ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Econometrics ■ (■■■■) ■■■–■■■

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Information measures for generalized gamma family

Ali Dadpay[a,1], Ehsan S. Soofi[b,*], Refik Soyer[c]

[a]*Department of Economics, University of Wisconsin-Milwaukee, P.O. Box 431, Milwaukee, WI 53201, USA*
[b]*Sheldon B. Lubar School of Business and Center for Research on International Economics, University of Wisconsin-Milwaukee, P.O. Box 742, Milwaukee, WI 53201, USA*
[c]*Department of Decision Sciences, George Washington University, Washington, DC 20052, USA*

Abstract

The objective of this paper is to integrate the generalized gamma (*GG*) distribution into the information theoretic literature. We study information properties of the *GG* distribution and provide an assortment of information measures for the *GG* family, which includes the exponential, gamma, Weibull, and generalized normal distributions as its subfamilies. The measures include entropy representations of the log-likelihood ratio, AIC, and BIC, discriminating information between *GG* and its subfamilies, a minimum discriminating information function, power transformation information, and a maximum entropy index of fit to histogram. We provide the full parametric Bayesian inference for the discrimination information measures. We also provide Bayesian inference for the fit of *GG* model to histogram, using a semi-parametric Bayesian procedure, referred to as the maximum entropy Dirichlet (MED). The *GG* information measures are computed for duration of unemployment and duration of CEO tenure.

JEL classification: C11; C14; C15; C16; C41; C52

Keywords: Entropy; Exponential; Gamma; Kullback–Leibler information; Model fitting; Weibull

1. Introduction

The generalized gamma (*GG*) distribution offers a flexible family in the varieties of shapes and hazard functions for modeling duration. It was introduced by Stacy (1962).

*Corresponding author.
E-mail addresses: adadpay@uwm.edu (A. Dadpay), esoofi@uwm.edu (E.S. Soofi), soyer@gwu.edu (R. Soyer).
[1]Currently at. Rapp Collins Worldwide, 1660 North Westridge Circle, Irving, TX 75062.

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.05.010

Difficulties with convergence of algorithms for maximum likelihood estimation (Hager and Bain, 1970) inhibited applications of the *GG* model. Prentice (1974) resolved the convergence problem using a nonlinear transformation of *GG* model. However, despite its long history and growing use in various applications, the *GG* family has been remarkably absent in the information theoretic literature. Thus far a maximum entropy (ME) derivation of *GG* is given in Kapur (1989), where it is referred to as generalized Weibull distribution, and only recently the entropy of *GG* has appeared in the context of flexible families of distributions (Nadarajah and Zografos, 2003). The *GG* family has not been included in information studies such as the existing ME distributional fitting of the parametric families (see, e.g., Soofi and Retzer, 2002 and references therein), the discrimination information statistics analysis of the parametric families (Alwan et al., 1998), and the entropy orderings of the parametric families (Ebrahimi et al., 1999). The main objective of this paper is to fill this void and integrate the *GG* family into the information theoretic literature. For this purpose, we develop information criteria for discriminating between the *GG* and its subfamilies and for assessing the fit of *GG* to the data. We also present Bayesian inference about the discrimination and the fit.

Analysis of duration data is increasingly used in various areas of economics and related fields (Keifer, 1988). In labor economics, examples include studies of the duration of unemployment, (Lancaster, 1979; Kiefer, 1984; McDonald and Butler, 1987; Yamaguchi, 1992), turnover in labor market (Kiefer et al., 1985), length of contract (Gronberg, 1994), and duration of strike (Jaggia, 1991). Examples in other areas include studies of firms survival (Audretsch and Mahmoud, 1995), duration that firms spend under Chapter 11 (Orbe et al., 2002), duration that a property is on the market (Genesove and Mayer, 1997), duration of schooling at higher education (Diaz, 1999), duration of stages of oilfield exploration (Favero et al., 1994), household interpurchase time (Vakratsas and Bass, 2002), interpurchase time in financial markets (Allenby et al., 1999), and length of the time that new movies stay on screens (Blumenthal, 1988).

Distributions that are used in duration analysis in economics include exponential (Kiefer, 1984; Diebold and Rudebusch, 1990), lognormal (Eckstein and Wolpin, 1995), gamma (Lancaster, 1979), and Weibull (Favero et al., 1994). The *GG* family, which encompasses exponential, gamma, and Weibull as subfamilies, and lognormal as a limiting distribution, has been used in economics by Jaggia (1991), Yamaguchi (1992), and Allenby et al. (1999). Some authors (e.g., Jaggia, 1991; Allenby et al., 1999) have argued that the flexibility of *GG* makes it suitable for duration analysis, while others have been using simpler models and avoiding the estimation difficulties caused by the complexity of *GG* parameter structure. Obviously, there would be no need to endure the costs associated with the application of a complex *GG* model if the data do not discriminate between the *GG* and members of its subfamilies, or if the fit of a simpler model to the data is as good as that for the complex *GG*. The question therefore is: Do the data necessitate use of a *GG* model? From the information theoretic perspective, this question is dealt with derivation of probability models based on partial information in the form of a set of constraints, measuring the incremental information content of additional constraints, and thereby assessing compatibility of models with the data. The *GG* information measures, presented in this paper, offer tools, with axiomatic basis and intuitive appeals, for *GG* as a general class of duration models.

The paper is organized as follows. Section 2 discusses information properties of the *GG* family and presents several discrimination information measures for the *GG* and its subfamilies. Section 3 gives entropy representations of the likelihood statistic, AIC, and BIC measures. Section 4 discusses Bayesian inference about the *GG* parameters and discrimination information measures. Section 5 presents an information index of fit of the *GG* model to the histogram and Bayesian inference about the fit. Section 6 illustrates application of the *GG* information criteria to the duration of unemployment and duration of CEO tenure. Section 7 gives some brief concluding remarks.

2. Information properties of GG family

The probability density function of the *GG* distribution, $GG(\alpha, \tau, \lambda)$, is

$$f_{GG}(y|\alpha,\tau,\lambda) = \frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)} y^{\alpha\tau-1} \mathrm{e}^{-(y/\lambda)^{\tau}}, \quad y \geqslant 0, \ \ \alpha, \tau, \lambda > 0, \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function, α and τ are shape parameters, and λ is the scale parameter.

The *GG* family is flexible in that it includes several well-known models as subfamilies (see, Johnson et al., 1994). The subfamilies of *GG* thus far considered in the literature are exponential ($\alpha = \tau = 1$), gamma for ($\tau = 1$), and Weibull for ($\alpha = 1$). The lognormal distribution is also obtained as a limiting distribution when $\alpha \to \infty$. By letting $\tau = 2$ we obtain a subfamily of *GG* which is known as the generalized normal distribution, $GN(2\alpha, \lambda)$. The *GN* is itself a flexible family and includes Half-normal ($\alpha = 1/2$), Rayleigh ($\alpha = 1$), Maxwell–Boltzmann ($\alpha = 3/2$), and Chi ($\alpha = k/2, k = 1, 2, \ldots$). Moreover, the *GG* family is more flexible than gamma and Weibull distributions in terms of hazard rate function. It allows for nonmonotonicity in the form of single-peaked hazard functions (but that it would not be able to "handle" multi-peaked hazard functions).

An important property of *GG* family for information analysis is that the family is closed under power transformation. That is, if $Y \sim GG(\alpha, \tau, \lambda)$, then

$$Z = Y^s \sim GG(\alpha, \tau/s, \lambda^s), \quad s > 0. \tag{2}$$

In particular,

$$X = Y^{\tau} \sim G(\alpha, \lambda^{\tau}), \tag{3}$$

where $G(\alpha, \lambda^{\tau})$ denotes the gamma density with shape parameter α and scale λ^{τ}.

