Objective Bayesian analysis for exponential power regression models

Esther Salazar

SAMSI and Duke University, USA

Marco A. R. Ferreira

University of Missouri, USA

Helio S. Migon

Federal University of Rio de Janeiro, Brazil

Abstract

We develop objective Bayesian analysis for the linear regression model with random errors distributed according to the exponential power distribution. More specifically, we derive explicit expressions for three different Jeffreys priors for the model parameters. We show that only one of these Jeffreys priors leads to a proper posterior distribution. In addition, we develop fast posterior analysis based on Laplace approximations. Moreover, we show that our proposed Bayesian analysis compares favorably to a posterior analysis based on a competing noninformative prior. Finally, we illustrate our methodology with applications of the exponential power regression model to two different datasets.

Key words: Bayesian inference; exponential power errors; frequentist properties; Jeffreys prior; robustness.

1 Introduction

The exponential power is a flexible distribution for errors of regression models that may have tails either lighter (platykurtic) or heavier (leptokurtic) than Gaussian. In addition, the use of the exponential power distribution reduces the influence of outliers and consequently increases the robustness of the analysis (Box and Tiao, 1962; West, 1984; Liang et al., 2007). While leptokurtic distributions automatically protect against outliers, platykurtic distributions may occur as a result of truncation. Finally, the EP distribution is especially attractive because it includes the normal distribution as a special case and allows continuous variation from normality to nonnormality. Despite the importance of the EP distribution, there is no literature on objective priors for regression models with EP errors.

In this work we develop objective Bayesian analysis for regression models with independent EP errors. More specifically, we derive explicit expressions for three different Jeffreys priors and we show that only one of these Jeffreys priors leads to a proper posterior distribution. Moreover, a Monte Carlo study shows that our proposed Bayesian approach compares favorably to a posterior analysis based on a competing noninformative prior in terms of coverage of credible intervals and relative mean squared error. This good frequentist behavior of our objective Bayesian procedure has also been found in objective Bayesian analyses for other models such as, for example, in the analysis of elapsed times in continuous-time Markov chains (Ferreira and Suchard, 2008).

The EP distribution has been studied and popularized by Box and Tiao (1992) in the context of robustness studies. The EP density is given by

$$f(y|\mu,\sigma_p,p) = \left[2p^{1/p}\sigma_p\Gamma(1+1/p)\right]^{-1}\exp\left[-(p\sigma_p^p)^{-1}|y-\mu|^p\right],\ -\infty < y < \infty, \tag{1}$$

where $p > 1$, $-\infty < \mu < \infty$ and $\sigma_p > 0$. The EP distribution is characterized by three parameters: (i) $\mu = E(y)$, the location parameter, (ii) $\sigma_p = [E(|y-\mu|^p)]^{1/p}$, the scale parameter and (iii) p, the shape parameter. The scale parameter σ_p is also called power deviation of order p (Vianelli, 1963) and can be seen as a variability index that generalizes the standard deviation. Moreover, the kurtosis, denoted here by κ, is directly related to p since we have $\kappa = \Gamma(1/p)\Gamma(5/p)/(\Gamma(3/p))^2$ and therefore the shape parameter is linked to the thickness of the tails. For example, the EP distribution describes leptokurtic distributions if $p < 2$ ($\kappa > 3$) and platykurtic distributions if $p > 2$ ($\kappa < 3$). Some especial cases of the EP distribution are the Laplace distribution ($p = 1$), the normal distribution ($p = 2$) and, when $p \to \infty$, the uniform distribution on the interval $(\mu - \sigma_p, \mu + \sigma_p)$ (e.g., see Box and Tiao, 1992).

Bayesian procedures for regression models with EP errors have received little attention in the literature. In particular, no noninformative priors for the model parameters have been published to date. Box and Tiao (1992) studied the EP model from a Bayesian perspective in the context of robustness of regression models. Achcar and Pereira (1999) have considered mixtures of regression models with EP errors. In a different context, Choy and Smith (1997) used the EP distribution as a prior for a location parameter of a Gaussian likelihood model. Finally, when using Markov chain Monte Carlo (MCMC) to implement posterior analysis, one can explore representations of the EP distribution as a scale mixture of normals (West, 1987) or as a scale mixture of uniforms (Walker and Gutiérrez-Peña, 1999). Conversely, here we implement posterior analysis by using Laplace approximations and Newton-Cotes rules. As a result, when compared to MCMC alternatives our posterior analysis is much faster.

The remainder of the paper is organized as follows. Section 2 presents the EP regression model. Section 3 derives the Jeffreys-rule prior and two forms of the independence Jeffreys priors, and shows that only one of these priors leads to a proper posterior distribution. Section 4 studies frequentist properties of the Bayesian inferences based on our proposed prior and on a competing noninformative prior. Two applications illustrating leptokurtic and platykurtic behavior of the regression errors are presented in Section 5. Section 6 presents closing remarks and discussion.

2 The EP regression model

Consider the linear regression model in which an n-vector of observations $y = (y_1, \ldots, y_n)'$ satisfies

$$y = x\beta + \epsilon, \tag{2}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is the error vector such that $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed according to the exponential power distribution with location zero, scale parameter σ_p and shape parameter p. Here $x = (x_1, \ldots, x_n)'$ is the $n \times k$ matrix of explanatory variables and $\beta = (\beta_1, \ldots, \beta_k)' \in \mathbb{R}^k$ is the vector of regression coefficients.

Here we reparameterize the model in a fashion similar to Zhu and Zinde-Walsh (2009). More specifically, we define $\sigma = p^{1/p}\sigma_p\Gamma(1 + 1/p)$ and obtain the likelihood function

$$L(\theta; y, x) = \frac{1}{2^n\sigma^n} \exp\left[-\sum_{i=1}^{n}\left(\frac{\Gamma(1+1/p)|y_i - x_i'\beta|}{\sigma}\right)^p\right]. \tag{3}$$

Thus, the log-likelihood function is given by

$$l(\theta; y, x) \quad = \quad -n\log 2 - n\log\sigma - \sum_{i=1}^{n}\left[\frac{\Gamma(1+1/p)|y_i - x_i'\beta|}{\sigma}\right]^p. \tag{4}$$

We denote the model parameters by $\theta = (\beta, \sigma, p) \in \mathbb{R}^k \times (0, \infty) \times (1, \infty)$.

