Brazilian Journal of Probability and Statistics (2006), 20, pp. 67–91.

The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models

Ralph S. Silva[1], **Hedibert F. Lopes**[2] and **Helio S. Migon**[1]
[1] *Universidade Federal do Rio de Janeiro*
[2] *University of Chicago*

Abstract: We examine the class of *extended generalized inverse Gaussian* (EGIG) distributions. This class of distributions, which appeared briefly in a monograph by Jørgensen (1982), is more deeply and broadly studied in this paper. We start by deriving its probabilistic properties. Furthermore, we use the EGIG family in two popular and important statistical problems, one from survival analysis and the other from financial econometrics. In the first one, a survival model, we compare our results with those obtained by Achcar and Bolfarine (1986) and found out that the EGIG explains better the data at the cost of one extra parameter. In the second case, we extend the traditional univariate stochastic volatility model by allowing the errors of the squared returns of the BOVESPA index to be driven by a EGIG distribution. We compared our results with those found by Lopes and Migon (2002) and find once again that the EGIG is able to identify and estimate the quantities usually searched by standard univariate stochastic volatility models, such as highly persistent and skewed log-volatilities. Markov chain Monte Carlo are specifically tailored for both applications and a variant of the slice sampler is used to sample from some of the full conditionals.

Key words: Extended generalized inverse Gaussian distribution, Gibbs sampling, log-linear model, overrelaxation slice sampler, stochastic volatility model.

1 Introduction

In this paper we introduce and examine the class of *extended generalized inverse Gaussian* (EGIG) distributions. This class of distributions briefly appeared in Jørgensen (1982). We show that the Weibull and the generalized gamma distributions are particular cases of the EGIG family. We start by deriving its probabilistic properties and then use it in both survival analysis and financial econometrics contexts.

The EGIG distribution is obtained as a power transformation of the *generalized inverse Gaussian* (GIG) distribution Jørgensen (1982). Therefore, it is worthwhile starting with a few remarks about the GIG distribution. More precisely, if W

follows a GIG distribution with parameters λ, δ and γ, denote here by $W \sim GIG(\lambda, \delta, \gamma)$, then its probability density function is given by

$$p_{gig}(w|\lambda,\delta,\gamma) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{1}{2K_\lambda(\gamma\delta)} w^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\delta^2 w^{-1} + \gamma^2 w\right)\right\}, \quad \text{for} \quad w > 0, \tag{1.1}$$

where $-\infty < \lambda < \infty$, $\delta > 0$, $\gamma > 0$ and K_ν is the modified Bessel function of the third kind with index ν. The dependence of the parameters is as follows: $\delta \geq 0, \gamma > 0$ if $\lambda > 0$, $\delta > 0, \gamma > 0$ if $\lambda = 0$ and $\delta > 0, \gamma \geq 0$ if $\lambda > 0$. When $\lambda > 0$ and $\delta \to 0$, the density in (1.1) reduces to the pdf of a gamma distribution, while the pdf of an inverse gamma distribution is obtained when $\lambda < 0$ and $\gamma \to 0$. The following theorem relates the GIG to the EGIG and computes its moments.

Theorem 1.1 *If W has a GIG distribution with parameters λ, δ and γ, then $Y = W^{1/\theta}$, $\theta > 0$, has an EGIG distribution with parameters λ, δ, γ and θ, and probability density function,*

$$p_{egig}(y|\lambda,\delta,\gamma,\theta) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{\theta}{2K_\lambda(\gamma\delta)} y^{\lambda\theta-1} \exp\left\{-\frac{1}{2}\left(\delta^2 y^{-\theta} + \gamma^2 y^{\theta}\right)\right\}, \quad \textit{for} \quad y > 0, \tag{1.2}$$

where $-\infty < \lambda < \infty$, $\delta > 0$, $\gamma > 0$ and $\theta > 0$. Additionally, the moments $\mu^p = E(Y^p|\lambda,\delta,\gamma,\theta)$ are given by

$$\mu^p = \frac{K_{\lambda+\frac{p}{\theta}}(\gamma\delta)}{K_\lambda(\gamma\delta)} \left(\frac{\delta}{\gamma}\right)^{p/\theta}, \qquad p \in \mathbb{R}.$$

Proof. *Since $Y = W^{1/\theta}$, it is straightforward to check that $p_{egig}(y|\lambda,\delta,\gamma,\theta) = p_{gig}(y^\theta|\lambda,\delta,\gamma)y^{\theta-1}\theta$. Jørgensen (1982) gives all the moments, $E(X^k)$ with $X \sim GIG(\lambda,\delta,\gamma)$, for $k \in \mathbb{R}$, so we have $E(Y^r)$=$E(X^{r/\theta})$, then we replace $k = r/\theta$ in the moments of a GIG distribution.* □

Corollary 1.1 *The variance of the EGIG distribution is given by*

$$Var(Y|\lambda,\delta,\gamma,\theta) = \left(\frac{\delta}{\gamma}\right)^{2/\theta} \left[\frac{K_{\lambda+\frac{2}{\theta}}(\gamma\delta)}{K_\lambda(\gamma\delta)} - \left(\frac{K_{\lambda+\frac{1}{\theta}}(\gamma\delta)}{K_\lambda(\gamma\delta)}\right)^2\right].$$

We introduce the following alternative parameterization, which will prove useful in our two applications, because we want to compare with the log-linear model presented in Achcar and Bolfarine (1986). Considering, $\delta^2/2 = \xi^\theta$ and $\gamma^2/2 = \phi^\theta$, we say that $Y \sim EGIG(\lambda,\xi,\phi,\theta)$ when its probability density function is given by

$$p_{egig}(y|\lambda,\xi,\phi,\theta) = \left(\frac{\phi}{\xi}\right)^{\lambda\theta/2} \frac{\theta}{2K_\lambda(2(\xi\phi)^{\theta/2})} y^{\lambda\theta-1} \exp\left\{-(\xi/y)^\theta - (\phi y)^\theta\right\}, \tag{1.3}$$

for $y > 0$, $-\infty < \lambda < \infty$, $\xi > 0$, $\phi > 0$ and $\theta > 0$. As $\delta = (2\xi^\theta)^{1/2}$ and $\gamma = (2\phi^\theta)^{1/2}$, then we have

$$\mu^p = \frac{K_{\lambda+\frac{p}{\theta}}(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})}\left(\frac{\xi}{\phi}\right)^{p/2}, \quad p \in \mathbb{R} \tag{1.4}$$

and

$$\mathrm{Var}(Y|\lambda,\xi,\phi,\theta) = \left(\frac{\xi}{\phi}\right)\left[\frac{K_{\lambda+\frac{2}{\theta}}(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} - \left(\frac{K_{\lambda+\frac{1}{\theta}}(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})}\right)^2\right]. \tag{1.5}$$

Some well-known distributions are particular cases of the EGIG distribution given by (1.3) (see Appendix A). For $\lambda > 0$ and $\xi \to 0$, Y follows a *generalized gamma* (GG) distribution with parameters ϕ, λ and θ (Achcar and Bolfarine, 1986), denoted here by $Y \sim GG(\phi, \lambda, \theta)$ and with probability density function given by

$$p(y|\phi,\lambda,\theta) = \frac{\theta\phi^{\theta\lambda}}{\Gamma(\lambda)}y^{\theta\lambda-1}\exp\{-(\phi y)^\theta\}\ . \tag{1.6}$$

Now it is straightforward to see that (i) $\mathcal{E}(\phi)$, (ii) $\mathcal{W}(\phi,\theta)$, (iii) $G(\lambda,\phi)$ and (iv) $TN_{\mathbb{R}^+}(0,\phi^{-2})$ are particular cases of the $GG(\phi,\lambda,\theta)$ when (i) $\theta = \lambda = 1$, (ii) $\lambda = 1$, (iii) $\theta = 1$ and (iv) $\theta = 2, \lambda = 1/2$, respectively[1].

The rest of the article is organized as follows. Section 2 adapts the EGIG to two commonly used statistical models. The survival model introduced by Achcar and Bolfarine (1986) is extended in Section 2.1, while Section 2.2 extends the standard univariate stochastic volatility model. Markov chain Monte Carlo algorithms are designed for posterior inference under both model structures. Section 4 revisits the survival data from Achcar and Bolfarine (1986) and the financial data (BOVESPA index) from Lopes and Migon (2002). Our final comments and future directions are listed in Section 5.

2 Two statistical models

As we mentioned at the beginning of the previous section, our main objective is to use the rather general and flexible EGIG distribution to make two statistical well-known problems more flexible. In the first case, which is a survival model, we compare our results with those obtained by Achcar and Bolfarine (1986). In the second case, we extend the traditional univariate stochastic volatility model by allowing the errors of the squared returns of the BOVESPA index to be driven by a EGIG distribution. We compare our results with those found by Lopes and Migon (2002).

[1]Here $\mathcal{E}(\phi)$, $\mathcal{W}(\phi,\theta)$, $G(\lambda,\phi)$ and $TN_A(\mu,\sigma^2)$ denote the Exponential distribution with parameter ϕ, the Weibull distribution with parameters ϕ and θ, the gamma distribution with parameters λ and ϕ, and the truncated normal distribution with parameters μ and σ^2 and truncation region A. For $\lambda < 0$ and $\phi \to 0$, Y follows an inverse gamma distribution and its particular cases.