The power and logarithmic moments of *GG* distribution are given by

$$\begin{aligned} \mu_s(\alpha,\tau,\lambda) &= E_{GG}(Y^s|\alpha,\tau,\lambda) = \frac{\lambda^s \Gamma(\alpha + s/\tau)}{\Gamma(\alpha)}, \quad s > 0, \\ v_s(\alpha,\tau,\lambda) &= E_{GG}(\log Y^s) = \log \lambda^s + \frac{s}{\tau}\psi(\alpha), \end{aligned} \tag{4}$$

where $\psi(\alpha) = \mathrm{d} \log \Gamma(\alpha)/\mathrm{d}\alpha$ is the digamma function.

For studying the information properties of *GG* family, we consider the class of distribution functions

$$\Omega_{\theta} = \{F(y|\boldsymbol{\theta}) : E_f[T_j(y)|\theta] = \theta_j, \ j = 0, 1, 2\}, \tag{5}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$, and $\theta_0 = T_0(y) = 1$ normalizes the density. For a given τ, $T_1(y) = y^{\tau}$,

$$\theta_1 = \mu_{\tau}(\alpha,\tau,\lambda) = E_{GG}(Y^{\tau}|\alpha,\tau,\lambda) = \lambda^{\tau}\alpha, \tag{6}$$

$T_2(y) = \log y$ and

$$\theta_2 = v(\alpha, \tau, \lambda) = E_{GG}(\log Y) = \log \lambda + \frac{1}{\tau}\psi(\alpha) \tag{7}$$

is the geometric mean.

2.1. Entropy properties

The entropy of a distribution F in Ω_θ is given by

$$H(F) = -\int_0^\infty f(y|\alpha, \tau, \lambda) \log f(y|\alpha, \tau, \lambda)\, \mathrm{d}y.$$

The ME model in (5) is $F^* = GG^* = GG(\alpha, \tau, \lambda)$ with density (1). (The ME distribution in (5) is obtained when F_0 in (12) is uniform, which is not a proper distribution over an infinite support). Kapur (1989) gives a proof for a different parameterization of (1) refers to it as generalized Weibull distribution.

The GG entropy is

$$H(GG^*) = \max_{F \in \Omega_\theta}[H(F)] = \log \lambda + \log \Gamma(\alpha) + \alpha - \log \tau + \left(\frac{1}{\tau} - \alpha\right)\psi(\alpha). \tag{8}$$

Nadarajah and Zografos (2003) includes $H(GG)$ among their list of entropies of flexible classes of distributions. For specific values of the parameters, (8) gives entropy expressions for gamma, Weibull, exponential, and half-normal distributions. Applications of ME characterizations of GG include developing ME fit indices discussed in the Section 5.

Entropy ordering of distributions within many parametric families is studied in Ebrahimi et al. (1999), but GG is not included. It is clear that the entropy of GG family is ordered by scale parameter λ. For the entropy orderings in terms of the shape parameters, we have

$$\frac{\partial H(GG)}{\partial \alpha} \geqslant 0 \quad \text{for } (\tau\alpha - 1)\psi'(\alpha) \leqslant \tau,$$

$$\frac{\partial H(GG)}{\partial \tau} \geqslant 0 \quad \text{for } \tau \leqslant -\psi(\alpha).$$

The first inequality holds for all values of τ, and hence $H(GG)$ is increasing in α. Since $\psi(\alpha) < 0$ for $\alpha < 1.5$ approximately, $H(GG)$ can be increasing in τ only when $\alpha < 1.5$.

2.2. Discrimination information properties

Suppose that we wish to examine if a distribution $F \in \Omega_\theta$ can be approximated by a given model F_0. The measure of information discrepancy between F and F_0 is the Kullback–Leibler discrimination information function

$$K(F : F_0) = \int f(y) \log \frac{f(y)}{f_0(y)} \mathrm{d}y. \tag{9}$$

It is well-known that $K(F : F_0) \geqslant 0$; the equality holds if and only if $f(y) = f_0(y)$ for all y in the support of the distributions. $K(F : F_0)$ is not symmetric and is a measure of directed divergence between F and F_0, where F_0 is referred to as the *reference distribution*. Symmetric versions of $K(F : F_0)$ includes Jeffreys divergence, $J(F, F_0) = K(F : F_0) +$

$K(F_0 : F)$ (Jeffreys, 1946), and $\min\{K(F : F_0), K(F_0 : F)\}$ referred to as the intrinsic information by Bernardo and Rueda (2002).

Let $F_0 = GG_0 = GG(\alpha_0, \tau_0, \lambda_0)$ be a given GG distribution. It can be shown that the discrimination information function between $F = GG(\alpha, \tau, \lambda)$ and F_0 is given by

$$K(GG : GG_0) = \log \frac{\phi_\tau}{\phi_\lambda^{\alpha\phi_\tau}} - \log \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)} - \alpha + \mu(\alpha, \phi_\tau, \phi_\lambda) + (\alpha\phi_\tau - \alpha_0)\nu(\alpha, \phi_\tau, \phi_\lambda), \tag{10}$$

where $\phi_\tau = \tau/\tau_0$, $\phi_\lambda = (\lambda/\lambda_0)^{\tau_0}$, $\mu(\alpha, \phi_\tau, \phi_\lambda)$ is the first moment and $\nu(\alpha, \phi_\tau, \phi_\lambda)$ is the geometric mean of a GG distribution with parameters $(\alpha, \phi_\tau, \phi_\lambda)$. The discrimination information $K(GG : GG_0)$ is a function of ratio of the scales ϕ_λ, the ratio of the Weibull shape parameters ϕ_τ, and a complicated function of the gamma shape parameters α and α_0.

Although $K(GG : GG_0)$ is a complicated function of the parameters, (10) is a general representation that encompasses discrimination information functions between the GG and its subfamilies, between distributions within each subfamily, and between members of different subfamilies. The discrimination information between $GG(\alpha, \tau, \lambda)$ and a gamma $G(\alpha_0, \lambda_0)$ is given by (10) with $\phi_\tau = \tau$. The discrimination information between $GG(\alpha, \tau, \lambda)$ and Weibull $W(\tau_0, \lambda_0)$ is given by (10) with $\alpha_0 = 1$. The discrimination information between $GG(\alpha, \tau, \lambda)$ and exponential $\mathscr{E}(\lambda_0)$ is given by (10) with $\phi_\tau = \tau$ and $\alpha_0 = 1$. The discrimination information between $GG(\alpha, \tau, \lambda)$ and generalized-normal $GN(\alpha_0, \lambda_0)$ is given by (10) with $\phi_\tau = \tau/2$ and $\alpha_0 = 2\alpha$.

The minimum discrimination information (MDI), also referred to as the minimum cross-entropy distribution, in Ω_θ, relative to F_0 is defined by the solution $F^* \in \Omega_\theta$ of

$$K(F^* : F_0) = \min_{F \in \Omega_\theta} K(F : F_0). \tag{11}$$

By the MDI Theorem of Kullback (1959), the MDI model F^*, if it exists, has a density in the form of

$$f^*(y|\boldsymbol{\theta}) = \eta_0 f_0(y) y^{\eta_2} \mathrm{e}^{\eta_1 y^\tau}, \tag{12}$$

where $\eta_0 = \eta_0(\boldsymbol{\theta})$ is the normalizing constant, $\eta_1 = \eta_1(\boldsymbol{\theta})$ and $\eta_2 = \eta_2(\boldsymbol{\theta})$ are Lagrange multipliers for the moment constraints $E(X^\tau) = \theta_1$ and $E(\log X) = \theta_2$, respectively.

From (12) we note that the MDI distributions in Ω_θ reference to the exponential and gamma are not members of the GG family. The MDI distribution with reference to generalized normal is not a GG when $\tau \neq 2$, and with reference to a Weibull with shape parameter τ_0 is not a GG if $\tau \neq \tau_0$. From (12) and the invariance of Kullback–Leibler function under one-to-one transformations we obtain the following MDI properties of GG distribution.

