

Some Flexible Parametric Models for Partially Adaptive Estimators of Econometric Models

Christian B. Hansen

Graduate School of Business at the University of Chicago

James B. McDonald

Department of Economics at Brigham Young University

Panayiotis Theodossiou

School of Business at Rutgers University

Abstract:

This paper provides a survey of three families of flexible parametric probability density functions (the skewed generalized t, the exponential generalized beta of the second kind, and the inverse hyperbolic sine distributions) which can be used in modeling a wide variety of econometric problems. A figure, which can facilitate model selection, summarizing the admissible combinations of skewness and kurtosis spanned by the three distributional families is included. Applications of these families to estimating regression models demonstrate that they may exhibit significant efficiency gains relative to conventional regression procedures, such as ordinary least squares estimation, when modeling non-normal errors with skewness and/or leptokurtosis, without suffering large efficiency losses when errors are normally distributed. A second example illustrates the application of flexible parametric density functions as conditional distributions in a GARCH formulation of the distribution of returns on the S&P500. The skewed generalized t can be an important model for econometric analysis.

JEL: C13, C14, C15

Keywords: Partially Adaptive Estimation, Econometric Models

www.economics-ejournal.org/economics/journalarticles

1. Introduction

Assumptions about the distributions of economic variables are useful for much of economic modeling; however, it is important that the assumed models are consistent with the stylized facts. For example, selecting a normal distribution permits modeling two data characteristics—the mean and variance, but is not appropriate for data which are skewed or have thick tails. Similarly the use of other distributions, such as the lognormal or Weibull distributions, is restricted to applications with admissible data characteristics. Efforts to model more diverse data characteristics have led to a rapid development of alternative methodological approaches in economics. Semiparametric procedures provide one approach which reduces the structure imposed in the modeling process. Because semiparametric procedures impose relatively little structure on the data, they have desirable large sample properties under quite general conditions. However, in specific applications, the use of semiparametric procedures requires the specification of user specified objects, such as a kernel and window width in kernel regression, and since little structure is assumed, the resulting models may not be parsimonious. In addition, if the assumed structure in a parametric model is approximately correct, the resulting estimator will typically have superior properties to a semiparametric estimator. Pagan and Ullah (1999) provide an excellent summary of these and related issues.

In this paper, we explore an intermediate position between the specification of a simple parametric form for the probability density function and semi-parametric estimation. This approach is based on “flexible” parametric density functions that involve few parameters but can accommodate a wider range of data characteristics than are available with such commonly used distributions as the normal, lognormal, or the student t distribution. Section 2 summarizes three alternative families of flexible probability density functions, some basic distributional characteristics, important special and limiting cases, and a visual representation of skewness and kurtosis combinations spanned by the respective distributions. Section 3 considers two applications of these distributions in economics: quasi maximum likelihood estimation of regression models and GARCH modeling. Concluding remarks are offered in Section 4.

2. Alternative Models

The normal and Laplace distributions are two of the first probability density functions to have been considered for model building in economics and statistics. They are both symmetric and have kurtosis of 3 and 6 respectively and provide good models for many economic series, with the Laplace being able to model thicker-tailed distributions than the normal. However, it is not uncommon to encounter data which is both skewed and heavy tailed in economics and finance applications. In the following, we summarize three alternative families of distributions that may be used as models for possibly skewed and thick-tailed distributions.

2.1 Skewed Generalized T distribution (SGT)

The skewed generalized t distribution (SGT) was obtained by Theodossiou (1998) and can be defined by

$$SGT(y; m, \lambda, \phi, p, q) = \frac{p}{\left[2\phi q^{1/p} B(1/p, q) \left(1 + \frac{|y-m|^p}{q(1 + \text{sign}(y-m)\lambda)^p \phi^p} \right)^{q+1/p} \right]}$$

where $B(\cdot, \cdot)$ is the beta function, m is the mode of y and the parameters p and q are both positive and control the height and tails of the density. The parameter ϕ is a positive scale parameter and λ , $-1 < \lambda < 1$, determines degree of skewness with the area to the left of the mode equal to $(1-\lambda)/2$. Setting $\lambda = 0$ in the SGT yields the generalized t (GT) of McDonald and Newey (1988). Similarly, setting $p=2$ yields the skewed t (ST) of Hansen (1994) which includes the student t distribution when $\lambda = 0$. The h^{th} order moments of $(Y-m)$ can be shown to be given by

$$\left(\frac{\phi^h}{2} \right) \left(\frac{q^{h/p} B\left(\frac{h+1}{p}, q - \frac{h}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right) \left((1+\lambda)^{h+1} + (-1)^h (1-\lambda)^{h+1} \right)$$

for $h < pq$ ¹; hence, the SGT defines moments of order less than the degrees of freedom ($n = pq$).

¹ The results for the h^{th} order moments about the mode can be used to derive expressions for moments about the mean using the binomial expansion.

The standardized values for skewness and kurtosis² in the ranges $(-\infty, \infty)$ and $(1.8, \infty)$, respectively, can be modeled with the SGT. Thus, the SGT allows for significantly more flexibility in modeling skewness and kurtosis than the student t distribution which is symmetric and has kurtosis $3 + 6/(n - 4)$ for degrees of freedom exceeding four.