3 Jeffreys priors

We derive here three possible Jeffreys priors (Jeffreys, 1961) for the parameters of the exponential power regression model. The first is the Jeffreys-rule prior, given by $\pi(\theta) \propto \sqrt{\det I(\theta)}$, where $I(\theta)$ is the Fisher information matrix with (i,j) entry given by

$$\{I(\theta)\}_{ij} = E_{Y|\theta}\Big[-\frac{\partial^2}{\partial\theta_i \partial\theta_j} l(\theta; y, x)\Big], \qquad \theta_1 = \beta, \quad \theta_2 = \sigma, \quad \theta_3 = p,$$

where $l(\theta; y, x)$ is given by (4).

The other priors are two forms of the independence Jeffreys prior. Jeffreys (1961) noted that for multi-parameter problems the independence Jeffreys prior may provide better results than the Jeffreys-rule prior. More specifically, an independence Jeffreys prior is obtained by partitioning the parameter vector in groups of parameters. Then, for each group of parameters we compute a prior by applying Jeffreys-rule prior as if the other parameters were fixed. Finally, the joint independence Jeffreys prior is the product of the priors for the different groups. It is important to note that different groupings may lead to different independence Jeffreys priors. Here we consider independence Jeffreys priors corresponding to two groupings: $\pi^{I_1}(\theta)$ associated with the grouping $\{(\beta), (\sigma, p)\}$; and $\pi^{I_2}(\theta)$ associated with the grouping $\{(\beta), (\sigma), (p)\}$.

As we show below, the three priors we consider belong to the class of improper prior distributions given by

$$\pi(\theta) \propto \frac{\pi(p)}{\sigma^a}, \tag{5}$$

where $a \in \mathbb{R}$ is a hyperparameter and $\pi(p)$ is the 'marginal' prior of the shape parameter p.

The following proposition gives the Fisher information matrix for the EP regression model.

Proposition 1 *For $y = (y_1, \ldots, y_n)$, the Fisher information matrix $I(\theta)$ with elements ϕ_{ij} given by $\phi_{ij} = E_{y|\theta}\left[-\frac{\partial^2}{\partial\theta_i \partial\theta_j'} l(\theta; y, x)\right]$, with $\phi_{ij} = \phi_{ji}$ and θ_j the jth element of $\theta = (\beta, \sigma, p)$, is:*

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2}\Gamma(\frac{1}{p})\Gamma(2-\frac{1}{p})\sum_{i=1}^n x_i x_i' & 0 & 0 \\ 0 & \frac{np}{\sigma^2} & -\frac{n}{\sigma p} \\ 0 & -\frac{n}{\sigma p} & \frac{n}{p^3}(1+\frac{1}{p})\Psi'(1+\frac{1}{p}) \end{bmatrix}.$$

where $\Psi(\alpha) \equiv \Gamma'(\alpha)/\Gamma(\alpha)$ and $\Psi'(\alpha) \equiv \partial\Psi(\alpha)/\partial\alpha$ are the digamma and trigamma functions, respectively.

Proof. This follows as a consequence of Proposition 5 of Zhu and Zinde-Walsh (p. 90, 2009). $\square$

The following theorem provides the Jeffreys-rule prior and two independence Jeffreys priors for the parameters of the EP regression model.

Theorem 1 *Consider the EP regression model given in (4). Then, the independence Jeffreys priors based on the groupings* $\{(\beta), (\sigma, p)\}$ *and* $\{(\beta), (\sigma), (p)\}$*, and the Jeffreys-rule prior for* θ *denoted by* $\pi^{I_1}(\theta)$*,* $\pi^{I_2}(\theta)$*, and* $\pi^{J}(\theta)$*, respectively, are of the form (5) with*

$$a = 1, \qquad \pi^{I_1}(p) \propto p^{-1}\left[\left(1+\frac{1}{p}\right)\Psi'\left(1+\frac{1}{p}\right)-1\right]^{1/2}, \tag{6}$$

$$a = 1, \qquad \pi^{I_2}(p) \propto p^{-3/2}\left[\left(1+\frac{1}{p}\right)\Psi'\left(1+\frac{1}{p}\right)\right]^{1/2}, \tag{7}$$

$$a = k+1, \qquad \pi^{J}(p) \propto \left[\Gamma\left(\frac{1}{p}\right)\Gamma\left(2-\frac{1}{p}\right)\right]^{k/2}\pi^{I_1}(p). \tag{8}$$

Proof. See the Appendix. $\square$

A prior of the form (5) leads to a proper posterior distribution if and only if

$$\int_1^{\infty} L^I(p; y)\pi(p)dp < \infty, \tag{9}$$

where $L^I(p; y)$, the integrated likelihood for p, is given by

$$L^I(p; y) \;\propto\; \int_{\mathbb{R}^k}\int_0^{\infty} L(\beta, \sigma, p; y)\sigma^{-a}d\sigma d\beta.$$

The following two lemmas give the tail behavior of the marginal priors for p given in Theorem 1 and of the integrated likelihood for p, providing the key to determining whether the corresponding posterior distributions are proper.

Lemma 1 *The marginal priors for* p *given in Theorem 1 are continuous functions in* $[1, \infty)$ *and are such that* $\pi^{I_1}(p) = O(p^{-1})$*,* $\pi^{I_2}(p) = O(p^{-3/2})$*, and* $\pi^{J}(p) = O(p^{(k-2)/2})$ *as* $p \to \infty$*.*

Proof. Direct inspection shows that $\pi^{I_1}(p)$, $\pi^{I_2}(p)$, and $\pi^{J}(p)$ are continuous functions in $[1, \infty)$. Their behavior when $p \to \infty$ follows from the fact that $\Psi'\left(1+\frac{1}{p}\right) \to 1.6449$ and $\Gamma\left(\frac{1}{p}\right) = O(p)$ as $p \to \infty$. $\square$

Lemma 2 *Provided that $n > k + 1 - a$, the integrated likelihood for p under the class of priors (5) is a continuous function in $[1, \infty)$ and is such that $L^I(p; y) = O(1)$ as $p \to \infty$.*

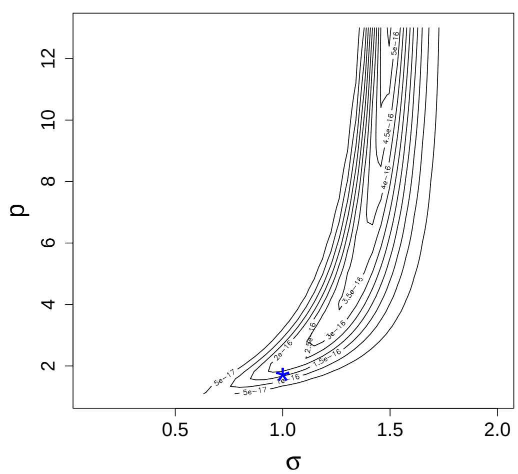
Proof. See the Appendix. □

The following proposition establishes that among the three Jeffreys priors considered above, only the independence Jeffreys prior π^{I_2} yields a proper posterior distribution.