(a) The MDI distribution in Ω_θ relative to the reference distribution $F_0 = GG_0 = GG(\alpha_0, \tau_0, \lambda_0)$, $\alpha_0 \neq \alpha$, $\tau_0 = \tau$, $\lambda_0 \neq \lambda$ is $F^* = GG^*(\alpha, \tau_0, \lambda)$, and the MDI function is given by

$$\begin{aligned} K(GG^* : GG_0) &= \min_{F \in \Omega_\theta} K(F : GG_0) = K(G : G_0) \\ &= -\log \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)} + (\alpha - \alpha_0)\psi(\alpha) + \alpha(\phi_\lambda - \log \phi_\lambda - 1) \\ &= K(G : G_0; \alpha) + \alpha K(\mathscr{E}_1 : \mathscr{E}_2; \phi_\lambda), \end{aligned} \tag{13}$$

where $K(G:G_0)$ is the discrimination information between two gamma distributions with shape parameters α, α_0 and scale ratio ϕ_λ, and $K(G:G_0;\alpha)$ is the information discrepancy due to the different shapes and $K(\mathscr{E}_1:\mathscr{E}_2;\phi_\lambda)$ is the discrimination information between two exponential distributions $\mathscr{E}_1$ and $\mathscr{E}_2$ with mean ratio ϕ_λ.

(b) The MDI distribution in Ω_θ, relative to a Weibull reference distribution $F_0 = W_0 = W(\tau_0, \lambda_0)$, $\tau_0 = \tau$, $\lambda_0 \neq \lambda$ is $F^* = GG^*(\alpha, \tau, \lambda)$, and the MDI function,

$$\begin{aligned} K(GG^*:W_0) &= \min_{F \in \Omega_\theta} K(F:W) \\ &= -\log \Gamma(\alpha) + (\alpha - 1)\psi(\alpha) + \alpha(\phi_\lambda - \log \phi_\lambda - 1) \\ &= K(G:\mathscr{E};\alpha) + \alpha K(\mathscr{E}_1:\mathscr{E}_2;\phi_\lambda), \end{aligned} \tag{14}$$

where $K(G:\mathscr{E};\alpha)$ is the discrimination information between a gamma and exponential due to the gamma shape parameter.

(c) Let $F_\tau \in \Omega_\theta \equiv \Omega_{\mu,v,\tau}$ and $F_1 \in \Omega_{\mu,v,1}$, where $\Omega_{\mu,v,1}$ denotes the class of distributions with $E(Y) = \mu_1$ and $E(\log Y) = v_1$. Then

$$\begin{aligned} K(GG^*:W_\tau) &= \min_{F_\tau \in \Omega_{\mu,v,\tau}} K(F_\tau:W_\tau) = \min_{F_1 \in \Omega_{\mu,v,1}} K(F_1:\mathscr{E}) = K(G^*:\mathscr{E}) \\ &= \log \frac{\alpha}{\Gamma(\alpha)} + (\alpha - 1)[\psi(\alpha) - 1], \end{aligned} \tag{15}$$

where W_τ is the Weibull reference distribution with $E_W(Y^\tau) = E_{GG}(Y^\tau)$ and $E_{\mathscr{E}}(X) = E_G(X)$, are the Weibull and exponential reference distributions implied by the data scaling condition, and $G^* = G(\alpha, \lambda^\tau)$ implied by the transformation.

The condition $\tau_0 = \tau$ is needed for the application of (12) and the parameter distinction conditions $\alpha_0 \neq \alpha$ and $\lambda_0 \neq \lambda$ are for avoiding the trivial case of zero Lagrange multipliers.

Property (a) is along the lines of Alwan et al. (1998) who developed an information theoretic framework for statistical process control where the same types of moments are used for the in-control and monitoring stages. Property (b) extends that framework to the case of flexible families. Property (c) is insightful about information analysis and data transformation,where the scaling by moment, $E_{GG}(X^\tau) = E_W(X^\tau)$, is needed.

2.3. Data transformation

Information analysis of the GG family provides some interesting measures in terms of data transformation. Since the GG family is closed under power transformation, by (2) we can assess the effect of power transformation $Z = Y^s$ by the discrimination information between $Y \sim GG(\alpha, \tau, \lambda)$ and $Y^s \sim GG_s(\alpha, \tau/s, \lambda^s)$. In this case, $\phi_\tau = s$ and $\phi_\lambda = \lambda^{\tau/s - \tau}$ in (10). After some simplifications, we find that the information effect of transformation is given by

$$K_{GG}(Y:Y^s) \equiv K(GG:GG_s) = \log s + \alpha \left[\frac{\mu_{\tau/s}(\alpha, \tau, \lambda)}{\mu_\tau(\alpha, \tau, \lambda)} - [v_{\tau/s}(\alpha, \tau, \lambda) - v_\tau(\alpha, \tau, \lambda)] - 1 \right]. \tag{16}$$

Thus, the effect of power transformation is captured through the ratio of the power means and the difference between the geometric means of the transformed and original variables.

The information function (16) is a general representation of some important power transformation information measures for the GG family and subfamilies.

As a measure of information disparity between the distributions of the real data (prior to transformation) and the transformed data, $K_{GG}(Y : Y^s)$ may be interpreted as the loss of information due transformation. A large $K_{GG}(Y : Y^s)$ indicates the effect of transformation on the distribution is pronounced.

The information function $K_{GG}(Y : Y^\tau)$ measures the effect of transformation (3); i.e., discrepancy between $GG(\alpha, \tau, \lambda)$ and gamma $G(\alpha, \lambda^\tau)$. A GN variable Z can be obtained from a GG variable Y by $Z = Y^{\tau/2}$. For $s = \tau/2$, (16) gives the effect of this transformation. A gamma variable X can be obtained by the square transformation of a GN variable Z. The effect of this transformation is measured by (16) with $\tau = 2$ and $s = 2$. However, there is no simple relationship between $K_{GG}(Y : Y^{\tau/2})$, $K_{GG}(Y^{\tau/2} : Y^\tau)$, and $K_{GG}(Y : Y^\tau)$.

The simplest information theoretic model in the GG family is the exponential distribution $\mathscr{E}$. The exponential model can be obtained from GG sequentially in two ways: the MDI derivation of GG in Ω_θ reference to Weibull, followed by the power transformation of the Weibull variable Y to the exponential $X = Y^\tau$; or the power transformation of GG to gamma $X = Y^\tau$, followed by the MDI derivation of the gamma in the subclass of Ω_θ with $\alpha = 1$ and reference to exponential. For $s = \tau$ and $\alpha = 1$, (16) gives the transformation information from Weibull to the exponential. However, the transformation information for the two routes are different, in general.

The ME distribution subject to a single mean constraint is the exponential. Under the data scaling by the mean, a result of Soofi et al. (1995) gives

$$K(G^* : \mathscr{E}) = H(\mathscr{E}) - H(G^*). \tag{17}$$

The gamma distribution is ME in the class of distributions with the addition of the geometric mean constraint to the same mean constraint for exponential distribution being the ME model. By (17), the MDI function $K(G^* : \mathscr{E})$ quantifies the amount of entropy reduction (information gain) due to the use of geometric mean $E(\log X)$ in addition to the mean $E(X)$. In light of the MDI property (c) of the preceding section, the same interpretation holds for $K(GG^* : W)$ using $E(\log Y)$ and $E(X)$, where $X = Y^\tau$. That is, $K(G^* : \mathscr{E}) = K(GG^* : W)$ measures the information content of additional geometric constraint. The parameter α in the GG and gamma densities is due to the geometric mean constraint and the MDI function (15) may be interpreted as a measure of the "marginal" effect of α.

The information measure for the simultaneous effects of the transformation and the inclusion of geometric mean is $K(GG : \mathscr{E})$, obtained by (10) with $\alpha_0 = \tau_0 = 1$. There is no simple relationship between the discrimination information measures for the simultaneous and sequential cases.

