ELSEVIER Journal of Econometrics 66 (1995) 133–152

JOURNAL OF Econometrics

A generalization of the beta distribution with applications

James B. McDonald*[,a], Yexiao J. Xu[b]

[a]*Brigham Young University, Provo, UT 84602, USA*
[b]*Princeton University, Princeton, NJ 08544, USA*

(Received January 1993; final version received December 1993)

Abstract

This paper introduces a five-parameter beta distribution (GB) which nests the generalized beta and gamma distributions and includes more than thirty distributions as limiting or special cases. The generalized beta leads to an exponential generalized beta (EGB) distribution which includes generalized forms of the logistics, exponential, Gompertz, and Gumbell distributions, and the normal as special cases. The EGB family of distributions provides the basis for partially adaptive estimation of econometric models with possibly skewed and leptokurtic error distributions. Applications of the models to investigating the distribution of income, stock returns and in regression analysis are considered.

Key words: Generalized beta; Generalized exponential; Generalized gamma; Generalized logistics; Partially adaptive estimation
JEL classification: C13; C19; C29

1. Introduction

Statistical distributions provide the foundation for the analysis of empirical data and for many statistical procedures. Empirical results can be sensitive to the degree to which distributional characteristics such as the mean, variance, skewness, and kurtosis of the data can be modeled by the assumed statistical

* Corresponding author.

The authors appreciate comments received from Barry Arnold, Whitney K. Newey, Rulon Pope, and two referees.

SSDI 0304407694016124

distribution. The beta of the first kind (B1) and second kind (B2), which are very flexible distributions for positive random variables, are two of the most widely used in statistics. The B1 or B2 include the power, uniform, gamma, Lomax, *F*, chi-square, and exponential distributions as special cases (see Johnson and Kotz, 1970; Patil et al., 1984). Various generalizations of the B1 and B2 have been considered. Essentially equivalent generalizations of the B2 have been referred to as the generalized *F* by Kalbfleish and Prentice (1980), the Feller–Pareto by Arnold (1983), the generalized beta prime distribution by Patil et al. (1984), the transformed beta by Venter (1984), and as a generalized beta of the second kind (GB2) by McDonald (1984). In addition to the B2, the GB2 includes two Burr distributions, the generalized gamma, Fisk, Weibull, and lognormal distributions as special cases. A similar generalization of the B1, the generalized beta of the first kind (GB1), also includes the generalized gamma as a limiting case, which includes the gamma, Weibull, lognormal, Rayleigh, exponential, and chi-square. Pham-Gia and Duong (1989) summarize two other generalizations of the B1 and B2 which they refer to as three-parameter generalized beta (G3B) and generalized *F* (G3F) distributions (also see Libby and Novic, 1982). They note that the G3B is a special case of the GB2; however, neither of their generalizations includes both the B1 and B2.

This paper introduces a generalized beta distribution (GB) which nests the B1, B2, GB1, *and* GB2 as well as the generalizations considered by Pham-Gia and Duong. The GB distribution is defined in Section 2, along with a discussion of related properties and special and limiting cases. Section 3 considers the distribution of the logarithm of a generalized beta random variable, an exponential generalized beta (EGB), which can model positive and negative random variables. Special cases of the EGB include the generalized logistics (EGB2), the generalized exponential (EGB1) discussed by Johnson and Kotz (1970, Vol. 2), and generalizations of the Gompertz and Gumbell distributions. The exponential generalized beta of the second kind (EGB2) provides the basis for partially adaptive estimation in regression and time series models to accommodate possibly thick-tailed and skewed error distributions. Some applications of the GB and EGB are considered in Section 4.

2. Generalized beta distributions

We first review the definition of the generalized beta of the first and second kind (GB1 and GB2, respectively) and several important special cases. We then define the generalized beta (GB) family (which includes *both* GB1 and GB2) and give expressions for the defined moments.

The generalized beta of the first kind is defined by the probability density function (pdf):

$$\mathrm{GB1}(y; a, b, p, q) = \frac{|a|\, y^{ap-1}(1-(y/b)^a)^{q-1}}{b^{ap}\,\mathrm{B}(p,q)} \quad \text{for} \quad 0 < y^a < b^a, \tag{2.1}$$

where the parameters b, p, and q are positive. The defined hth-order moments of GB1 random variables are given by

$$\mathrm{E}_{\mathrm{GB1}}(y^h) = \frac{b^h\,\mathrm{B}(p + h/a, q)}{\mathrm{B}(p,q)} \quad \text{for} \quad p + h/a > 0 \tag{2.2}$$

(McDonald, 1984). This four-parameter pdf is very flexible and includes the beta of the first kind (B1), Pareto, generalized gamma (GG), and others as the following special cases:

$$\begin{aligned} \mathrm{B1}(y; b, p, q) &= \mathrm{GB1}(y; a = 1, b, p, q) \\ &= \frac{y^{p-1}(1 - y/b)^{q-1}}{b^p\mathrm{B}(p,q)}, \end{aligned} \tag{2.3}$$

$$\begin{aligned} \mathrm{PARETO}(y; b, p) &= \mathrm{GB1}(y; a = -1, b, p, q = 1) \\ &= \frac{pb^p}{y^{p+1}} \quad \text{for} \quad b < y, \end{aligned} \tag{2.4}$$

$$\begin{aligned} \mathrm{GG}(y; a, \beta, p) &= \lim_{q\to\infty} \mathrm{GB2}(y; a, b = q^{1/a}\beta, p, q) \\ &= \frac{|a|\, y^{ap-1}\, \mathrm{e}^{-(y/\beta)^a}}{\beta^{ap}\,\Gamma(p)}. \end{aligned} \tag{2.5}$$

The standard beta of the first kind corresponds to (2.3) with $b = 1$. The power (P) and uniform (U) distributions are obvious special cases of the GB1. A special case that is not so obvious is the unit gamma (UG):

$$\begin{aligned} \mathrm{UG}(y; b, \delta, q) &= \lim_{a\to 0} \mathrm{GB1}(y; a, b\, p = \delta/a, q) \\ &= \frac{\delta^q y^{\delta-1}}{b^\delta\Gamma(q)}[\ln(b/y)]^{q-1} \quad \text{for} \quad 0 < y < b, \end{aligned} \tag{2.6}$$

and zero otherwise. The UG with $b = 1$ is mentioned in Patil et al. (1984) and has been used as a mixing distribution for the parameter p in the binomial (Grassia, 1977). The authors are not aware of any previous references to the relationship between the UG and the GB1. Additional special cases include the inverse beta of the first kind (IB1) with density given by $\mathrm{IB1}(y; b, p, q) = \mathrm{GB1}(y; a = -1, b, p, q)$, from which it readily follows that the

$\text{PARETO}(y;b,p) = \text{IB1}(y;b,p,q=1)$. The flexibility of the GB1 should be apparent from these examples.