2.1 Log-linear model

Let $y_1, \ldots, y_n$ be responses to levels $x_1, \ldots, x_n$ of a certain covariate. In the application below, x_t is either 0 or 1, representing the placebo and the treatment group, respectively, while y_t are survival times. Achcar and Bolfarine (1986) model $y_t|\phi_t, \lambda, \theta \sim GG(\lambda, \phi_t, \theta)$, where $ln(\phi_t^{-1}) = \mu + \beta x_t$. We extend their idea by replacing the GG distribution with our EGIG distribution, ie.

$$y_t|\phi_t, \Theta_y \sim \text{EGIG}(\lambda, \xi, \phi_t, \theta) \tag{2.7}$$

with

$$h_t = \ln(\phi_t^{-1}) = \mu + \beta x_t \tag{2.8}$$

and $\Theta_y = (\lambda, \xi, \theta)$. When $\xi \to 0$ we return to the log-linear model from Achcar and Bolfarine (1986). The likelihood function for $\Theta = (\Theta_y, \mu, \beta)$ is given by

$$\begin{aligned} L(\Theta|\boldsymbol{y}, \boldsymbol{x}) &= \frac{\theta^n}{2^n \xi^{n\lambda\theta/2}} \exp\left\{ -\frac{n\lambda\theta}{2}(\mu + \beta\bar{x}) - \xi^\theta \sum_{t=1}^n y_t^{-\theta} \right\} \\ &\times \exp\left\{ -e^{-\theta\mu} \sum_{t=1}^n (e^{-\beta x_t} y_t)^\theta \right\} \prod_{t=1}^n \frac{y_t^{\theta\lambda-1}}{K_\lambda\left(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2)\right)}, \end{aligned}$$

with $\boldsymbol{x} = (x_1, \ldots, x_n), \boldsymbol{y} = (y_1, \ldots, y_n)$ and $n\bar{x} = \sum_{t=1}^n x_t$. We complete the model with independent and fairly vague prior distributions on the components of Θ. More precisely, $\pi(\Theta) = \pi(\lambda)\pi(\xi)\pi(\theta)\pi(\mu)\pi(\beta)$, with $\lambda \sim N(\mu_\lambda, \sigma_\lambda^2)$, $\xi \sim G(a_\xi, b_\xi)$, $\theta \sim G(a_\theta, b_\theta)$, $\mu \sim N(\mu_\mu, \sigma_\mu^2)$ and $\beta \sim N(\mu_\beta, \sigma_\beta^2)$. The hyperparameters $\mu_\lambda, \sigma_\lambda^2, a_\xi, b_\xi, a_\theta, b_\theta, \mu_\mu, \sigma_\mu^2, \mu_\beta$ and σ_β^2 are kept fixed at values that represent prior vagueness[2].

2.1.1 Posterior inference

It is easy to see that the posterior distribution of Θ, $\pi(\Theta|\boldsymbol{y}, \boldsymbol{x}) \propto L(\Theta|\boldsymbol{y}, \boldsymbol{x})\pi(\Theta)$, has no known form and, as it has become traditional in the literature, posterior inference is accessed by devising an Markov chain Monte Carlo sampler (Gamerman and Lopes, 2006) that, after convergence, samples ordinates from the posterior distribution. By letting $\Theta_\lambda = (\xi, \theta, \mu, \beta)$, with $\Theta_\xi, \Theta_\theta, \Theta_\mu, \Theta_\beta$ similarly defined, the log-full conditional distributions for the components of Θ are give as follows:

$$\begin{aligned} \log \pi(\lambda \mid \Theta_\lambda, \boldsymbol{x}, \boldsymbol{y}) &= -\frac{n\lambda\theta}{2}(\mu + \beta\bar{x} + \log\xi) - \frac{1}{2\sigma_\lambda^2}(\lambda - \mu_\lambda)^2 + \lambda\theta \sum_{t=1}^n \log y_t \\ &- \sum_{t=1}^n \log K_\lambda(2\xi^{\theta/2} \exp(-\theta(\mu + \beta x_t)/2) \end{aligned}$$

[2]Here, $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

$$\begin{aligned}\log\pi(\xi \mid \Theta_\xi, \boldsymbol{x}, \boldsymbol{y}) &= (a_\xi - 1 - n\lambda\theta/2)\log\xi - \xi^\theta \sum_{t=1}^{n} \log y_t^{-\theta} - b_\xi \xi \\ &- \sum_{t=1}^{n} \log K_\lambda(2\xi^{\theta/2}\exp(-\theta(\mu+\beta x_t)/2)\end{aligned}$$

$$\begin{aligned}\log\pi(\theta \mid \Theta_\theta, \boldsymbol{x}, \boldsymbol{y}) &= -\frac{n\lambda\theta}{2}(\mu + \beta\bar{x} + \log\xi) + \lambda\theta\sum_{t=1}^{n}\log y_t + (n - a_\theta - 1)\log\theta \\ &- b_\theta\theta - \xi^\theta\sum_{t=1}^{n} y_t^{-\theta} - e^{-\theta\mu}\sum_{t=1}^{n}(e^{-\beta x_t}y_t)^\theta \\ &- \sum_{t=1}^{n}\log K_\lambda(2\xi^{\theta/2}\exp(-\theta(\mu+\beta x_t)/2)\end{aligned}$$

$$\begin{aligned}\log\pi(\mu \mid \Theta_\mu, \boldsymbol{x}, \boldsymbol{y}) &= -\frac{n\lambda\theta\mu}{2} - e^{-\theta\mu}\sum_{t=1}^{n}(e^{-\beta x_t}y_t)^\theta - \frac{1}{2\sigma_\mu^2}(\mu - \mu_\mu)^2 \\ &- \sum_{t=1}^{n}\log K_\lambda(2\xi^{\theta/2}\exp(-\theta(\mu+\beta x_t)/2)\end{aligned}$$

$$\begin{aligned}\log\pi(\beta \mid \Theta_\beta, \boldsymbol{x}, \boldsymbol{y}) &= -\frac{n\lambda\theta\beta\bar{x}}{2} - e^{-\theta\mu}\sum_{t=1}^{n}(e^{-\beta x_t}y_t)^\theta - \frac{1}{2\sigma_\beta^2}(\beta - \mu_\beta)^2 \\ &- \sum_{t=1}^{n}\log K_\lambda(2\xi^{\theta/2}\exp(-\theta(\mu+\beta x_t)/2).\end{aligned}$$

Since the full conditional distributions have no known form, we use the *overrelaxation slice sampler* proposed by Neal (2003). See Appendix C for further details.

2.2 Stochastic volatility model

Let z_t be the observed value of a (financial) time series at time t, for $t = 1, \ldots, n$. In the standard stochastic volatility model, $z_t \sim N(0, e^{\kappa_t})$, where κ_t is the log-volatility at time t. Therefore, $y_t = z_t^2 \sim G(1/2, e^{-\kappa_t}/2)$. From the relations between the gamma, the generalized gamma and the extended generalized inverse gamma distributions (Section 1), it can be easily seen that $G(1/2, e^{-\kappa_t}/2) \equiv EGIG(1/2, \xi = 0, e^{-\kappa_t}/2, \theta = 1)$. We propose to let λ, ξ and θ free and model y_t by an unconstrained EGIG distribution, i.e. $y_t|\phi_t, \Theta_y \sim EGIG(\lambda, \xi, \phi_t, \theta)$ with $\Theta_y = (\lambda, \xi, \theta)$ and $\phi_t = e^{-h_t}$, as before. However, we replace equation (2.8) by the notoriously used first-order autoregression,

$$h_t|h_{t-1}, \Theta_h \sim N(\alpha + \beta(h_{t-1} - \alpha), \tau^2), \tag{2.9}$$

where $\Theta_h = (\alpha, \beta, \tau^2)$. As expected, the standard stochastic volatility model is the particular case where $\Theta_y = (1/2, 0, 1)$ and $h_t = \kappa_t + \ln 2$ (see Barndorff-Nielsen, 1997) for a similar development based on the normal inverse Gaussian distributions). We use the following prior for the seven-dimensional parameter vector $\Theta = (\Theta_y, \Theta_h, h_0)$: $\lambda \sim N(\mu_\lambda, \tau_\lambda^2)$, $\xi \sim G(a_\xi, b_\xi)$, $\theta \sim G(a_\theta, b_\theta)$, $\alpha \sim N(\alpha_0, \tau_\alpha^2)$, $\beta \sim N(\beta_0, \tau_\beta^2)$, $\tau^{-2} \sim G(\nu_0/2, \nu_0 s_0^2/2)$ and $h_0 \sim N(m_0, C_0)$. The hyperparameters μ_λ, τ_λ^2, a_ξ, b_ξ, a_θ, b_θ, α_0, τ_α^2, β_0, τ_β^2, ν_0, $\nu_0 s_0^2$, m_0 and C_0 are kept fixed at values that represent prior vagueness.