3. Likelihood-based measures

The likelihood function based on a set of observations $\boldsymbol{y} = (y_1, \ldots, y_n)$ from $y \sim f(y|\alpha, \tau, \lambda) = GG(\alpha, \tau, \lambda)$ is

$$f(\boldsymbol{y}|\alpha, \tau, \lambda) = \left(\frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)}\right)^n \exp\left\{n\left[(\alpha\tau - 1)\,\overline{\log y} - \frac{\overline{y^\tau}}{\lambda^\tau}\right]\right\}, \tag{18}$$

where $\overline{y^\tau} = (1/n)\sum_{i=1}^n y_i^\tau$ and $\overline{\log y} = (1/n)\sum_{i=1}^n \log y_i$.

The likelihood equations for the derivatives of the log-likelihood function $L(\alpha, \tau, \lambda) = \log f(\boldsymbol{y}|\alpha, \tau, \lambda)$ with respect to α and λ are the two moment equations (6) and (7) with $\theta_1 = \overline{y^\tau}$ and $\theta_2 = \overline{\log y}$. These equations give $L(\hat{\alpha}, \hat{\tau}, \hat{\lambda}) = -n\hat{H}_{GG}$, where $\hat{H}_{GG}$ is given by (8) with the MLE estimates $\hat{\alpha}$, $\hat{\tau}$, and $\hat{\lambda}$ of the parameters. We therefore have the following entropy representation of the log-likelihood ratio statistic for the GG family:

$$-2\log\left[\frac{\max_{\alpha_0,\tau_0,\lambda_0} f_{GG_0}(\boldsymbol{y}|\alpha_0, \tau_0, \lambda_0)}{\max_{\alpha,\tau,\lambda} f_{GG}(\boldsymbol{y}|\hat{\alpha}, \hat{\tau}, \hat{\lambda})}\right] = 2n(\hat{H}_{GG_0} - \hat{H}_{GG}),$$

where $\alpha_0 = 1$ for Weibull, $\tau_0 = 1$ for gamma, $\alpha_0 = \tau_0 = 1$, for exponential, $\alpha_0 = 2\alpha$, $\tau_0 = 2$ for the GN, and $\hat{H}_{GG_0}$ is the respective entropy estimate by the MLE of the subfamily.

The AIC and BIC for the GG and its subfamilies may be represented as:

$$AIC = 2n\hat{H}_{GG} + 2k,$$

$$BIC = 2n\hat{H}_{GG} + k\log n,$$

where $k = 1$ for exponential and GN, $k = 2$ for gamma and Weibull, and $k = 3$ for GG.

4. Bayesian inference for discrimination information

Given data $y_1, \ldots, y_n$, discrimination information statistics for the GG family are obtained by estimating the Kullback–Leibler functions presented in the preceding section. We may estimate the discrimination information measures by estimating the parameters using the maximum likelihood, the methods of moments, generalized method of moments, and Bayesian procedures. These estimates of information provide descriptive statistics which are useful diagnostic measures for quantifying data information for discriminating between two GG models and between a GG and a model in its subfamilies. We provide Bayesian inference for these measures and mention in passing that the asymptotic properties of discrimination information statistics based on a set of consistent estimates are well known (Kullback, 1959).

4.1. Bayesian inference

Given a prior distribution $\pi(\alpha, \tau, \lambda)$, the Bayes Theorem gives posterior distribution

$$\pi(\alpha, \tau, \lambda|\boldsymbol{y}) \propto \pi(\alpha, \tau, \lambda) f(\boldsymbol{y}|\alpha, \tau, \lambda). \tag{19}$$

We use an inverse gamma prior for $\beta = \lambda^\tau \sim IG(a, b)$; i.e., an inverse GG prior for λ. Assuming that α, τ, and β parameters are independent, a priori, we obtain the following conditional posterior distributions:

$$\pi(\alpha|\tau, \lambda, \boldsymbol{y}) \propto [\lambda^{\alpha\tau}\Gamma(\alpha)]^{-n}\exp\{n(\alpha\tau - 1)\overline{\log y}\}\pi(\alpha), \tag{20}$$

$$\pi(\tau|\alpha, \lambda, \boldsymbol{y}) \propto \tau^n \exp\left\{n\left[(\alpha\tau - 1)\overline{\log y} - \frac{\overline{y^\tau}}{\lambda^\tau}\right]\right\}\pi(\tau), \tag{21}$$

$$\pi(\beta|\tau, \alpha, \boldsymbol{y}) = IG(a + n\alpha, b + n\overline{y^\tau}), \tag{22}$$

where $\pi(\alpha)$ and $\pi(\tau)$ are the priors for α and τ, respectively.

For any reasonable prior for $\pi(\alpha)$ and $\pi(\tau)$, the conditional posteriors (20) and (21) cannot be obtained in a familiar form. However, the conditional likelihoods in (20) and (21), are log concave. Thus, the conditional posteriors in (20) and (21) are both log concave densities if the priors $\pi(\alpha)$ and $\pi(\tau)$ are log concave. Consequently, the adaptive rejection sampling method of Gilks and Wild (1992) can be implemented to sample from (20) and (21) and a Gibbs sampler can be used to obtain samples from the joint posterior distribution (19). We will use uniform priors for α and τ over finite intervals. Whence samples from (19) are generated via the Gibbs sampler, Bayesian inference for the GG entropy (8) and discrimination information measures can be obtained as functions of the parameters.

4.2. Elaboration information

Another information measure of interest for Bayesian analysis of *GG* family is obtained via the elaboration approach of Carota et al. (1996). Let $\boldsymbol{\omega} = (\tau, \alpha)$ be the vector of elaboration parameters in the *GG* family. Then

$$K(\pi_{\omega|y} : \pi_{\omega}) = \int \pi(\boldsymbol{\omega}|\boldsymbol{y}) \log \frac{\pi(\boldsymbol{\omega}|\boldsymbol{y})}{\pi(\boldsymbol{\omega})} \mathrm{d}\boldsymbol{\omega}$$

is a measure of information in the data $\boldsymbol{y}$ about $\boldsymbol{\omega}$. As such, it is an elaboration information measure for embedding the exponential model in *GG* family.

It is easy to see that $K(\pi_{\omega|y} : \pi_{\omega})$ can be written as

$$K(\pi_{\omega|y} : \pi_{\omega}) = E_{\omega|y}\left[\log \frac{f(\boldsymbol{y}|\boldsymbol{\omega})}{f(\boldsymbol{y})}\right].$$

A linearized version of this expression can be found along the lines of Carota et al. (1996). However, we proceed with a Markov chain Monte Carlo (MCMC) inference for $K(\pi_{\omega|y} : \pi_{\omega})$. For this purpose, we write

$$K(\pi_{\omega|y} : \pi_{\omega}) = \sum_{i=1}^{n} E_{\omega|y}[\log f(y_i|\boldsymbol{\omega})] - \sum_{i=1}^{n} \log f(y_i). \tag{23}$$

It can be shown that under the IG prior for β,

$$\begin{aligned} f(y_i|\boldsymbol{\omega}) &= \int f(y_i|\boldsymbol{\omega}, \beta)\pi(\beta)\,\mathrm{d}\beta \\ &= \frac{1}{Beta(a+\alpha)} \frac{b^a \tau y_i^{\alpha\tau-1}}{(b+y_i^{\tau})^{a+\alpha}}, \end{aligned} \tag{24}$$

where $Beta(a+\alpha) = \Gamma(a+\alpha)/\Gamma(a)\Gamma(\alpha)$ is the beta function. For $x = y^{\tau}$, (24) gives the density of beta prime distribution (Zellner, 1971, p. 375). We therefore refer to (24) as the *generalized beta prime* distribution.