The generalized beta of the second kind is defined by the pdf

$$\text{GB2}(y;a,b,p,q) = \frac{|a|\,y^{ap-1}}{b^{ap}\,\text{B}(p,q)(1+(y/b)^a)^{p-q}} \quad \text{for} \quad 0 < y < \infty, \tag{2.7}$$

and zero otherwise. Variations of the GB2 are also known as the generalized Feller–Pareto by Arnold (1983), the generalized beta prime by Patil et al. (1984), the generalized F by Kalbfleisch and Prentice (1980) and the transformed beta by Venter (1984). Letting $a = 1$ in (2.7) gives the beta of the second kind (B2). The hth-order moments of the GB2 are given by

$$\text{E}_{\text{GB2}}(y^h) = \frac{b^h\,\text{B}(p+h/a, q-h/a)}{\text{B}(p,q)} \quad \text{for} \quad -p < h/a < q. \tag{2.8}$$

The GB2 nests many important distributions as special or limiting cases, including the generalized gamma (GG), Burr types 3 and 12 (BR3 and BR12), lognormal (LN), Weibull (W), gamma (GA), Lomax (L), F statistic (F), Fisk or Rayleigh (R), chi-square (χ^2), half-normal ($\frac{1}{2}$N(0, σ2)), half-Student's $t(\frac{1}{2}t)$, exponential (EXP), and the log-logistic. The details of many of these relationships are summarized in McDonald (1984).

Since neither the GB1 or GB2 (B1 or B2) includes the other as a special case they have often been considered separately, with any comparisons being rather informal. Each of these distributions can be shown to be special cases of a more general distribution, the generalized beta (GB) distribution defined by the pdf

$$\text{GB}(y;a,b,c,p,q) = \frac{|a|\,y^{ap-1}(1-(1-c)(y/b)^a)^{q-1}}{b^{ap}\,\text{B}(p,q)(1+c(y/b)^a)^{p+q}} \quad \text{for } 0 < y^a < b^a/(1-c), \tag{2.9}$$

and zero otherwise with $0 \leqslant c \leqslant 1$, and b, p, and q positive. The moments of (2.9) can be shown (see Section A.1 of the Appendix for the derivation) to be

$$\text{E}_{\text{GB}}(y^h) = \frac{b^h\,\text{B}(p+h/a,q)}{\text{B}(p,q)}\,{}_2F_1\begin{bmatrix} p+h/a, h/a; c \\ p+q+h/a; \end{bmatrix}, \tag{2.10}$$

where ${}_2F_1[\ ]$ denotes the hypergeometric series which converges for all h if $c < 1$ or for $h/a < q$ if $c = 1$ (Rainville, 1960). Substituting $h = 0$ into (2.10) verifies that (2.9) integrates to one.

Comparing (2.9) and (2.10) with (2.1) and (2.2) and (2.7) and (2.8), respectively, we can easily verify that

$$\text{GB1}(y;a,b,p,q) = \text{GB}(y;a,b,c=0,p,q) \tag{2.11}$$

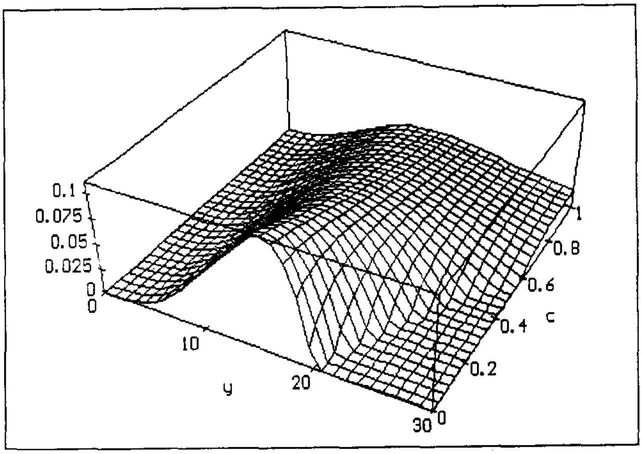

Fig. 1. $\mathrm{GB}(y; a = 2, b = 20, c, p = 2, q = 2]$.

and

$$\mathrm{GB2}(y; a, b, p, q) = \mathrm{GB}(y; a, b, c = 1, p, q), \tag{2.12}$$

i.e., the GB includes the GB1 and GB2 corresponding to $c = 0$ and $c = 1$.

Fig. 1 depicts the shape of the generalized beta density as the parameter c changes for representative values for a, b, and q: $\mathrm{GB}(y;\ a = 2, b = 20, c, p = 2, q = 2)$. For $c = 0$, we observe the density $\mathrm{GB1}(y;\ a = 2,\ b = 20,\ p = 2,\ q = 2)$ and that the relative frequency for values of y greater than 20 is zero. As the value of c increases, the 'upper bound', $b/(1-c)^{1/a}$, increases until for $c = 1$ the associated density is the $\mathrm{GB2}(y;\ a = 2,\ b = 20,\ p = 2,\ q = 2)$. Similar figures could be constructed for changes in the other parameters. The parameter a impacts the peakedness of the density, whereas b is basically a scale parameter, and p and q control the shape and skewness.

A *beta distribution* can be defined in terms of the pdf

$$\begin{aligned} \mathrm{B}(y; b, c, p, q) &= \mathrm{GB}(y; a = 1, b, c, p, q) \\ &= \frac{y^{p-1}(1-(1-c)(y/b))^{q-1}}{b^p \mathrm{B}(p,q)(1+c(y/b))^{p+q}} \quad \text{for} \quad 0 < y < b/(1-c), \end{aligned} \tag{2.13}$$

and zero otherwise. If $b = 1$ in (2.13), the corresponding pdf is

$$\mathrm{GB}(y; a = 1, b = 1, c, p, q) = \frac{y^{p-1}(1-(1-c)y)^{q-1}}{\mathrm{B}(p,q)(1+cy)^{p+q}}, \tag{2.13$'$}$$

and will be referred to the *standard form* of the beta distribution. The beta family includes the B1 and B2 distributions as members corresponding to $c = 0$ and $c = 1$, respectively. The G3B and G3F of Pham-Gia and Duong (1989) are also members of the beta family: $\mathrm{B}(y; b = 1/\lambda, c = 1 - 1/\lambda, p, q)$ and $\mathrm{B}(y; b = 1/\lambda, c = 0, p, q) = \mathrm{B2}(y; b = 1/\lambda, p, q)$, respectively. The GB distribution nests *both* the GB1 and GB2; hence, the B1 and B2 are also special cases of the beta distribution (B) as well as of the generalized beta distribution (GB). The introduction of the GB and B distributions allows a comparison of the GB1 with the GB2 and of the B1 with the B2 distributions based on a nested framework.