2.2.1 Posterior inference

Once again, assessment from the posterior distribution of the unknown parameters is done by an Markov chain Monte Carlo sampler. Because of the linearity and normality of $p(h_t|\Theta_h, h_{t-1})$, it is trivial to sample from the full conditional distributions of α, β, τ^2 and h_0. First, let us assume that $h = (h_0, \ldots, h_n)$ and $y = (y_1, \ldots, y_n)$. Therefore,

- $\alpha|\beta, \tau^2, h \sim N(m; C)$ for $C = [\tau_\alpha^{-2} + \tau^{-2} n(1-\beta)^2]^{-1}$ and $m = C[\tau_\alpha^{-2}\alpha_0 + \tau^{-2}(1-\beta)\sum(h_t - \beta h_{t-1})]$,
- $\beta|\alpha, \tau^2, h \sim N(m; C)$ for $C = [\tau_\beta^{-2} + \tau^{-2}\sum(h_{t-1} - \alpha)^2]^{-1}$ and $m = C[\tau_\beta^{-2}\beta_0 + \tau^{-2}\sum(h_t - \alpha)(h_{t-1} - \alpha)]$,
- $h_0|h_1, \Theta_h \sim N(m, C)$ for $C = (C_0^{-1} + \beta^2\tau^{-2})^{-1}$ and $m = C[m_0 C_0^{-1} + \beta(h_1 - \alpha + \beta\alpha)\tau^{-2}]$
- $\tau^{-2}|\alpha, \beta, h \sim G(\nu_1/2, \nu_1 s_1^2/2)$ for $\nu_1 = \nu_0 + n$ and $\nu_1 s_1^2 = \nu_0 s_0^2 + \sum(h_t - \alpha - \beta(h_t - \alpha))^2$,

where the summations are from $t = 1$ to $t = n$. For λ, ξ and θ, a natural and simple, but not necessarily efficient, possibility is to sample them, respectively, from univariate proposal densities $q_\lambda(\cdot)$, $q_\xi(\cdot)$ and $q_\theta(\cdot)$ with some degree of proximity, respectively, to $\pi(\lambda|\xi, \theta, \Theta_h, h, y)$, $\pi(\xi|\lambda, \Theta_h, h, y)$ and $\pi(\theta|\lambda, \xi, \Theta_h, h, y)$. Instead, we use Neal's (2003) overrelaxation slice sampler, as in the log-linear model presented in Section 2.1. For $t = 1, \ldots, n$ and $h_{-t} = (h_0, \ldots, h_{t-1}, h_{t+1}, \ldots, h_n)$, the full conditional of h_t is proportional to $p(y_t|h_t, \Theta_y)p(h_t|h_{t-1}, \Theta_h)\ p(h_{t+1}|h_t, \Theta_h)$, ie.

$$\begin{aligned} p(h_t|h_{-t}, \Theta, \boldsymbol{y}) &\propto e^{-h_t\lambda\theta/2}\frac{1}{K_\lambda(2(\xi e^{-h_t})^{\theta/2})}\exp\left\{-y_t^\theta e^{-h_t\theta}\right\} \\ &\times \exp\left\{-0.5\tau^2(1-\beta^2)\left[h_t^2 - 2h_t\left(\alpha + 2\beta\frac{\bar{h}_t - \alpha}{1-\beta^2}\right)\right]\right\}, \end{aligned}$$

where $\bar{h}_t = (h_{t+1} + h_{t-1})/2$. Either single-move Metropolis-type steps (Carlin, Polson and Stoffer, 1997) or single-move slice steps (Neal, 2003) could be used here.

3 Model criticism

In this section we briefly overview alternative measures of model adequacy, determination and/or selection. Without loss of generality, we will focus on the stochastic volatility context presented in Section 2.2. Therefore, let z_t be the observed value of a (financial) time series at time t, for $t = 1, \ldots, n$ and $z_t = y_t^2$, such that the two competing models are $\mathcal{M}_1 : z_t|h_{1t} \sim e^{h_{1t}}\chi_1^2$ and $\mathcal{M}_2 : z_t|h_{2t}, \lambda, \xi, \theta \sim EGIG(\lambda, \xi, e^{-h_{2t}}, \theta)$, with $h_{it}|h_{i,t-1}, \Delta_i \sim N(\alpha_i + \beta_i(h_{i,t-1} - \alpha_i), \tau_i^2)$. Model $\mathcal{M}_1$ is a particular case of model $\mathcal{M}_2$, since $e^{h_t}\chi_1^2$ is equivalent to $G(0.5, 0.5e^{-h_t})$ which turns out to be equivalent to $EGIG(0.5, 0, e^{-h_t}, 1)$, with $\lambda = 0.5, \xi = 0$, and $\theta = 1$. Let $\Theta_1 = (\boldsymbol{h}_1, \Delta_1)$ and $\Theta_2 = (\boldsymbol{h}_2, \Delta_2, \Xi)$, for $\Xi = (\lambda, \xi, \theta)$, $\Delta_i = (\alpha_i, \beta_i, \tau_i^2)$, and $\boldsymbol{h}_i = (h_{i0}, \ldots, h_{in})$ for $i = 1, 2$. Additionally, let $\boldsymbol{z} = (z_1, \ldots, z_n)'$, so that $p_1(\boldsymbol{z}|\Theta_1) = \prod_{t=1}^n p_1(z_t|h_{1t})$ and $p_2(\boldsymbol{z}|\Theta_2) = \prod_{t=1}^n p_2(z_t|h_{2t}, \Xi)$ be the likelihood functions for both models, which are conditional, of course, on all unknown parameters in Θ_1 and Θ_2. Finally, $p(\boldsymbol{z}|\mathcal{M}_i) = \int_{\Theta_i} p_i(\boldsymbol{z}|\Theta_1)\pi(\Theta_i|\mathcal{M}_i)d\Theta_i$ is the predictive ordinate or marginal likelihood averaged across the prior $\pi(\Theta_i|\mathcal{M}_i)$ and under model $\mathcal{M}_i$.

3.1 AIC and BIC

The choice of the best amongst competing models is traditionally made by using information criteria that penalized the likelihood function, with the AIC (Akaike, 1974),

$$AIC(\mathcal{M}_i) = -2\log p_i(\boldsymbol{z}|\widehat{\Theta}_i) + 2d_i \tag{3.10}$$

and the BIC (Schwarz, 1978),

$$BIC(\mathcal{M}_i) = -2\log p_i(\boldsymbol{z}|\widehat{\Theta}_i) + \log(n)d_i \tag{3.11}$$

the most popular ones. Here, d_i and $\widehat{\Theta}_i$ are, respectively, the number of parameters and the maximum likelihood estimator under model i, for $i = 1, 2$. One of the major problems with AIC/BIC is that defining p_i is not trivial, mainly in Bayesian Hierarchical models, where the prior distributions implicitly reduce the *effective number of parameters* through their interdependencies. Nonetheless, for model $\mathcal{M}_1$, the standard stochastic volatility, we will assume that the parameters comprise the volatilities, h_{1t}, and the parameters governing the log-volatilities' dynamics, α_1, β_1 and τ_1^2. For model $\mathcal{M}_2$, additional parameters are λ, ξ and θ. Therefore, $d_1 = n + 3$ and $d_2 = n + 6$. If the volatilities are not included, then $d_1 = 3$ and $d_2 = 6$.

3.2 DIC

Spiegelhalter, Best, Carlin and Linde (2002) developed an information criterion where the effective number parameter is defined as $d_i = \overline{D}_i - D(\widetilde{\Theta}_i)$, where $D(\Theta_i) = -2\ln p(\boldsymbol{z}|\Theta_i)$ is a function of the *deviance*, $\widetilde{\Theta}_i = E(\Theta_i|\boldsymbol{z}, \mathcal{M}_i)$ and $\overline{D}_i = E(D(\Theta_i)|\boldsymbol{z}, \mathcal{M}_i)$. The *Deviance Information Criterion*, or simply DIC,

is defined as $DIC(\mathcal{M}_i) = -2\log p_i(\boldsymbol{z}|\widetilde{\Theta}_i) + 2d_i$, with $\overline{D}_i$ and d_i measuring the model's fit and complexity, respectively. DIC generalize AIC and short values of DIC means a better fit. Computationally, DIC is more attractive than Bayes Factor (see below) because can be easily incorporated in MCMC routines. For computational purposes, the DIC can be rewritten as $DIC(\mathcal{M}_i) = 2\overline{D}_i - D(\widetilde{\Theta}_i)$. Under model $\mathcal{M}_i$, suppose that $\{\Theta_i^{(1)}, \ldots, \Theta_i^{(M)}\}$ correspond to a MCMC sample from $\pi(\Theta_i|\boldsymbol{z}, \mathcal{M}_i)$, ie. the parameters' joint posterior distribution. Then, the DIC can be estimated as,

$$DIC(\mathcal{M}_i) \approx -\frac{4}{M}\sum_{j=1}^{M}\log p_i(\boldsymbol{z}|\Theta_i^{(j)}) + 2\log p_i(\boldsymbol{z}|\widehat{\Theta}_i), \tag{3.12}$$

with $\widehat{\Theta}_i = \frac{1}{M}\sum_{j=1}^{M}\Theta_i^{(j)}$ an estimate of $\widetilde{\Theta}_i$.

