

Discussion Paper 2007-13
March 26, 2007

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

Please cite the corresponding journal article:

<http://www.economics-ejournal.org/economics/journalarticles/2007-7>

Abstract:

This paper discusses three families of flexible parametric probability density functions: the skewed generalized t , the exponential generalized beta of the second kind, and the inverse hyperbolic sin distributions. These families allow quite flexible modeling the first four moments of a distribution and could be considered in modeling a wide variety of economic problems. We illustrate their use in a simple regression model with a simulation study that demonstrates that the use of the flexible distributions may result in significant efficiency gains relative to more conventional regression procedures, such as ordinary least squares or least absolute deviations regression, without a suffering from a large efficiency loss when errors are Gaussian.

JEL: C13, C14, C15

Keywords: Partially Adaptive Estimation, Econometric Models

**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

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 ground 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 defines the alternative probability density functions, discusses important special and limiting cases, and provides a characterization of their moments. Section 3 explores the use of these models in providing a basis for quasi maximum likelihood estimators (QMLE) of the slope and intercept parameters in a simple linear regression model using Monte Carlo simulations. We offer some concluding remarks 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 is 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 parameters p and q control the height and tails of the density. The parameter ϕ is a scale parameter and λ 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 $\lambda = 0$. Standardized values for skewness and kurtosis¹ 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 $3 + 6/(\nu-4)$ where ν ($\nu=pq$) denotes the degrees of freedom parameter. The SGT defines moments of order less than the degrees of freedom.

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

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

¹ The standardized skewness and kurtosis correspond to $K_h = E(Z - \mu)^h / [E(Z - \mu)^2]^{h/2}$ for $h=3$ and 4 , respectively, where μ denotes the mean of Z .

The parameter p in the SGED controls the height and tails of the density and λ controls the skewness. The SGED is symmetric for $\lambda = 0$ and positively (negatively) skewed for positive (negative) values of λ . The symmetric SGED 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.

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 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 sign, 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

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

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 $w = \sinh(\lambda + z/k)$.² 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. The IHS allows skewness and kurtosis in the range $(3, \infty)$ and $(-\infty, \infty)$, 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 1. 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. An Application to Regression Models: A Simulation Example

We provide a simple example that illustrates the potential usefulness 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

² The mean and variance of y are related to the corresponding moments of w by $\mu = a + b\mu_w$ and $\sigma = b\sigma_w$.

data from the model

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

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 unit normal, another is a thick tailed variance mixture or contaminated distribution, and the third corresponds to a skewed error distribution.³ We consider $T = 50$ and use 500 replications for all results. For each model, we estimate the slope and intercept parameters using ordinary least squares (OLS) and least absolute deviations (LAD) as benchmarks and also estimate the parameters⁴ using QMLE based on the error distributions summarized in Section 2.

The root mean squared errors (RMSE) for the estimating the intercept and 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 that the QMLE would 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 observe that there appears to be relatively little efficiency loss for the QML intercept and slope estimators relative to the OLS estimator for the data generating process with normally distributed error terms.

Not surprisingly, OLS has the largest RMSE of any of the estimators considered for the mixture (thick tailed and symmetric) distribution considered. In this case, LAD appears to be the optimal estimator, which is again unsurprising given the symmetry and value of kurtosis in the underlying error distribution. However, as before, the QMLE tend to do quite well, especially for the slope coefficient, where LAD, EGB2, and IHS all have nearly identical RMSE's.

In the case of the skewed and thick tailed error distribution, OLS again performs the worst for estimating the slope, and LAD performs the worst for estimating the intercept. In addition, both are dominated by all of the QMLE for both the slope and intercept. In this case, all of the partially adaptive estimators offer substantial gains relative to OLS or LAD, especially in

³ 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.

estimating the slope coefficient. The performance of the EGB2 and IHS is particularly impressive. The strong performance of the EGB2 is 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 1. In this sense, it appears that accounting for the 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.

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. They allow one to quite flexibly model the first four moments of a distribution while maintaining the parsimony of a completely specified parametric model.

These models can be used as the basis for partially adaptive or QML estimation of many economic models. To illustrate the potential usefulness of these models, we performed a simulation study where we estimated the parameters of a simple linear regression model. In the simulations, we found that the efficiency loss in the standard linear model with normally distributed errors was modest and that the use of the partially adaptive procedures significantly improved estimation performance when the error distribution was skewed or leptokurtotic. The use of the flexible distributions could readily be extended from the simple regression case to modeling univariate time series, for example by ARCH or GARCH, or to other more general settings.