Fig. 2 provides a convenient visual summary of some limiting and special cases of the GB and their interrelationships. The GB includes many members of the Pearson family as special cases and some additional important distributions such as the generalized gamma and Burr distributions (types 3 and 12) which are not in the Pearson family. The generalized gamma is a limiting case of the GB1 and GB2, i.e., the GB with $c = 0$ and $c = 1$. It is also interesting to note that

$$\lim_{q \to \infty} \mathrm{GB}(y; a, b = \beta q^{1/a}, c, p, q) = \mathrm{GG}(y; a, \beta, p).$$

Hence, the generalized beta includes the generalized gamma as a limiting case for all admissible values of c. Still another example is obtained by taking the following limit:

$$\begin{aligned} &\lim_{c \to 1} \mathrm{GB}(y; a, b = \beta(1 - c)^{1/a}, c, p = 1/\delta(1 - c), q) \\ &\quad = \frac{|a|}{\delta^q y \Gamma(y)} \left(\frac{\beta}{y}\right)^{q(a-1)+1} \mathrm{e}^{((\beta/y)^{a} - 1)/\delta} \\ &\quad \text{for } 0 < y < \beta. \end{aligned} \tag{2.14}$$

This distribution might be thought of as a translated inverse generalized gamma (TIGG) since $U = [(\beta/Y)^a - 1]$ is distributed as $\mathrm{GA}(U; q, \delta)$. (2.14) is related to the unit gamma (2.6) by

$$\lim_{a \to 0} \mathrm{TIGG}(y; a, \beta, a/\delta, q) = \mathrm{UG}(y; \delta, \beta, q).$$

Fig. 2 not only gives an overview of the relationships between many parametric distributions for positive random variables, but also suggests other distributions that might be considered. For example, the Burr distributions (types 3 and 12) have been widely used because they have closed forms for the cumulative distribution functions. They correspond to the GB with $c = 1$ and p or $q = 1$. It is easily shown that the GB with p or q selected to be 1 has a closed form distribution for any value of c, thus providing another generalization of the

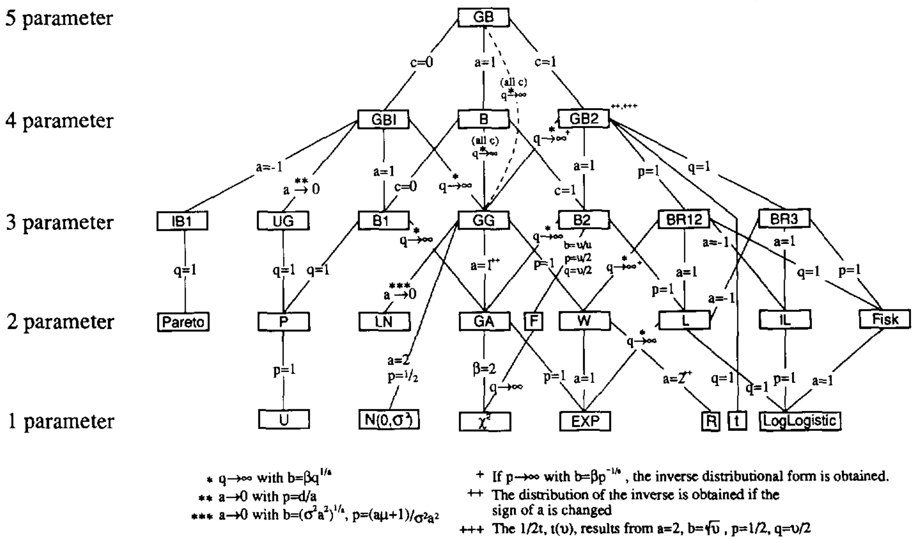

Fig. 2. Distribution tree.

Burr distributions. The flexibility suggested by Fig. 2 is impressive, but does not illustrate all of the possibilities.

Generally speaking, a one-parameter distribution can model one independent empirical characteristic such as the mean. Greater flexibility for fitting empirical data is obtained as we introduce additional parameters and move up the distribution tree. The computation involved in estimating additional parameters can generally be accommodated by recent advances in computational capability. It is often important to test whether the extra computations significantly improve our ability to model empirical data. The hypothesis that there are no additional improvements involving nested distributions can be tested using Wald (W), Lagrangian multiplier (LM), and likelihood ratio (LR) tests based on asymptotic chi-square distributions. However, distributions that involve parameters on the boundary of the parameter space violate regularity conditions used in deriving the asymptotic distribution for the W and LR tests. A direct consequence of violating this regularity condition is that the size of a test based on one of these test statistics need not be accurately approximated by the 'asymptotic' chi-square distribution. Reparameterization can sometimes circumvent this problem. In some other cases involving parameters on the boundary of the parameter space, the authors have found that the use of the likelihood ratio test can be too conservative (McDonald and Xu, 1992). Gouriéroux, Holly, and Monfort (1982) find that the LR test statistic is distributed as a mixture of chi-square variables for hypotheses involving inequalities corresponding to linear regression models. These theoretical results are consistent with the empirical results obtained by McDonald and Xu. In either case, Fig. 2 identifies situations in which various tests can be performed and situations in which the size of the LR test might be expected to be smaller than that suggested by 'conventional' chi-square critical values. Even in these cases it might be useful to think of the chi-square value as providing a useful model selection criterion such as with the MSE or AIC.

Distributional parameters can be estimated using conventional methods of maximum likelihood (MLE), method of moments or maximum product spacing (MPS) estimation, in situations that involve parameters on the boundary of the parameter space. Maximum likelihood estimators are obtained by maximizing $\sum_t \ln f(y_t;\Theta)$ over Θ, where y_t denotes the tth observed value and $f(y_t;\Theta)$ represents the pdf with unknown parameters Θ. For grouped data with intervals $0 = Y_0 < Y_1 < \cdots < Y_{G-1} < Y_G = \infty$, maximum likelihood estimators based on a multinomial with an underlying density $f(y;\Theta)$ are obtained by maximizing $\ln[N!] + \sum_i [n_i \ln p_i(\Theta) - \ln(n_i!)]$ over Θ where $p_i(\Theta) = F(Y_i;\Theta) - F(Y_{i-1};\Theta)$, $F(\ )$ denotes the cumulative distribution function, n_i is the number of observations in the ith inteval, and $N = \sum_i n_i$.

