Information Criteria for Specification Search in Generalized Gamma Family

Ali Dadpay
Department of Economics
University of Wisconsin-Milwaukee
P.O.Box 431, Milwaukee, WI 53201
adadpay@uwm.edu

Ehsan S. Soofi
School of Business Administration
University of Wisconsin-Milwaukee
P.O.Box 742, Milwaukee, WI 53201
esoofi@uwm.edu

Refik Soyer
Department of Management Science
George Washington University
Washington D.C. 20052
soyer@gwu.edu

March 31, 2004

Abstract

Analysis of duration data is increasingly growing in business and economic research. The distributional assumptions often used for the duration variables are exponential, gamma, Weibull, and lognormal. Recently, the generalized gamma (GG) model, which is more complex than these models and includes them as sub-families, has been gaining popularity. This paper addresses the following basic modeling question. Do the data necessitate use of a complex GG model vis a vis its simpler subfamilies? We propose three types of information criteria for this specification search. First, we obtain information criteria for discriminating among GG models and its sub-families. We also discuss the full parametric Bayesian inference for the model parameters and these information measures. Then, we develop an information index for the fit of a GG distribution to the data. The inference about the fit uses a semi-parametric Bayesian procedure, referred to as the Maximum Entropy Dirichlet (MED). The MED procedure uses the data moments to derive a GG as a tentative model and incorporates uncertainty about the model via the Dirichlet process prior. The MED generates prior and posterior distributions of an information index for assessing the fit of the GG model. As byproducts, MED also produces prior and posterior distributions for the model parameters. Furthermore, we obtain an information index for the regression fit when the dependent variable is GG distributed. Applications of the GG information indices are demonstrated using data on the duration of licensing for extraction of oil in North Sea. The results show that the data do not discriminate between the GG and its sub-families, the Weibull model attains the best fit, and the parametric Bayes inference is sensitive but the MED procedure is robust with respect to the prior specification.

Keywords: Entropy, Exponential, Gamma, Kullback-Leibler information, Model fitting, Weibull.

1 Introduction

Analysis of duration data is increasingly used in various areas of economics and related fields. In labor economics, examples include studies of the duration of unemployment, (Lancaster 1979, Kiefer 1984, McDonald and Butler, 1987, Steiner 1991, Yamaguchi 1992, Plasman 1993, van den Berg 1994, Steiner and Puhani 1997, Winter-Ember 1998, Abrevaya and Hausman 1999), turnover in labor market (Kiefer, Burdett, Sharma 1985), length of contract (Gronberg 1994), and duration of strike (Kennan and Wilson 1989, Jaggia 1991, Ondrich and Schnell 1993). Examples in other areas include studies of firms survival (Audretsch and Mahmoud 1995), duration that firms spend under Chapter 11 (Orbe, Ferreira and

Nunez-Anton 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, Pesaran, and Sharma 1994), household interpurchase time (Vakratsas and Bass 2002), interpurchase time in financial market (Allenby, Leone and Jen 1999), and length of the time that new movies stay on screens (Blumenthal 1988, Throsby 1994, De Vany and Walls 1996).

The main interest in economics is developing models for a duration variables in terms of other economic factors. The assumption used for the distribution of duration, therefore, is of prime importance for the analysis. Distributions that are used in economic problems are exponential (Kiefer 1984, Kennan and Wilson 1989, Diebold and Rudebusch 1990, Diaz 1999), lognormal (Eckstein and Wolpin 1995, Diaz 1999, Vakratsas and Bass 2002), gamma (Lancaster 1979, Kooreman and Ridder 1983, Steiner 1991, Plasman 1993, van der Berg 1994, Hernaes and Storm 1997, Winter-Ember 1998), Weibull (Atkinson 1984, Favero et al 1994, Engle and Russell 1995), generalized gamma (Jaggia 1991, Yamaguchi 1992, Bergrstom and Edin 1992, van den Berg and van Ours 1996, Allenby, Leone and Jen 1999), and inverse Gaussian (van der Linden 2000).

Significance of the distributional assumption and its consequences in duration analysis have been recognized by several authors (Heckman and Singer 1984, Han and Hausman 1990, and Engle and Russell 1995). Although some authors have proposed nonparametric approaches (Han and Hausman 1990, Huh and Sickles 1994, Imbens 1994, Anton, Sainz, Rodrigues-Poo 2001), several others have used a flexible family of distributions. McDonald and Butler (1987) used a mixture distribution for modeling duration of unemployment. Engle and Russell (1995) used Weibull family which includes the exponential and compared the two in an autoregressive conditional context. Several authors have used generalized gamma model (McDonald and Butler 1987, Jaggia 1991, Yamaguchi 1992, Bergrstom and Edin 1992, van den Berg and van Ours 1996, Allenby, Leone and Jen 1999).

The generalized gamma distribution (GG), introduced by Stacy (1962), is flexible in variety of shapes and implied hazard functions. It encompasses exponential, gamma, and Weibull as subfamilies, and lognormal as a limiting distribution. 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 non linear transformation of GG model. Some authors (e.g. Jaggia, 1991) have argued that the flexibility of GG makes it suitable for duration analysis, while some other authors (e.g. Allenby et al, 1999) have advocated use of simpler models because of 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? This is a specification search question, which is addressed in this paper. For this purpose, we develop information criteria that serve as diagnostic tools for discriminating between the

GG and its subfamilies, for assessing the fit of GG to the data, and for the regression fit of GG.

This paper is organized as follows. Section 2 discusses information theoretic properties of the GG family. Section 3 develops three types of information criteria: information criteria for discriminating between a GG distribution and its subfamilies; an information index for the fit of a GG model to the data; and an information index for assessing the regression fit of GG distributed variables. Section 4 discusses Bayesian inference for the GG information criteria. Section 5 illustrates application of the GG information criteria on the duration of oil exploration considering the choice between the generalized gamma, exponential, gamma, and Weibull models.

2 Information Properties GG Family

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

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

where $\Gamma(\alpha)$ is the gamma function. The parameters α and τ determine the shape of the distribution and λ is the scale parameter. Other scale parameterization found in the literature are λ^{τ}, λ^{-1}, and $\lambda^{-\tau}$.

The subfamilies of GG are exponential when $\alpha = \tau = 1$, gamma when $\tau = 1$, Weibull when $\alpha = 1$, Half-normal when $\alpha = \frac{1}{2}, \tau = 2$, and lognormal as a limiting distribution when $\alpha \to \infty$.

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,

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

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

The moments of GG distribution are given by

$$\mu_s(\alpha,\tau,\lambda) = E_{GG}(Y^s|\alpha,\tau,\lambda) = \frac{\lambda^s\Gamma(\alpha + s/\tau)}{\Gamma(\alpha)}, \quad s > 0. \tag{4}$$

The geometric mean is given by

$$\nu(\alpha,\tau,\lambda) = E_{GG}(\log Y) = \log\lambda + \frac{1}{\tau}\psi(\alpha), \tag{5}$$

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

The entropy of GG distribution is given by

$$\begin{aligned} H(GG) \equiv H[f_{GG}] &= -\int_0^\infty f_{GG}(y|\alpha,\tau,\lambda)\log f_{GG}(y|\alpha,\tau,\lambda)dy \\ &= \log\lambda + \log\Gamma(\alpha) + \alpha - \log\tau + \left(\frac{1}{\tau}-\alpha\right)\psi(\alpha). \end{aligned} \tag{6}$$

For specific values of the parameters, (6) gives entropy expressions for gamma, Weibull, exponential, and half-normal distributions.