The first term in (23) can be easily approximated by a Monte Carlo integral using

$$E_{\omega|y}[\log f(y_i|\boldsymbol{\omega})] \approx \frac{1}{G}\sum_{g=1}^{G} \log f(y_i|\boldsymbol{\omega}^{(g)}).$$

Noting that we have posterior samples available (as generated by MCMC) and for each posterior realization $\boldsymbol{\omega}^{(g)}$, we evaluate $\log f(y_i|\boldsymbol{\omega}^{(g)})$ using the observed value of the y_i.

Evaluation of the marginal density $f(\boldsymbol{y})$ in (23) using prior samples to calculate $E_\omega[\log f(y_i|\boldsymbol{\omega})]$ is not very efficient as noted by many. A better but still computationally cheap alternative is to use the harmonic mean estimator of Newton and Raftery (1994) given by

$$f(\boldsymbol{y}) \approx \left[\frac{1}{G}\sum_{g=1}^{G}(f(\boldsymbol{y}|\boldsymbol{\omega})^{(g)})^{-1}\right]^{-1}.$$

This approach uses samples from the posterior $\pi(\boldsymbol{\omega}|y)$ and thus is computationally simple. Although it may be unstable, as noted by Kass and Raftery (1995), "... it often gives results that are accurate enough...".

Computation of elaboration information measures for α and τ individually, for embedding gamma and Weibull models in the GG family is more involved.

5. Bayesian inference for maximum entropy index

Maximum entropy fit indices and tests are constructed based on properties of the parametric family of the model. Consider the distributions in the moment class (5). If $GG^* \in \Omega_\theta$ is the ME model, then for any $F \in \Omega_\theta$, by the information distinguishability (ID) relation (Soofi et al., 1995), we have

$$K(F : GG^*|\boldsymbol{\theta}) = H(GG^* \in \Omega_\theta) - H(F \in \Omega_\theta). \tag{25}$$

That is, the discrepancy between ME distribution GG^* and any other distribution in Ω_θ is given by the difference between entropies of the two models.

Given observations $y_1, \ldots, y_n$ from an unknown distribution F, we can assess if $F \in \Omega_\theta$ using an ID statistic,

$$K(F_n : GG^*|\boldsymbol{\theta}_n) = H(GG^* \in \Omega_{\theta_n}) - H(F_n \in \Omega_{\theta_n}), \tag{26}$$

where $F_n \in \Omega_{\theta_n}$ is a nonparametric distribution estimate with entropy $H(F_n \in \Omega_{\theta_n})$ and moments $\boldsymbol{\theta_n} = (\mu_{1,n}, v_{1,n})$ and $GG^* \in \Omega_{\theta_n}$ is the ME model in the estimated moment class Ω_{θ_n}.

A normalized ID index is constructed as

$$\begin{aligned} ID(F_n : GG^*|\boldsymbol{\theta}_n) &= 1 - e^{-K(F_n:GG^*|\boldsymbol{\theta}_n)} \\ &= 1 - e^{-[H(GG^*\in\Omega_{\theta_n}) - H(F_n\in\Omega_{\theta_n})]}. \end{aligned} \tag{27}$$

This ID index is in the interval $0 < ID(F_n : GG^*|\boldsymbol{\theta}_n) < 1$, where $ID(F_n : GG^*|\boldsymbol{\theta}_n) \approx 0$ indicates that GG^* is a good fit.

In order to ensure the non-negativity of $K(F_n : GG^*|\boldsymbol{\theta}_n)$, the parameters of the ME model GG^* must be computed by the moments of F_n. Unlike the maximum likelihood and the full parametric Bayesian method, where all three GG parameters can be estimated, in the method of moments one can estimate (6) and (7) by $\theta_{1,n} = \mu_{\tau,n}$ and $\theta_{2,n} = v_n$ for a given τ and proceed with the moment class Ω_{θ_n}. In this case the moments are functions of τ and Ω_{θ_n} is indexed by τ. In the ME fitting procedure, one can use a grid search on τ and find parameter estimates $(\alpha_n, \tau_n, \lambda_n)$ that yield the best fit according to $ID(F_n : GG^*|\boldsymbol{\theta}_n)$.

Bayesian inference about the ME model parameters and ID fit (25) is obtained by the maximum entropy Dirichlet (MED) procedure developed by Mazzuchi et al. (2000, 2006).

For application of the MED procedure to the GG family, we consider the data-generating distribution F as an unknown member of Ω_θ. We specify a Dirichlet process prior for the unknown F,

$$F|GG^*, \mathscr{B} \sim \mathscr{D}(GG^*, \mathscr{B}),$$

where $GG^* \in \Omega_\theta$ is the ME model and $\mathscr{B}$ reflects the strength of belief about GG^*. As such, the ME model serves as the prior expected distribution (an initial guess) about which we can infer through updating the Dirichlet prior.

For any partition of the real line, $-\infty \leqslant \xi_0 < \xi_1 < \cdots < \xi_q \leqslant \infty$, the increments $\Delta F_k = F_k - F_{k-1}$, $k = 1, \ldots, q$, provide a quantized distribution $\boldsymbol{F} = (\Delta F_1, \ldots, \Delta F_q)$ which has a Dirichlet prior

$$\pi(\boldsymbol{F}) \propto (\Delta F_1)^{\alpha(\xi_1)-1} (\Delta F_2)^{\alpha(\xi_2)-\alpha(\xi_1)-1} \cdots (\Delta F_q)^{\mathscr{B}-\alpha(\xi_{q-1})-1}, \tag{28}$$

where $\alpha(\xi_k)$ is a measurable function defined over $\Re$ such that $\lim_{\xi\to\infty} \alpha(\xi) = \mathscr{B}$ and $\alpha(\xi_k) \equiv \alpha((-\infty, \xi_k]) = \mathscr{B} GG^*(\xi_k)$.

It is well-known that the posterior distribution of F based on a complete sample $\boldsymbol{y} = (y_1, \ldots, y_n)$ from F is also a Dirichlet process with the parameters updated as

$$\tilde{\mathscr{B}} = \mathscr{B} + n, \quad \text{and} \quad \tilde{\alpha}(\xi_k) = \alpha(\xi_k) + \sum_{i=1}^{q} \delta[y_i \leqslant \xi_k], \tag{29}$$

where $\delta[\cdot]$ is the indicator function of the set. The posterior for the quantized distribution $\boldsymbol{F}$ is Dirichlet with parameters (29). For each partition, the posterior mean of F_k is given by

$$\tilde{F}_k \equiv E[F_k | GG^*, \mathscr{B}, \boldsymbol{y}] = \frac{\mathscr{B}}{\mathscr{B}+n} GG^*(\xi_k) + \frac{n}{\mathscr{B}+n} \hat{F}(\xi_k).$$

Note that $\tilde{F}_k$ is a weighted average of the prior best guess GG_k^* and the empirical distribution $\hat{F}_k$. As $\mathscr{B} \to 0$, $\tilde{F}_k \to \hat{F}_k$. In the limit the prior is improper, but the posterior is proper with the mean being the empirical distribution. Also as $n \to \infty$, $\tilde{F}_k \to \hat{F}_k$.

The MED prior and posterior for $H(F \in \Omega_\theta)$ in (25) are derived via a quantized entropy. Mazzuchi et al. (2006) developed a general class of quantized entropies that can be used for this purpose. We use a special case constructed based on the partition of bin width h, referred to as the histogram partition: $\xi_k = (k-1)h$, $k = 1, \ldots, q$. For this case, the quantized entropy is given by

$$H_h^q(F) = -\sum_{k=1}^{q} \Delta F_k \log \frac{\Delta F_k}{h}. \tag{30}$$

This measure is analogous to a histogram entropy estimate developed by Hall and Morton (1993), which uses histogram probabilities $\Delta \hat{F}_k$ in (30). The bin width may be chosen according to an optimal rule suggested by Hall and Morton (1993). However, in the present context it is more apt to select h such that the histogram moments are about the same as the data moments.