Maximum product spacing (MPS) estimators share the same asymptotic distribution as MLE when the regularity conditions are satisfied, but may dominate MLE when the boundary conditions are violated with an estimated parameter being on the boundary of the parameter space (Cheng and Amin, 1983). MPS estimators are obtained by maximizing $\sum_i [1/N] \ln p_i(\Theta)$ over Θ, corresponding to ordered individual observations $\{y_i\}$ and augmented by $y_0 = 0$ and $y_{N+1} = b$ or ∞ with $p_i(\Theta) = F(y_i;\Theta) - F(y_{i-1};\Theta)$, $i = 1, 2, \ldots, N+1$.

Estimation involving expressions for the cumulative distribution function can be easily performed using computer programs for the incomplete beta function. This relationship is formalized by observing that if $Y \sim \mathrm{GB}(y;a,b,c,p,q)$, then $X_1 = (Y/b)^a/(1 + c(Y/b)^a) \sim \mathrm{B1}(x_1; b = 1, p, q)$; hence, the incomplete beta function can be used to evaluate the cumulative distribution function.

3. Another generalized distribution: The exponential generalized beta (EGB)

If $Y \sim \mathrm{GB}(y;a,b,c,p,q)$, then the random variable $Z = \ln(Y)$ will be said to be distributed as an exponential generalized beta (EGB), with pdf defined by

$$\mathrm{EGB}(z;\delta,\sigma,c,p,q) = \mathrm{GB}(\mathrm{e}^z; a = 1/\sigma, b = \mathrm{e}^\delta, c, p, q)\,\mathrm{e}^z$$

$$= \frac{\mathrm{e}^{p(z-\delta)/\sigma}(1-(1-c)\,\mathrm{e}^{(z-\delta)/\sigma})^{q-1}}{|\sigma|\,\mathrm{B}(p,q)(1+c\,\mathrm{e}^{(z-\delta)/\sigma})^{p+q}}$$

$$\text{for} \quad -\infty < \frac{z-\delta}{\sigma} < \ln\left(\frac{1}{1-c}\right), \tag{3.1}$$

and zero otherwise. A case could be made for saying that Z is distributed as a log-GB2 (LGB2). However, the convention introduced by the relationship between the lognormal and normal distribution will be adopted. Since the EGB and GB distributions are related by the logarithmic transformation, similar 'exponential' distributions can be obtained by transforming each of the distributions shown in Fig. 2. Several of these distributions are of special interest. Consider, for example, the exponential generalized beta of the first and second kind (EGB1 and EGB2) and the exponential generalized gamma (EGG) defined by the probability density functions:

$$\begin{aligned}\mathrm{EGB1}(z;\delta,\sigma,p,q) &= \mathrm{EGB}(z;\delta,\sigma,c=0,p,q)\\ &= \frac{\mathrm{e}^{p(z-\delta)/\sigma}(1-\mathrm{e}^{(z-\delta)/\sigma})^{q-1}}{|\sigma|\,\mathrm{B}(p,q)},\end{aligned} \tag{3.2}$$

$$\begin{aligned}\mathrm{EGB1}(z;\delta,\sigma,p,q) &= \mathrm{EGB}(z;\delta,\sigma,c=1,p,q)\\ &= \frac{\mathrm{e}^{p(z-\delta)/\sigma}}{|\sigma|\,\mathrm{B}(p,q)(1+\mathrm{e}^{(z-\delta)/\sigma})^{p+q}},\end{aligned} \tag{3.3}$$

$$\begin{aligned}\mathrm{EGG}(z;\delta,\sigma,p,q) &= \lim_{q\to\infty}\mathrm{EGB}(z;\delta* = \sigma\ln q+\delta,c,p,q)\\ &= \frac{\mathrm{e}^{p(z-\delta)/\sigma}\,\mathrm{e}^{-\mathrm{e}^{(z-\delta)/\sigma}}}{|\sigma|\,\Gamma\,\Xi(p)}.\end{aligned} \tag{3.4}$$

The EGB1, EGB2, and EGG are merely alternative representations of the generalized exponential, logistic, and Gompertz distributions introduced in Ahuja and Nash (1967) and reviewed in Johnson and Kotz (1970, Vol. 2) and Patil et al. (1984). If (δ,σ,p,q) in (3.2)–(3.4) is replaced with $(\sigma\ln(\rho), -\sigma,\phi,\theta)$ the notation found in Johnson and Kotz (1970, Vol. 2, p. 271) is obtained. Prentice (1975) considers an alternative development that leads to a different parameterization of (3.3). The terminology for these and closely related distributions differs in the literature. For example, Patil et al. (1984) refer to generalized exponential and logistics distributions that are different from those referred to by Johnson and Kotz (1970). The notation used here may help clarify some of the interrelationships. The generalized Gumbell corresponds to (3.3) with $p=q$. The EBR3 is merely the Burr type 2 distribution. The exponential Weibull (EW) is the extreme value type I distribution. A figure similar to Fig. 2 could be constructed for the EGB distribution, with each distributional type in Fig. 2 being preceded with an 'E' to denote 'exponential', and the parameters a and b replaced with $1/\sigma$ and e^{δ}, respectively. Thus the EGB would appear in place of GB. In the case of the log-normal, $\mathrm{LN}(y;\mu,\sigma^2)$, ELN would correspond to the normal, $\mathrm{N}(z;\mu,\sigma^2)$. The structure of this distributional tree would be the same as Fig. 2, but would include generalized forms of the exponential, logistics, Gompertz, Gumbell, and extreme value distributions.

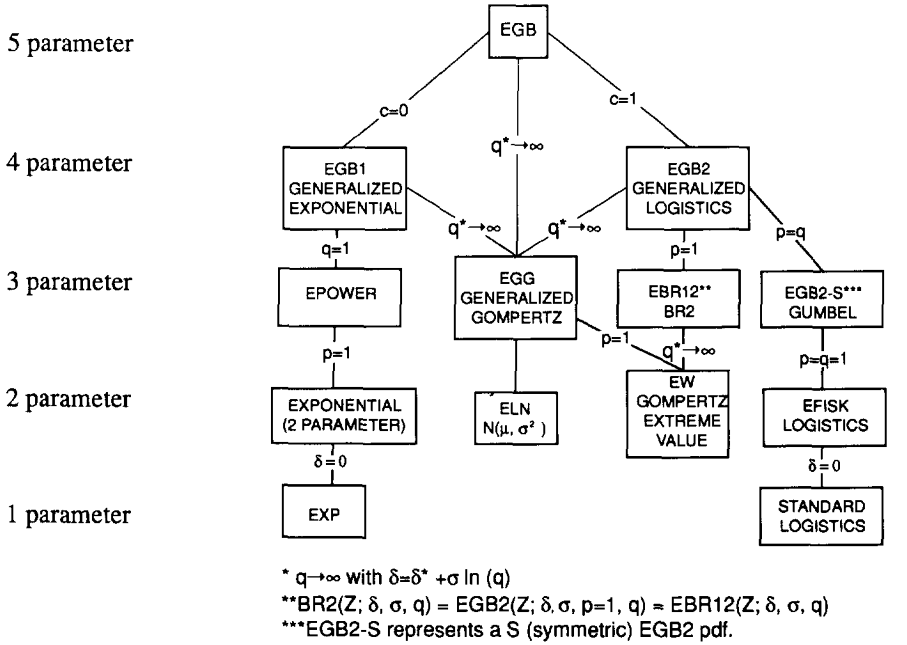

Fig. 3. Partial EGB distribution tree.