Another important class of flexible density functions corresponds to a limiting case of the SGT. Letting $q \rightarrow \infty$ yields the skewed generalized error distribution (SGED) defined by

$$SGED(y; m, \lambda, \phi, p) = \frac{p e^{-\left(\frac{|y-m|^p}{((1+\text{sign}(y-m)\lambda)^p \phi^p)}\right)}}{2\phi \Gamma(1/p)}.$$

The parameter p in the SGED controls the height and tails of the density and λ controls the skewness. The h^{th} order moments of $(Y - m)$ for the SGED can be shown to be given by

$$\left(\frac{\phi^h}{2}\right) \left(\frac{\Gamma\left(\frac{h+1}{p}\right)}{\Gamma\left(\frac{1}{p}\right)}\right) \left((1+\lambda)^{h+1} + (-1)^h (1-\lambda)^{h+1}\right)$$

for all positive values of h . The SGED is symmetric for $\lambda = 0$ and positively (negatively) skewed for positive (negative) values of λ . The symmetric SGED, GED, is also known as the generalized power (Subbotin (1923)) distribution or the Box-Tiao (Box and Tiao (1962)) distribution. The SGED can easily be seen to include the skewed ($\lambda \neq 0$) or symmetric ($\lambda = 0$) Laplace or normal corresponding to $p = 1$ or 2 , respectively. Figure 1 provides a visual summary of the interrelationships between some of the pdf's in the SGT family of distributions where "S" denotes the skewed generalization of the indicated pdf. As the parameter p grows larger, the SGT pdf approaches the uniform pdf, for example SLaplace denotes the skewed Laplace pdf.

² The standardized skewness and kurtosis correspond to $K_h = E(Y - \mu)^h / [E(Y - \mu)^2]^{h/2}$ for $h=3$ and 4 , respectively, where μ denotes the mean of Y .

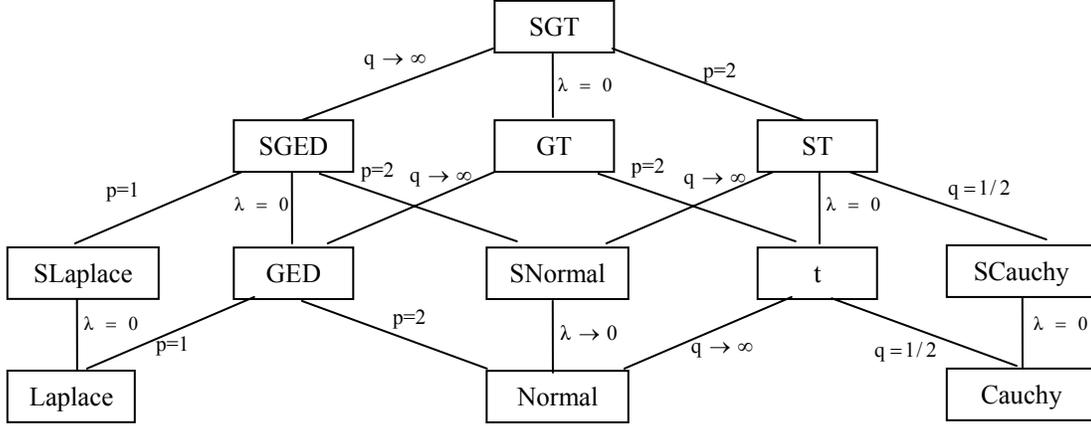


Figure 1: The SGT Distribution Family

2.2 Exponential generalized beta of the second kind (EGB2)

The four parameter EGB2 distribution is defined by the probability density function

$$EGB2(y; m, \phi, p, q) = \frac{e^{p(y-m)/\phi}}{\left[\phi B(p, q) \left(1 + e^{(y-m)/\phi} \right)^{p+q} \right]}$$

where the parameters ϕ , p , and q are assumed to be positive, cf. McDonald and Xu (1995). m and ϕ are respectively location and scale parameters. The parameters p and q are shape parameters. The EGB2 pdf is symmetric if and only if p and q are equal. The normal distribution is a limiting case of the EGB2 where the parameters p and q are equal and grow indefinitely large. The moment generating function for the EGB2 is

$$M_{EGB2}(t) = \frac{e^{\delta t} B(p + t\sigma, q - t\sigma)}{B(p, q)}$$

from which the first four moments can be readily derived as

$$\text{Mean: } \delta + \sigma [\psi(p) - \psi(q)]$$

$$\text{Variance: } \sigma^2 [\psi'(p) + \psi'(q)]$$

$$\text{Skewness: } \sigma^3 [\psi''(p) - \psi''(q)]$$

$$\text{Kurtosis: } \sigma^4 [\psi'''(p) + \psi'''(q)]$$

where $\psi(\cdot)$ denotes the derivative of the log of the gamma function. The EGB2 may accommodate standardized values for skewness in the range (-2.0, 2.0) and standardized values of kurtosis in the range (3.0, 9.0).