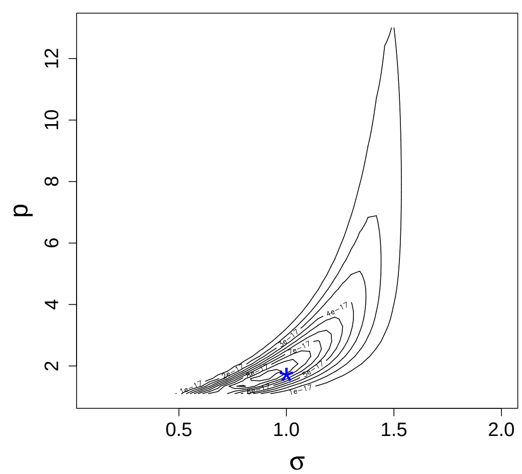
Proposition 2 *The independence Jeffreys prior π^{I_1} given in (6) and the Jeffreys-rule prior π^J given in (8) lead to improper posterior distributions. Moreover, provided that $n > k + 1 - a$, the independence Jeffreys prior π^{I_2} given in (7) yields a proper posterior distribution.*

Proof. These follow directly from condition (9), and Lemmas 1 and 2. □

Figure 1 shows the effect of the proposed independence Jeffreys prior π^{I_2} in the analysis of a simulated sample of size 30. As Figure 1 shows, the proposed prior calibrates the likelihood function and leads to a better behaved posterior density.



(a)



(b)

Figure 1: (a) Contour plot of the likelihood function for (σ, p) considering a data set of size $n = 30$ with parameters $\beta = 0$, $\sigma = 1$ and $p = 1.8$. (b) Contour plot of the joint posterior distribution of (σ, p) based on the proposed prior. The star symbol $*$ indicates the position of the true value.

We have implemented posterior analysis based on the independence Jeffreys prior $\pi^{I_2}(\theta)$. More specifically, we obtain marginal posterior densities, posterior expectations, and credible intervals using a combination of Laplace approximations and Newton-Cotes integration. Our deterministic numerical integration implementation allows posterior analysis that is fast and precise.

4 Computational details

This section presents the computational details of our implementation of a fast posterior analysis for the EP regression model. More specifically, we base our implementation on the Laplace approximation of integrals (Tierney and Kadane, 1986) and on Newton-Cotes integration (Press et al., 2007). For simple notation, density approximations obtained via Laplace and Newton-Cotes methods are indicated by the superscripts LA and NC, respectively. Throughout this section we consider a prior of the form (5).

First, we note that σ can be integrated out analytically. More specifically, integrating out σ we obtain the integrated likelihood function for (β, p)

$$\begin{aligned} L^I(\beta, p; y) &= \int_0^\infty L(\beta, \sigma, p; y)\pi(\sigma)d\sigma \\ &\propto p^{-1}\Gamma\left(\frac{n+a-1}{p}\right)\left\{\Gamma\left(1+\frac{1}{p}\right)\right\}^{-(n+a-1)}\left\{\sum_{i=1}^{n}|y_i - x_i'\beta|^p\right\}^{-\frac{n+a-1}{p}}. \end{aligned} \tag{10}$$

Based on the integrated likelihood function (10), the marginal posterior densities for p, β_l ($l = 1, \ldots, k$) and σ are given by:

$$\pi(p|y) \propto \int L^I(\beta, p; y)\pi(p)d\beta, \tag{11}$$

$$\pi(\beta_l|y) \propto \int L^I(\beta_l, \beta^{(-l)}, p; y)\pi(p)d\beta^{(-l)}\, dp, \tag{12}$$

$$\pi(\sigma|y) \propto \int \pi(\sigma|p, \beta, y)L^I(\beta, p; y)\pi(p)d\beta\, dp, \tag{13}$$

where $\beta^{(-l)}$ is a $(k-1)$-dimensional vector without the element β_l.

We can apply Laplace's approximation to (11) if we are able to compute first and second derivatives and maximize the integrand. More specifically, for a given p, let $\hat{\beta} = \hat{\beta}(p)$ maximize the function $L^I(\beta, p; y)\pi(p)$, and let $\hat{\Sigma}(p)$ be minus the inverse of the Hessian of $h(\hat{\beta}, p) = (\log L^I(\hat{\beta}, p; y) + \log \pi(p))/n$. Hence, applying Laplace's approximation to the integral of expression (11) we obtain

$$\pi^{LA}(p|y) \propto \left(|\hat{\Sigma}(p)|\right)^{1/2} \exp\{nh(\hat{\beta}, p)\}. \tag{14}$$

The computation of $\hat{\Sigma}(p)$ becomes numerically unstable for values of p close to 1. In this case, we approximate (11) using the Newton-Cotes method, which yields

$$\pi^{NC}(p|y) \;\; \propto \;\; \sum_j L^I(\beta_j, p; y)\pi(p)\Delta_j^{\beta}, \tag{15}$$

where the sum is over values of β with volume weights Δ_j^{β}. Usually, approximation (14) is much faster to compute then approximation (15). Thus, when $\hat{\Sigma}(p)$ can be computed the approximation (14) is preferred.