Rather than reconstructing Fig. 2 for the 'exponential' distributions, we present an abbreviated form in Fig. 3, which includes some of the most well-known members of the EGB family. A comparison of Figs. 2 and 3 suggests many other distributions that could have been included.

It can be shown that the moment generating function for the EGB is given by

$$M_{\mathrm{EGB}}(t) = \mathrm{E}(\mathrm{e}^{tz})$$

$$= \frac{\mathrm{e}^{\delta t}\mathrm{B}(p+t\sigma, q)}{\mathrm{B}(p,q)}\ {}_2F_1\left[\begin{matrix} p+t\sigma, & t\sigma; & c \\ p+q+t\sigma & & \end{matrix}\right], \tag{3.5}$$

where the ${}_2F_1[\,]$ denotes a hypergeometric function. Closed form solutions for the hypergeometric series in (3.5) for $c = 0$ and $c = 1$ yield the moment generating functions for the EGB1 and EGB2. Section A.2 of the Appendix contains the derivations of the results in this section and expressions for $M_{\mathrm{EGB1}}(t)$, $M_{\mathrm{EGB2}}(t)$, and $M_{\mathrm{EGG}}(t)$. The mean of the EGB can be shown to be

$$\mathrm{E}_{\mathrm{EGB}}(Z) = \delta + \sigma[\psi(p) - \psi(p+q)] + \frac{c\sigma p}{(p+q)}\ {}_3F_2\left[\begin{matrix} 1, & 1, & p+1; & c \\ 2, & p+q+1 & & \end{matrix}\right], \tag{3.6}$$

Table 1

Estimated distribution: 1985 nominal family income

	GB	GB1	GB2	B	B1	B2	GG	BR3	BR12	GA	LN
a	2.488	1.306	2.489	1.000	1.000	1.000	1.306	3.399	1.582	1.000	$\mu = 10.12$
$b(\beta)$	5.419×10^8	3.011×10^4	5.419×10^4	3.108×10^5	3.108×10^5	2.274×10^{11}	2.932×10^4	4.382×10^4	1.242×10^5	1.705×10^4	$\sigma = 0.8421$
p	0.5732	1.235	0.5732	1.754	1.754	1.899	1.235	0.3996	1.000	1.903	
q	1.887	1.734×10^5	1.887	15.135	15.166	1.330×10^7		1.000	7.870		
c	1.0000			0.0000							
Mean[a]	32.748	32.250	32.748	32.226	32.227	32.461	32.254	33.274	32.397	32.439	35.350
SSE[b]	0.00011	0.00024	0.00011	0.00029	0.00029	0.00034	0.00024	0.00016	0.00017	0.00034	0.00413
SAE[b]	0.0425	0.0574	0.0425	0.0600	0.0600	0.0711	0.0574	0.0491	0.0503	0.0705	0.2379
χ^2[b]	154.5	330.8	154.5	423.4	423.4	592.8	330.8	217.8	252.5	593.9	9198.1
$-\ln L$	173.9	261.3	173.9	305.4	305.4	378.9	261.3	206.0	221.5	378.8	3370.9

[a] Census estimate: 32.944 (in thousands).

[b] The *SSE*, *SAE*, and χ^2 values are obtained by evaluating $\sum_i (n_i/N - p_i(\hat{\Theta}))^2$, $\sum_i |n_i/N - p_i(\hat{\Theta})|$, $N\sum_i (n_i/N - p_i(\hat{\Theta}))^2/p_i(\hat{\Theta})$.

estimator in a regression model, based on an EGB2 error distribution. The asymptotic distributions of distributional parameters arising when regularity conditions are violated can be investigated using bootstrap methods. All optimizations reported were performed using GQOPT (obtained from Richard Quandt) with a convergence criterion of 10^{-8}. When the parameter c is estimated, it is estimated freely in the interval [0, 1] and not just at the endpoints. The income data are included in Section A.3 of the Appendix; the other data are available from the authors.

4.1. Distribution of income

The first example is based on 1985 nominal family income obtained from the Census Population Reports. The data correspond to a sample of 63,558 families combined into 21 income groups. Since the data are reported in a grouped format, maximum likelihood estimates are obtained by maximizing the multinomial log-likelihood function discussed in Section 2. The GB must provide at least as large a log-likelihood (l_{GB}) value as any of its special or limiting cases; however, l_{GB} need not be larger than the corresponding log-likelihood value for a special case if the data were generated by that distribution. The results of fitting several representative distributions to this data are reported in Table 1. The reported means are obtained by evaluating the expressions for the mean summarized in Section 2, which corresponds to the estimated parameters. The sum of squared errors (SSE), sum of absolute errors (SAE), and chi-square goodness of fit values (χ^2) compare the estimated and observed cell frequencies, $p_i(\hat{\Theta})$ and n_i/N. The freely estimated value for c is 1.000; thus, the income data provides an example in which the GB and GB2 are observationally equivalent and yield the best fit of any of the reported distributions. This is seen by comparing parameter estimates, goodness of fit measures (SSE, SAE, χ^2), and log-likelihood values. The more general (generalized beta) distribution 'selects' the GB2 in this case, and based on a comparison with the GB1, appears to improve the fit significantly. A closer inspection of the table reveals several other cases of near observational equivalence: the GB1 and GG, the B and B1, and the B2 and GA. It is interesting to note that while the GB is virtually identical to the GB2, the B distribution yields results that are similar to the B1 but not the B2. This is because the B distribution is associated with a unitary value for parameter a, which is not in close agreement with the freely estimated value of a obtained with the GB, i.e., $\hat{a} = 2.488$. All the fitted distributions except the log-normal give reasonably similar and accurate estimates of mean income; however, there is great variation in the goodness of fit measures, with the lognormal clearly providing the worst fit. There is a statistically significant improvement in fit (log-likelihood) in moving from the gamma to the generalized gamma, but not in going from B1 to B or from GB2 to GB. The observational equivalence of the GB and GB2 or of the B and B1 distributions makes a test of