The entropy ordering in GG family in terms of its parameters are easily obtained. 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 note that

$$\frac{\partial H(GG)}{\partial\alpha} \geq 0 \quad \text{for} \quad \left[\alpha - \frac{1}{\psi'(\alpha)}\right]^{-1} \geq \tau$$

$$\frac{\partial H(GG)}{\partial\tau} \geq 0 \quad \text{for} \quad \tau \geq -\psi(\alpha),$$

where $\psi'(\alpha)$ is the trigamma function. Since $\left[\alpha - \frac{1}{\psi'(\alpha)}\right]^{-1}$ is decreasing in α with a limiting value of 2, the entropy is increasing for all α when $\tau \leq 2$. However, since $\psi(\alpha) < 0$ for $\alpha < 1.5$ approximately, the entropy can be increasing in τ only when $\alpha < 1.5$. These findings include the entropy orderings of gamma and Weibull distributions obtained by Ebrahimi, Maasoumi, and Soofi (1999) as special cases.

The GG distribution is maximum entropy (ME) in the moment class of distributions

$$\Omega_\theta = \{F(y|\boldsymbol{\theta}) : E_f[T_j(y)|\theta] = \theta_j,\ j = 0,1,2\} \tag{7}$$

where $\boldsymbol{\theta} = (\theta_0,\theta_1,\theta_2),\ \theta_0 = 1,\ \theta_1 = \mu_\tau(\alpha,\tau,\lambda),\ \theta_2 = \nu(\alpha,\tau,\lambda)$, $T_0(y) = 1,\ T_1(y) = y^\tau,\ T_2(y) = \log y$, and τ is given; see Kapur (1989) who referred to (1) as generalized Weibull distribution.

The Kullback-Leibler information between two GG models $GG_1 = GG(\alpha_1,\tau_1,\lambda_1)$ and $GG_2 = GG(\alpha_2,\tau_2,\lambda_2)$ is given by:

$$\begin{aligned} K(GG_1 : GG_2) &= \int f_{GG_1}(y)\log\frac{f_{GG_1}(y)}{f_{GG_2}(y)}dy \\ &= \log\frac{\phi_\tau}{\phi_\lambda^{\alpha_1\phi_\tau}} - \log\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_2)} - \alpha_1 + \mu_1(\alpha_1,\phi_\tau,\phi_\lambda) + (\alpha_1\phi_\tau - \alpha_2)\nu(\alpha_1,\phi_\tau,\phi_\lambda), \end{aligned} \tag{8}$$

where $\phi_\tau = \tau_1/\tau_2,\ \phi_\lambda = (\lambda_1/\lambda_2)^{\tau_2},\ \mu_1(\alpha_1,\phi_\tau,\phi_\lambda)$ is the moment of order $s = 1$ defined in (4) and $\nu(\alpha_1,\phi_\tau,\phi_\lambda)$ is the geometric mean defined in (5); proof is given in the Appendix.

Thus the discrimination information between two GG models 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_1 and α_2. Although $K(GG_1 : GG_2)$ is a complicated function, it is a general representation for discrimination information functions between all combinations of GG, gamma, Weibull, exponential, and half-normal distributions.

3 Information Criteria for GG Modeling

Several modeling diagnostics are derived from the information properties of GG family.

3.1 Information Criteria for Selecting Among GG Models

It is well-known that $K(GG_1 : GG_2) \geq 0$; equality holds if and only if $f_{GG_1}(y) = f_{GG_2}(y)$ for all y in the support of the distributions. Since $K(GG_1 : GG_2)$ is not symmetric, GG_2 is the reference distribution; $K(GG_1 : GG_2)$ is also referred to as cross-entropy and relative entropy. Estimates of (8) with specified GG_2 provide Information Criteria for selecting among the GG models and the models in the subfamilies (GGIC).

The information criteria for selecting a GG model with specified parameter values is obtained by letting $GG_2 = GG_0 = GG(\alpha_0, \tau_0, \lambda_0)$ and estimating the parameters of GG_1 by $\alpha_1 = \hat{\alpha}, \tau_1 = \hat{\tau}, \lambda_1 = \hat{\lambda}$ in expression (8). That is,

$$\hat{K}(GG : GG_0) = \log \frac{\hat{\phi}_\tau}{\hat{\phi}_\lambda^{\hat{\alpha}\hat{\phi}_\tau}} - \log \frac{\Gamma(\hat{\alpha})}{\Gamma(\alpha_0)} - \hat{\alpha} + \mu_1(\hat{\alpha}, \hat{\phi}_\tau, \hat{\phi}_\lambda) + (\hat{\alpha}\hat{\phi}_\tau - \alpha_0)\nu(\hat{\alpha}, \hat{\phi}_\tau, \hat{\phi}_\lambda). \tag{9}$$

For any set of consistent estimates $\hat{\alpha}$, $\hat{\tau}$ and $\hat{\lambda}$, the information statistic $2n\hat{K}(GG : GG_0)$ is approximately distributed as chi-square with three degrees of freedom Kullback (1959, p. 98). That is, the information statistic $2n\hat{K}(GG : GG_0)$ is asymptotically equivalent to the likelihood ratio test where the data is used instead of the expectation and the parameters are estimated by the MLE. However, for the discrimination information test (9) the use of MLE as well as other parameter estimates such as the methods of moments, generalized method of moments, and Bayes are allowed.

Some specific cases of (9) that are of interest in data analysis are testing for the subfamilies of the GG family. A GGIC for the fit of a gamma model $G(\alpha, 1, \lambda)$ in place of a GG can be obtained by letting $\alpha_1 = \alpha_2 = \hat{\alpha}, \tau_1 = \hat{\tau},\ \tau_0 = 1$, and $\lambda_1 = \lambda_2 = \hat{\lambda}$. This gives the information diagnostics for discrimination between the GG and gamma models:

$$\hat{K}(GG : G) = \log \hat{\tau} - \hat{\alpha} + \mu_1(\hat{\alpha}, \hat{\tau}, 1) + \hat{\alpha}(\hat{\tau} - 1)\nu(\hat{\alpha}, \hat{\tau}, 1). \tag{10}$$

As shown in Figure 1, the seemingly simple specification of $\tau = 1$ can have profound distributional consequences. Figure 1a shows the density of $G(1.5, 1) = GG(1.5, 1, 1)$ (solid) and the densities of $GG(1.5, 1.5, 1)$ (dashed) and $GG(1.5, .5, 1)$ (dotted). We note that although the τ parameter of the two GG distributions differ the same amount (.5) from the specified value of $\tau = 1$, the shapes of the GG densities are quite different. The discrepancy between each GG and the gamma distribution is depicted in Figure 1b in terms of the log-ratio of GG to gamma densities. The information diagnostic (10) summarizes these discrepancies in terms of their averages with respect to each GG distribution: $K(GG : G) = 1.05$ for $GG(1.5, .5, 1)$ and $K(GG : G) = .008$ for $GG(1.5, 1.5, 1)$.