3.3 Bayes and pseudo-Bayes factors

The Bayes factor is defined as the ratio of marginal likelihood (or predictive) densities, ie. $BF(\mathcal{M}_1, \mathcal{M}_2) = p(\boldsymbol{z}|\mathcal{M}_1)/p(\boldsymbol{z}|\mathcal{M}_2)$, recalling that $p(\boldsymbol{z}|\mathcal{M}_i) = \int_{\Theta_i} p_i(\boldsymbol{z}|\Theta_i)\pi(\Theta_i|\mathcal{M}_i)d\Theta_i$. Unfortunately, predictive ordinates, also well known as normalizing constants, is rarely available in closed form so analytical or Monte Carlo approximation must be used. Gelfand and Dey (1994) proposed the *harmonic mean estimator* of $p(\boldsymbol{z}|\mathcal{M}_i)$,

$$\widehat{p}(\boldsymbol{z}|\mathcal{M}_i)^{-1} = \frac{1}{M}\sum_{j=1}^{M}\left(\pi(\Theta_i^{(j)}|\mathcal{M}_i)p_i(\boldsymbol{z}|\Theta_i^{(j)})\right)^{-1} \tag{3.13}$$

derived from the identity $p(\boldsymbol{z}|\mathcal{M}_i)^{-1} = \int[\pi(\Theta_i|\mathcal{M}_i)p_i(\boldsymbol{z}|\Theta_i)]^{-1}\pi(\Theta_i|\boldsymbol{z}, \mathcal{M}_i)d\Theta_i$, where, again, $\{\Theta_i^{(1)}, \ldots, \Theta_i^{(M)}\}$ is a MCMC sample from $\pi(\Theta_i|\boldsymbol{z}, \mathcal{M}_i)$.

The pseudo-Bayes factor is defined as

$$PBF(\mathcal{M}_1, \mathcal{M}_2) = \prod_{t=1}^{n} p(z_t|\boldsymbol{z}_{(t)}, \mathcal{M}_1)/p(z_t|\boldsymbol{z}_{(t)}, \mathcal{M}_2)\ ,$$

where $\boldsymbol{z}_{(t)} = (z_1, \ldots, z_{t-1}, z_{t+1}, \ldots, z_n)$ and $p(z_t|\boldsymbol{z}_{(t)}, \mathcal{M}_i) = p(\boldsymbol{z}|\mathcal{M}_i)/p(\boldsymbol{z}_{(t)}|\mathcal{M}_i)$ (Gelfand, 1996). One could argue that $p(\boldsymbol{z}|\mathcal{M}_i)$ and $\prod_{t=1}^{n} p(z_t|\boldsymbol{z}_{(t)}, \mathcal{M}_i)$ are very similar quantities, therefore the pseudo-Bayes factor could be thought of as a natural surrogate to the Bayes factor. When $\{\Theta_i^{(1)}, \ldots, \Theta_i^{(M)}\}$ is a MCMC sample from $\pi(\Theta_i|\boldsymbol{z}, \mathcal{M}_i)$, $p(z_t|\boldsymbol{z}_{(t)}, \mathcal{M}_i)$ can be estimated as,

$$\widehat{p}(z_t|\boldsymbol{z}_{(t)}, \mathcal{M}_i)^{-1} = \frac{1}{M}\sum_{j=1}^{M} p_i(z_t|\Theta_i^{(j)})^{-1}. \tag{3.14}$$

The estimator from equation (3.14) is more stable than that from equation (3.13), since it is based on only one observation at each evaluation.

4 Applications

In this section we use the EGIG distribution to reanalyze both Achcar and Bolfarine's (1986) two-treament survival times and BOVESPA's squared returns (Lopes and Migon, 2002).

4.1 Comparison of two treatments

Let us assume a special case of the log-linear model (2.7) considering an indicator (group-membership) variable x such that $x = 0$ for treatment 1 and $x = 1$ for treatment 2. Usually, in this two sample problem, we have application of standard survival models. Alternatively, we are proposing this more general class with a great flexibility to fit the data. Let $y_1, \ldots, y_{n_1}$ be the survival times of n_1 individuals from treatment 1 and $y_{n_1+1}, \ldots, y_n$ be the survival times of the $n - n_1$ individuals from treatment 2. Consider the survival times of the patients in the two treatment groups in Table 1 (from Achcar and Bolfarine, 1986).

Table 1 *Survival times of patients in two treatment groups.*

Treatment 1	5	10	17	32	32	33	34	36
	43	44	44	48	48	61	64	65
	65	66	67	68	82	85	90	92
	92	102	103	106	107	114	114	116
	117	124	139	142	143	151	158	195
Treatment 2	20.9	32.2	33.2	39.4	40.0	46.8	54.3	57.3
	58.0	59.7	61.1	61.4	66.0	66.3	67.4	68.5
	69.9	72.4	73.0	73.2	88.7	89.3	91.6	93.1
	94.2	97.7	101.6	101.9	107.6	108.0	109.7	110.8
	114.1	117.5	119.2	120.3	133.0	133.8	163.3	165.1

We consider $\mu_\lambda = \mu_\mu = \mu_\beta = 0$, $\sigma^2_\lambda = 100$, $\sigma^2_\mu = \sigma^2_\beta = 1000$ and $a_\xi = b_\xi = a_\theta = b_\theta = 0.01$ in the way to have vague prior distribution. Now, with all the log-full conditional well defined we ran the overrelaxation slice sampler within Gibbs sampling - written in FORTRAN 90 with IMSL library - with start points $\lambda^{(0)} = 1$, $\xi^{(0)} = 1$, $\mu^{(0)} = 0$, $\beta^{(0)} = 0$ and $\theta^{(0)} = 1$. We ran 100000 iterations, considering convergence at 10000 iteration and taken one observation at every 15 sampled, resulting a sample of size 6000 of the posterior distribution. Some of the parameter still had some non-negligible lag-correlations, but not so strong. Figure 1 and Table 2 summarizes of the posterior distribution. The parameter $\sigma = 1/\theta$ is just to compare with the results obtained in Achcar and Bolfarine (1986).

The posterior mean of the parameters' correlation matrix is presented in Table 3, from which it can be seen that λ, μ and θ are highly and nonlinearly correlated (see also Figures 2, 3 and 4). The over-relaxed slice sampling algorithm has proven to be more efficient than standard random-walk steps when dealing with such nonlinear dependencies.

Table 2 *Summary of the posterior distribution. StDev,* 2.5% *and* 97.5% *are the standard deviation and percentiles from the posterior distribution of the parameters, respectively.*

Parameter	Mean	StDev	2.5%	Median	97.5%
λ	1.456	1.505	0.349	1.022	5.096
ξ	0.312	1.490	0.000	0.000	3.341
μ	4.313	0.777	2.346	4.535	4.996
β	-0.018	0.112	-0.233	-0.021	0.212
θ	2.252	0.963	0.802	2.133	4.441
σ	0.541	0.280	0.225	0.469	1.247

Table 3 *Correlation matrix.*

Parameter	λ	ξ	μ	β	θ
λ	1.000	0.102	-0.985	0.247	-0.639
ξ	0.102	1.000	-0.126	0.050	-0.042
μ	-0.985	-0.126	1.000	-0.324	0.693
β	0.247	0.050	-0.324	1.000	-0.287
θ	-0.639	-0.042	0.693	-0.287	1.000

The results suggests that Achcar and Bolfarine's log-linear model with generalized gamma distribution and our extended model are relatively similar, as far as parameter estimation is concerned (see Table 4). However, when the two competing models are checked against the criteria proposed in Section 3, we see that the particular model is slightly preferable by all of them, but the DIC which turns out to favor, also slightly, our extended model (see Table 5).

Table 4 *Summary of the posterior distribution. StDev,* 2.5% *and* 97.5% *are the standard deviation and percentiles from the posterior distribution of the parameters, respectively.*

Parameter	Mean	StDev	2.5%	Median	97.5%
λ	1.450	1.550	0.340	1.012	5.142
μ	4.319	0.788	2.334	4.538	4.996
β	-0.019	0.112	-0.231	-0.022	0.215
θ	2.268	0.970	0.796	2.139	4.558
σ	0.537	0.281	0.219	0.467	1.256

Table 5 *Model comparison: $\mathcal{M}_1$: EGIG and $\mathcal{M}_2$: GG. The Bayes factor is BF($\mathcal{M}_1, \mathcal{M}_2$)=0.667 and the pseudo-Bayes factor is PBF($\mathcal{M}_1, \mathcal{M}_2$)=0.984.*

Criterion	EGIG ($\mathcal{M}_1$)	GG ($\mathcal{M}_2$)
AIC	835.74	834.06
BIC	847.65	843.59
DIC	4109.1	4119.7

4.2 IBOVESPA revisited

For comparison reasons, let us use the IBOVESPA index, as presented in Lopes and Migon (2002). The series corresponds to the index returns, for about 1500 consecutive days, daily measured by the Sao Paulo Stock Exchange (see the top part of Figure 5). For comparisons, we have fit a standard AR(1) model for the log-volatilities, ie. $y_t|h_t \sim N(0, e^{h_t})$ and $h_t|h_{t-1}, \alpha, \beta, \tau^2 \sim N(\alpha + \beta(h_{t-1} - \alpha), \tau^2)$, where $t = 1, \ldots, n$ with $n = 1481$. We chose relatively vague prior distributions for α, β and τ^2. More specifically, $\alpha \sim N(0.6546, 10)$, $\beta \sim N(0.9797, 10)$, $\tau^{-2} \sim G(a, b)$ with $a/b = 0.0167^{-1}$ and $a/b^2 = 10$ and $h_0 \sim N(0, 10)$. The initial values for α, β, τ^{-2} and h_0 were $-0.6546, 0.9797, 0.0244$ and 0.0, respectively. After running the MCMC chain for 10000 iterations, we kept every 20^{th} of the next 100000 draws, creating a total of 5000 draws from the posterior distribution. Posterior summary for α, β and τ^2 are presented on Table 6. The bottom part of Figure 5 exhibits the posterior mean of e^{h_t}, ie. $E(e^{h_t}|y_1, \ldots, y_n)$, for $t = 1, \ldots, n$.

