed against the hypothesis that the data are skewed GT distributed. The two hypotheses are H_0: $\lambda = 0$, $k = 2$, and $n = \infty$, and H_1: $k \neq 2$, $n \neq \infty$, and $\lambda \neq 0$. The statistic is calculated using the formula $LR = -2[L(\beta_0) - L(\beta_a)]$, where $L(\beta_0)$ and $L(\beta_a)$ are the maximum values of log-likelihood functions under the normal and skewed GT density specifications, respectively. The LR statistic is asymptotically distributed as a chi-square, $\chi^2(3)$, with three degrees of freedom. All LR values are greater than their critical value at the one-percent level of significance of 11.34, rejecting the null hypothesis of normality.

The last two rows give the Kolmogorov-Smirnov (KS) statistics and their critical values at the one-percent level of significance for testing the null hypothesis that the data follow the skewed GT with parameters k, n, and λ. The statistic is $KS = \max|F_E(y_t) - F_G(y_t)|$, for $t = 1, 2, \ldots, T$, where $F_E(y_t) = t/(T-1)$ and $F_G(y_t)$ are the empirical and skewed GT cumulative distributions for y_t. All KS values are lower than their respective critical values, providing strong support to the null hypothesis. Thus, the estimated skewed GT distributions appear to provide a good fit to the empirical distribution of the data.

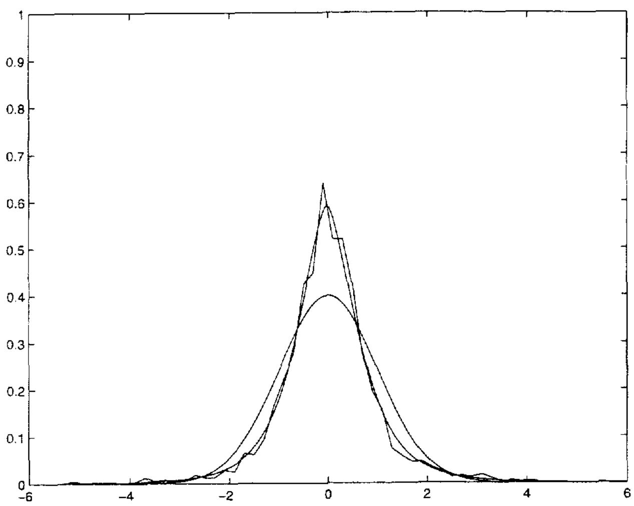
Figures (1)–(6) present graphical illustrations of the estimated skewed GT and non-parametric density functions

Figure 1 **U.S. Stock Market Returns**



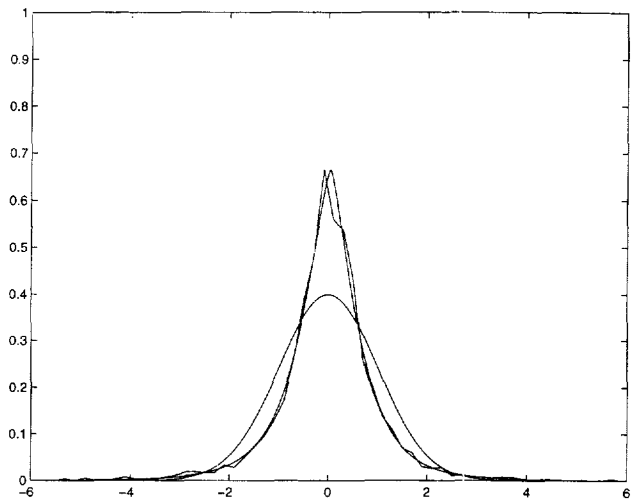
Skewed GT with k= 1.4497, n= 9.6622 and λ= 0.0152