Using Laplace's method to approximate integrals with respect to p is much more delicate. In particular, it is often the case that the marginal posterior density of p is not log-concave as illustrated by Figure 4b in the application of Section 6.1. Thus, to obtain an approximation to the marginal posterior density of each element of β, i.e. β_l $(l = 1, \ldots, k)$, we use Newton-Cotes method. Applying this method to expression (12) we obtain

$$\pi^{NC}(\beta_l|y) \;\; \propto \;\; \sum_j L^I(\beta_l, \beta_j^{(-l)}, p_j; y)\pi(p_j)\Delta_j^{\beta^{(-l)},p},$$

where the sum is over values of $(\beta^{(-l)}, p)$ with volume weights $\Delta_j^{\beta^{(-l)},p}$.

Finally, we note the fact that the full conditional density of σ is

$$\pi(\sigma|p,\beta,y) = p\sigma^{p-1} IG\left(\sigma^p \;\middle|\; (n+a-1)p^{-1}, [\Gamma(1+p^{-1})]^p \sum_{i=1}^n |y_i - x_i'\beta|\right),$$

where $IG(s|a,b)$ denotes the density function of an Inverse-Gamma distribution with parameters a and b evaluated at s. Using this full conditional density, we approximate (13) by

$$\pi^{NC}(\sigma|y) \;\; \propto \;\; \sum_j \pi(\sigma|p_j, \beta_j, y) L^I(\beta_j, p_j; y)\pi(p_j)\Delta_j^{\beta,p},$$

where the sum is over values of (β, p) with volume weights $\Delta_j^{\beta,p}$.

5 Frequentist properties

This section compares frequentist properties of Bayesian procedures based on the objective prior π^{I_2} and based on a competing noninformative prior that we denoted by π^U. More specifically,

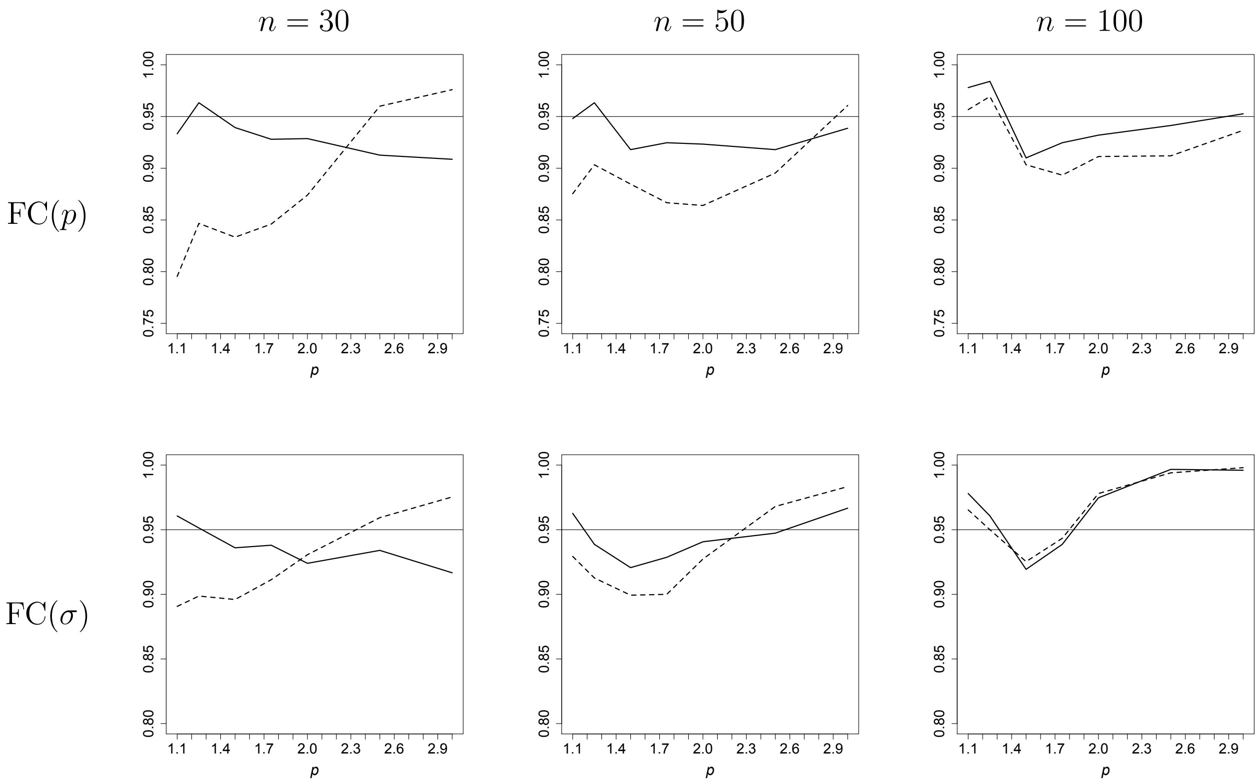
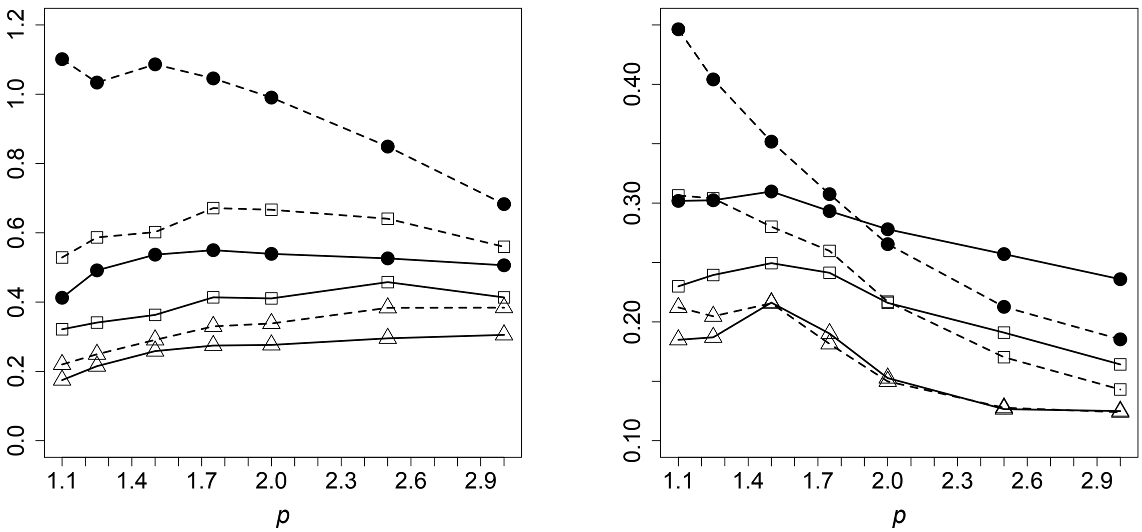