2.3 Inverse hyperbolic sine (IHS)

Johnson (1949, 1994) proposed three families of distributions of random variables that are transformations of normal variables. These transformations allow modeling a wide range of values of skewness and kurtosis. We consider the inverse hyperbolic sine (IHS) transformation which allows unbounded random variables. For this paper we use a slightly different parameterization than used by Johnson (1949). Specifically, we consider $y = a + b \sinh(\lambda + z/k) = a + bw$ where \sinh is the hyperbolic sine, z is a standard normal, and a , b , λ , and k are scaling constants related respectively to the mean (μ), variance (σ^2), skewness, and kurtosis of the random variable Y . The pdf of y is given by

$$IHS(y; \mu, \sigma, k, \lambda) = \frac{ke^{-\left(\frac{k^2}{2} \left(\ln(u/\sigma + \sqrt{\theta^2 + u^2/\sigma^2}) - (\lambda + \ln \theta) \right)^2\right)}}{\sqrt{2\pi(\theta^2 + u^2/\sigma^2)}\sigma^2}$$

where

$u = y - \mu + \delta\sigma$, $\theta = 1/\sigma_w$, $\delta = \mu_w/\sigma_w$, $\mu_w = .5(e^\lambda - e^{-\lambda})e^{5k^{-2}}$, $\sigma_w = .5(e^{2\lambda+k^{-2}} + e^{-2\lambda+k^{-2}} + 2)^{.5}(e^{k^{-2}} - 1)^{.5}$, and μ_w and σ_w denote the mean and standard deviation of $w = \sinh(\lambda + z/k)$.³ Negative values of λ generate negative skewness, positive values of λ generate positive skewness, and zero corresponds to symmetry. Smaller values of k result in more leptokurtic distributions. Given that the mean and variance for the IHS are given by mean μ and σ^2 , respectively, the standardized skewness and kurtosis are

$$SK = A_3 - 3\delta - 3\delta^3$$

$$\text{KU} = \frac{\theta^4}{16} \left(1 - 4e^{-2|\lambda| - 6k^{-2}} + 6e^{-4|\lambda| - 8k^{-2}} - 4e^{-6|\lambda| - 6k^{-2}} + e^{-8|\lambda|} \right) e^{4|\lambda| + 8k^{-2}} - 4A_3\delta + 6\delta^2 + 3\delta^4$$

$$\text{where } A_3 = \frac{\theta^3 \text{sign}(\lambda)}{8} \left(1 - 3e^{-2|\lambda| - 4k^{-2}} + 3e^{-4|\lambda| - 4k^{-2}} - e^{-6|\lambda|} \right) e^{3|\lambda| + 4.5k^{-2}}.$$

The IHS allows standardized skewness and kurtosis values in the range $(-\infty, \infty)$ and $(3, \infty)$, respectively. The IHS includes the normal as a limiting case where $k \rightarrow \infty$ with $\lambda = 0$.

2.4 Partitioning the skewness-kurtosis space

While the SGT, EGB2, and IHS are all flexible distributions which can potentially accommodate a wide variety of skewed and leptokurtic data, they do not cover all cases which could arise in practice. To illustrate the data characteristics consistent with each distribution, we plot the admissible skewness-kurtosis combinations in figure 2. The solid and dotted U-shaped curves provide respectively the lower bound for the SGT and IHS permissible combinations of skewness and kurtosis, and the smile-like space provides the lower and upper bounds of the permissible skewness-kurtosis combinations for the EGB2. As might be expected, the SGT clearly admits a larger range of skewness-kurtosis combinations than the other two distributions. However, the coverage of the IHS is remarkably close to that of the SGT; and while the EGB2 space is limited in coverage in comparison to the other two distributions, it does cover many skewness-kurtosis combinations encountered in practice.

³ The mean and variance of y are related to the corresponding moments of w by $\mu = a + b\mu_w$ and $\sigma = b\sigma_w$.

3. Applications

The flexible pdf's summarized in Section 3 have many applications in economic modeling where a normality assumption may be unnecessarily restrictive. We consider two applications in this section: (1) partially adaptive or robust estimation of regression models and (2) estimating models characterized by generalized autoregressive conditional heteroskedasticity (GARCH).

3.1 Regression Models: A Simulation Example

We use a Monte Carlo simulation to illustrate the potential usefulness and efficiency gains available from the application of the flexible distributions discussed above in regression modeling. Following Hsieh and Manski(1984), Newey(1988), McDonald and White(1994), and Ramirez, Misra, and Nelson (2003), we simulate data from the model or data generating process (DGP)

$$Y_t = -1 + X_t + u_t \quad \text{for } t = 1, \dots, T$$

where the X_t 's are drawn from a Bernoulli distribution with $\text{Prob}(X=1) = 0.5$. We consider three different error distributions, each with a zero mean and unit variance. One error distribution is the standard normal, another is a thick tailed variance mixture or contaminated distribution, and the third corresponds to a skewed error distribution.⁴ We consider samples of size of fifty and one hundred with one thousand replications. For each model, we estimate the slope and intercept parameters using ordinary least squares (OLS)⁵ and least absolute deviations (LAD)⁶ as benchmarks and also estimate the parameters⁷ using partially adaptive estimation based on the error distributions summarized in Section 2. Comparing the standardized skewness and kurtosis

⁴ Thus, the first error distribution is merely the unit normal, $Z_1 = N[0,1]$. The thick-tailed variance contaminated distribution is generated as a mixture by $Z_2 = U*N[0, 1/9] + (1-U)*N[0,9]$ where U is 1 with probability .9 and 0 otherwise. Z_2 is symmetrically distributed with kurtosis of 24.3. The skewed distribution is generated by $Z_3 = (Y - e^{-5}) / \sqrt{e(e-1)}$ where Y is $\text{LN}[0,1]$. Z_3 has standardized skewness and kurtosis values of 6.185 and 113.94, respectively.