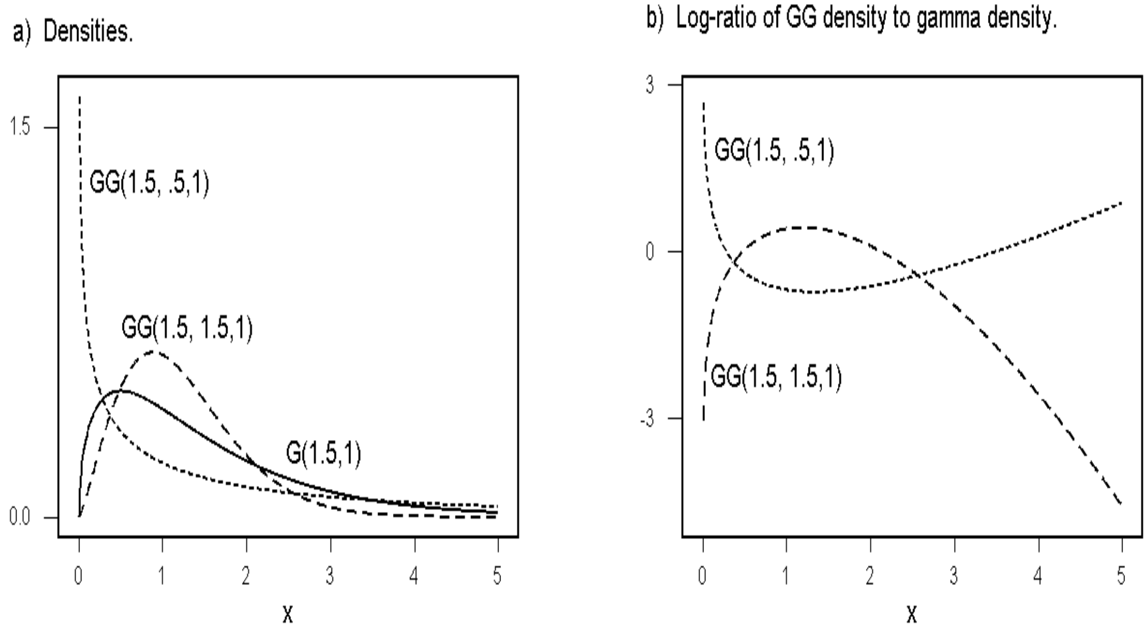

Figure 1: A gamma and two generalized gamma densities and their log-ratios.

A diagnostic for the fit of a Weibull model $W(\tau, \lambda)$ in place of a GG can be obtained by letting $\tau_1 = \tau_2 = \hat{\tau}, \alpha_1 = \hat{\alpha}, \alpha_2 = 1$, and $\lambda_1 = \lambda_2 = \hat{\lambda}$. This gives the information diagnostics for discrimination between the GG and Weibull models:

$$\hat{K}(GG : W) = -\log \Gamma(\hat{\alpha}) + (\hat{\alpha} - 1)\psi(\hat{\alpha}). \tag{11}$$

We note that here the discrepancy measure is independent from τ. Specification of $\alpha = 1$ can also have profound distributional consequences as discussed for the gamma case.

A diagnostic for the fit of an exponential model in place of a GG can be obtained by letting $\tau_1 = \hat{\tau}, \alpha_1 = \hat{\alpha}, \alpha_2 = \tau_2 = 1$, and $\lambda_1 = \lambda_2 = \hat{\lambda}$. This gives the information diagnostics for discrimination between the GG and exponential models:

$$\begin{aligned} \hat{K}(GG : E) &= \log \hat{\tau} - \log \Gamma(\hat{\alpha}) - \hat{\alpha} + \mu_1(\hat{\alpha}, \hat{\tau}, 1) + (\hat{\alpha}\hat{\tau} - 1)\nu(\hat{\alpha}, \hat{\tau}, 1) \\ &= \hat{K}(GG : G) + \hat{K}(GG : W) - (\hat{\alpha} - 1)(\hat{\tau} - 1)\nu(\hat{\alpha}, \hat{\tau}, 1). \end{aligned} \tag{12}$$

The second expression indicates that $K(GG : E)$ is not a simple additive composition of $K(GG : G)$ and $K(GG : W)$. The last term in the second expression can be positive or negative. That is, depending on the values of the two shape parameters, the information discrepancy between a GG and exponential may be less or more that the sum of information discrepancies between a GG and gamma and a GG and Weibull.

Diagnostics for the fit for half-normal and models with the subfamilies can be obtained similarly.

3.2 Maximum Entropy Fit Index

Maximum entropy fit indices and tests are constructed based property of the parametric family of the model. Consider the distributions in the moment class (7). If $GG^* \in \Omega_\theta$ is the ME model, then for any $F \in \Omega_\theta$, by the information distinguishability (ID) relation (Soofi, Ebrahimi, and Habibullah 1995), we have

$$K(F : GG^*|\boldsymbol{\theta}) = H(GG^* \in \Omega_\theta) - H(F \in \Omega_\theta). \tag{13}$$

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.

The ID relationship (13) provides information diagnostics for assessing the fit of GG^* to the data. Given observations $y_1, \cdots, y_n$ from an unknown distribution F, we can assess if $F \in \Omega_\theta$ using an ID statistic

$$K(F_n : GG^*|\boldsymbol{\theta}_n) = H(GG^* \in \Omega_{\theta_n}) - H(F_n \in \Omega_{\theta_n}), \tag{14}$$

where $F_n \in \Omega_{\theta_n}$ is a nonparametric distribution estimate with entropy $H(F_n \in \Omega_{\theta_n})$ and moments $\boldsymbol{\theta_n} = (\mu_{1,n}, \nu_{1,n})$ 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

$$\begin{aligned} ID(F_n : GG^*|\boldsymbol{\theta}_n) &= 1 - e^{-K(F_n:GG^*|\boldsymbol{\theta}_n)} \\ &= 1 - e^{-[H(GG^* \in \Omega_{\theta_n}) - H(F_n \in \Omega_{\theta_n})]}. \end{aligned} \tag{15}$$

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.

Several procedures for computing the nonparametric entropy are available, see, e.g., Soofi and Retzer (2002). A simple procedure, developed by Hall and Morton (1993), is the entropy of a histogram,

$$H_h^q(\hat{F}) = -\sum_{k=1}^{q} \Delta\hat{F}_k \log \frac{\Delta\hat{F}_k}{h}, \tag{16}$$

where $\Delta\hat{F}_k = \dfrac{n_k}{n},\ k = 1, \cdots, q$ are the relative frequencies in the bin width h. The histogram entropy is a consistent nonparametric entropy estimate.