Table 6 *Univariate stochastic volatility model: posterior summary.*

Parameter	Mean	2.5%	Median	97.5%	
α	-0.591	0.391	-1.144	-0.586	-0.009
β	0.976	0.007	0.961	0.976	0.988
τ^2	0.055	0.010	0.036	0.053	0.077

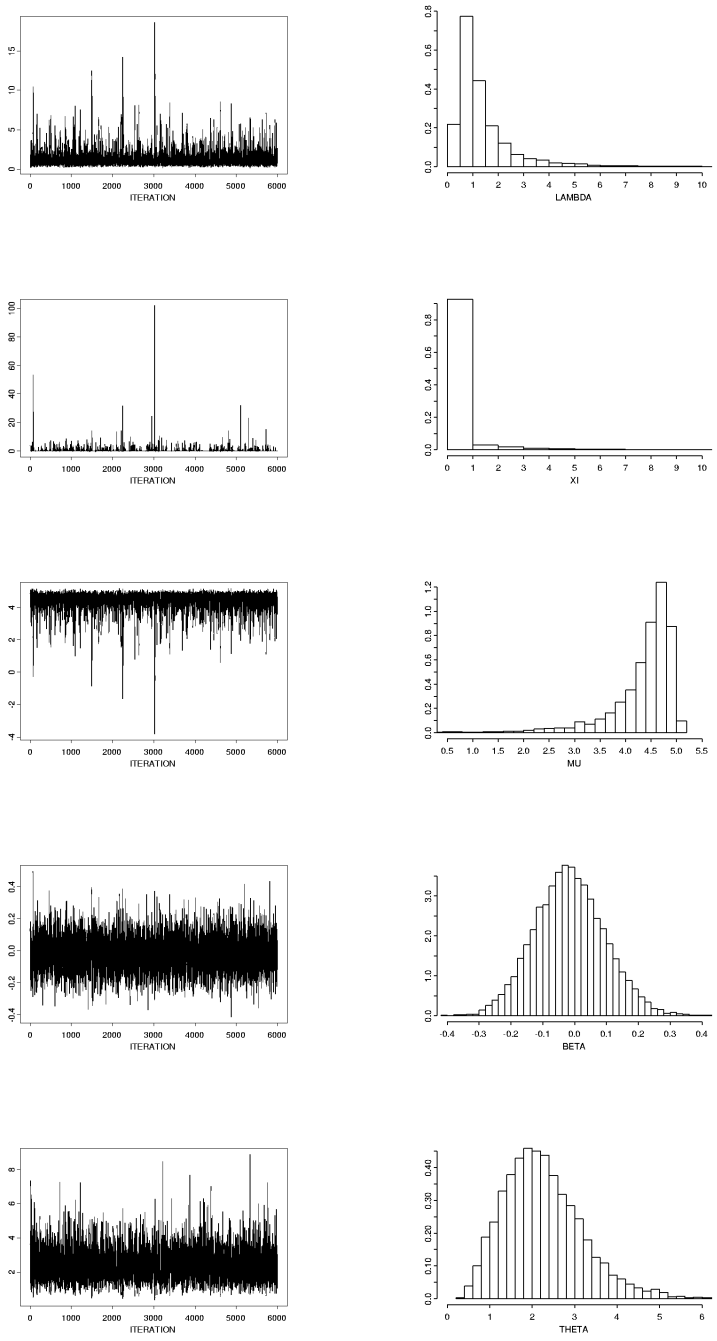

Figure 1 *Log-linear model: posterior distribution of λ (top), ξ, μ, β and θ (bottom).*

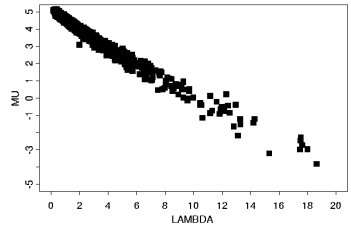

Figure 2 *Log-linear model: Joint posterior of* (λ, μ)*.*

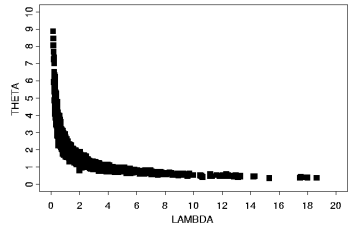

Figure 3 *Log-linear model: Joint posterior of* (λ, θ)*.*

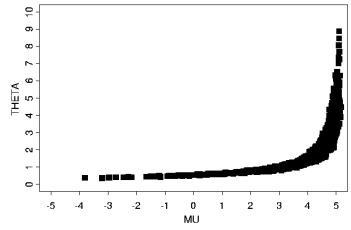

Figure 4 *Log-linear model: Joint posterior of* (μ, θ)*.*

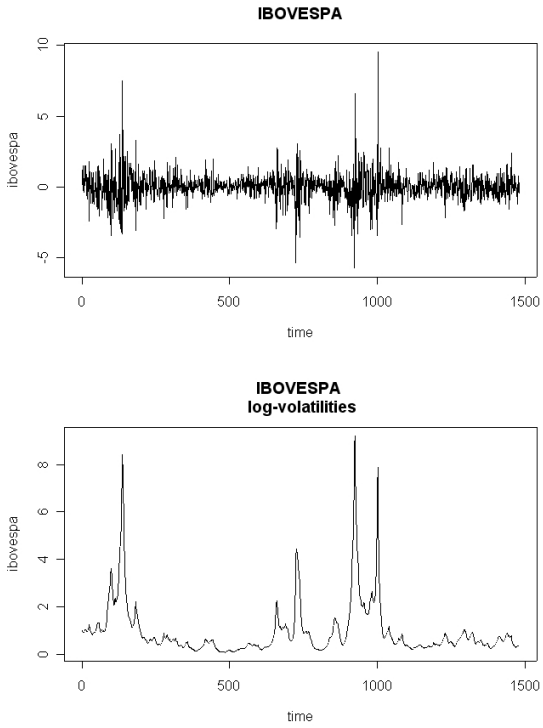

Figure 5 *Top: Bovespa index daily returns (from the Sao Paulo Stock Exchange). Bottom: Standard univariate stochastic volatility model: $E(e^{h_t}|y_1,\ldots,y_n)$, for $t = 1,\ldots,n$. Stochastic volatility model: Posterior summary for the log-volatility parameters. After running the chain for 10000 iterations, we kept every 20^{th} of the next 100000 draws, for a total of 5000 draws from the posterior distribution.*

We now apply the univariate stochastic volatility model with EGIG distribution to the square of the index returns defined above. We consider vague prior: $\alpha \sim N(0.0, 10^4)$, $\beta \sim N(0.0, 10^4)$, $\tau^{-2} \sim G(0.01, 0.01)$, and $h_0 \sim N(0, 10^4)$. Additionally, $\lambda \sim N(0, 100)$, $\xi \sim G(0.01, 0.01)$, and $\theta \sim G(0.01, 0.01)$. The initial values for α, β, τ^2 and h_0 were $0.0, 0.5, 1.0$ and 0.0, respectively. After running four parallel MCMC chain for 20000 iterations, we kept every 10^{th} of the next 10000 draws, creating a total of 4000 draws from the posterior distribution. Posterior summary for α, β and τ^2 are presented on Table 7 and Figure 6. Figure 7 exhibits the posterior mean of h_t, ie. $E(h_t|y_1,\ldots,y_n)$, for $t = 1,\ldots,n$ for every one of the four chains, it is indistinguishable among them. Besides it has the same feature as in Lopes and Migon (2002), i.e, high volatilities near time 100, 700 and 1000.

Table 7 *Univariate stochastic volatility model: posterior summary.*

Parameter	Mean	S.D	2.5%	Median	97.5%
λ	0.439	0.108	0.282	0.419	0.720
$10^{-6}\xi$	0.149	0.351	0.00014	0.012	1.18
θ	1.140	0.203	0.758	1.136	1.550
α	0.206	0.356	-0.530	0.238	0.778
β	0.967	0.012	0.942	0.968	0.988
τ^2	0.070	0.022	0.031	0.068	0.119

Since the posterior mean of λ, ξ and θ are not significantly away from $0.5, 0$ and 1.0, respectively, one could argue that we have recovered the traditional univariate stochastic volatility model. In fact, when the two competing models are checked against the criteria proposed in Section 3, we found out that the particular model is picked by all of them as the best model (see Table 8).