⁵ The expected standard deviation for the OLS slope estimator for this data generating process is $2/\sqrt{T}$ which is approximately .28 and .20 for $T=50, 100$, respectively.

⁶ The asymptotic variances for the LAD estimator are given by $(X'X)^{-1} / (2f(\text{median}))^2$ which yields slope RMSE's of .354, .130, and .164 for the normal, mixture, and lognormal error distributions with $T = 50$. Corresponding results for $T = 100$ are obtained by multiplying the $T = 50$ results by $\sqrt{(50/100)} = .707$.

⁷ For the possibly asymmetric QMLE the intercept term was adjusted so that the expected error was zero.

(SK,KU) for the normal (0, 3), mixed normal (0, 24.3), and lognormal (6.185, 113.9) error distributions with Figure 2 suggests the IHS and SGT-estimation would be compatible with the data; whereas the EGB2 estimators would not. Ramirez, Misra, and Nelson (2003) and McDonald and White(1994) applied the IHS and GT and EGB2 partially adaptive estimators to this DGP. Adaptive maximum likelihood (AML) estimation, based on a normal kernel, following Hsieh and Manski (1984), and generalized method of moments (GMM) estimation, as outlined in Newey (1988)⁸, are performed to provide additional benchmarks which are independent of the assumption of any particular distribution or family of distributions. In this Monte Carlo study, the performance of the ST, SGED, and SGT are considered and compared with results of previous research.

The root mean squared errors (RMSE) for the estimated slope parameters, using each of the previously mentioned methods, are reported in Table 1. Since each of the flexible pdf's considered includes the normal as a special or limiting case, one would expect partially adaptive estimators to perform similarly to OLS for normally distributed errors, but not necessarily for the mixture or skewed error distributions. This intuition seems to be confirmed based on the results reported in Table 1 where we also observe that there appears to be relatively little efficiency loss for the partially adaptive estimators relative to the OLS estimator for the data generating process with normally distributed error terms. The DGP for each error distribution satisfies the Gauss Markov assumptions with OLS yielding the most efficient linear unbiased estimators, but with the normality assumption OLS is also the most efficient of all unbiased (linear and nonlinear) estimators. The contaminated normal and lognormal error distributions are examples in which the OLS estimator will be the minimum variance unbiased linear estimator, but there will be nonlinear estimators which provide significant improvements in estimator efficiency. Comparing the two panels (T=50 and T=100) in Table 1, the results appear to be generally consistence with \sqrt{T} convergence.

Not surprisingly, OLS has the largest RMSE of any of the estimators considered for the mixture (thick-tailed and symmetric) distribution considered. However, AML, GMM, and the

⁸ The normal kernel estimate of the error density is $f(u) = \left(\frac{1}{sT}\right) \sum_{i=1}^T \phi\left(\frac{u - \hat{u}}{s}\right)$ where ϕ and \hat{u} , respectively,

denote the standard normal density function and the least squares residuals. Bootstrapping was used to select the value of the smoothing parameter (s) to minimize the RMSE of the slope parameter. The GMM estimators are the transformed GMM estimators based minimizing the correlations between $\left[u / (1 + |u|)\right]^j$ and the regressors for moments, $j=1, 2, \dots, J$. $J=4$ was used in this simulation.

partially adaptive estimators (GED, SGED, ST, GT, SGT, EGB2, and IHS) do much better than OLS and have very similar RMSE's.⁹

In the case of the skewed and thick tailed error distribution, OLS again performs the worst for estimating the slope. The partially adaptive estimators offer substantial efficiency gains relative to both OLS and LAD. The AML and GMM estimators perform similarly to the partially adaptive estimators for thick-tailed and symmetric distribution GT; however, the RMSE for the possibly skewed partially adaptive estimators (SGED, ST, SGT, EGB2, and IHS) appear to outperform AML and GMM for both sample sizes. The performance of the EGB2 and IHS for the skewed error distributions is particularly impressive. The strong performance of the EGB2 may be surprising since the moments of the true underlying error distribution do not lie in the portion of the moment space covered by the EGB2 as illustrated in Figure 2. In this sense, it appears that accounting for the some potential skewness and kurtosis may be more important than capturing it exactly when estimating parameters characterizing the mean. Of course, if we were interested in estimating other features of the distribution, we would expect the performance of the EGB2 to deteriorate.

3.2 GARCH Models

Many economic time series involving financial or macroeconomic data are characterized by volatility clustering, that is, the tendency of large residuals (deviations of the series from the mean) to be followed by large residuals and small residuals by small residuals of unpredictable sign. Engle's (1982) Autoregressive Conditional Heteroskedastic (ARCH) models attempt to model this behavior by expressing the conditional variance of the error as a function of the squares of previous errors

$$\text{var}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sigma_t^2 = \omega + \alpha(L) \varepsilon_{t-1}^2,$$

where $\alpha(L)$ denotes a polynomial in the lag operator (L), $\varepsilon_t = y_t - \mu$ is the residual and μ is the mean of the data series y_t . Bollerslev (1986) generalizes Engle's ARCH model with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model defined by

$$\sigma_t^2 = \omega + \alpha(L) \varepsilon_{t-1}^2 + \beta(L) \sigma_{t-1}^2,$$

⁹ Some convergence problems existed for a few of the GED, SGED, and SGT cases; hence, the reported results correspond to constrained optimization with $p > 1$.

where $\beta(L)$ is a polynomial in the lag operator. The GARCH model can often model high order ARCH models with fewer parameters.