Figure 2: Frequentist coverage, as a function of p, of 95% HPD credible intervals for p and σ based on the objective prior π^{I_2} (solid line) and the noninformative prior π^U (dashed line). Sample sizes: $n = 30$, 50 and 100. Horizontal line indicates the 95% nominal level.

the prior π^U is of the form (5) with $a = 1$ and $\pi^U(p) \propto 1$ such that $1 < p < 10$. The uniform prior $\pi^U(p)$ reflects lack of information about p, and the resulting joint prior $\pi^U(\theta)$ yields a proper posterior distribution. We use as Bayesian estimators the posterior modes and compare them using the square root of the frequentist relative mean squared error. In addition, we investigate the frequentist coverage of the 95% highest posterior density (HPD) credible intervals.

To compute the frequentist properties of the several procedures, we have simulated 1500 datasets for each set of parameter and sample size specifications. More specifically, we have considered three sample sizes: $n = 30$, $n = 50$ and $n = 100$. Moreover, for p equal to one of several values ranging from 1 to 3 we have generated samples considering $k = 2$, $x_i = (1, x_{1i})$, $x_{1i} \sim N(2, 1)$, $\beta = (1.5, -3)$, and $\sigma = 1$.

Figure 2 shows, as a function of p, the frequentist coverage (FC) of 95% HPD credible intervals



(a) $\sqrt{MSE(\widehat{p})}/p$ (b) $\sqrt{MSE(\widehat{\sigma})}/\sigma$

Figure 3: Square root of the relative mean squared error, as a function of p, of estimators of p (left panel) and σ (right panel) based on the independence Jeffreys prior π^{I_2} (solid line) and prior π^U (dashed line) for $n = 30$ (circle), $n = 50$ (square) and $n = 100$ (triangle).

for p and σ. As expected, when the sample size increases the behavior of the credible intervals based on the two priors becomes more similar in terms of frequentist coverage. Moreover, when we consider credible intervals for σ with sample sizes equal to 30 or 50 the π^{I_2}-based credible intervals have frequentist coverage slightly closer to the nominal level. Finally, for all the sample sizes we consider the π^{I_2}-based credible intervals for p have frequentist coverage overall closer to the nominal level.

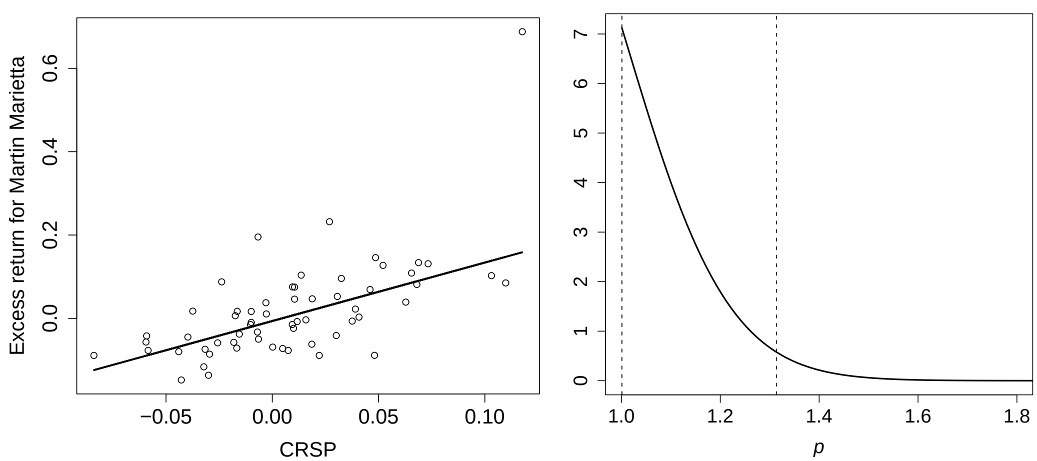
Figure 3 shows, as a function of p, the square root of the relative mean squared error (RMSE), $\sqrt{MSE(\widehat{\theta})}/\theta$, for estimators of p and σ. As expected, the RMSE decreases as the sample size increases. In addition, the behavior of the π^{I_2}- and π^U-based estimators becomes more similar in terms of RMSE as the sample size increases. For the estimation of σ, the π^{I_2}-based estimator is better when $p < 2$ and the π^U-based estimator is better when $p > 2$. For the estimation of p, in the range of values we consider the π^{I_2}-based estimator provides uniformly better results.

6 Applications

We illustrate our objective Bayesian methodology for exponential power regression models with two applications. The first application considers a dataset previously analyzed by Butler et al. (1990) on excess rate of return for the Martin Marietta company. The second application uses a dataset previously analyzed by Levine et al. (2006) to study the relationship between profits at the box office and number of sold home videos.

6.1 Excess returns for Martin Marietta company

Here we consider 60 observations from the Martin Marietta company collected over a period of 5 years on a monthly basis, from January 1982 to December 1986. This dataset has been previously analyzed by Butler et al. (1990) and DiCiccio and Monti (2004). The variables of interest are the excess rate of return for the Martin Marietta company (y) and the index for the excess rate returns (x) for the New York stock exchange (CRSP). The scatterplot of the data (Figure 4(a)) shows one very extreme observation and thus indicates that it may not be appropriate to assume Gaussian errors.



(a) (b)

Figure 4: Martin and Marietta data set. (a) Scatterplot of the data and fitted EP regression model. (b) Marginal posterior densities for p (Vertical dashed lines indicate the 95% HPD credible interval).

We have fitted the exponential power regression model to these data. Table 1 shows posterior summaries for each parameter: posterior mode, posterior median and 95% HPD credible intervals. In particular, the 95% credible interval for p, equal to $(1.000, 1.314)$, is completely contained in the interval $(1.0, 2.0)$ and thus strongly indicates a leptokurtic distribution for the errors. This is further confirmed by the marginal posterior density for p presented in Figure 4(b): the density is a decreasing function of p and the posterior mode of p is equal to 1.