Table 8 *Model comparison: $\mathcal{M}_1$: EGIG and $\mathcal{M}_2$: GG. The Bayes factor is BF($\mathcal{M}_1, \mathcal{M}_2$)=0.071 and the pseudo-Bayes factor is PBF($\mathcal{M}_1, \mathcal{M}_2$)=0.069.*

Criterion	EGIG ($\mathcal{M}_1$)	GG ($\mathcal{M}_2$)
AIC	4579.4	834.06
BIC	12461.2	11933.4
DIC	5857.7	5351.8

4.2.1 Mean square error

In Table 9 the means of the posterior distribution of the EGIG model. Conditionally on the parameters in Table 9, we calculated the mean square errors (MSE), $MSE = \sum_{i=1}^{n} r_i^2/n$, where $r_i = (y_i - \mu_i)/\sigma_i$ for mean μ_i and variance σ_i^2 functions of the posteriori mean. The first row in the table presents the mode of the posteriori distribution given by Achcar and Bolfarine (1986) and the corresponding MSE. The second row of the table presents the means of the posterior distribution given in Table 2. The last three rows are the posterior means for the *Sampling Importance Resampling* (SIR) algorithm. In SIR1, the parameters are generated by uniforms proposals with unit length centered on the posterior mean found in Achcar and Bolfarine (1986) that is also considered a priori. In SIR2, the parameters are generated by uniforms proposals with unit length centered on the posterior mean of the Table 2. SIR3 is as SIR2, but the priors are the same as that presented in the previous section. The samples are of size 100000 for both the importance sampling and resampling (of the posterior). The posterior means of the parameters are closed.

Table 9 *Mean square errors for the log-linear model*

	λ	ξ	μ	β	θ	MSE
AB	1.0000	0.0001	4.578	-0.03400	2.500	1.209
SLM	1.4450	0.3350	4.318	-0.01600	2.263	1.674
SIR1	0.7876	0.2500	4.682	-0.03457	2.506	0.974
SIR2	1.1210	0.4159	4.482	-0.01822	2.045	0.992
SIR3	1.1410	0.1147	4.470	-0.03715	2.004	0.991

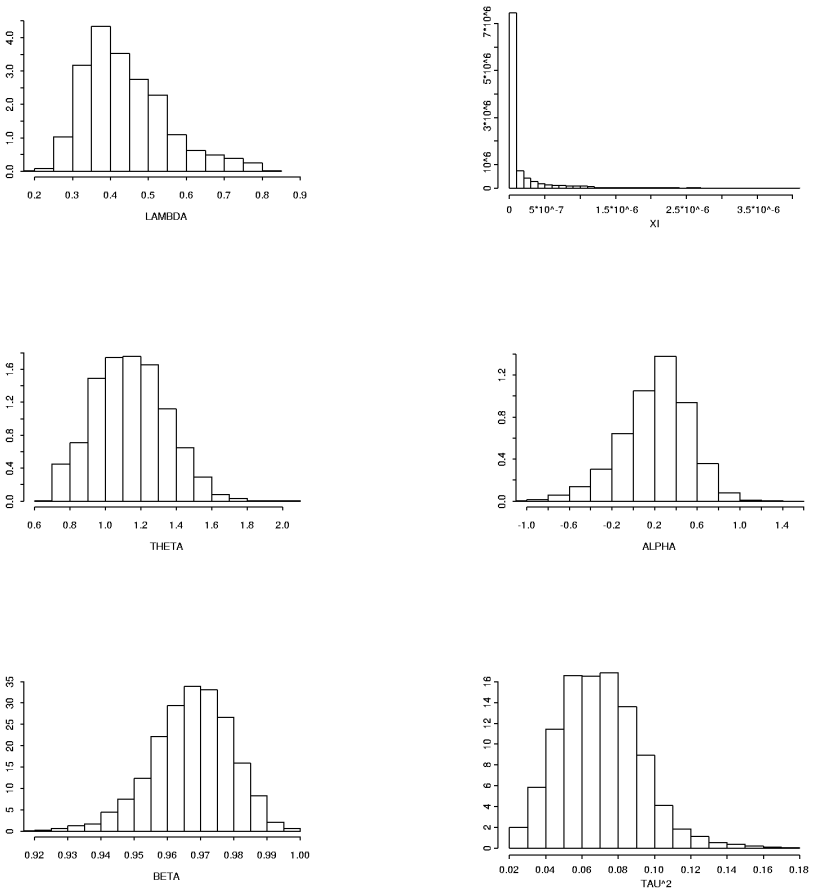

Figure 6 *Stochastic volatility model with EGIG: Posterior distribution of λ, ξ (top), θ, α (center), and β, τ^2 (bottom).*

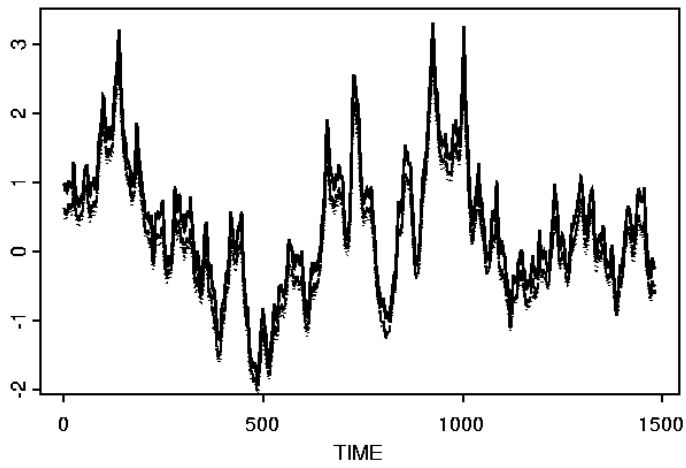

Figure 7 *Stochastic volatility model with EGIG: $E(h_t|y_1,\ldots,y_n)$, for $t = 1,\ldots,n$. Normal model (solid line) and EGIG model (dotted line).*

5 Conclusions

A class of distribution was introduced throughout the generalized inverse Gaussian distribution Jørgensen (1982). This class is very flexible and include many particular cases. A crucial question to be further investigated is how the dependence of the parameters affects or is affected by the MCMC algorithms under use. The log-linear model presented here reproduced the results found in Achcar and Bolfarine (1986), but further investigation is needed. Moreover, we are implementing other possibilities of the log-linear model in the parameter ξ, possibility accomodating heavy tails. Similarly, the combined stochastic volatility model with EGIG performed reasonably well and exhibited results similar to those found by Lopes and Migon (2002). For the applications, model criticism was performed agains a variety of criteria. We used the well known and stablished AIC and BIC criteria, along with their newest cousin, the DIC. Bayes factor and pseudo Bayes factor are also computed in our examples.

Our current agenda includes, among other things, modelling the returns directly as a log-EGIG distribution and investigating the extent to which competing models differ. To that end we are currently modelling the main EGIG parameters (λ, ξ, θ) with point masses at special cases. For instance, the posterior distribution of ξ seemed to be reasonable closed to zero in both log-linear and stochastic volatility models. Moreover, the stochastic volatility model with EGIG opens new

areas of research since multiple parameters of the EGIG distribution are related to the data variability.

We are also currently investigating the class of Jeffreys' priors for the parameters of the EGIG distributions and how they affect the propriety of the posterior densities. Appendix B introduces the derivations of the Jeffreys' prior for the i.i.d data case. Even in this apparently simple case, several numerical derivatives and integrals ought to be evaluated numerically. Results will appear elsewhere.

A Some proofs

First, let us remember some aspects of the Bessel function K_ν of the third kind with index ν. Among the many integral representations of $K_\nu(\omega)$, $\omega > 0$, one has Jørgensen (1982)

$$K_\nu(\omega) = \frac{1}{2}\int_0^\infty x^{\nu-1} exp\left\{-\frac{1}{2}\omega(x+x^{-1})\right\} dx.$$

The Bessel functions K_ν, $\nu \in \mathbb{R}$, satisfy the relations

$$K_{-\nu}(\omega) = K_\nu(\omega)$$

then we will need the asymptotic expansions for $x \to 0$

$$\begin{aligned} & K_\nu(\omega) \simeq \frac{\Gamma(|\nu|)}{2}\left(\frac{\omega}{2}\right)^{-|\nu|} \\ & K_0(\omega) \simeq -\ln(\omega) \\ \text{and} \quad & K_\nu(\omega) \simeq \omega^{-1/2}e^{-\omega}\sqrt{\pi/2} \qquad \text{for} \quad \omega \to \infty. \end{aligned} \tag{A.1}$$

For $\lambda = n + 1/2$ and $n = 0, 1, 2, \ldots$ one has

$$K_{n+\frac{1}{2}}(\omega) = \sqrt{\frac{\pi}{2\omega}}e^{-\omega}\left(1 + \sum_{i=1}^{n}\frac{(n+i)!}{(n-i)!i!}(2\omega)^{-i}\right).$$

So, given EGIG density function

$$g(y|\lambda,\xi,\phi,\theta) = \left(\frac{\phi}{\xi}\right)^{\lambda\theta/2}\frac{\theta}{2K_\lambda(2(\xi\phi)^{\theta/2})}y^{\lambda\theta-1}\exp\left\{-\xi^\theta y^{-\theta} - \phi^\theta y^\theta\right\}, \tag{A.2}$$

for $y > 0$. Then for $\lambda > 0$ and $\xi \to 0$, we have with (A.1)

$$\frac{1}{\xi^{\lambda\theta/2}K_\lambda(2(\xi\phi)^{\theta/2})} \simeq \frac{2\phi^{\lambda\theta/2}}{\Gamma(\lambda)}.$$

Inserting this in (A.2) we get

$$g(y|\lambda,\phi,\theta) = \frac{\theta}{\Gamma(\lambda)}\phi^{\lambda\theta}y^{\lambda\theta-1}\exp\left\{-\phi^\theta y^\theta\right\}, \quad \text{for} \quad y > 0,$$

which equals to the density of a generalized gamma distribution. Now, it is very easy to see that this distribution has Weibull, gamma and exponential as particular cases. Besides, take $\theta = 2$ and $\lambda = 1/2$ to get a truncated normal distribution on $\mathbb{R}^+$ with scale ϕ^{-2}.
When $\lambda < 0$ and $\phi \to 0$, one gets the reciprocal generalized inverse gamma and their particular cases. The proof is similar of that given above.