For a given partition, the prior and posterior distributions of the quantized distribution $\boldsymbol{F}$ induce the prior and posterior for $H_h^q(F)$. The Bayes entropy estimate is the mean of the posterior distribution of the quantized entropy $\tilde{H}_h^q(F) = E[H_h^q(F)|GG^*, \mathscr{B}, \boldsymbol{y}]$. A closed form expression for $\tilde{H}_h^q(F)$ is given in Mazzuchi et al. (2006), where it is shown that for large sample, $\tilde{H}_h^q(F) \approx H_h^q(\hat{F})$ and the Bayes entropy estimate $\tilde{H}_h^q(F)$ is consistent.

For computing the discrimination information function (26) and ID index (27), we find α and λ using the quantized moment equations:

$$\begin{cases} \theta_{1,q} \approx \overline{y_q^{\tau}} = \sum_{k=1}^{q} \Delta F_k([k-.5]h)^{\tau} = \lambda_q^{\tau}\alpha_q, \\ \theta_{2,q} \approx \overline{\log y}_q = \sum_{k=1}^{q} \Delta F_k \log([k-.5]h) = \log \lambda_q + \frac{1}{\tau}\psi(\alpha_q), \end{cases} \tag{31}$$

with a given τ. Then τ parameter is estimated by iteration to minimize (32).

The MED procedure produces prior and posterior distributions for the moments in (31), by drawing samples from the Dirichlet prior (28) and posterior (29). Then the priors and posteriors for the model parameters α and λ are obtained using (31). The prior and posterior distributions of $H(GG^*|\boldsymbol{\theta}_q)$ are obtained from the distributions of the model parameters. The prior and posterior distributions of $K(F : GG^*|\boldsymbol{\theta})$ are estimated by $K_q(F : GG^*|\boldsymbol{\theta}_q)$ via (25) using the distributions of $H(GG^*|\boldsymbol{\theta}_q)$ and $H_h^q(F)$. The partition (particularly the endpoint) must be selected such that $K_q(F : GG^*|\boldsymbol{\theta}_q) \geqslant 0$.

The Bayes estimate of the $K(F : GG^*|\boldsymbol{\theta})$ is the posterior mean of $K_q(F : GG^*|\boldsymbol{\theta}_q)$, given by

$$\tilde{K}_q(F : GG^*|\boldsymbol{\theta_q}) = \tilde{H}(GG^*;\boldsymbol{\theta_q}) - \tilde{H}_h^q(F), \tag{32}$$

where $\tilde{H}(GG^*|\boldsymbol{\theta_q})$ is the Bayes estimate for the entropy of the parametric ME model, and $\tilde{H}_h^q(F)$ is the semiparametric (nonparametric) Bayes entropy estimate. If the ME model GG^* is not a suitable approximation for the true data-generating distribution F, then for large n, we can generally expect large values of $\tilde{K}_q(F : GG^*|\boldsymbol{\theta}_q)$; details are given in Mazzuchi et al. (2006).

For the index of fit, we compute the MED prior and posterior distributions of the *ID* index $ID_q(F : GG^*|\boldsymbol{\theta}_q)$. The Bayes estimate of the ID index is given by the posterior mean of $ID_q(F : GG^*|\boldsymbol{\theta}_q)$ and is referred to as *BID Index*.

The ME indices of fit for gamma $K_q(F : G^*|\boldsymbol{\theta}_q)$ and Weibull $K_q(F : W^*|\boldsymbol{\theta}_q)$ models are obtained similarly using $\tau = 1$ and $\alpha = 1$ in (31), respectively. The ME index for the exponential $K_q(F : \mathscr{E}^*|\theta_{1,q})$ is obtained by using the quantized mean constraint only.

6. Examples

We illustrate applications of the discrimination information measures and ME fit indices using two data sets. The first data set pertains to unemployment duration, drawn from the Bureau of Labor Statistics 2001. We studied unemployment data for females and males in rural and urban areas, and will report the results for female workers in the urban areas. The results of information analyses for other categories were all remarkably similar to those reported here. The second data set pertains to the tenure of CEO in their positions, drawn from *Standard and Poors ExecumComp*.

Table 1 gives the MLE entropy estimates, AIC, BIC, and the log-likelihood ratio statistics for the two data sets. For the unemployment data, AIC gives a slight edge to the *GG* model, BIC gives a slight edge to the exponential model, and the likelihood ratios are significant at 5% due to the sample size. For the CEO data, all measures are against exponentiality, AIC gives slight edges to the *GG* and gamma, as compared with the Weibull, and BIC gives a slight edge to gamma. The likelihood ratio for the gamma is negligible, for the Weibull is significant, and for the exponential is highly significant. Here, we should note that the MLE of *GG* parameters are computed based on Prentice (1974) approach, which does not work well for estimating the models from the subfamilies, and hence they are estimated using the standard MLE procedure. The use of different algorithms can create some difficulties in computation of the likelihood ratio when a shape parameter is close to one; e.g., giving a negative likelihood ratio statistic. We ran into that situation with the gamma likelihood ratio for the CEO data, and were able to correct the problem manually by increasing the number of decimal point of the MLE estimates.

Table 2 summarizes the MLE and posterior results for *GG* parameters, moments, and discrimination information measures. The posterior results are obtained by MCMC using the parametric Bayesian inference described in the previous section. These results are based on independent uniform priors for α and τ in the interval $[.5, 5]$, and $IG(3, .1)$ prior for $\theta = \lambda^{\tau}$. The posterior correlations for unemployment parameters are

Table 1
MLE estimates of entropy, AIC, BIC, and likelihood ratio statistic for two data sets

	Unemployment ($n = 802$)				CEO tenure ($n = 940$)			
	Entropy	AIC	BIC	LR statistic	Entropy	AIC	BIC	LR statistic
GG	3.7121	5960.27	5974.33		5.7899	10,890.94	10,905.48	
Gamma	3.7150	5962.85	5972.23	4.58	5.7905	10,890.20	10,899.89	1.26
Weibull	3.7161	5964.67	5974.05	6.40	5.7951	10,898.86	10,908.55	9.92
Exponential	3.7162	5962.73	5967.41	6.46	5.8935	11,081.78	11,086.63	194.84

Table 2
Bayesian posterior results for parameters and information measures

	Unemployment				CEO tenure			
	Parametric posterior				Parametric posterior			
	MLE	Mean	Median	95%	MLE	Mean	Median	95%
GG parameters								
α	1.17	1.31	1.31	1.43	2.44	2.56	2.59	2.92
τ	.94	.81	.81	.84	.87	.86	.84	.91
λ	12.48	11.83	11.76	13.40	46.50	43.49	41.50	60.43
Moments								
$E(Y)$	15.12	18.02	18.00	19.16	133.89	133.32	133.30	138.70
$E(Y^{\tau})$	12.55	9.66	9.58	10.43	68.88	65.87	60.13	104.40
$E(\log Y)$	2.18	2.17	2.23	2.38	4.61	4.61	4.62	4.66
Discrimination information (*Eq.* (10))								
Gamma	.003	.007	.005	.011	.001	.006	.002	.007
Weibull	.003	.009	.007	.017	.010	.015	.013	.019
Exponential	.001	.007	.005	.011	.104	.109	.108	.128
Geom. mean information (Eq. (15))	.007	.021	.021	.033	.178	.194	.199	.242
Transformation information (Eq. (16))	.025	.361	.361	.418	.700	.866	1.120	1.920
Elaboration information (Eq.(23))		13.46				404.76		

$Corr(\alpha, \tau) = -.27$, $Corr(\alpha, \lambda) = -.76$, $Corr(\tau, \lambda) = .52$, and for CEO parameters are $Corr(\alpha, \tau) = -.89$, $Corr(\alpha, \lambda) = -.96$, $Corr(\tau, \lambda) = .95$. Inference about discrimination between the *GG* and its subfamily based on (10) requires a pair of independent realizations for each parameter involved.