In order to ensure the non-negativity of $K(F : GG^*|\boldsymbol{\theta})$, the parameters of the ME model GG^* must by computed by the moments of the histogram. In the moment class of distributions (7), τ must be specified. Then Ω_θ provides two moment equations from which α and λ can be estimated. For a given τ, the following histogram moment equations give solutions for α and λ:

$$\begin{cases} \theta_{1,q} = \overline{y_q^\tau} = \displaystyle\sum_{k=1}^{q} \Delta\hat{F}_k([k - .5]h)^\tau = \lambda_q^\tau \alpha_q \\ \theta_{2,q} = \overline{\log y}_q = \displaystyle\sum_{k=1}^{q} \Delta\hat{F}_k \log([k - .5]h) = \log \lambda_q + \dfrac{1}{\tau}\psi(\alpha_q). \end{cases} \tag{17}$$

Then the maximum entropy $H(GG^* \in \Omega_{\theta_q})$ in (13) is computed using the solutions to (17), and an ID statistic is obtained by

$$K_q(F : GG^*|\boldsymbol{\theta}_q) = H(GG^*; \boldsymbol{\theta}_q) - H_h^q(F).$$

The parameter values $(\alpha_q, \tau_q, \lambda_q)$ that yield the best fit (minimum $K_q(F : GG^*|\boldsymbol{\theta}_q)$) are chosen by a grid search for τ.

The ME fit index (15) encompasses the ME fit indices previously used for gamma, Weibull, and exponential models. The moment equations for gamma, Weibull, and exponential models can be obtained from (17). For gamma, the moment equations are obtained by letting $\tau = 1$ in (17),

$$\begin{cases} \theta_{1,q} = \bar{y}_q = \sum_{k=1}^{q} \Delta\hat{F}_k([k-.5]h) = \lambda_q \alpha_q \\ \theta_{2,q} = \overline{\log y}_q = \sum_{k=1}^{q} \Delta\hat{F}_k \log([k-.5]h) = \log \lambda_n + \psi(\alpha_q) \end{cases}$$

For the Weibull ME characterization the moment equations are obtained by using $\alpha = 1$ in (17),

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

Similarly, for the exponential ME characterization a single moment equation is needed, obtained by using $\alpha = \tau = 1$ in the first equation in (17),

$$\theta_{1,q} = \bar{y}_q = \sum_{k=1}^{q} \Delta\hat{F}_k([k-.5]h) = \lambda_q.$$

3.3 Covariate Information

A covariate information index can be obtained for GG regression via the information R^2 index of the exponential family regression (Cameron and Windmeijer 1997). A distribution is said to be in the exponential family if its density can be written in the following equivalent forms:

$$\begin{aligned} f_\mu(y) &= a(\mu)g(y)e^{-b(\mu)y} \\ f_\eta(y) &= c(\boldsymbol{\eta})g(y)e^{\eta_1 d_1(y)+\cdots+\eta_J d_J(y)} \end{aligned} \tag{18}$$

where $\mu = E(Y)$, $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_J)$ is the vector of natural parameters, $a(\mu)$, $b(\mu)$, and $c(\boldsymbol{\eta})$ are all free from y, and $g(y)$ and $T_j(y)$, $j = 1, \cdots, J$ are all free from the parameters.

The GG density (1) can not be written in the forms shown in (18). However, we can utilize the facts that $Z = Y^\tau$ has a gamma distribution which is in the exponential family. This power transformation enables us to utilize all the methodological advantages of the exponential family. In particular, for a given τ, the information decompositions that are available for the exponential family (Kullback 1971, Simon 1973) are applicable to the transformation of GG data $z_i = y_i^\tau$.

In GG regression, $Z_i,\ i = 1, \cdots, n$ are independent and have gamma distribution with mean $\mu_\tau = E(Z_i) = E(Y_i^\tau) = \alpha\lambda^\tau$. The regression is given by $\mu_\tau = \varphi_\tau^{-1}$, where

$$\varphi_\tau = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi},$$

is the canonical link between μ_τ and the covariates $\boldsymbol{X_i} = (X_{1i} + \cdots + \beta_p X_{pi})$.

Hastie (1987) applied an information decomposition of exponential family to the generalized linear model (GLIM) and referred to it as "analysis of deviance". Decomposition of information for GLIM gives

$$\sum_{i=1}^{n} K(g_{z_i} : g_{\bar{z}}) = \sum_{i=1}^{n} K(g_{z_i} : g_{\hat{\mu}_{\tau_i}}) + \sum_{i=1}^{n} K(g_{\hat{\mu}_{\tau_i}} : g_{\bar{z}}), \tag{19}$$

where g_z is the gamma density, $z_i,\ i = 1, \cdots, n$ are the transformed data, $\bar{z} = \overline{y^\tau} = \frac{1}{n}\sum_{i=1}^{n} y_i^\tau$, and $\hat{\mu}_{\tau_i} = \hat{\varphi}_{\tau_i}^{-1}$ is computed from the regression estimate $\hat{\varphi}_{\tau_i} = \hat{\beta}' X_i$, (e.g., MLE). For the normal regression (19) gives the analysis of variance decomposition of the least squares regression.

The covariate information index for GG regression is given by

$$IR_{GG}^2 = 1 - \frac{\sum_{i=1}^{n}[(y_i^\tau - \hat{\mu}_{\tau_i})/\hat{\mu}_{\tau_i} - \log(y_i^\tau/\hat{\mu}_{\tau_i})]}{\sum_{i=1}^{n} \log(y_i^\tau/\overline{y^\tau}\,)}. \tag{20}$$

A high value of IR_{GG}^2 indicates strong explanatory power of the covariates.

4 Bayesian Inference

This section describes procedures for Bayesian inference about the GG parameters, the parametric modeling diagnostics (10)-(12), the GG entropy (6), and the ID index (15).

4.1 Parametric Inference

Given a set of observations $\boldsymbol{y} = (y_1, \cdots, y_n)$ from $y \sim f(y|\alpha, \tau, \lambda) = GG(\alpha, \tau, \lambda)$, the likelihood function is

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

where $\overline{y^\tau}$ is defined in (19) and $\overline{\log y} = \frac{1}{n}\sum_{i=1}^{n} \log y_i$.