B Jeffreys' prior distribution for i.i.d EGIG data

In this appendix, the Jeffreys' prior is derived from both the score function $U = U(\lambda, \xi, \phi, \theta)$ and the expected value of UU', $\mathrm{E}(UU')$.

Let $(X_1, \ldots, X_n)$ be a random sample from the EGIG distribution with probability density function given by (1.3). Therefore, the log-likelihood function is given by

$$\begin{aligned}\ell(\lambda, \xi, \phi, \theta) = {} & \frac{n\lambda\theta}{2}\left(\ln(\phi) - \ln(\xi)\right) + n\ln(\theta) - n\ln(2) - n\ln\left(K_\lambda(2(\xi\phi)^{\theta/2})\right) \\ & + (\lambda\theta - 1)\sum_{i=1}^{n}\ln(x_i) - \xi^\theta\sum_{i=1}^{n}x_i^{-\theta} - \phi^\theta\sum_{i=1}^{n}x_i^\theta. \end{aligned} \tag{B.1}$$

First, we define $k_\lambda(x) = \partial K_\lambda(x)/\partial\lambda$ that must be evaluated numerically. $K'(z)$ is the derivative with respect to z, also evaluated numerically.

$$U_1(X;\lambda) = \frac{\partial}{\partial\lambda}\ell(\lambda,\xi,\phi,\theta) = \theta\left[\ln(X) - \left(\frac{k_\lambda(2(\xi\phi)^{\theta/2})}{\theta K_\lambda(2(\xi\phi)^{\theta/2})} + \frac{1}{2}(\ln(\xi) - \ln(\phi))\right)\right]$$

$$U_2(X;\xi) = \frac{\partial}{\partial\lambda}\ell(\lambda,\xi,\phi,\theta) = -\frac{\lambda\theta}{2\xi} - \theta\phi^{\theta/2}\xi^{\theta/2-1}\frac{K'_\lambda(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} - \theta\xi^{\theta-1}X^{-\theta}$$

$$U_3(X;\phi) = \frac{\partial}{\partial\lambda}\ell(\lambda,\xi,\phi,\theta) = \frac{\lambda\theta}{2\phi} - \theta\xi^{\theta/2}\phi^{\theta/2-1}\frac{K'_\lambda(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} - \theta\phi^{\theta-1}X^{\theta}$$

$$\begin{aligned} U_4(X;\theta) = \frac{\partial}{\partial\lambda}\ell(\lambda,\xi,\phi,\theta) = {} & \frac{\lambda}{2}\left(\ln(\phi) - \ln(\xi)\right) + \frac{1}{\theta} + \lambda\ln(X) \\ & - \xi^{\theta/2}\phi^{\theta/2}\left(\ln(\xi) + \ln(\phi)\right)\frac{K'_\lambda(2(\xi\phi)^{\theta/2})}{K_\lambda(2(\xi\phi)^{\theta/2})} \\ & - \xi^\theta\ln(\xi)X^{-\theta} + \xi^\theta X^{-\theta}\ln(X) - \phi^\theta\ln(\phi)X^\theta - \phi^\theta X^\theta\ln(X). \end{aligned}$$

As $E(U_1(X;\lambda)) = 0$, then

$$E\left(\ln(X)\right) = \frac{k_\lambda(2(\xi\phi)^{\theta/2})}{\theta K_\lambda(2(\xi\phi)^{\theta/2})} + \frac{1}{2}(\ln(\xi) - \ln(\phi)). \tag{B.2}$$

Let us define the following relations:

$$\begin{aligned}
\rho &= 2(\xi\phi)^{\theta/2};\\
a_1 &= \frac{\theta}{2}(\ln(\phi)-\ln(\xi)) - \frac{k_\lambda(\rho)}{K_\lambda(\rho)};\\
a_2 &= -\frac{\lambda\theta}{2\xi} - \theta\phi^{\theta/2}\xi^{\theta/2-1}\frac{K'_\lambda(\rho)}{K_\lambda(\rho)};\\
b_2 &= -\theta\xi^{\theta-1};\\
a_3 &= \frac{\lambda\theta}{2\phi} - \theta\xi^{\theta/2}\phi^{\theta/2-1}\frac{K'_\lambda(\rho)}{K_\lambda(\rho)};\\
b_3 &= -\theta\phi^{\theta-1};\\
a_4 &= \frac{\lambda}{2}(\ln(\phi)-\ln(\xi)) + \frac{1}{\theta} - \xi^{\theta/2}\phi^{\theta/2}(\ln(\xi)+\ln(\phi))\frac{K'_\lambda(\rho)}{K_\lambda(\rho)};\\
b_4 &= -\xi^\theta\ln(\xi); \quad \text{e}\\
c_4 &= -\phi^\theta\ln(\phi).
\end{aligned}$$

Then,

$$\begin{aligned}
U_1(X;\lambda) &= a_1 + \theta\ln(X);\\
U_2(X;\xi) &= a_2 + b_2X^{-\theta};\\
U_3(X;\phi) &= a_3 + b_3X^{\theta}; \quad \text{e}\\
U_4(X;\theta) &= a_4 + \lambda\ln(X) + b_4X^{-\theta} + \xi^\theta X^{-\theta}\ln(X) + c_4X^\theta - \phi^\theta x^\theta\ln(X).
\end{aligned}$$

Definition: The Jeffreys' prior distribution is defined by

$$p^J(\lambda,\xi,\phi,\theta) \propto \sqrt{|E(UU')|},$$

where $U = U(\lambda,\xi,\phi,\theta) = (U_1(X;\lambda), U_2(X;\xi), U_3(X;\phi), U_4(X;\theta))'$.

One needs to calculate the expected values of $E(U_iU_j)$ for $i,j = 1,2,3,4$ and

$i \leqslant j$. It follows that

$$
\begin{aligned}
E(U_1^2) &= a_1^2 + 2\theta a_1 E(\ln(X)) + \theta^2 E(\ln(X))^2 \\
E(U_2^2) &= a_2^2 + 2a_2 b_2 E(X^{-\theta}) + b_2^2 E(X^{-2\theta}) \\
E(U_3^2) &= a_3^2 + 2a_3 b_3 E(X^{\theta}) + b_3^2 E(X^{2\theta}) \\
E(U_4^2) &= a_4^2 + \lambda^2 E(\ln(X))^2 + b_4^2 E(X^{-2\theta}) + \xi^{2\theta} E(X^{-2\theta}(\ln(X))^2) \\
&\quad + c_4^2 E(X^{2\theta}) + \phi^{2\theta} E(X^{2\theta}(\ln(X))^2) + 2a_4\lambda E(\ln(X)) \\
&\quad + 2a_4 b_4 E(X^{-\theta}) + 2a_4\xi^{\theta} E(X^{-\theta}\ln(X)) + 2a_4 c_4 E(X^{\theta}) \\
&\quad - 2a_4\phi^{\theta} E(X^{\theta}\ln(X)) + 2b_4\lambda E(X^{-\theta}\ln(X)) + 2\lambda\xi^{\theta} E(X^{-\theta}(\ln(X))^2) \\
&\quad + 2c_4\lambda E(X^{\theta}\ln(X)) - 2\lambda\phi^{\theta} E(X^{\theta}(\ln(X))^2) \\
&\quad + 2b_4\xi^{\theta} E(X^{-2\theta}\ln(X)) + 2b_4 c_4 - 2b_4\phi^{\theta} E(\ln(X)) \\
&\quad + 2c_4\xi^{\theta} E(\ln(X)) - 2\xi^{\theta}\phi^{\theta} E(\ln(X))^2 - 2c_4\phi^{\theta} E(X^{2\theta}\ln(X)) \\
E(U_1U_2) &= a_1 a_2 + a_1 b_2 E(X^{-\theta}) + a_2\theta E(\ln(X)) + b_2\theta E(X^{-\theta}\ln(X)) \\
E(U_1U_3) &= a_1 a_3 + a_1 b_3 E(X^{\theta}) + a_3\theta E(\ln(X)) + b_3\theta E(X^{\theta}\ln(X)) \\
E(U_1U_4) &= a_1 a_4 + a_1\lambda E(\ln(X)) + a_1 b_4 E(X^{-\theta}) + a_1\xi^{\theta} E(X^{-\theta}\ln(X)) \\
&\quad + a_1 c_4 E(X^{\theta}) - a_1\phi^{\theta} E(X^{\theta}\ln(X)) + a_4\theta E(\ln(X)) \\
&\quad + \lambda\theta E(\ln(X))^2 + b_4\theta E(X^{-\theta}\ln(X)) + \theta\xi^{\theta} E(X^{-\theta}(\ln(X))^2) \\
&\quad + c_4\theta E(X^{\theta}\ln(X)) - \theta\phi^{\theta} E(X^{\theta}(\ln(X))^2) \\
E(U_2U_3) &= a_2 a_3 + a_2 b_3 E(X^{\theta}) + a_3 b_2 E(X^{-\theta}) + b_2 b_3 \\
E(U_2U_4) &= a_2 a_4 + a_2\lambda E(\ln(X)) + a_2 b_4 E(X^{-\theta}) + a_2\xi^{\theta} E(X^{-\theta}\ln(X)) \\
&\quad + a_2 c_4 E(X^{\theta}) - a_2\phi^{\theta} E(X^{\theta}\ln(X)) + a_4 b_2 E(X^{-\theta}) + b_2 b_4 E(X^{-2\theta}) \\
&\quad + b_2\xi^{\theta} E(X^{-2\theta}\ln(X)) + b_2 c_4 - b_2\phi^{\theta} E(\ln(X)) + b_2\lambda E(X^{-\theta}\ln(X)) \\
E(U_3U_4) &= a_3 a_4 + a_3\lambda E(\ln(X)) + a_3 b_4 E(X^{-\theta}) + a_3\xi^{\theta} E(X^{-\theta}\ln(X)) \\
&\quad + a_3 c_4 E(X^{\theta}) - a_3\phi^{\theta} E(X^{\theta}\ln(X)) + a_4 b_3 E(X^{\theta}) + b_3\lambda E(X^{\theta}\ln(X)) \\
&\quad + b_3 b_4 + b_3\xi^{\theta} E(\ln(X)) + b_3 c_4 E(X^{2\theta}) - b_3\phi^{\theta} E(X^{2\theta}\ln(X)).
\end{aligned}
$$