As seen in Table 2, for the unemployment data, both of the *GG* shape parameters are close to one. The discrimination information measures are all about the same and close to zero. Thus, the unemployment data do not discriminate between *GG* and its simpler subfamilies. For the CEO data, the estimates for α is about 2.5 and for τ is about .9. The discrimination information clearly indicates gamma is very close to *GG*. The measures for Weibull is ten-fold and exponential is hundred-fold of the measure for gamma, indicating that they are not so close to *GG*. The information measures for the geometric mean constraint and transformation are substantial for the CEO data, as compared with the unemployment data. The elaboration information is also quite large (it is nearly 30 times larger than for the unemployment).

Table 3 summarizes the results for the ME fit index. In the MED prior specification we set $\mathscr{B} = 5$ to reflect a weak degree of belief in the ME distribution; (i.e., the relative weights of the prior and sample are 5 and 802 for the unemployment, and 5 and 940 for CEO data, respectively). By this, we let the data dominate the posterior. The posterior means are practically the same as the results that we obtained using histogram entropy as a

Table 3
The MED posterior statistics for two data sets and their surrogates

	Unemployment ($\tau = .90$)			CEO tenure ($\tau = .85$)		
	Mean	Median	95%	Mean	Median	95%
Actual data						
GG parameters						
α	1.23	1.23	1.29	2.73	2.72	2.92
λ	11.38	11.33	12.38	39.65	39.59	43.48
Entropy						
H_{GG}	3.70	3.70	3.76	5.77	5.77	5.81
$H_h^q(F)$	3.61	3.61	3.67	5.74	5.74	5.78
Information measure *ID*						
GG	.086	.085	.108	.031	.029	.043
Gamma	.087	.087	.108	.050	.049	.064
Weibull	.085	.085	.105	.163	.163	.184
Exponential	.092	.092	.114	.142	.142	.163
Surrogate data						
GG parameters						
α	1.41	1.41	1.49	2.46	2.46	2.63
λ	10.11	10.94	11.17	42.42	42.38	46.23
Entropy						
H_{GG}	3.72	3.72	3.78	5.75	5.75	5.79
$H_h^q(F)$	3.70	3.70	3.76	5.74	5.74	5.78
Information measure *ID*						
GG	.015	.015	.027	.011	.010	.021
Gamma	.021	.020	.031	.030	.029	.041
Weibull	.025	.024	.034	.118	.118	.138
Exponential	.027	.026	.037	.099	.099	.117

descriptive measure. In the ME fit procedure, a value for the shape parameter τ must be selected. The τ parameters shown in Table 3 are found based on an ID analysis using various values of τ, for the best fit of *GG* and the best fit for Weibull.

As seen in Table 3, the MED procedure confirms the results of parametric Bayesian procedure for the unemployment data, reported above. These results lead to inferring that there is no need for a more complex model than the exponential. The MED results for CEO shed some lights on the results found by the parametric Bayesian procedure. The ME indices rank the fit of *GG* as the best, followed by gamma, exponential, and Weibull. The indices for *GG* and gamma are rather close.

The MED prior and posterior distributions for the ID indices of *GG* and its subfamilies for the CEO Tenure data are shown in Fig. 1. We note that the posterior distributions of ID indices for *GG* and gamma are concentrated near zero, and for the Weibull and exponential are concentrated near .2. Fig. 2 shows the MED prior and posterior distributions for the parameters α and λ of the *GG* distribution for the CEO Tenure data. These posteriors can be used for Bayesian inferential purposes.

The lower section of Table 3 shows the results for surrogate data, in which data are simulated using the parameter estimates from the actual data ($\alpha = 1.2$, $\tau = .9$, $\lambda = 11.4$ for the unemployment and $\alpha = 2.7$, $\tau = .85$, $\lambda = 40.0$ for CEO tenure). Surrogate analysis is common in physical sciences for evaluating performance of a methodology. The results of surrogate analysis show the same pattern as those found for the actual data, thereby providing additional confidence for application of the MED procedure.

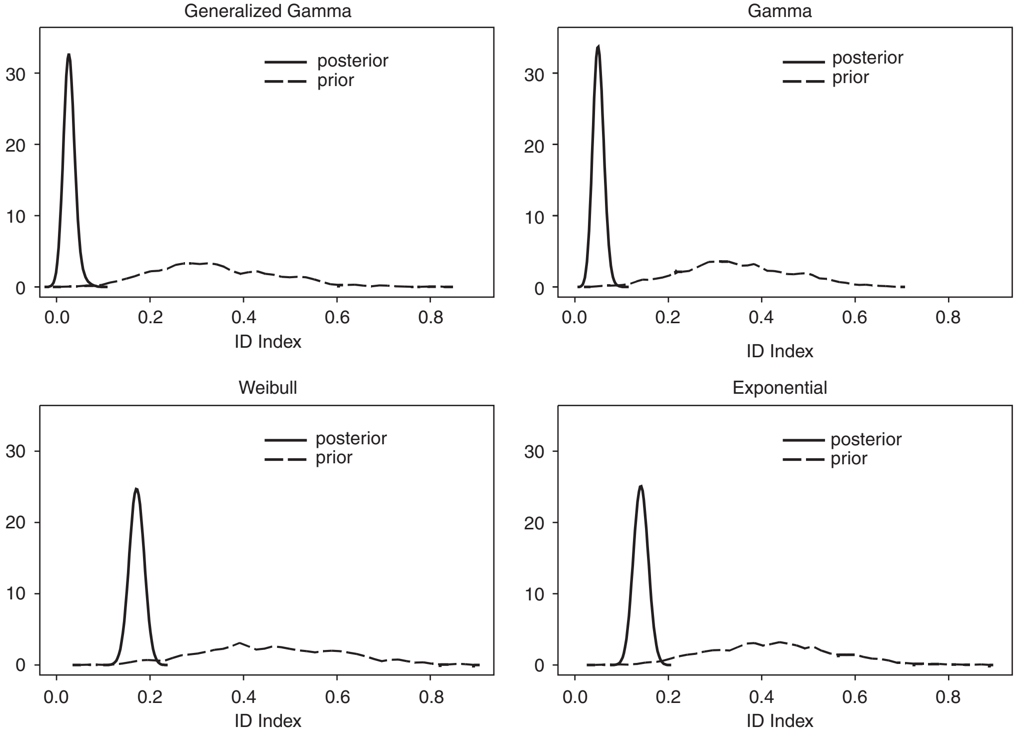

Fig. 1. Prior and posterior distributions of the ID indices for *GG* and its subfamilies.