Using a prior distribution $\pi(\alpha, \tau, \lambda)$ the inference about the parameters will be made by the posterior distribution given by the Bayes Theorem

$$\pi(\alpha, \tau, \lambda|\boldsymbol{y}) \propto \pi(\alpha, \tau, \lambda) f(\boldsymbol{y}|\alpha, \tau, \lambda). \tag{22}$$

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

$$\pi(\alpha|\tau,\lambda,\boldsymbol{y}) \quad \propto \quad [\lambda^{\alpha\tau}\Gamma(\alpha)]^{-n} \exp\left\{n(\alpha\tau-1)\overline{\log y}\right\} \; \pi(\alpha) \tag{23}$$

$$\pi(\tau|\alpha,\lambda,\boldsymbol{y}) \quad \propto \quad \tau^n \exp\left\{n\left[(\alpha\tau-1)\overline{\log y} - \frac{\overline{y^\tau}}{\lambda^\tau}\right]\right\} \; \pi(\tau) \tag{24}$$

$$\pi(\phi|\tau,\alpha,\boldsymbol{y}) \quad = \quad IG\left(a+n\alpha,\; b+n\overline{y^\tau}\;\right) \tag{25}$$

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 (23) and (24) can not be obtained in a familiar form. However the conditional likelihoods in (23) and (24), are log concave. Thus an attractive feature of the conditional posteriors in (23) and (24) is that they are both log concave densities if the priors $\pi(\alpha)$ and $\pi(\tau)$ are log concave. That is, even for vague priors $\pi(\alpha) \propto \frac{1}{\alpha}$ and $\pi(\tau) \propto \frac{1}{\tau}$ the conditional posteriors (23) and (24) will be log concave. Consequently, the adaptive rejection sampling method of Gilks and Wild (1992) can be implemented to sample from the posterior densities (23) and (24). Therefore Gibbs sampler is applicable based on the full conditional distributions (23)-(25). Whence samples from (23)-(25) are generated via MCMC, Bayesian inference for the GG entropy (6) and discrimination information diagnostics (10)-(12) can be obtained as functions of the parameters.

4.2 Maximum Entropy Dirichlet Inference

Bayesian inference about the ME model parameters and ID fit (13) can be obtained using the Maximum Entropy Dirichlet (MED) procedure developed by Mazzuchi, Soofi and Soyer (2000, 2003). In the MED procedure, the Dirichlet process prior for the unknown data distribution F is specified as

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

where the ME model GG^* is the prior expected distribution (an initial guess) and $\mathcal{B}$ reflects the strength of belief about GG^*. For any partition of the real line, $-\infty \le \xi_0 < \xi_1 < \cdots < \xi_q \le \infty$, the increments $\Delta F_k = F_k - F_{k-1},\ k = 1,\cdots,q$, provide a quantized distribution $\boldsymbol{F} = (\Delta F_1,\cdots,\Delta F_q)$ which has a Dirichlet prior

$$\pi(\boldsymbol{F}) \propto (\Delta F_1)^{\alpha(\xi_1)-1}(\Delta F_2)^{\alpha(\xi_2)-\alpha(\xi_1)-1}\cdots(\Delta F_q)^{\mathcal{B}-\alpha(\xi_{q-1})-1}, \tag{26}$$

where $\alpha(\xi_k)$ is a measurable function defined over $\Re$ such that

$$\lim_{\xi\to\infty}\alpha(\xi) = \mathcal{B} \quad \text{and} \quad \alpha(\xi_k) \equiv \alpha((-\infty,\xi_k]) = \mathcal{B}GG^*(\xi_k).$$

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

$$F|GG^*, \mathcal{B}, \boldsymbol{y} \sim \mathcal{D}(\tilde{F}, \tilde{\mathcal{B}}), \tag{27}$$

with the parameters updated by

$$\tilde{\mathcal{B}} = \mathcal{B} + n, \qquad \text{and} \qquad \tilde{\alpha}(\xi_k) = \alpha(\xi_k) + \sum_{i=1}^{q} \delta[y_i \leq \xi_k] \tag{28}$$

where $\delta[\cdot]$ is the indicator function of the set. The posterior for the quantized distribution $\boldsymbol{F}$ is Dirichlet with parameters (28).

For each partition, the posterior mean of F_k is given by

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

Note that $\tilde{F}_k$ is a weighted average of the prior best guess GG^*_k and the empirical distribution $\hat{F}_k$. As $\mathcal{B} \to 0$, $\tilde{F}_k \to \hat{F}_k$. At 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 (13) are derived via a quantized entropy. Mazzuchi, et al (2003) 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: $\xi_k = (k-1)h, \; k = 1, \cdots, q$. For this case, the quantized entropy is given by

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

This quantized entropy is analogous with the histogram entropy (16).

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 given by

$$\begin{aligned} \tilde{H}_h^q(F) &\equiv E[H_h^q(F)|GG^*, \mathcal{B}, \boldsymbol{y}] \\ &= \psi(\tilde{\mathcal{B}}+1) - \log h - \sum_{k=1}^{q} (\Delta\tilde{F}_k)\psi[\tilde{\mathcal{B}}(\Delta\tilde{F}_k) + 1] \end{aligned} \tag{29}$$

It is shown by Mazzuchi, et al (2003) 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 (14) and ID index (15), we use the quantized moments obtained by replaying $\Delta\hat{F}_k$ in (17) with ΔF_k. In addition to ensuring positivity of the Kullback-Leibler function, by using (17) the MED procedure also produce prior and posterior for the moments via drawing samples from the Dirichlet prior (26) and posterior (27). Then the priors and posteriors for the model parameters α and λ are obtained using (17). The distributions of the maximum entropy $H(GG^*|\boldsymbol{\theta}_q)$ are obtained from the distributions of the model parameters. The prior and posterior distributions of $K(F : GG^*|\boldsymbol{\theta})$ are estimated via (13) using the distributions of $H(GG^*|\boldsymbol{\theta}_q)$ and $H_h^q(F)$.

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

$$\tilde{K}_q(F : GG^*|\boldsymbol{\theta_q}) = \tilde{H}(GG^*; \boldsymbol{\theta_q}) - \tilde{H}_h^q(F) \tag{30}$$

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 (29). 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. (2003).

For the index of fit, we compute the MED prior and posterior distributions of the ID index by

$$ID_q(F : GG^*|\boldsymbol{\theta}_q) = 1 - \exp\{H_h^q(F) - H(GG^*|\boldsymbol{\theta}_q)\}. \tag{31}$$

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*.

5 Application to Oil Exploration Data

This example uses data from Favero et al (1994) who studied duration of development stages in oil field explorations. They used Weibull and Cox's proportional hazard models for estimating the relationship between the duration and a number of economic factors. Hurn and Wright (1994) also investigated this problem using a logit model.

We consider developing a duration model for the first stage, which is exploring the capacity of the oil field. This stage ends by applying for a license to extract oil from the field. The data set includes 57 observations measured in months. The sample mean is 63.21 and median is 51.00 months. The coefficient of skewness is .93, indicating that the distribution of duration variable is skewed.

Table 1 summarizes the MLE and parametric Bayes results. The MLE for the GG parameters lead to the moment estimates that are close to the sample moments. Both of the GG shape parameters are close to one. The GGIC measures are $\hat{K}(GG : G) = .03$, $\hat{K}(GG : W) = .02$, and $\hat{K}(GG : E) = .03$. The corresponding information statistics are $2n\hat{K}(GG : G) = 3.31$, $2n\hat{K}(GG : W) = 1.94$, and $2n\hat{K}(GG : E) = 3.31$ which are not significant as compared with the upper percentiles of the chi-square distribution.