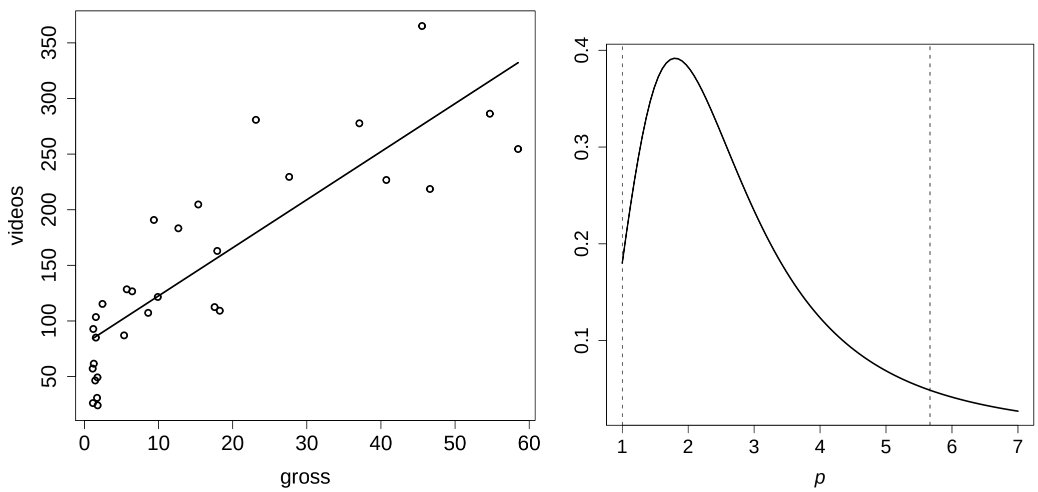
Table 1: Martin and Marietta data set. Posterior summaries based on the independence Jeffreys prior π^{I_2}.

Parameter	Median	Mode	95% C.I.
β_1	-0.006	-0.006	(-0.027, 0.014)
β_2	1.327	1.295	(0.891, 1.844)
σ	0.064	0.062	(0.047, 0.085)
p	1.092	1.000	(1.000, 1.314)

6.2 Sold home videos vs. profits at the box office

Figure 5(a) shows the scatterplot of a dataset previously analyzed by Levine et al. (2006) to study the relationship between the number of sold home videos in thousands (videos: y) and the profits at the box office in million of dollars (gross: x) for a sample of 30 movies. A preliminary analysis with a Gaussian regression model yields residuals with kurtosis equal to 2.25 and thus suggests a platykurtic behavior.

We have fitted a linear regression model assuming the EP distribution for the errors. Figure 5(b) shows the marginal posterior density for p. While the *prior* probability of p being greater than 2 is about 0.33, the *posterior* probability increases to 0.68. Thus, the data analysis provides some evidence of platykurtic behavior. Moreover, as the prior for p is proper we can use the Bayes Factor for model comparison (Kass and Raftery, 1995). The Bayes factor for the regression model with EP errors against the Gaussian regression model is 190.44 and indicates strong evidence in favor of the EP regression model.



(a) (b)

Figure 5: Videos data set: (a) Scatterplot of the data and fitted EP regression model. (b) Marginal posterior densities for p (Vertical dashed lines indicate the 95% HPD credible interval).

7 Discussion

We have developed objective Bayesian analysis for the exponential power regression model. In addition, we have developed computational methods based on Laplace's method and Newton-Cotes integration to approximate marginal posterior densities of the EP regression model parameters. This computational methodology allows fast and precise posterior analysis. Finally, when compared with procedures based on a uniform noninformative prior for the shape parameter, our proposed Bayesian procedures have better frequentist properties.

Our finding that for the EP regression model the *Jeffreys-rule prior* leads to an improper posterior distribution whereas an *independence Jeffreys prior* leads to a proper posterior distribution is quite surprising. When there is any posterior propriety issue, usually either the *independence Jeffreys prior* leads to an improper posterior while the *Jeffreys-rule prior* leads to a proper posterior (e.g., Berger et al., 2001; Ferreira and De Oliveira, 2007), or both lead to an improper posterior (e.g., Wasserman, 2000). In contrast, and to the best of our knowledge, the EP regression model we consider is the only known example to date where the *Jeffreys-rule prior* leads to an improper posterior distribution whereas an *independence Jeffreys prior* leads to a proper posterior distribution.

Acknowledgement

The work of Salazar was supported in part by a FAPERJ grant. The work of Ferreira was supported in part by National Science Foundation Grant DMS-0907064. The work of Migon was supported in part by CNPq and CAPES Grants. Part of this research was done while Salazar was a postdoctoral fellow at the Federal University of Rio de Janeiro.

Appendix

Proof of Theorem 1. Using the results of Proposition 1 we have that

- For independence Jeffreys prior $\pi^{I_1}(\theta)$: marginal priors for β and (σ, p) are independent a priori such that $\pi^{I_1}(\beta, \sigma, p) = \pi^{I_1}(\beta)\pi^{I_1}(\sigma, p)$ and

$$\begin{aligned} \pi^{I_1}(\sigma, p) &\propto \sqrt{I_{\sigma\sigma}I_{pp} - I_{\sigma p}^2} \propto \frac{1}{\sigma p}\left[\left(1+\frac{1}{p}\right)\Psi'\left(1+\frac{1}{p}\right) - 1\right]^{1/2}, \\ \pi^{I_1}(\beta) &\propto \sqrt{\det[I_{\beta\beta}]} \propto 1. \end{aligned}$$

- For independence Jeffreys prior $\pi^{I_2}(\theta)$: let us consider independence Jeffreys prior such that $\pi^{I_2}(\theta) \propto \pi^{I_2}(\beta)\pi^{I_2}(\sigma)\pi^{I_2}(p)$, that is, taking each of these parameters as independent. Then, from the Fisher information matrix, $\pi^{I_2}(\beta) \propto 1$, $\pi^{I_2}(\sigma) \propto \sigma^{-1}$ and $\pi^{I_2}(p) \propto p^{-3/2}(1 + p^{-1})^{1/2}\{\Psi'(1 + p^{-1})\}^{1/2}$. Thus,

$$\pi^{I_2}(\beta, \sigma, p) \propto \frac{1}{\sigma p^{3/2}}\left[\left(1+\frac{1}{p}\right)\Psi'\left(1+\frac{1}{p}\right)\right]^{1/2}.$$