ARCH and GARCH models build on the standardized process z_t defined by

$$z_t = \frac{\varepsilon_t}{\sigma_t}$$

which is assumed to have a conditional mean zero and unitary conditional variance. The distribution of z_t determines the conditional distribution of the original series y_t . The most common conditional distribution used in the estimation of GARCH models has been the standard normal.

The standard normal leads to an unconditional distribution with leptokurtosis, but does not capture all the leptokurtosis present in high frequency speculative prices (Bollerslev, Engle, and Nelson 1994, p. 2979). In an attempt to better model leptokurtosis Bollerslev (1987) proposed using the standardized t-distribution. Nelson (1991) used the GED and Bollerslev et al. (1994) applied the GT. It was found that the “generalized t-distribution is a marked improvement over the GED, though perhaps not over the usual student’s t distribution. Nevertheless the generalized t is not entirely adequate, as it does not account for the fairly small skewness in the fitted z_t ’s and also appears to have insufficiently thick tails for the S&P500 data.” Wang et al. (2001) and Ramirez (2001) apply the EGB2 and IHS to GARCH models, respectively.

To illustrate the application of the SGT distribution family in estimating GARCH models, we report the results of fitting representative SGT–GARCH, EGB2-GARCH, and IHS-GARCH specifications¹⁰ to the Standard and Poor’s 500 stock market index (S&P500). The data analyzed were computed using $z_t = 100[\ln(S_t/S_{t-1})]$, where S_t denotes the daily level for each series at time t . The data cover the period from January 1992 through December 2001 (10 years) and Table 2 reports sample characteristics. The Dickey-Fuller tests provide support for the hypothesis that the series of logarithmic changes are stationary. The S&P500 data are characterized by thick tails and negative skewness. Based on the estimated skewness, kurtosis, Jarque-Bera, and chi-square goodness of fit statistics, the assumption of normality is rejected. Comparing the sample skewness and kurtosis values with the feasible combinations in Figure 1 we note that for the unconditional

¹⁰ The conditional distributions are constrained to have a zero mean and unitary variance.

distributions the IHS and SGT are consistent with the data, but the EGB2 is not. This does not imply that the same could be said for the GARCH specifications.

MATLAB was used to obtain partially adaptive or quasi-maximum likelihood estimates (QMLE) of the unknown distributional and GARCH parameters for each of the models indicated in Table 3, both with and without GARCH effects. Given that QMLE was used rather than method of moments, sample and theoretical moments may differ. Since the GARCH formulations provide a statistically significant improvement over a non-GARCH specification only the estimated GARCH formulations, their estimated parameters and robust standard errors, and conditional distributions are reported. The sample standardized skewness and kurtosis for the estimated GARCH residuals and corresponding theoretical skewness and kurtosis are also reported.

The last six rows of Table 3 present goodness-of-fit statistics for the models. Specifically, *LR-GARCH* tests the significance of the GARCH effects in each of the models. It is calculated using the formula $LR-GARCH (= 2(\ell - \ell^*))$, where ℓ and ℓ^* , respectively, denote the optimized log-likelihood values for the GARCH and non-GARCH formulations. This test statistic has an asymptotic chi square (χ^2) distribution with 2 degrees of freedom (d.f.), where the d.f. are equal to the number of additional estimated GARCH parameters of α and β in the GARCH(1,1) specification, which is a common specification in these applications. The LR-Normal statistic is calculated as $2(\ell_A - \ell_N)$ where ℓ_N and ℓ_A represent the log-likelihood values for the GARCH-normal distribution and the alternative GARCH specification.¹¹ Since the normal distribution is a limiting/special case of each of the alternative specifications, a chi-square distribution with m d.f. equal to the number of additional parameters will be used to test the hypothesis. Specifically, for T, GED, EGB2* (symmetric EGB2 with $p = q$), $m=1$; for GT, SGED, EGB2, and IHS, $m=2$; and for SGT, $m=3$.

Since the LR-GARCH statistic is significant at the 5% level for all models, there is strong support for the GARCH specification. Similarly, all of the LR-Normal test statistics are significant at the 5% level; hence, providing the basis for rejecting the normal specification for each model for the S&P500 data. The last row in Table 4 reports LR tests comparing several skewed pdf's with

¹¹ The asymptotic χ^2 may not be appropriate in comparing nested models in which the parameters of the constrained model lie on the boundary of the parameter space, where the asymptotic distribution may be a mixture of χ^2 's.

their corresponding symmetric special cases, GT vs. SGT, T vs. SGT, GED vs. SGED, and the EGB2* vs. EGB2. Statistical significance arises in the case of the EGB2; hence, compelling evidence of asymmetry in the conditional pdf's for the S&P500 is lacking.