The posterior results shown in Table 1 are obtained by MCMC using the parametric Bayesian inference described in the previous section. These results are based on uniform priors for α and τ in the interval $[.5, 5]$, and $IG(1, .1)$ prior for $\phi = \lambda^\tau$. The posterior statistics of the discrimination information between GG and its subfamilies indicate that any of the three simpler models can be used

Table 1. MLE and parametric Bayesian posterior statistics for Oil Exploration Data

	MLE	Parametric Posterior Mean	Median	95%
Model parameters				
α	.87	1.55	1.54	2.38
τ	1.22	.89	0.85	1.29
λ	76.51	40.18	35.77	87.73
Moments				
$E(Y)$	63.28	63.36	62.68	81.47
$E(\log Y)$	3.66	3.67	3.68	3.95
Entropy				
H_{GG}	5.14	4.57	4.50	5.23
Information measure $GGIC$				
Gamma	.03	.07	.05	.23
Weibull	.02	.21	.10	.69
Exponential	.03	.24	.21	.64

in place of GG, with the gamma being the best. However, we should note that the parametric Bayes results are very sensitive to the prior specification. For example using uniform priors for α and τ in the interval $[.5, 10]$, lognormal, or loggamma priors for these parameters led to some convergence problems, bimodal posteriors, and/or substantially different results.

Table 2 summarizes the results for the ME fit index and the MED posteriors based on the histogram partition. In the MED prior specification we set $\mathcal{B} = 5$ to reflect a week degree of belief in the ME distribution; (i.e., the relative weights of the prior and sample are 5 and 57, respectively). In the ME fit procedure, a value for the shape parameter τ must be selected. Based on an ID analysis using various values of τ, for the data histogram, the best fit GG is obtained when $\tau = 1.25$. The best fit Weibull is obtained when $\tau = 1.50$. The histogram moments are $\bar{y}_q = 63.89$, $\overline{\log y}_q = 3.71$, and $\overline{y_q^\tau} = 159.20$, which are close to those obtained by the MLE. The other two GG parameters, α and λ are computed by solving the ME moment equations (17). By the MED analysis the best fit GG is obtained when $\tau = 1.20$. The best fit Weibull is obtained when $\tau = 1.60$.

As seen in Table 2, both the histogram and MED procedures choose the Weibull as the best fit model. Thus, both procedures formally confirm the choice of distributional assumption made by Favero et al (1994). According to the histogram ID indices, the second best model is gamma, followed by the exponential, then by GG. However, according to the MED results, GG and gamma are tied for the second best and the fit index for the exponential is also close.

Table 2. Histogram ME fit and MED posterior statistics for Oil Exploration Data

	Histogram ($\tau = 1.25$)	MED Posterior ($\tau = 1.20$) Mean	Median	95%
Model parameters				
α	.88	.94	.92	1.12
λ	76.95	70.81	70.01	85.55
Entropy				
H_{GG}	5.16	5.10	5.11	5.28
$H_h^q(F)$	5.04	4.95	4.95	5.10
Information measure ID				
GG	.121	.144	.140	.239
Gamma	.025	.146	.140	.243
Weibull	.002	.089	.084	.184
Exponential	.113	.160	.155	.246

Figure 2 shows the MED prior and posterior distributions for the ID indices of GG and its subfamilies. We note that the prior distributions of the ID index give nonnegligible probabilities to ID values in the entire interval $(0, 1)$. For all four models the posterior distributions of the index shift to the left and concentrate in an interval near zero, indicating that all of these are acceptable models for the data. The posterior distribution for the Weibull is slightly to the left of the posteriors for the other three models.

Figure 3 shows the MED prior and posterior distributions of the shape parameter α and scale parameter λ of the GG distribution. Both figures indicate that there is lot of learning from data as implied by the highly concentrated posterior distributions. We note that the posterior distribution of the shape parameter α is concentrated around one, which confirms the best fit of the Weibull distribution for the data. The scale parameter λ is concentrated around the MLE. Unlike the parametric Bayes procedure, the MED inference is robust with respect to the specification of the Dirichlet prior parameters.

Application of the information index for the regression fit is illustrated through examining the relationship between the duration of licensing and the size of the oil field determined by the amount of recoverable oil reserves. The size of an oil field determines the net present value of the investment for given oil price. Consequently, the size has been used as and economic factor for explaining variations of duration of development stages across different oil fields. Since for a large amount of recoverable oil reserves, the net present value of investments becomes positive for lower oil price, an inverse