- For Jeffreys-rule prior: $\pi^J(\beta, \sigma, p) \propto \sqrt{\det[I(\theta)]} = \sqrt{I_{\sigma\sigma}I_{pp} - I_{\sigma p}^2}\sqrt{\det[I_{\beta\beta}]}$ where $\det[I_{\beta\beta}] \propto \left[\frac{1}{\sigma^2}\Gamma\left(\frac{1}{p}\right)\Gamma\left(2-\frac{1}{p}\right)\right]^k$. Therefore,

$$\pi^J(\beta, \sigma, p) \propto \frac{1}{\sigma^k}\left[\Gamma\left(\frac{1}{p}\right)\Gamma\left(2-\frac{1}{p}\right)\right]^{k/2}\pi^{I_1}(\sigma, p).$$

$\square$

Proof of Lemma 2.

Considering the integrated likelihood in (10), the integrated likelihood for p is

$$\begin{aligned} L^I(p;y) &= \int_{\mathbb{R}^k} L^I(\beta,p;y)\pi(\beta)d\beta \\ &\propto p^{-1}\Gamma\left(\frac{n+a-1}{p}\right)\left\{\Gamma\left(1+\frac{1}{p}\right)\right\}^{-(n+a-1)} \int_{\mathbb{R}^k}\left\{\sum_{i=1}^n |y_i - x_i'\beta|^p\right\}^{-\frac{n+a-1}{p}} d\beta. \end{aligned}$$

Let us define the following two functions:

$$\begin{aligned} h(\beta,p) &= n\left(\max|y_i| + \sum_{l=1}^k |\tilde{x}_l||\beta_l|\right)^p \\ g(\beta,p) &= \begin{cases} n\,|\bar{y} - \bar{x}'\beta^*|^p\,, & \text{if } \beta \in C_1, \\ n\,|\bar{y} - \bar{x}'\beta|^p\,, & \text{if } \beta \in C_2, \end{cases} \end{aligned}$$

where $|\tilde{x}_l| = \max|x_{il}|$, $C_1 = \{\beta \in \mathbb{R}^k : \sum_{l=1}^k |\bar{x}_l||\beta_l| \le |\bar{y}|\}$, $C_2 = \{\beta \in \mathbb{R}^k : \sum_{l=1}^k |\bar{x}_l||\beta_l| > |\bar{y}|\}$ and $\beta^* = \arg\min_\beta \sum_{i=1}^n |y_i - x_i'\beta|^p$. Note that $\sum_{i=1}^n |y_i - x_i'\beta^*|^p > 0$ with probability 1 and, for example, if $p = 2$ then $\beta^* = (x'x)^{-1}x'y$.

Note the following:

(i)

$$\begin{aligned} \sum_{i=1}^n |y_i - x_i'\beta|^p &\le \sum_{i=1}^n (|y_i| + |x_i'\beta|)^p \le \sum_{i=1}^n \left(\max|y_i| + \sum_{l=1}^k |x_{il}|\,|\beta_l|\right)^p \\ &\le \sum_{i=1}^n \left(\max|y_i| + \sum_{l=1}^k |\tilde{x}_l|\,|\beta_l|\right)^p = h(\beta,p). \end{aligned}$$

(ii) For $p \ge 1$ the function $|\cdot|^p$ is convex. Thus, by Jensen's inequality

$$\sum_{i=1}^n |y_i - x_i'\beta|^p \ge n\left|\frac{1}{n}\sum_{i=1}^n (y_i - x_i'\beta)\right|^p = n|\bar{y} - \bar{x}'\beta|^p.$$

(iii) Consider $\beta \in C_1$. Using the definition of β^* and result (ii) we have

$$\sum_{i=1}^n |y_i - x_i'\beta|^p \ \ge\ \sum_{i=1}^n |y_i - x_i'\beta^*|^p \ge n|\bar{y} - \bar{x}'\beta^*|^p = g(\beta,p).$$

(iv) Consider $\beta \in C_2$. Then by result (ii) we have

$$\sum_{i=1}^n |y_i - x_i'\beta|^p \ge n|\bar{y} - \bar{x}'\beta|^p = g(\mu,p).$$

(v) Thus, by results (iii) and (iv),

$$\sum_{i=1}^{n} |y_i - x_i'\beta|^p \geq g(\beta, p).$$

Therefore, by results (i) and (v) above,

$$\int_{\mathbb{R}^k} \{h(\beta,p)\}^{-\frac{n+a-1}{p}} d\beta \leq \int_{\mathbb{R}^k} \left\{ \sum_{i=1}^{n} |y_i - x_i'\beta|^p \right\}^{-\frac{n+a-1}{p}} d\beta \leq \int_{\mathbb{R}^k} \{g(\beta,p)\}^{-\frac{n+a-1}{p}} d\beta.$$

Let us now compute the left and right integral above. For the left integral and considering $\beta_{(-j)} = (\beta_{j+1}, \ldots, \beta_k)' \in \mathbb{R}^{k-j}$ for $j = 1, \ldots, k-1$ we have:

$$\begin{aligned}
\int_{\mathbb{R}^k} \{h(\beta,p)\}^{-\frac{n+a-1}{p}} d\beta &= 2 \int \int_0^{\infty} \{n^{1/p}(\max|y_i| + \sum_{l=2}^{k} |\tilde{x}_l||\beta_l| + |\tilde{x}_1|\beta_1)\}^{-(n+a-1)} d\beta_1 d\beta_{(-1)} \\
&= 2 \int \left. \frac{n^{-\frac{n+a-1}{p}} \{max|y_i| + \sum_{l=2}^{k} |\tilde{x}_l||\beta_l| + |\tilde{x}_1|\beta_1\}^{-(n+a-2)}}{|\tilde{x}_1|\{-(n+a-2)\}} \right|_0^{\infty} d\beta_{(-1)} \\
&= 2 \int \frac{n^{-\frac{n+a-1}{p}} \{\max|y_i| + \sum_{l=2}^{k} |\tilde{x}_l||\beta_l|\}^{-(n+a-2)}}{|\tilde{x}_1|(n+a-2)} d\beta_{(-1)}.
\end{aligned}$$

The integral above can be computed recursively for each element of the vector $\beta_{(-1)}$. The result is convergent as long as $n > k + 1 - a$, thus

$$\int_{\mathbb{R}^k} \{h(\beta,p)\}^{-\frac{n+a-1}{p}} d\beta = n^{-\frac{n+a-1}{p}} m_1(y),$$

where $m_1(y) = \frac{2^k \Gamma(n+a-k-1)\{\max|y_i|\}^{-(n+a-1-k)}}{\Gamma(n+a-1)\prod_{l=1}^{k}|\tilde{x}_l|} > 0$, does not depend on p.