The sum of squared errors (SSE), sum of absolute errors (SAE), and chi-squared goodness of fit (χ^2) provide a basis for comparing non-nested specifications. The χ^2 statistic is asymptotically distributed as chi-square with degrees of freedom one less than the difference between the number of groups and estimated parameters.¹² The χ^2 goodness of fit statistic fails to reject the symmetric EGB2 (EGB2*), GT, EGB2, IHS, and SGT at the 5% level of significance. Looking across all specifications, based on the SSE, SAE, and χ^2 criteria, the GT appears to provide the best fit.

In summary, the S&P500 data provides an example of the importance of using parametric specifications which are flexible enough to accommodate observed data characteristics. In each case considered the GARCH effects are statistically significant and yield conditional distributions for the standardized residuals with different characteristics than for the unconditional residuals. That is, the distributions for the standardized residuals exhibit higher (lower) levels of skewness (leptokurtosis) than the non-standardized ones. Nevertheless, the conditional heteroskedasticity itself cannot fully account for the non-normality of the log-return series.

4. Summary and conclusions

This paper has reviewed three families of flexible parametric probability density functions: the skewed generalized t distribution, the exponential generalized beta of the second kind, and the inverse hyperbolic sine distribution. These distributional families include as limiting or special cases many common parametric distributions, including the normal. They allow one to quite flexibly model the first four moments of a distribution while maintaining the parsimony of a completely specified parametric model as is summarized in Figure 2.

These models can provide the basis for partially adaptive or QML estimation of many

¹² The SAE, SSE, and χ^2 are calculated by comparing the observed frequencies (n_i / n) with the predicted frequencies where n_i and n , respectively, denote the number of observations in the i^{th} of g groups and n is the sample size. Twenty equal probability intervals were used in calculating the goodness of fit measures.

economic models. To illustrate the potential usefulness of these methods, a simulation study was performed in which the parameters of a simple linear regression model were estimated using partially adaptive estimators based on the flexible parametric error distributions. In the simulations, partially adaptive estimators were associated with modest efficiency loss with normal errors and demonstrated the potential for significantly improved estimator performance when the error distribution was skewed and thick-tailed. The simulations for the SGED, ST, and SGT add to previous results for the IHS, EGB2, and GT.

The use of the flexible distributions to model GARCH specifications was illustrated using S&P500 stock return data. For this data set, the SGT, IHS, and EGB2-GARCH specifications provided statistically significant improvements relative to normal-GARCH specifications and with respect to non-GARCH specifications.

References

- Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-321.
- Bollerslev, T. 1987. A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics* 69, 542-547.
- Bollerslev, T., R. F. Engle, and D. B. Nelson, 1994. *ARCH Models*. Handbook of Econometrics, Vol. IV. Edited by R. F. Engle and D. L. McFadden, Amsterdam: Elsevier Science B. V., 2959-3038.
- Box, G. E. P. and Tiao, G. C., 1962. A Further Look at Robustness Via Bayes Theorem. *Biometrika* 49, 419-432. [Further information](#)
- Cox, D. R. and D. V. Hinkley, 1974. *Theoretical Statistics*, London: Chapman and Hall.
- Engle, R. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 987-1008.
- Hansen, B. E., 1994. Autoregressive Conditional Density Estimation. *International Economic Review* 35 (3), 705-730. [Further information in IDEAS/RePEc](#)
- Hsieh, D. A. and Manski, C. F., 1987. Monte Carlo Evidence on Adaptive Maximum Likelihood Estimation of a Regression. *Annals of Statistics* 15, 541-551.
- Johnson, N. L., 1949. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika* 36, 149-176. [Further information](#)
- Johnson, N. L., S. Kotz, S. Balakrishnan, N., 1994. *Continuous Univariate Distributions*, Volume 1, Second Edition. New York: John Wiley & Sons, New York. [Further information](#)
- McDonald, J. B. and Newey, W. K., 1988. Partially Adaptive Estimation of Regression Models Via the Generalized t Distribution. *Econometric Theory* 4, 428-457.
- McDonald, J. B. and White, S. B., 1993. A Comparison of Some Robust, Adaptive, and Partially Adaptive Estimators of Regression Models. *Econometric Reviews* 12 (1), 103-124.
- McDonald, J. B. and Xu, Y.J., 1995. A Generalization of the Beta Distribution with Applications," *Journal of Econometrics* 66, 133-152. Errata 69(1995), 427-428. [Further information in IDEAS/RePEc](#)
- Nelson, D. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347-370.

Newey, W. K., 1988. Adaptive Estimation of Regression Models Via Moment Restrictions. *Journal of Econometrics* 38, 301-339. [Further information in IDEAS/RePEc](#)

Pagan, A. and A. Ullah, 1999. *Nonparametric Econometrics*. Cambridge University Press, Cambridge. [Further information](#)

Ramirez, O. A. 2001. Autoregressive Conditional Heteroskedasticity under Error Term Non-Normality, working paper, Texas Tech University, Lubbock Texas.

Ramirez, O. A., Misra, S. K., and Nelson, J., 2003. Efficient Estimation of Agricultural Time Series Models with Nonnormal Dependent Variables. *American Journal of Agricultural Economics* 85 (4), 1029-1040. [Further information in IDEAS/RePEc](#)

Subbotin, M. T., 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. [Further information](#)

Wang, K. L., C. Fawson, C. Barrett, and J. B. McDonald 2001. A Flexible Parametric GARCH Model with an Application to Exchange Rates. *Journal of Applied Econometrics* 16, 521-526.