Table 1

Estimated distribution: 1985 nominal family income

	GB	GB1	GB2	B	B1	B2	GG	BR3	BR12	GA	LN
a	2.488	1.306	2.489	1.000	1.000	1.000	1.306	3.399	1.582	1.000	$\mu = 10.12$
$b(\beta)$	5.419×10^8	3.011×10^4	5.419×10^4	3.108×10^5	3.108×10^5	2.274×10^{11}	2.932×10^4	4.382×10^4	1.242×10^5	1.705×10^4	$\sigma = 0.8421$
p	0.5732	1.235	0.5732	1.754	1.754	1.899	1.235	0.3996	1.000	1.903	
q	1.887	1.734×10^5	1.887	15.135	15.166	1.330×10^7		1.000	7.870		
c	1.0000			0.0000							
Mean[a]	32.748	32.250	32.748	32.226	32.227	32.461	32.254	33.274	32.397	32.439	35.350
SSE[b]	0.00011	0.00024	0.00011	0.00029	0.00029	0.00034	0.00024	0.00016	0.00017	0.00034	0.00413
SAE[b]	0.0425	0.0574	0.0425	0.0600	0.0600	0.0711	0.0574	0.0491	0.0503	0.0705	0.2379
χ^2[b]	154.5	330.8	154.5	423.4	423.4	592.8	330.8	217.8	252.5	593.9	9198.1
$-\ln L$	173.9	261.3	173.9	305.4	305.4	378.9	261.3	206.0	221.5	378.8	3370.9

[a] Census estimate: 32.944 (in thousands)

[b] The *SSE*, *SAE*, and χ^2 values are obtained by evaluating $\sum_i (n_i/N - p_i(\hat{\Theta}))^2$, $\sum_i |n_i/N - p_i(\hat{\Theta})|$, $N\sum_i (n_i/N - p_i(\hat{\Theta}))^2/p_i(\hat{\Theta})$.

their relationship unnecessary. The GB2, BR3, and gamma are the best fitting four-, three-, and two-parameter distributions, respectively. The relative fits of the distributions were not altered by aggregating the 21 income groups into eleven groups. The chi-square goodness of fit statistic has an asymptotic chi-square distribution, but all models must be rejected at conventional levels of significance. This result, not uncommon in applications involving large sample sizes, raises questions about the power of these tests.

4.2. Daily stock returns

The second example analyzes 505 daily returns based on a value-weighted index (including dividends) of the New York Stock Exchange for 1988 and 1989. The results of fitting selected distributions are reported in Table 2. The sample mean and variance of daily returns (ratio of value-weighted indices, including dividends) are 1.00087 and 7.16×10^{-5}, respectively, over this two-year period. Dubin–Watson (1.97), LM, and ARCH tests show no signs of either autocorrelation or ARCH problems. Returns are also frequently defined as the logarithm of the ratio of prices. Using MLE to fit the log data would give comparable results to fitting the GB distribution to the price ratio data reported in this paper. The data are negatively skewed, with a skewness coefficient of -1.316 and a kurtosis coefficient of 13.46. The freely estimated value of the parameter c in the GB distribution is 1; consequently, the GB and GB2 appear again to be observationally equivalent, with the GB providing significant improvements relative to the GB1 or B distribution. The large estimate for q in the GB1 [see Eq. (2.5) or Fig. 2] suggests that the estimated GB1 may agree closely with its limiting case, the GG, and this is indeed what happens. The B and B1 provide another case of 'observational equivalence'. It is interesting to note again that maximum likelihood estimation of the B distribution 'selects' B1, while the GB distribution 'selects' GB2. The distribution of income data (grouped) and daily return data share this result. The B2 and gamma distributions provide very similar fits to the data; however, this might be expected due to the large estimate for the parameter q in the B2 distribution. The log-normal, widely used in empirical and theoretical finance literature, cannot model the negative skewness or large kurtosis observed in this data set. In addition, the log-normal also provides a statistically inferior fit relative to the GB, GB1, GB2, and GG (which include it as a limiting case). In summary, we again find the GB2, BR3, and gamma to be the best fitting four-, three- and two-parameter distributions, respectively.

4.3. The market model and beta

The beta for a particular stock is the estimated slope in a regression model of the form:

$$(R_i - R_f) = \alpha + \beta(R_m - R_f) + \varepsilon_t,$$

Table 2

Estimated distribution: Daily returns on New York stock exchange

	GB	GB1	GB2	B	B1	B2	GG	BR3	BR12	GA	LN
a	1804	39.32	1804	1.000	1.000	1.000	39.33	269.5	233.3	1.000	$\mu = 8.361 \times 10^{-4}$
$b(\beta)$	1.0016	1.293	1.0016	1.0714	1.0714	15.47	0.9453	1.003	1.002	7.210×10^{-5}	$\sigma = 0.08505$
p	0.09291	9.944	0.09291	952.8	953.3	1.4766×10^4	9.941	0.7456	1.000	1.389×10^4	
q	0.1059	2.227×10^5	0.1059	67.13	67.20	2.283×10^5		1.000	1.091		
c	1.0000			0.0000							
Mean	1.0009	1.0009	1.0009	1.0009	1.0009	1.0009	1.0009	1.0008	1.0010	1.0009	1.0009
Variance	6.39×10^{-5}	6.84×10^{-5}	6.39×10^{-5}	6.913×10^{-5}	6.91×10^{-5}	7.22×10^{-5}	6.84×10^{-5}	5.79×10^{-5}	5.69×10^{-5}	7.22×10^{-5}	7.25×10^{-5}
Skewness	−0.211	−0.296	−0.211	−0.219	−0.219	−0.020	−0.299	−0.307	−0.057	0.017	0.026
Kurtosis	5.92	2.58	5.92	3.07	3.07	2.79	3.18	4.45	4.15	3.01	3.00
$\ln L$	1756.6	1708.8	1756.6	1704.3	1704.3	1691.3	1708.8	1746.7	1744.1	1691.3	1690.4

Sample mean = 1.00087, skewness = −1.316, variance = 7.160×10^{-5}, kurtosis = 13.46.