Figure 2 **Canadian Stock Market Returns**



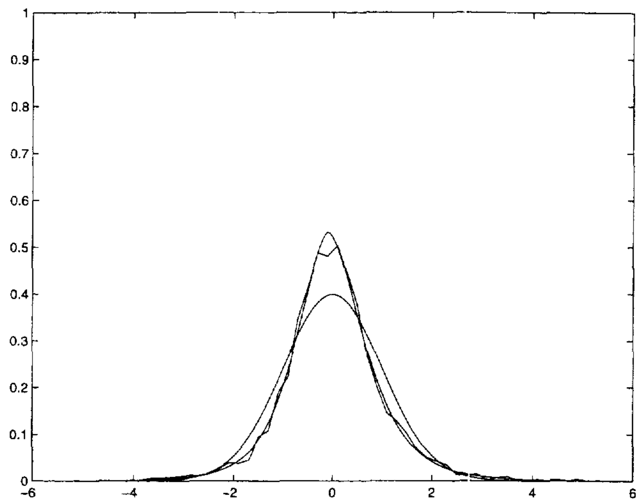
Skewed GT with k= 1.6162, n= 4.6639 and λ= 0.0172

Figure 3 **Japanese Stock Market Returns**



Skewed GT with k= 1.4273, n= 4.6706 and λ= -0.0234

Figure 4 **Percentage Changes in Canadian to U.S. Dollar Exchange Rates**



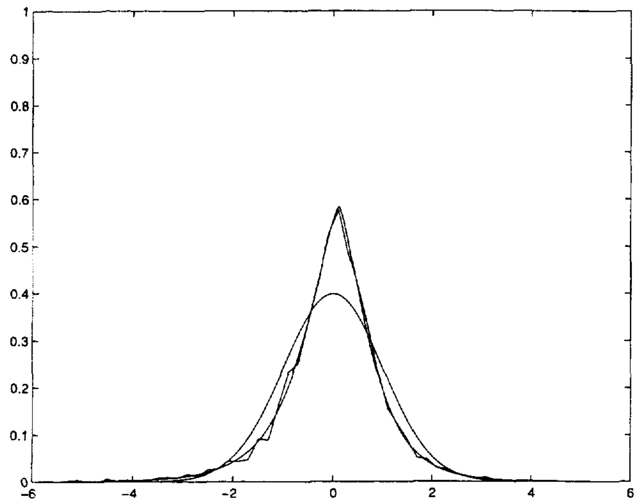
Skewed GT with k= 1.6771, n= 6.0192 and λ= 0.0648

for each series. The non-parametric density functions are constructed using the histogram (nonparametric) estimator, e.g., Tapia and Thompson (1978, §2.5). In addition, the figures provide graphical illustrations of the standard normal density function. For comparison purposes, all three densities are scaled to have zero mean and unit variance. Unlike the estimated normal densities, the estimated skewed GT and nonparametric probability density curves are almost indistinguishable, indicating that the skewed GT provides a good representation of the empirical distribution of each series.

5. Conclusions

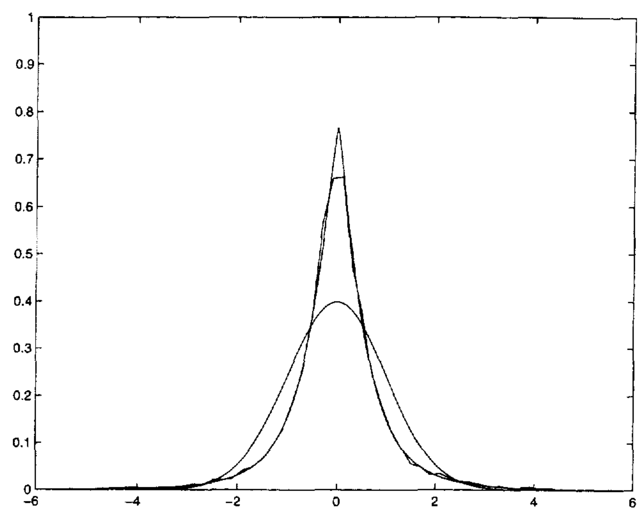
This paper developed a skewed extension of the generalized t (GT) distribution introduced by McDonald

Figure 5 **Percentage Changes in Japanese to U.S. Dollar Exchange Rates**



Skewed GT with k= 1.4170, n= 7.8277 and λ= −0.0753

Figure 6 **Percentage Changes in Gold Prices**



Skewed GT with k= 1.2504, n= 4.7325 and λ= 0.0094

and Newey (1988). The paper derived the mathematical moments and other properties of the skewed GT distribution and presented six applications of the distribution to financial data characterized by excess kurtosis and skewness. In all cases, the skewed GT provided an excellent fit to the empirical distribution of the data.