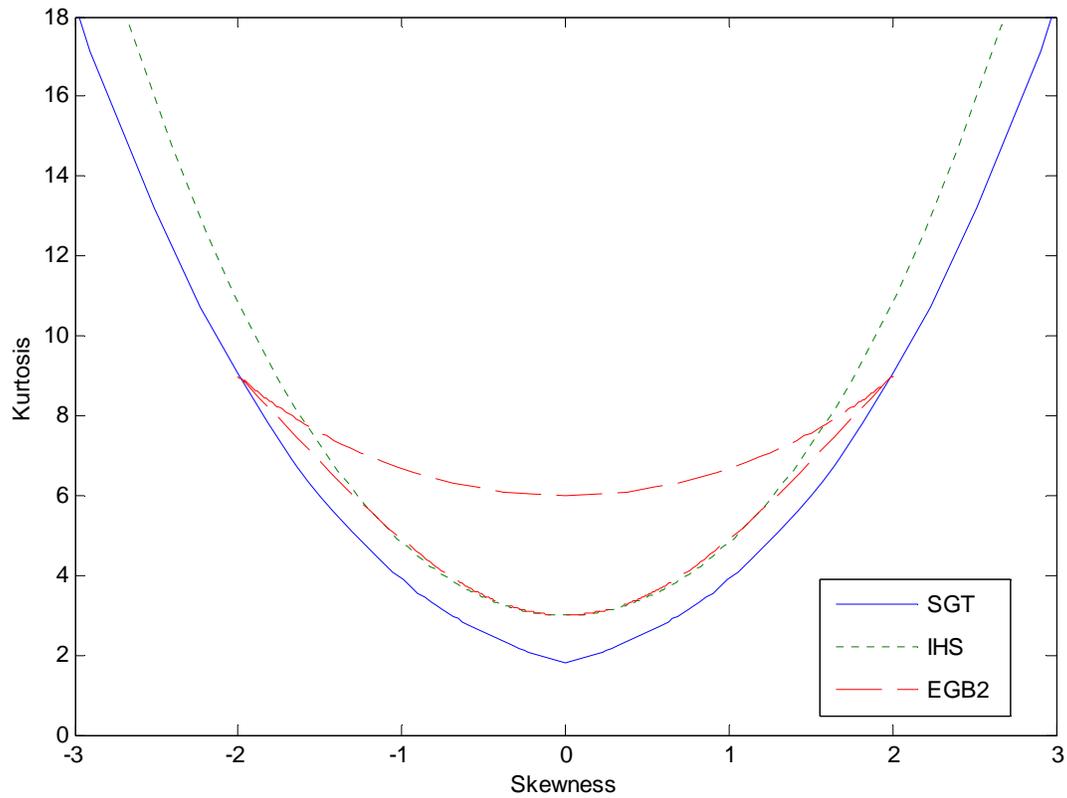


Figure 2. Admissible combinations of skewness and kurtosis

This figure illustrates the admissible combinations of the standardized values of skewness and kurtosis for the SGT, IHS, and EGB2. The regions for the SGT and IHS are given by the areas above corresponding curves (solid line for SGT, - for IHS). For the EGB2 the set of admissible values is given by the area within the dashed curve. The horizontal axis corresponds to values for skewness, and the vertical axis corresponds to values for kurtosis.

Table 1. RMSE Slope Simulation Results

	T=50			T=100		
	Normal	Mixture	Skewed	Normal	Mixture	Skewed
OLS	.275	.287	.280	.201	.210	.201
LAD	.332	.122	.159	.248	.091	.117
GED	.301	.122	.158	.207	.091	.118
SGED	.335	.128	.060	.213	.090	.038
ST	.293	.112	.054	.204	.080	.037
GT	.314	.133	.135	.211	.085	.085
SGT	.335	.125	.073	.216	.085	.041
EGB2	.287	.125	.049	.203	.088	.033
IHS	.285	.119	.054	.202	.085	.034
AML	.285	.114	.128	.201	.081	.087
GMM	.319	.115	.088	.209	.082	.060

This table gives root mean squared errors for estimates of the slope parameters from the simulation example defined in Section 3.1. Columns labeled normal are results for the model where the errors are drawn from a standard normal distribution. Columns labeled mixture have errors drawn from a symmetric, leptokurtotic mixture of normals. Columns labeled skewed have errors drawn from a LN(0,1) distribution which are scaled and centered to have mean zero and variance one. Each row corresponds to a different estimation procedure as discussed in the text.