Finally, one needs to calculate expected values of products of powers of X and the logarithm of X. Expected values of the form $E(X^p)$ are given by (1.4) while

$E(\ln(X))$ by (B.2), while the others are as follows, for $R_\lambda(z) = K_{\lambda+1}(z)/K_\lambda(z)$,

$$
\begin{aligned}
E(\ln(X))^2 &= \beta^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\ln(x))^2 x^{\lambda\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^\theta \ln(X)) &= \beta^{\frac{-\theta}{2}} R_\lambda(\rho) \left[\frac{k_{\lambda+1}(\rho)}{\theta K_{\lambda+1}(\rho)} - \frac{1}{2} \ln \beta \right] \\
E(X^{-\theta} \ln(X)) &= \beta^{\frac{\theta}{2}} \frac{1}{R_{\lambda-1}(\rho)} \left[\frac{k_{\lambda-1}(\rho)}{\theta K_{\lambda-1}(\rho)} - \frac{1}{2} \ln \beta \right] \\
E(X^\theta (\ln(X))^2) &= \beta^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\ln(x))^2 x^{(\lambda+1)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^{-\theta} (\ln(X))^2) &= \beta^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\ln(x))^2 x^{(\lambda-1)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^{2\theta} \ln(X)) &= \beta^{-\theta} \frac{K_{\lambda+2}(\rho)}{K_\lambda(\rho)} \left[\frac{k_{\lambda+2}(\rho)}{\theta K_{\lambda+2}(\rho)} - \frac{1}{2} \ln \beta \right] \\
E(X^{-2\theta} \ln(X)) &= \beta^{\theta} \frac{K_{\lambda-2}(\rho)}{K_\lambda(\rho)} \left[\frac{k_{\lambda-2}(\rho)}{\theta K_{\lambda-2}(\rho)} - \frac{1}{2} \ln \beta \right] \\
E(X^{2\theta} (\ln(X))^2) &= \beta^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\ln(x))^2 x^{(\lambda+2)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx \\
E(X^{-2\theta} (\ln(X))^2) &= \beta^{\frac{\lambda\theta}{2}} \frac{\theta}{2K_\lambda(\rho)} \int_0^\infty (\ln(x))^2 x^{(\lambda-2)\theta-1} \exp\{-\xi^\theta x^{-\theta} - \phi^\theta x^\theta\} dx,
\end{aligned}
$$

where $\beta = \phi/\xi$. All the above quantities are used to fill the Fisher information matrix and (numerically) obtain the Jeffreys' prior.

C Slice sampling

The basic idea of the slice sampler is to sample from a distribution, simply sampling from the uniform region under the density function and consider only the horizontal coordinates. One way to do this is to introduce auxiliary variables, then use Gibbs sampling (Gelfand and Smith, 1990) on the area beneath the density. Suppose we wish to sample from $f(X)$ where $x \in \mathbb{F} \subseteq \mathbb{R}$. To do this we sample uniformly from the 2-dimensional region under $f(x)$ or $g(x) = cf(x)$. This is implemented as follows:

- Introduce a auxiliary variable z with $z \mid x \sim U(0, g(x))$;
- This defines a uniform distribution for z and x over the region $\{(z, x) : 0 \leq z \leq g(x)\}$ with density

$$
\begin{aligned}
f(z,x) &= f(z \mid x) f(x) \\
&= \begin{cases} \dfrac{1}{g(x)} f(x) & , \text{se } 0 \le z \le g(x) \\ 0 & , \text{otherwise.} \end{cases} \\
&= \begin{cases} \dfrac{1}{c} & , \text{se } 0 \le z \le g(x) \\ 0 & , \text{otherwise.} \end{cases}
\end{aligned}
$$

- The conditional distribution for $x \mid z$ has density

$$
\begin{aligned}
f(x \mid z) &\propto f(z,x) \\
&\propto \begin{cases} \dfrac{1}{c} & , \text{se } 0 \le z \le g(x) \\ 0 & , \text{otherwise,} \end{cases}
\end{aligned}
$$

  that is, $(x \mid z) \sim U(S(z))$, where $S(z) = \{x : g(x) \ge z\}$.

Here $S(z)$ is the union of intervals that constitute the slice through the density defined by z. To obtain a sample from x, we first sample (x_i, z_i) from $f(x,z)$, then ignore the z_i's. The structure of this model leads us naturally to simulate using Gibbs sampling. The ith iteration of the algorithm is:

- `simulate` $z^{(i)} \sim U(0, g(x^{(i-1)}))$
- `simulate` $x^{(i)} \sim U(S(z^{(i)}))$ where $S(z^{(i)}) = \{x : g(x) \ge z^{(i)}\}$.

A key aspect to slice sampling is that only uniform random variates need be simulated, but determining the slice $S(y)$ may be trick. An alternative method to determine the slice $S(y)$ is given in Neal (2003).
However, when updates used do not account for the dependencies between variables, many updates will be needed to move from one part to the distribution to another. Sampling efficiency can be improved in this context by suppressing random walk behavior characteristic of simple schemes such as Gibbs sampling. One way of achieving this is by using 'overrelaxed' updates. Like Gibbs sampling, overrelaxation methods update each variable in turn, but rather than drawing a new value for a variable from its conditional distribution independently of the current value, the new value is instead chosen to be on the opposite side of the mode from the current value. A overrelaxation method within slice sampling is given in Neal (2003).

(Received January, 2005. Accepted February, 2006.)

References

Achcar, J. and Bolfarine, H. (1986). The log-linear model with a generalized gamma distribution for the error: a Bayesian approach. *Statistics and Probability Letters*, **4**, 325–332.

Akaike, H. (1974). New look at the statistical model identification. *IEEE Transactions in Automatic Control*, **AC-19**, 716–723.

Barndorff-Nielsen, O. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, **24**, 1–13.

Carlin, B. P., Polson, N. G. and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, **87**, 493–500.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.).

Gelfand, A. (1996). Model determination using sampling-based methods. In W. R. Gilks, S. Richardson and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (p. 145–161). London: Chapman and Hall.

Gelfand, A. and Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, **56**, 501–514.

Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Jørgensen, S. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. New York: Springer-Verlag.

Lopes, H. F. and Migon, H. S. (2002). Comovements and contagion in emergent markets: stock indexes volatilities. In A. Carriquiry, C. Gatsonis, A. Gelman, D. Higdon, R. Kass and D. Pauler. (Eds.), *Case Studies in Bayesian Statistics, Volume VI* (p. 285-300). New York: Springer-Verlag.

Neal, R. (2003). Slice sampling (with discussion). *Annals of Statistics*, **31**, 705–767.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. V. D. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*, **64**, 583–639.

Hedibert F. Lopes
Graduate School of Business
University of Chicago
5807 South Woodlawn Avenue, Chicago, IL 60637, USA
Telephone: (773) 834-5458
Fax: (773) 702-0458
E-mail: hlopes@ChicagoGSB.edu

Ralph S. Silva and **Helio dos Santos Migon**
Instituto de Matemática
Universidade Federal do Rio de Janeiro
Caixa Postal 68507
21945-970, Rio de Janeiro/RJ, Brazil
Tel./Fax: (21) 2280-7438
E-mails: ralph@dme.ufrj.br and migon@im.ufrj.br