The probabilistic properties of financial data may have important implications for market equilibrium models such as portfolio models; the capital asset pricing model; and models for pricing futures and options contracts for currencies, stocks, and precious metals (e.g., Corrado and Sue 1995). Many of the existing models are based on the assumption that the distribution of financial data is symmetric and normal. The analytical and empirical results for the skewed GT presented in this paper could provide the foundation for developing economic and financial models that provide better approximations of reality. Moreover, the skewed GT distribution could be used to calculate value at risk (VaR) measures for the assets of a financial company.

In statistics, the skewed GT could be used within the GARCH (e.g., Engle 1982 and Bollerslev 1986), or the EGARCH frameworks (e.g., Nelson 1991), to model the conditional distribution and sample log-likelihood function of financial data. Under the correct specification of the conditional distribution, the ML estimators of the standard errors are more efficient than the quasi-ML standard errors, e.g., Engle and Gonzales-Rivera (1991). Note that, due to Jensen's inequality, the conditional distribution of the data is expected to be less skewed and leptokurtic than the unconditional distribution, e.g., Hsieh (1989, p. 310).[1]

[1] A portion of this article was written while the author was visiting the University of Cyprus.

Appendix

I. Proof of Theorem 1

The r^{th} noncentered moment of x for integer values of r is

$$
\begin{aligned}
M_r &= \int_{-\infty}^{+\infty} x^r f(x)\,dx = \int_{-\infty}^{0} x^r f_1\,dx + \int_{0}^{+\infty} x^r f_2\,dx \\
&= (-1)^r \int_0^{+\infty} x^r f_1\,dx + \int_0^{+\infty} x^r f_2\,dx \\
&= (-1)^r C \int_0^{+\infty} x^r (1 + (k/(n-2))\theta^{-k}(1-\lambda)^{-k}|x/\sigma|^k)^{-(n+1)/k}\,dx \\
&\quad + C \int_0^{+\infty} x^r (1 + (k/(n-2))\theta^{-k}(1+\lambda)^{-k}|x/\sigma|^k)^{-(n+1)/k}\,dx.
\end{aligned}
$$

Gradshteyn and Ryzhik (1994, p. 341, §3.241.4) show that

$$
\int_0^{+\infty} x^r (1 + qx^k)^{-m}\,dx = k^{-1} q^{-(r+1)/k} \Gamma((r+1)/k) \times \Gamma(m - (r+1)/k) \Gamma(m)^{-1},
$$

where $0 < (r+1)/k < q$ and $n \neq 0$. Letting $m = (n+1)/k$, $q = (k/(n-2))\theta^{-k}(1-\lambda)^{-k}\sigma^{-k}$, for $x < 0$, or $q = (k/(n-2))\theta^{-k}(1+\lambda)^{-k}\sigma^{-k}$, for $x > 0$, and substituting the above integral into M_r gives

$$M_r = (-1)^r Ck^{-1}((n-2)/k)^{(r+1)/k}\theta^{r+1}(1-\lambda)^{r+1}\sigma^{r+1}\Gamma((r+1)/k)$$
$$\times \Gamma((n-r)/k)\Gamma((n+1)/k)^{-1}$$
$$+ Ck^{-1}((n-2)/k)^{(r+1)/k}\theta^{r+1}(1+\lambda)^{r+1}\sigma^{r+1}\Gamma((r+1)/k)$$
$$\times \Gamma((n-r)/k)\Gamma((n+1)/k)^{-1}$$
$$= C((-1)^r(1-\lambda)^{(r+1)} + (1+\lambda)^{(r+1)})k^{-1}((n-2)/k)^{(r+1)/k}B$$
$$\times ((r+1)/k, (n-r)/k)\theta^{r+1}\sigma^{r+1}, \quad \text{(A1)}$$

where $B((r+1)/k, (n-r)/k) = \Gamma((r+1)/k)\Gamma((n-r)/k)\Gamma((n+1)/k)^{-1}$ is the beta function. For $f(x)$ to be a proper probability density function,

$$M_0 = C((1-\lambda) + (1+\lambda))k^{-1}((n-2)/k)^{1/k}B(1/k, n/k)\theta\sigma$$
$$= 2Ck^{-1}((n-2)/k)^{1/k}B(1/k, n/k)\theta\sigma = 1,$$

thus,

$$C = 0.5k(k/(n-2))^{1/k}B(1/k, n/k)^{-1}\theta^{-1}\sigma^{-1}. \quad \text{(A2)}$$