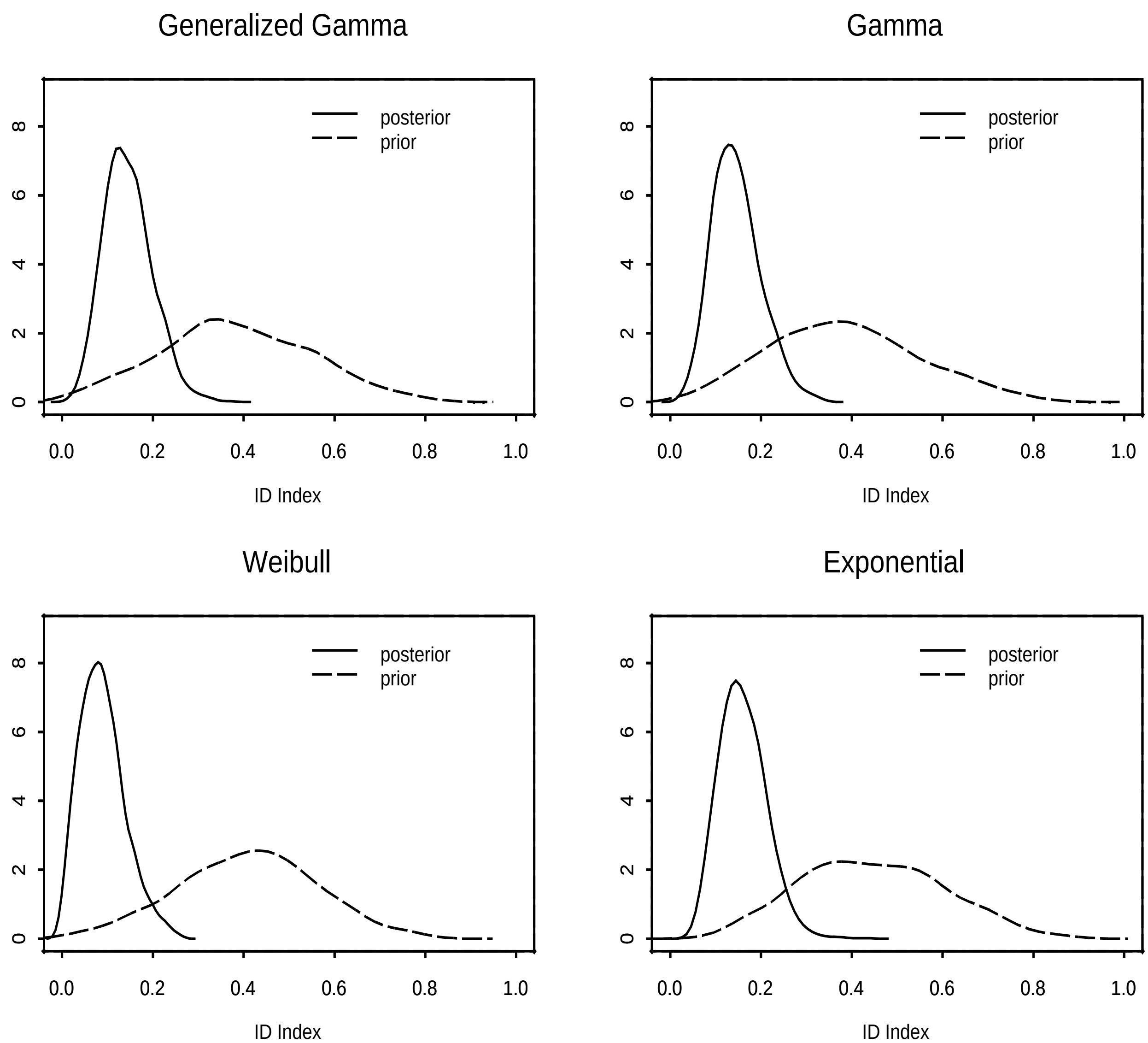

Figure 2: Prior and posterior distributions of the ID indices for GG and its subfamilies.

relationship between duration and the size is expected.

Table 3 summarizes the regression results. The results are shown for a normal, a lognormal, a gamma, and two GG regression models. The residual plots and tests of normality rejected the normal and lognormal assumptions. The results for these models are included for the purpose of comparison. As seen in Table 3, the fit index of normal regression is much lower than that for the other models and the fit of lognormal regression is relatively lower than the gamma and GG regression models. The gamma assumption is justified by the $GGIC$ statistic and leads to the best regression fit. The power transformation $Y^{1.2}$ is based on the best fit GG and $Y^{1.6}$ is based on the best fit Weibull obtained by

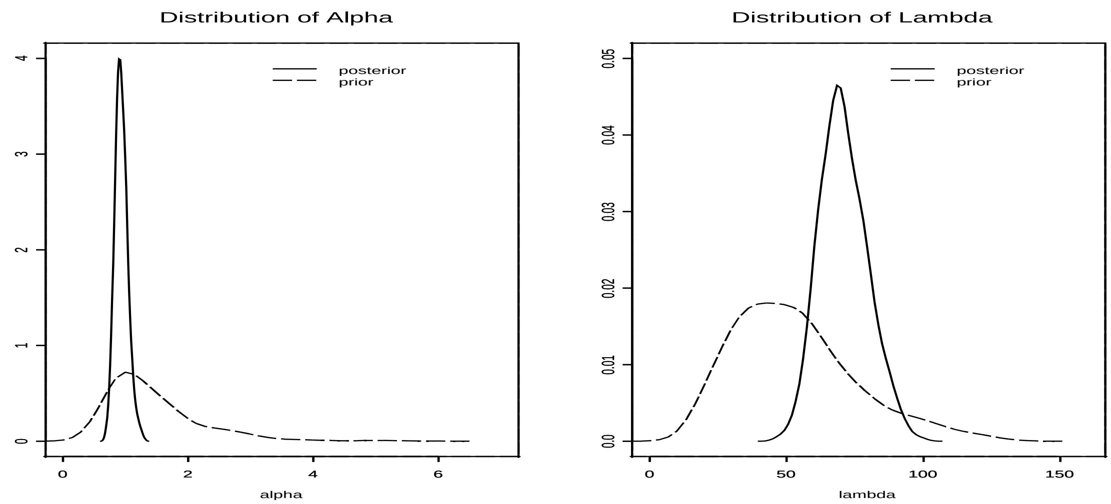

Figure 3: Prior and posterior of the shape (α) and scale (λ) parameters for GG distribution

the MED inference. The fit indices for both transformations are close to the case of gamma regression.

Table 3. Regression Results for the Oil Exploration Data

Link function	Normal & Lognormal $\varphi = \mu$		Gamma & GG $\varphi_\tau = \mu_\tau^{-1}$		
Response Transformation	Y	$\log Y$	Y	$Y^{1.2}$	$Y^{1.6}$
Coefficient	-.240	-.006	3×10^{-4}	2×10^{-4}	4×10^{-5}
F-ratio (Chi-sq)	7.52	15.09	(10.58)	(9.93)	(8.48)
Information R^2	12.8%	22.8%	28.6%	28.0%	25.6%
Distribution Fit	No	No	Yes	Yes	Yes

Appendix

Derivation of $K(GG_1 : GG_2)$

Note that

$$
\begin{aligned}
K(GG_1 : GG_2) &= E_1\left[\log \frac{f_{GG_1}(Y)}{f_{GG_2}(Y)}\right] \\
&= \log \frac{\tau_1}{\tau_2} - \log \frac{\lambda_1^{\alpha_1\tau_1}}{\lambda_2^{\alpha_2\tau_2}} - \log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_2)} + (\alpha_1\tau_1 - \alpha_2\tau_2)E_1(\log Y) \\
&\quad -\frac{1}{\lambda_1^{\tau_1}} E_1\left(Y^{\tau_1}\right) + \frac{1}{\lambda_2^{\tau_2}} E_1\left(Y^{\tau_2}\right),
\end{aligned}
$$

where E_1 denotes the expectation with respect to GG_1. By (4), (5), and the relation $\Gamma(\alpha_1 + 1) = \alpha_1\Gamma(\alpha_1)$, we obtain

$$
\begin{aligned}
K(GG_1 : GG_2) &= \log \frac{\tau_1}{\tau_2} - \log \frac{\lambda_1^{\alpha_1\tau_1}}{\lambda_2^{\alpha_2\tau_2}} - \log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_2)} + (\alpha_1\tau_1 - \alpha_2\tau_2)\left(\log \lambda_1 - \frac{1}{\tau_1}\psi(\alpha_1)\right) \\
&\quad -\alpha_1 + \frac{1}{\lambda_2^{\tau_2}} \frac{\lambda_1^{\tau_2}\Gamma(\alpha_1 + \tau_2/\tau_1)}{\Gamma(\alpha_1)}.
\end{aligned}
$$

Combining the log terms that involve the scale parameters and letting $\phi_\tau = \tau_1/\tau_2$, $\phi_\lambda = (\lambda_1/\lambda_2)^{\tau_2}$, give

$$
\begin{aligned}
K(GG_1 : GG_2) &= \log \frac{\phi_\tau}{\phi_\lambda^{\alpha_1\phi_\tau}} - \log \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_2)} + (\alpha_1\phi_\tau - \alpha_2)\left(\log \phi_\tau - \frac{1}{\phi_\tau}\psi(\alpha_1)\right) \\
&\quad -\alpha_1 + \frac{\phi_\tau\Gamma(\alpha_1 + 1/\phi_\tau)}{\Gamma(\alpha_1)}.
\end{aligned}
$$

The expression (8) is obtained by (4) and (5).

References

Abrevaya, J., and Hausman, J. A., 1999. Rank Estimation of a Transformation Model with Observed Truncation. Econometrics Journal. 2, 292-305.