Table 2. S&P 500 Data Characteristics

Dates	1/2/1992 – 12/31/2001
Sample size	2523
Dickey-Fuller F-value for $\ln(S_t)$.703
Dickey-Fuller F-value for S_t	28.23*
SK-returns	-.294*
	(0.049)
KU-returns	7.823*
	(0.097)
Jarque-Bera	2469.6*
χ^2 goodness of fit	334.3*

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 with a trend. The Dickey-Fuller critical values at the five and one-percent levels of significance are 6.25 and 8.27, respectively. SK-returns and KU-returns are the sample skewness and kurtosis for the log-changes computed using the formulas $b_1 = m_3 / m_2^{3/2}$ and $b_2 = (m_4 / m_2^2)$, where m_i denotes the sample estimate of the i^{th} moment about the mean. The Jarque-Bera (JB) statistic is used for testing for normality and is asymptotically distributed as $\chi^2(2)$. The chi-square goodness of fit statistic also tests for

normality and is asymptotically distributed as $\chi^2(57)$. The five percent critical values for the JB and chi-square goodness of fit tests are 5.99 and 75.6, respectively

Table 3. Parameter Estimates for the S&P500 Stock Market Log-returns

Models	4 parameters		5 parameters			6 parameters			7 parameters
	NORMAL	T	GED	$EGB2^*$	GT	$SGED$	$EGB2$	IHS	SGT
μ	0.05551* (0.01458)	0.06046* (0.01349)	0.05268* (0.01481)	0.05825* (0.01347)	0.05812* (0.01473)	0.04334* (0.01473)	0.04705* (0.01445)	0.04906* (0.01442)	0.05053* (0.01576)
ω	0.00611* (0.00216)	0.00365* (0.00149)	0.00425* (0.00165)	0.00372* (0.00153)	0.00377* (0.00151)	0.00452* (0.00166)	0.00414* (0.00160)	0.00406* (0.00155)	0.00402* (0.00155)
α	0.07109* (0.00607)	0.05430* (0.00404)	0.05941* (0.00431)	0.05551* (0.00395)	0.05524* (0.00436)	0.06057* (0.00432)	0.05730* (0.00602)	0.05636* (0.00402)	0.05646* (0.00434)
β	0.92532* (0.00076)	0.94412* (0.00032)	0.93798* (0.00048)	0.94243* (0.00038)	0.94270* (0.00037)	0.93620* (0.00058)	0.93975* (0.00233)	0.94097* (0.00042)	0.94095* (0.00040)
$n=pq$	–	6.16687* (0.90891)	–	–	8.43547 (7.90869)	–	–	–	8.88999 (8.82280)
k or p	–	–	1.34526* (0.06729)	0.69446* (0.15912)	1.75634* (0.45958)	1.36871* (0.07057)	0.71637 (1.01237)	1.85989* (0.18781)	1.75400* (0.46069)
λ or q	–	–	–	–	–	–0.05036 (0.04238)	0.84344 (1.36552)	–0.18637* (0.08517)	–0.04184 (0.02725)
SK -Residuals	–0.48	–0.48	–0.48	–0.48	–0.48	–0.48	–0.48	–0.48	–0.48
SK	0.00	0.00	0.00	0.00	0.00	–0.14	–0.21	–0.22	–0.13
KU -Residuals	5.12	5.22	5.19	5.21	5.21	5.19	5.20	5.21	5.21
KU	3.00	5.77	4.18	4.62	4.99	4.12	4.54	4.94	4.85
ℓ	–3210.89	–3147.58	–3151.96	–3147.78	–3146.63	–3150.12	–3144.98	–3144.80	–3145.46
LR-GARCH	705.19*	444.31*	435.34*	437.23*	433.67*	439.03*	442.43*	441.27*	435.94*
LR-Normal	0.00	126.61*	117.85*	126.22*	128.52*	121.55*	131.81*	132.17*	130.86*
SSE	0.00128	0.00102	0.00049	0.00030	0.00028	0.00043	0.00035	0.00037	0.00033
SAE	0.1359	0.1163	0.0818	0.0662	0.0621	0.0702	0.0685	0.0748	0.0670
χ^2	64.0*	48.6*	24.4*	15.2	14.0	21.8	17.5	18.7	16.7

LR-pdf: LR(GT vs. SGT)= 2.35*, LR(T vs. SGT)= 4.25*, LR($SGED$ vs. GED)= 3.69#, LR($EGB2^*$ vs. $EGB2$)= 5.58

Notes: The estimated GARCH model is $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$, where $\varepsilon_t = y_t - \mu$. Parentheses include robust standard errors for the estimates. SK -Residuals and KU -Residuals are the standardized sample skewness and kurtosis for the GARCH residuals computed as in Table 1. SK and KU are the theoretical skewness and kurtosis computed using the formulas in the Appendix. LR-GARCH is a log-likelihood ratio statistics for testing the significance of the GARCH specifications. It follows the χ^2 distribution with 2 d.f. LR-normal is a log-likelihood ratio statistic for testing the normal against the remaining eight probability specifications. It follows the χ^2 distribution with m d.f. ($m = 1$ for the T , GED and $EGB2^*$, $m = 2$, for the GT , $SGED$, $EGB2$ and IHS , and $m = 3$, for the SGT). The last row presents the LR statistics for testing the T , GT and $SGED$ against the SGT , the GED against the $SGED$, and the $EGB2^*$ against the $EGB2$. SSE is the sum of squared errors, SAE is the sum of absolute errors, and χ^2 is the chi-square goodness of fit. ***Statistically significant at the 5% level.** Critical values for the χ^2 distribution at the 5% are for 3.84, 5.99, 7.81, 21.03, 22.36, 23.68, and 25.00 for 1, 2, 3, 12, 13, 14, and 15 d.f., respectively.