Substituting C into equations (A1) gives

$$M_r = 0.5((-1)^r(1-\lambda)^{(r+1)} + (1+\lambda)^{(r+1)})((n-2)/k)^{r/k}$$
$$\times B((r+1)/k, (n-r)/k)B(1/k, n/k)^{-1}\theta^r\sigma^r. \quad \text{(A3)}$$

The expected value of x is

$$\mu = M_1 = 0.5(-(1-\lambda)^2$$
$$+ (1+\lambda)^2)B(2/k, (n-1)/k)B(1/k, n/k)^{-1}\theta\sigma$$
$$= 2\lambda((n-2)/k)^{1/k}B(2/k, (n-1)/k)B(1/k, n/k)^{-1}\theta\sigma. \quad \text{(A4)}$$

The expected value of x^2 is

$$E(x^2) = M_2 = 0.5((1-\lambda)^3 + (1+\lambda)^3)$$
$$\times ((n-2)/k)^{2/k}B(3/k, (n-2)/k)B(1/k, n/k)^{-1}\theta^2\sigma^2$$
$$= (1+3\lambda^2)((n-2)/k)^{2/k}B(3/k, (n-2)/k)B(1/k, n/k)^{-1}\theta^2\sigma^2. \quad \text{(A5)}$$

The variance of x is

$$\mathrm{Var}(x) = E(x^2) - \mu^2$$
$$= (1+3\lambda^2)((n-2)/k)^{2/k}B(3/k, (n-2)/k)B(1/k, n/k)^{-1}\theta^2\sigma^2$$
$$- (2\lambda((n-2)/k)^{1/k}B(2/k, (n-1)/k)B(1/k, n/k)^{-1}\theta\sigma)^2$$
$$= (1 + 3\lambda^2 - 4\lambda^2 B(2/k, (n-1)/k)^2$$
$$\times B(1/k, n/k)^{-1}B(3/k, (n-2)/k)^{-1})$$
$$\times ((n-2)/k)^{2/k}B(3/k, (n-2)/k)B(1/k, n/k)^{-1}\theta^2\sigma^2$$
$$= S(\lambda)^2((n-2)/k)^{2/k}B(3/k, (n-2)/k)B(1/k, n/k)^{-1}\theta^2\sigma^2$$
$$= \sigma^2, \quad \text{(A6)}$$

where $S(\lambda) \equiv (1 + 3\lambda^2 - 4\lambda^2 B(2/k, (n-1)/k)^2 B(1/k, n/k)^{-1}B(3/k, (n-2)/k)^{-1})^{1/2}$. Therefore,

$$\theta = (k/(n-2))^{1/k}B(1/k, n/k)^{1/2}B(3/k, (n-2)/k)^{-1/2}S(\lambda)^{-1}.$$

Substituting θ into equations (A3) gives equation

$$M_r = 0.5((-1)^r(1-\lambda)^{(r+1)} + (1+\lambda)^{(r+1)})$$
$$\times ((n-2)/k)^{r/k}B((r+1)/k, (n-r)/k)B(1/k, n/k)^{-1}$$
$$\times (k/(n-2))^{r/k}B(1/k, n/k)^{r/2}B(3/k, (n-2)/k)^{-r/2}S(\lambda)^{-r}\sigma^r$$
$$= 0.5((-1)^r(1-\lambda)^{(r+1)} + (1+\lambda)^{(r+1)})S(\lambda)^{-r} \times B((r+1)/k,$$
$$\times (n-r)/k)B(1/k, n/k)^{-1+(r/2)}B(3/k, (n-2)/k)^{-r/2}\sigma^r. \quad \text{(A7)}$$