Allenby, G.M., Leone, R.P., and Jen, L., 1999. A Dynamic Model of Purchase Timing with Application to Direct Marketing. Journal of the American Statistical Association, 94, 365-74.

Anton, A.A., Sainz, A.F., Rodrigues-Poo, J., 2001. Semiparametric Estimation of a Duration Model. Oxford Bulletin of Economics and Statistics, 63, 517-33.

Atkinson, A. B., 1984. Unemployment Benefit, Duration and Incentives in Britain: How Robust Is the Evidence?. Journal of Public Economics. 23, 3-26.

Audretsch D.B. and Mahmoud T., 1995. New Firm Survival: New Results Using a Hazard Function. Review of Economics and Statistics. 77, 97-103.

Bergrstom R. and Edin P.A., 1992. Time Aggregation and the Distributional Shape of Unemployment Duration. Journal of Applied Econometrics. 7, 5-30.

Blumenthal, M. A., 1988 Auctions with Constrained Information: Blind Bidding for Motion Pictures. The Review of Economics and Statistics. 70, 191-198.

Cameron, A. C. and Windmeijer, F. A. G., 1997. An R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models. Journal of Econometrics. 77, 329-342.

De Vany, A. and Walls, A. D., 1996. Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry. The Economic Journal. 106, 1493-1514

Diaz, M. D. M., 1999. Extended Stay at University: An Application of Multinomial Logit and Duration Models. Applied Economics. 31, 1411-22.

Diebold, F.X. and 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., and Soofi, E. S., 1999. Ordering Univariate Distributions by Entropy and Variance. Journal of Econometrics. 90, 317-336.

Eckstein, Z. and Wolpin, K. I., 1995. Duration to First Job and the Return to Schooling: Estimates from a Search-Matching Model. Review of Economic Studies. 62, 263-86.

Engle, R. F. and Russell, J. R., 1994. Forecasting Transaction Rates: The Autoregressive Conditional Duration Model. National Bureau of Economic Research Working Paper: 4966.

Favero, C.A., Pesaran, M.H and Sharma, S. 1994. A Duration Model of Irreversible Oil Investment: Theory and Empirical Evidence. Journal of Applied Econometrics. 9, 95-112.

Genesove, D. and Mayer, C. J., 1997. Equity and Time to Sale in the Real Estate Market. American Economic Review. 87, 255-69.

Gilks, W. R. and 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, 911-31.

Hager, H. W. and 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. and Morton, S. C., 1993. On the Estimation of Entropy. Annals of Institute of Mathematical Statistics. 45, 69-88.

Han, A. and Hausman, J. A., 1990. Flexible Parametric Estimation of Duration and Competing Risk Models. Journal of Applied Econometrics. 5, 1-28.

Hastie, T. (1987). A Closer Look at the Deviance. The American Statistician. 41, 16-20.

Heckman, J. and Singer, B., 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. Econometrica. 52, 271-320.

Hernaes, E. and Strom, S., 1996. Heterogeneity and Unemployment Duration. Labour. 10, 269-96.

Huh, K. and Sickles, R. C., 1994. Estimation of the Duration Model by Nonparametric Maximum Likelihood, Maximum Penalized Likelihood, and Probability Simulators. Review of Economics and Statistics. 76, 683-94.

Hurn, A. S. and Wright, R. E., 1994. Geology or Economics? Testing Models of Irreversible Investment Using North Sea Oil Data. Economic Journal. 104, 363-71

Imbens, G. W., 1994. Transition Models in a Non-stationary Environment. Review of Economics and Statistics. 76, 703-20.

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.

Kapur, J. N., 1989. Maximum Entropy Models in Science and Engineering. N.Y.: Wiley.

Kennan, J. and Wilson, R., 1989. Strategic Bargaining Models and Interpretation of Strike Data. Journal of Applied Econometrics. 4, S87-S130.

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. and Sharma, S., 1985. Layoffs and Duration Dependence in a Model of Turnover. Journal of Econometrics. 28, 51-69.

Kooreman, P. and Ridder, G., 1983. The Effects of Age and Unemployment Percentage on the Duration of Unemployment: Evidence from Aggregate Data. European Economic Review. 20, 41-57.

Kullback, S (1971), "Marginal Homogeneity of Multidimensional Contingency Tables", *The Annals of Mathematical Statistics*, 42, 594-606.

Kullback, S. (1959), *Information Theory and Statistics*, N.Y.: Wiley (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. and 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. and Soyer, R., 2003. Bayes Estimate of Entropy and Information Index of Fit.

McDonald, J. and Butler, R. J. , 1987. Some Generalized Mixture Distributions with an Application to Unemployment Duration. The Review of Economics and Statistics. 69, 232-240.

Ondrich, J. I. and Schnell, J. F., 1993. Strike Duration and the Degree of Disagreement. Industrial Relations. 32, 412-31.

Orbe, J., Ferreira, E. and Nunez-Anton, V., 2002. Length of Time Spent in Chapter 11 Bankruptcy: A Censored Partial Regression Model. Applied Economics. 34, 1949-57.

Plasman, R., 1993. Times Are Changing: Working Time in Fourteen Industrialised Countries: Belgium; By Times are changing: Working time in 14 industrialised countries. Geneva: International Institute for Labour Studies, 87-103.

Prentice, R. L., 1974. A Log Gamma Model and Its Maximum Likelihood Estimation. Biometrika. 61, 539-544.

Simon, G., 1973. Additivity of Information in Exponential Family Probability Laws. Journal of the American Statistical Association. 68, 478-482.

Soofi, E. S., and Retzer, J.J., 2002. Information Indices: Unification and Applications. Journal of Econometrics. 107, 17-40.

Soofi, E. S., Ebrahimi, N, and 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.

Steiner, V., 1991. Long-term Unemployment, Heterogeneity, and State Dependence: New Microeconometric Evidence on Unemployment Persistence. Empirica, 17, 41-59.

Steiner, V. and Puhani, P. A., 1997. The Effectiveness and Efficiency of Active Labour Market Policies in Poland. Empirica. 24, 209-31.

Throsby, D., 1994. The Production and Consumption of the Arts: A View of Cultural Economics. Journal of Economic Literature. 32, 1-29.

Vakratsas, D. and Bass, F. M., 2002. A Segment-Level Hazard Approach to Studying Household Purchase Timing Decisions. Journal of Applied Econometrics. 17, 49-59.

van den Berg, G.J., 1994. The Effects of Changes of the Job Offer Arrival Rate on the Duration of Unemployment. Journal of Labor Economics. 12, 478-98.

van den Berg, G. J. and van Ours, J. C., 1996. Unemployment Dynamics and Duration Dependence. Journal of Labor Economics. 14, 100-125.

van der Linden, B., 2002. Is Basic Income a Cure for Unemployment in Unionized Economies? A General Equilibrium Analysis. Annales d'Economie et de Statistique. v. 0, 81-105.

Winter-Ember, R., 1998. Unknown Wage Offer Distribution and Job Search Duration. Economics Letters. 60, 237-42.

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, 284-92.