where R_i = return on the 'ith' security, R_f = risk-free rate, R_m = return on the market, and ε_t denotes a random disturbance term. R_f and R_m were respectively estimated by the returns on three-month T-bills and on the S&P index. Monthly return data for Valley National Corporation (VNCP) from the period 1986 to 1992 were used in estimation. Least squares estimates yielded $(\hat{\alpha}, \hat{\beta},) = (0.0871, 1.32)$ with an $R^2 = 0.29$. Autocorrelation was not a problem ($DW = 1.61$) and heteroskedasticity associated with an ARCH specification was rejected. Estimated skewness and kurtosis coefficients, obtained from OLS residuals, were -1.23 and 7.63 and yielded a Bera–Jarque normality statistic of 87.0 which is statistically significant. Thus, the OLS estimator can still be Best Linear Unbiased, but need not be efficient in a larger class of estimators. Since the error term in the market model can be any real number, we will assume that the underlying error distribution can be approximated by a member of the EGB2 family and perform maximum likelihood estimation. Selected results are reported in Table 3. The fit for the EGB2 and EBR3 error distributions were very similar and yielded an estimated beta of about 1.10. A likelihood ratio test for nested models, relative to the EGB2, only provided the basis for rejecting the EW (Gompertz or extreme value) and the normal distributions. The logistics or EFisk specification performed much better than the normal with results which did not differ significantly from the EGB2 specification. For comparative purposes, the beta parameter was also estimated using a normal Kernel (with a smoothing parameter of 0.5) and with generalized method of moments (GMM) (Newey, 1988) using two, three, and four moments in the estimation procedure. These estimated values are 1.23, 1.17, 1.17, and 1.07, respectively. The GMM estimator which takes account of possible skewness and kurtosis in the error distribution yields an estimate similar to that obtained from the EGB2.

Table 3

Estimated parameters: Valley National Corporation

	EGB2	EBR3	EBR12	EGB2-S	EGG	EW	EFISK	NORMAL (OLS)
α[a]	0.0984	0.0979	0.0999	0.0774	0.114	0.138	0.0812	0.0871
β	1.10	1.09	1.15	1.02	1.26	1.32	1.11	1.32
p	0.600	0.572	1.00	0.536	2.98	1.00	1.00	
q	1.07	1.00	2.03	0.536	—	—	1.00	—
σ	0.0469	0.0450	0.0689	0.0359	0.170	0.0951	0.0573	0.110
$\ln L$	65.7	65.7	65.5	64.7	64.3	62.6	64.5	59.8

OLS residuals: $s^2 = 0.0123$, skewness = -1.23, kurtosis = 7.63, Bera–Jarque test for normality = 87.0, $DW = 1.61$, $R^2 = 0.29$.

[a] α is adjusted to make the median of the residuals zero.

5. Summary and conclusions

The B1 and B2 and their special cases are two of the most widely used distributions in statistics. Often the decision to use a B1 or B2 distribution is an informal one. In this paper, a four-parameter beta distribution (B) is introduced; it nests both the beta of the first kind (B1) and the beta of the second kind (B2) and provides a basis for selecting between the two. A five-parameter generalized beta distribution (GB) is also introduced, which includes *both* the generalized beta of the first kind (GB1) and generalized beta of the second kind (GB2) (generalized F or generalized beta prime). These new distributions, the B and GB, provide an approach for comparing the B1 and B2 or the GB1 and GB2 distributions within a 'nested' framework; they also have the potential to yield significantly better fits than any special case. The GB distributions are fit to data for family income and stock returns. The GB 'selects' the GB2 for both the income distribution and stock return data. The natural logarithm of the GB is a five parameter distribution, exponential generalized beta (EGB), and includes generalizations of the Gompertz, Gumbell, extreme value type I, normal, and many other distributions. Members of the EGB family have application in partially adaptive estimation of regression and time series models. The EGB2 is used as an error distribution in the market model to estimate the beta for a stock in the market model characterized by skewed and leptokurtic errors.

Appendix

A.1. Moments of the GB distribution

The hth-order moments of the GB are defined by

$$\mathrm{E}_{\mathrm{GB}}(Y^h) = \int_{-\infty}^{\infty} y^h \mathrm{GB}(y; a, b, c, p, q)\,\mathrm{d}y. \tag{A.1}$$

Substituting (2.9) into (A.1), replacing $(1-(1-c)(y/b)^a)^{q-1}$ by its binomial expansion, making the change of variable $u = (1-c)(y/b)^a$ and collecting terms yields

$$\mathrm{E}_{\mathrm{GB}}(Y^h) = b^h \sum_{i=0}^{\infty} \frac{(p+q)_i(-c/(1-c))^i}{(1-c)^{p-h/a+i}} \int_0^1 u^{p-h/a+i-1}(1-u)^{q-1}\,\mathrm{d}u. \tag{A.2}$$

The integral in (A.2) is the beta function $\mathrm{B}(p+h/a+\mathrm{i}, q)$ which can be expressed in terms of gamma functions, $\Gamma(p+h/a+i)\Gamma(q)/\Gamma(p+q+h/a+i)$. After making this substitution in (A.2) and simplifying, the hth-order moments can be written in terms of a hypergeometric series as

$$\mathrm{E}_{\mathrm{GB}}(Y^h) = \frac{b^h \mathrm{B}(p+h/a, q)}{(1-c)^{p+h/a}\mathrm{B}(p,q)}\, {}_2F_1\begin{bmatrix} p+q, \quad q+h/a; & -c/(1-c) \\ p+q+h/a; & \end{bmatrix}. \tag{A.3}$$

It is convenient to rewrite (A.3) as

$$\mathrm{E}_{\mathrm{GB}}(Y^h) = \frac{b^h \mathrm{B}(p + h/a, q)}{\mathrm{B}(p,q)} \, {}_2F_1 \begin{bmatrix} p + h/a, & h/a; & c \\ p + q + h/a; & & \end{bmatrix}; \tag{A.4}$$

see Rainville (1960, p. 60). (A.4) is defined for all h if $c < 1$ or for $-p < h/a < q$ if $c = 1$. For $c = 0$, the hypergeometric series in (A.4) equals 1, and we obtain the moments for the GB1, $b^h \mathrm{B}(p + h/a, q)/\mathrm{B}(p,q)$. For $c = 1$, Gauss's theorem, Rainville (1960, p. 48) implies that the hypergeometric series in (A.4) equals $\Gamma(p + q + h/a)\Gamma(q - h/a)/\Gamma(q)\Gamma(p + q))$, and the resultant expression can be simplified to yield the moments for the GB2, $b^h \mathrm{B}(p + h/a, q - h/a)/\mathrm{B}(p,q)$. The hth-order moments for the GG are given by $\beta^h \Gamma(p + h/a)/\Gamma(p)$. The book by Gradshteyn and Ryzhik (1965) is a very helpful reference.