The expected of $|x|$ is

$$E(|x|) = \int_{-\infty}^{+\infty} |x| f(x)\,dx = \int_{-\infty}^{0} |x| f_1\,dx + \int_{0}^{+\infty} |x| f_2\,dx$$
$$= \int_0^{+\infty} x f_1\,dx + \int_0^{+\infty} x f_2\,dx$$
$$= C\int_0^{+\infty} x(1 + (k/(n-2))\theta^{-k}(1-\lambda)^{-k}|x/\sigma|^k)^{-(n+1)/k}\,dx$$
$$+ C\int_0^{+\infty} x(1 + (k/(n-2))\theta^{-k}(1+\lambda)^{-k}|x/\sigma|^k)^{-(n+1)/k}\,dx.$$

Using Gradshteyn and Ryzhik's (1994) integral and the results of equations (2)–(4), it can easily be shown that

$$E(|x|) = 0.5((1-\lambda)^2 + (1+\lambda)^2)((n-2)/k)^{1/k}$$
$$\times B(2/k, (n-1)/k)B(1/k, n/k)^{-1}$$
$$\times ((n-2)/k)^{-1/k}B(1/k, n/k)^{1/2}B(3/k, (n-2)/k)^{-1/2}S(\lambda)^{-1}\sigma$$
$$= (1+\lambda^2)S(\lambda)^{-1}B(2/k, (n-1)/k)$$
$$\times B(1/k, n/k)^{-1/2}B(3/k, (n-2)/k)^{-1/2}\sigma.$$

II. Proof of Theorem 2

For $r = 0$, equation (5) gives $M_0 = \int_{-\infty}^0 f_1\,dx + \int_0^{+\infty} f_2\,dx = (1-\lambda)/2 + (1+\lambda)/2 = 1$. It can easily be verified from the proof of equation (5) that $\int_{-\infty}^0 f_1\,dx = (1-\lambda)/2$ and $\int_0^{+\infty} f_2\,dx = (1+\lambda)/2$.

III. Proof of Theorem 3

This follows easily by direct substitution of $x = z + \mu$ into equation (1) and the fact that the Jacobean of this transformation is equal to one.

IV. Proof of Theorem 4

Using Theorem 2 and the fact that $\mu > 0$ for $\lambda > 0$ (see equation (7)),

$$E(|z|) = E|x| + \mu\int_{-\infty}^0 f_1\,dx - \mu\int_0^{+\infty} f_2\,dx - 2\int_0^{\mu} x f_2\,dx + 2\mu\int_0^{\mu} f_2\,dx$$
$$= E|x| + \mu(1-\lambda)/2 - \mu(1+\lambda)/2 - 2\int_0^{\mu} x f_2\,dx + 2\mu\int_0^{\mu} f_2\,dx$$
$$= E|x| - \lambda\mu - 2\int_0^{\mu} x f_2\,dx + 2\mu\int_0^{\mu} f_2\,dx.$$

For $\lambda < 0 \Rightarrow \mu < 0$ and

$$E|z| = E|x| + \mu\int_{-\infty}^{0} f_1\,dx - \mu\int_{0}^{+\infty} f_2\,dx + 2\int_{\mu}^{0} xf_1\,dx - 2\mu\int_{\mu}^{0} f_1\,dx$$

$$= E|x| + \mu(1-\lambda)/2 - \mu(1+\lambda)/2 + 2\int_{\mu}^{0} xf_1\,dx - 2\mu\int_{\mu}^{0} f_1\,dx$$

$$= E|x| - \lambda\mu + 2\int_{\mu}^{0} xf_1\,dx - 2\mu\int_{\mu}^{0} f_1\,dx$$

The proof for equations (18)–(21) follows easily from the definition of z.

References

Akgiray, V., G. G. Booth 1988. The stable-law of stock returns. *J. Business Econom. Statist.* **6** 51–57.

——, ——, J. J. Hatem, C. Mustafa 1991. Conditional dependence in precious metal prices. *Financial Rev.* **26** 367–386.