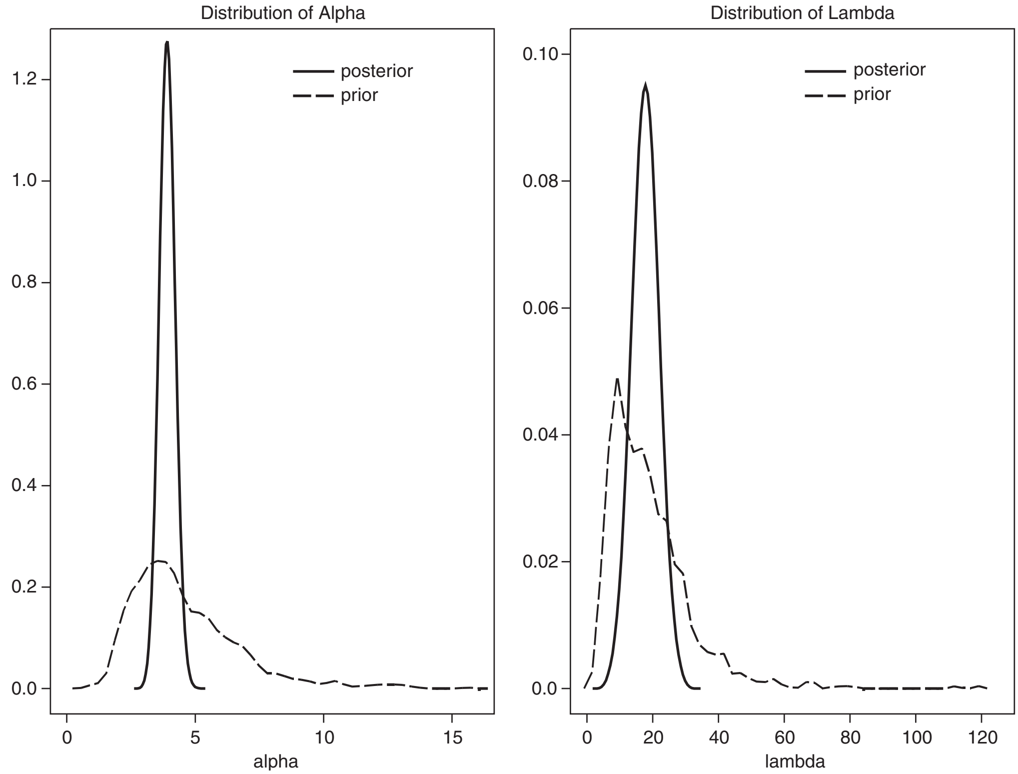

Fig. 2. Prior and posterior of the shape (α) and scale (λ) parameters for *GG* distribution.

7. Concluding remarks

This paper took the first major step toward closing the gap between the growing presence of *GG* model in duration analysis literature and its remarkable absence in the information studies. We presented some information properties of the *GG* distribution and showed that its flexibility leads to an assortments of information measures for the family. These information functions provide insights and can serve various data analysis purposes such as MDI modeling and data transformation. We gave entropy representations of the likelihood based measures for the *GG* family and presented Bayesian inference procedures for the *GG* information measures and for the fit of the *GG* model to a histogram.

As much as the flexibility of *GG* family provides a rich ground for information studies, its complexity provides challenges that yet to be taken. Examining the relationships between information functions for data transformation before and after MDI modeling (i.e., from exponential to Weibull and from gamma to *GG*) can provide new insights about the interrelationship among these distributions. Developing procedures for the components of the exponential elaboration to *GG* (i.e., gamma elaboration to *GG*, Weibull elaboration to *GG*) can shed light on the sequences of elaborations from exponential to *GG*. Computation of MLE and parametric Bayesian inference need to be improved so that all information measures involving *GG* distribution can be evaluated based on a common procedure. Finally, in order to provide a complete framework for applied econometric analysis, the *GG* distributional information measures developed in this paper should be complemented with information measures for *GG* regression. This issue is currently under study by the authors.

References

Allenby, G.M., Leone, R.P., Jen, L., 1999. A dynamic model of purchase timing with application to direct marketing. Journal of the American Statistical Association 94, 365–374.

Alwan, L.C., Ebrahimi, N., Soofi, E.S., 1998. Information theoretic framework for process control. European Journal of Operational Research 111, 526–542.

Audretsch, D.B., Mahmoud, T., 1995. New firm survival: new results using a hazard function. Review of Economics and Statistics 77, 97–103.

Bernardo, J.M., Rueda, R., 2002. Bayesian hypothesis testing: a reference approach. International Statistics Review 70, 351–372.

Blumenthal, M.A., 1988. Auctions with constrained information: blind bidding for motion pictures. The Review of Economics and Statistics 70, 191–198.

Carota, C., Parmigiani, G., Polson, N.G., 1996. Diagnostic measures for model criticism. Journal of the American Statistical Association 91, 753–762.

Diaz, M.D.M., 1999. Extended stay at university: an application of multinomial logit and duration models. Applied Economics 31, 1411–1422.

Diebold, F.X., Rudebusch, G.D., 1990. A nonparametric investigation of duration dependence in the American business cycle. Journal of Political Economy 98, 596–616.

Ebrahimi, N., Maasoumi, E., Soofi, E.S., 1999. Ordering univariate distributions by entropy and variance. Journal of Econometrics 90, 317–336.

Eckstein, Z., Wolpin, K.I., 1995. Duration to first job and the return to schooling: estimates from a search-matching model. Review of Economic Studies 62, 263–286.

Favero, C.A., Pesaran, M.H., Sharma, S., 1994. A duration model of irreversible oil investment: theory and empirical evidence. Journal of Applied Econometrics 9, 95–112.

Genesove, D., Mayer, C.J., 1997. Equity and time to sale in the real estate market. American Economic Review 87, 255–269.

Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41, 337–348.

Gronberg, T.J., 1994. Estimating workers' marginal willingness to pay for job attributes using duration data. Journal of Human Resources 29, 911–931.

Hager, H.W., Bain, L.J., 1970. Theory and methods inferential procedures for the generalized gamma distribution. Journal of the American Statistical Association 65, 1601–1609.

Hall, P., Morton, S.C., 1993. On the estimation of entropy. Annals of Institute of Mathematical Statistics 45, 69–88.

Jaggia, S., 1991. Specification tests based on the heterogeneous generalized gamma model of duration: with an application to Kennan's strike data. Journal of Applied Econometrics 6, 169–180.

Jeffreys, H., 1946. An invariant form for the prior probability in estimation problems. Proceedings of Royal Statistical Society (London) A 186, 453–461.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, vol. 1, second ed. Wiley, New York.

Kapur, J.N., 1989. Maximum Entropy Models in Science and Engineering. Wiley, New York.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.

Keifer, N.M., 1988. Economic duration data and hazard functions. Journal of Economic Literature 26, 646–676.

Kiefer, N.M., 1984. A simple test for heterogeneity in exponential models of duration. Journal of Labor Economics 2, 539–549.

Kiefer, N.M., Burdett, K., Sharma, S., 1985. Layoffs and duration dependence in a model of turnover. Journal of Econometrics 28, 51–69.

Kullback, S., 1959. Information Theory and Statistics. Wiley, New York (reprinted in 1968 by Dover).

Lancaster, T., 1979. Econometric methods for the duration of unemployment. Econometrica 47, 939–956.

Mazzuchi, T.A., Soofi, E.S., Soyer, R., 2000. Computations of maximum entropy Dirichlet for modeling lifetime data. Computational Statistics and Data Analysis 32, 361–378.

Mazzuchi, T.A., Soofi, E.S., Soyer, R., 2006. Bayes estimate of entropy and information index of fit. Submitted for publication.

McDonald, J., Butler, R.J., 1987. Some generalized mixture distributions with an application to unemployment duration. The Review of Economics and Statistics 69, 232–240.

Nadarajah, S., Zografos, K., 2003. Formulas for Rényi information and related measures for univariate distributions. Information Science 155, 119–138.

Newton, M.A., Raftery, A.E., 1994. Approximate bayesian inference with the weighted likelihood bootstrap. Journal of the Royal Statistical Society Series B 56, 3–48.

Orbe, J., Ferreira, E., Nunez-Anton, V., 2002. Length of time spent in Chapter 11 bankruptcy: a censored partial regression model. Applied Economics 34, 1949–1957.

Prentice, R.L., 1974. A log gamma model and its maximum likelihood estimation. Biometrika 61, 539–544.

Soofi, E.S., Retzer, J.J., 2002. Information indices: unification and applications. Journal of Econometrics 107, 17–40.

Soofi, E.S., Ebrahimi, N., Habibullah, M., 1995. Information distinguishability with application to analysis of failure data. Journal of the American Statistical Association 90, 657–668.

Stacy, E.W., 1962. A generalization of the gamma distribution. The Annals of Mathematical Statistics 33, 1187–1192.

Vakratsas, D., Bass, F.M., 2002. A segment-level hazard approach to studying household purchase timing decisions. Journal of Applied Econometrics 17, 49–59.

Yamaguchi, K., 1992. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of ‘‘permanent employment’’ in Japan. Journal of the American Statistical Association 87, 284–292.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York (reprinted in 1996).