A.2. Moment-generating and cumulative distribution functions for the EGB

A.2.1. Moment-generating function and moments

The moment-generating function of EGB variates can be obtained by making the substitution $u = \mathrm{e}^{((z-\delta)/\delta)}$ in $\mathrm{E}(\mathrm{e}^{tz})$ and combining terms:

$$\begin{aligned} \mathrm{M}_{\mathrm{EGB}}(Z) &= \mathrm{E}_{\mathrm{EGB}}(tZ) \\ &= \mathrm{e}^{\delta t} \mathrm{E}_{\mathrm{GB}}(Y^{\sigma t};\quad a = 1, b = 1, c, p, q) \\ &= \frac{\mathrm{e}^{\delta t}\mathrm{B}(p + t\sigma, q)}{\mathrm{B}(p,q)} \, {}_2F_1 \begin{bmatrix} p + t\sigma, & t\sigma; & c \\ p + q + t\sigma & & \end{bmatrix}. \end{aligned} \tag{A.5}$$

Substituting $c = 0$ and $c = 1$ (respectively) into (A.5) gives the moment-generating function for the EGB1 and EGB2,

$$\mathrm{M}_{\mathrm{EGB1}}(t) = \frac{\mathrm{e}^{\delta t}\mathrm{B}(p + \sigma t, q)}{\mathrm{B}(p,q)}, \qquad \mathrm{M}_{\mathrm{EGB2}}(t) = \frac{\mathrm{e}^{\delta t}\mathrm{B}(p + t\sigma, q - t\sigma)}{\mathrm{B}(p,q)}. \tag{A.6}$$

The moment-generating function for the EGG is $\mathrm{M}_{\mathrm{EGG}}(t) = \mathrm{e}^{\delta t}\Gamma(p + t\sigma)/\Gamma(p)$. The mean of EGB can be obtained by differentiating the natural logarithm of (A.5), the cumulant-generating function, and substituting $t = 0$ to yield

$$\mathrm{E}_{\mathrm{EGB}}(Z) = \delta + \sigma[\psi(p) - \psi(p + q)] + \frac{c\sigma p}{(p + q)} \, {}_3F_2 \begin{bmatrix} 1, & 1, & p + 1; & c \\ 2, & p + q + 1; & & \end{bmatrix}. \tag{A.7}$$

The means for the EGB1, EGB2, and EGG, respectively, are given by

$$\mathrm{E}_{\mathrm{EGB1}}(Z) = \delta + \sigma[(\Psi)(p) - \Psi(p + q)],$$

$$\mathrm{E}_{\mathrm{EGB2}}(Z) = \delta + \sigma[(\Psi)(p) - \Psi(q)],$$

$$\mathrm{E}_{\mathrm{EGG}}(Z) = \delta + \sigma\,\Psi(p),$$

where $\Psi(\ )$ denotes the digamma function. The mean of the EGB2 can be obtained directly from (A.5) or (A.6). It can also be obtained from (A.7), with $c = 1$ and using of the transformations found in Sneddon (1966) for ${}_3F_2[\ ]$ functions with a unit argument to another ${}_3F_2[\ ]$ with a known closed form expression.

Although higher-order moments for the EGB are quite involved, relatively simple expressions for the variance, skewness, and kurtosis for the EGB1, EGB2, and EGG can be easily obtained by differentiating the desired cumulant-generating functions obtained from (A.6).

A.2.2. Cumulative distribution function

The relationship between the random variables EGB, GB, and B1 facilitates the derivation of an expression for the cumulative distribution of EGB, $F_{EGB}(z) = \Pr(Z \leqslant z) = \Pr(e^Z \leqslant e^z) = F_{GB}(e^z)$, which involves the incomplete beta function (see Section 2).

A.3. Data

The data for 1985 Nominal Family Income Data are given below ($N = 63558$). The other data sets can be obtained from the first author.

Percentage	Upper-income interval
1.929	2500.0
2.886	4999.0
4.224	7499.0
4.317	9999.0
5.180	12499.0
5.040	14999.0
5.390	17499.0
5.106	19999.0
5.499	22499.0
4.781	24999.0
5.224	27499.0
4.539	29999.0
4.865	32499.0
3.957	34999.0
4.220	37499.0
3.402	39999.0
6.282	44999.0
4.882	49999.0
7.190	59999.0
5.543	74999.0
5.546	∞

References

Arnold, B.C., 1983, Pareto distributions (International Cooperative Publishing House, Burtonsville, MD).

Ahuja, J.C. and S.W. Nash, 1967, The generalized Gompertz–Verhulst family of distributions, Sanhkya 29 A, 141–156.

Cheng, R.C.H. and N.A.K. Amin, 1983, Estimating parameters in continuous univariate distributions with a shifted origin, Journal of the Royal Statistical Society 45 B, 394–403.

Gouriéroux, C., A. Holly, and A. Monfort, 1982, Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters, Econometrica 50, 63–80.

Gradshteyn, I.S. and I.M. Ryzhik, 1965, Tables of integrals, series, and products (Academic Press, New York, NY).

Grassia, A., 1977, On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions, Australian Journal of Statistics 19, 108–114.

Johnson, N.L. and S. Kotz, 1970, Continuous univariate distributions, Vols. 1 and 2 (Wiley, New York, NY).

Kalbfleisch, J.D. and R.L. Prentice, 1980, The statistical analysis of failure time data (Wiley, New York, NY).

Libby, D.L. and M.R. Novic, 1982, Multivariate generalized beta distributions with applications to utility assessment, Journal of Educational Statistics 7, 271–294.

McDonald, J.B., 1984, Some generalized functions for the size distributions of income, Econometrica 52, 647–663.

McDonald, J.B. and W.K. Newey, 1988, Partially adaptive estimation of regression models via the generalized t-distribution, Econometric Theory 4, 428–457.

McDonald, J.B. and S.B. White, 1993, A comparison of some robust, adaptive and partially adaptive estimators of regression models, Econometric Reviews 12, 103–124.

McDonald, J.B. and Y. Xu, 1992, An empirical investigation of the likelihood ratio test when the boundary condition is violated, Communications in Statistics 21, 879–892.

Newey, W.K., 1988, Adaptive estimation of regression models via moment restrictions, Journal of Econometrics 38, 301–339.

Patil, G.P., M.T. Boswell, and M.V. Ratnaparkhi, 1984, Dictionary and classified bibliography of statistical distributions in scientific work: Continuous univariate models (International Cooperative Publishing House, Burtonsville, MD).

Pham-Gia, T. and Q.P. Duong, 1989, The generalized beta and F distributions, statistical modeling, Mathematical Computational Modeling 12, 1613–1625.

Prentice, R.L., 1975, Discrimination among some parametric models, Biometrica 62, 607–614.

Rainville, E.D., 1960, Special functions (Macmillan, New York, NY).

Sneddon, I.N., 1966, Special functions of mathematical physics (Oliver and Boyd, London).

United States Bureau of the Census: Current population reports, Series P-60, no. 156, 40.

Venter, G., 1984, Transformed beta and gamma distributions and aggregate losses, Proceedings of the Casualty Actuarial Society, 156–193.