⁴ For the possibly asymmetric QMLE the intercept term was adjusted so that the expected error was zero.

References

- Box, G. E. P. and Tiao, G. C., 1962. A Further Look at Robustness Via Bayes Theorem. *Biometrika* 49, 419-432. [Further information](#)
- 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](#)
- 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., 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, 296301.
- Theodossiou, P., 1998. “Financial Data and the Skewed Generalized t Distribution,” *Management Science* 44, 1650-1661. [Further information](#)

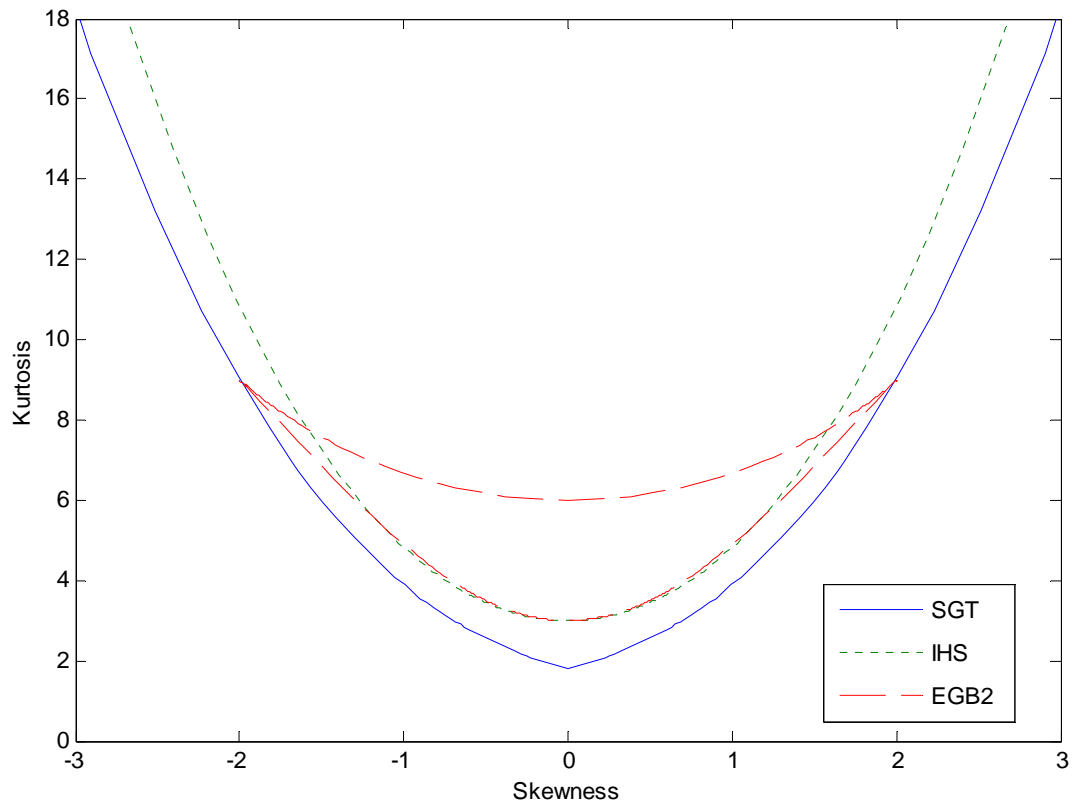


Figure 1. This figure illustrates the admissible skewness and kurtosis combinations for the SGT, IHS, and EGB2. The regions for the SGT and IHS are given by the areas above corresponding curves (solid line for SGT, - for IHS). For the EGB2 the set of admissible values is given by the area within the dashed curve. The horizontal axis corresponds to values for skewness, and the vertical axis corresponds to values for kurtosis.

Table 1. Simulation Results

	Intercept			Slope		
	Normal	Mixture	Skewed	Normal	Mixture	Skewed
OLS	.20	.21	.19	.28	.29	.29
LAD	.24	.10	.30	.33	.13	.17
SGED	.22	.13	.13	.33	.13	.07
SGT	.22	.13	.15	.32	.13	.08
EGB2	.20	.16	.14	.29	.13	.05
IHS	.21	.15	.17	.29	.12	.05

This table gives root mean squared errors for estimates of the slope and intercept parameters from the simulation example in Section 3. 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 and then scaled and centered to have mean zero and variance one. Each row corresponds to a different estimation method as discussed in the text.