Baillie, R. T., T. Bollerslev 1989. The message in daily exchange rates: a conditional variance tale. *J. Business Econom. Statist.* **7** 297–305.

Berndt, E. K., H. B. Hall, R. E. Hall, J. A. Hausman 1974. Estimation and inference in nonlinear structural models. *Ann. Econom. Social Measurement* **4** 653–666.

Bollerslev, T. 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. *Rev. Econom. Statist.* **69** 542–547.

——, R. F. Engle, D. B. Nelson 1994. ARCH models. *Handbook of Econometrics*. Vol. 4. R. F. Engle and D. L. McFadden, eds. Elsevier Science, Chapter 49, 2959–3038.

Box, G. E. P., G. C. Tiao 1962. A further look at robustness via Bayes's Theorem. *Biometrika* **49** 419–432.

Butler, R. J., J. B. Mcdonald, R. D. Nelson, S. B. White 1990. Robust and partially adaptive estimation of regression models. *Rev. Econom. Statist.* 321–327.

Corrado, C., T. Su 1995. Skewness and kurtosis in the S&P 500 index returns implied by option prices. *J. Financial Res.* **19** (2) 175–192.

Dickey D. A., W. A. Fuller 1979. Distribution of the estimates for autoregressive time series with unit root. *J. Amer. Statist. Assoc.* **74** June. 427–432.

——, —— 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** July. 1057–1072.

Engle, R. F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50** 987–1008.

——, G. Gonzalez-Rivera. 1991. Semiparametric ARCH models. *J. Business Econom. Statist.* **9** 345–359.

Gradshteyn, I. S., I. M. Ryzhik. 1994. *Table of Integrals, Series*, and *Products*. 5th edition. Alan Jeffrey, ed. Academic Press, New York.

Hansen, B. E. 1994. Autoregressive conditional density estimation. *Internat. Econom. Rev.* **35** (3) 705–730.

Hsieh, D. 1989. Modeling heteroskedasticity in daily foreign-exchange rates. *J. Business Econom. Statist.* **7** 307–317.

Johnson, N. L., S. Kotz, N. Balakrishnan 1995. *Continuous Univariate Distributions*. Volume 2, 2nd Ed. John Wiley, New York.

Jones, P. N., G. J. Mclachlan 1990. Laplace-normal mixtures fitted to wind shear data. *J. Appl. Statist.* **17** 271–276.

Koutmos, G., P. Theodossiou 1994. Time-series properties and predictability of exchange rates. *Managerial and Decision Econom.* **15** 159–167.

Lye, J. N., V. L. Martin 1993. Robust estimation, nonnormalities, and generalized exponential distributions. *J. Amer. Statist. Assoc.* **88** 261–267.

——, —— 1994. Non-linear time series modeling and distributional flexibility. *J. Time Series Anal.* **15** 65–84.

Markowitz, H. M., N. Usmen 1996. The likelihood of various stock market return distributions, part 2: empirical results. *J. Risk and Uncertainty* **13** 221–247.

Mcdonald, J. B., W. K. Newey 1988. Partially adaptive estimation of regression models via the generalized T distribution. *Econom. Theory* **4** 428–457.

——, R. D. Nelson 1989. Alternative beta estimation for the market model using partially adaptive techniques. *Comm. Statist.: Theory and Methods* **18** 4039–4058.

——, Y. J. Xu 1995. A generalization of the beta distribution with applications. *J. Econom.* **66** 133–152.

Miller, L. H. 1956. Table of percentage points of Kolmogorov statistics. *J. Amer. Statist. Assoc.* **51** 111–121.

Nelson, D. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59** 347–370.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery 1992. *Numerical recipes in C*, 2nd Edition. Cambridge University Press, New York.

Subbotin, M. T. 1923. On the law of frequency of error. *Math. Sbornik* **31** 296–301.

Tapia, R. A., J. R. Thompson 1978. *Nonparametric Probability Density Estimation*. Johns Hopkins University Press, Baltimore, MD.

Theodossiou, P. 1994. The stochastic properties of major Canadian exchange rates. *Financial Rev.* **29** (2) 193–221.

Accepted by Robert Heinkel; received June 6, 1996. This paper has been with the author for 2 revisions.