The right integral is

$$\begin{aligned}
\int_{\mathbb{R}^k} \{g(\beta,p)\}^{-\frac{n+a-1}{p}} d\beta &= \int_{C_1} \left\{ n\,|\bar{y} - \bar{x}'\beta^*|^p \right\}^{-\frac{n+a-1}{p}} d\beta \\
&\quad + \int_{C_2} \left\{ n\,|\bar{y} - \bar{x}'\beta|^p \right\}^{-\frac{n+a-1}{p}} d\beta \\
&= n^{-\frac{n+a-1}{p}} m_2(y),
\end{aligned}$$

where $m_2(y) = \int_{C_1} |\bar{y} - \bar{x}'\beta^*|^{-(n+a-1)} d\beta + \int_{C_2} |\bar{y} - \bar{x}'\beta|^{-(n+a-1)} d\beta$ does not depend on p.

Therefore, we have shown that for $n > k + 1 - a$

$$m_1(y) \leq n^{\frac{n+a-1}{p}} \int_{-\infty}^{\infty} \left\{ \sum_{i=1}^{n} |y_i - x_i'\beta|^p \right\}^{-\frac{n+a-1}{p}} d\beta \leq m_2(y).$$

The result above will allow us to study the tail behavior of the integrated likelihood for p. First, note that from the series expansion of $\{\Gamma(z)\}^{-1}$ (Abramowitz and Stegun, 1972, p.256) we obtain that $\{\Gamma(z)\}^{-1} \approx z$ from z close to 0. Thus, as p goes to ∞ we have $\Gamma(\frac{1}{p}) \approx p$. In addition, considering the first-order Taylor expansion of $\log \Gamma(1+z)$ around $z=0$, $\log \Gamma(1+z) \approx \Psi(1)z$, where $\Psi(1) \approx -0.5772$ is the digamma function evaluated at 1. Thus, $\Gamma(1+\frac{1}{p}) \approx \mathrm{e}^{-0.5772/p}$ for large p. Therefore, for $p \to \infty$,

$$\begin{aligned}
L^I(p;y) &\propto p^{-1}\Gamma\left(\frac{n+a-1}{p}\right)\left\{\Gamma\left(1+\frac{1}{p}\right)\right\}^{-(n+a-1)} \int_{\infty}^{\infty} \left\{\sum_{i=1}^{n} |y_i - x_i'\beta|^p\right\}^{-\frac{n+a-1}{p}} d\beta \\
&\approx p^{-1}\frac{p}{n+a-1}\mathrm{e}^{\Psi(1)(n+a-1)/p} O(n^{-(n+a-1)/p}) \\
&= O(\mathrm{e}^{-\frac{n+a-1}{p}\{\log n - \Psi(1)\}}) \\
&= O(1).
\end{aligned}$$

$\square$

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover.

Achcar, J. A. and Pereira, G. A. (1999). "Use of Exponential Power Distributions for Mixture Models in the Presence of Covariates." *Journal of Applied Statistics*, 26, 6, 669–679.

Berger, J. O., de Oliveira, V., and Sansó, B. (2001). "Objective Bayesian Analysis of Spatially Correlated Data." *Journal of the American Statistical Association*, 96, 456, 1361–1374.

Box, G. E. P. and Tiao, G. C. (1962). "A further look at robustness via Bayes's theorem." *Biometrika*, 49, 419–432.

— (1992). *Bayesian Inference in Statistical Analysis*. Wiley-Interscience.

Butler, R. J., McDonald, J. B., Nelson, R. D., and White, S. B. (1990). "Robust and partially adaptive estimation of regression models." *Review of Economic Statistics*, 72, 321–327.

Choy, S. T. B. and Smith, A. F. M. (1997). "On Robust Analysis of a Normal Location Parameter." *Journal of the Royal Statistical Society, Series B*, 59, 2, 463–474.

DiCiccio, T. J. and Monti, A. C. (2004). "Inferential aspects of the skew exponential power distribution." *Journal of the American Statistical Association*, 99, 466, 439–450.

Ferreira, M. A. R. and De Oliveira, V. (2007). "Bayesian reference analysis for Gaussian Markov Random Fields." *Journal of Multivariate Analysis*, 98, 789–812.

Ferreira, M. A. R. and Suchard, M. A. (2008). "Bayesian Analysis of Elapsed Times in Continuous-Time Markov Chains." *Canadian Journal of Statistics*, 36, 355–368.

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press.

Kass, R. E. and Raftery, A. E. (1995). "Bayes Factor." *Journal of the American Statistical Association*, 90, 773–795.

Levine, D. M., Krehbiel, T. C., and Berenson, M. L. (2006). *Business Statistics: A First Course*. Pearson Prentice Hall.

Liang, F., Liu, C., and Wang, N. (2007). "A robust sequential Bayesian method for identification of differentially expressed genes." *Statistica Sinica*, 17, 571–597.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (2007). *Numerical Recipes in C: The Art of Scientific Computing*. 3rd ed. Cambridge: Cambridge University Press.

Tierney, L. and Kadane, J. B. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81, 82–86.

Vianelli, S. (1963). "La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza." *Statistica*, 23, 447–474.

Walker, S. G. and Gutiérrez-Peña, E. (1999). "Robustifying Bayesian Procedures." In *Bayesian Statistics 6*. New York: Oxford University Press.

Wasserman, L. (2000). "Asymptotic inference for mixture models using data-dependent priors." *Journal of the Royal Statistical Society, Series B*, 62, 159–180.

West, M. (1984). "Outlier models and prior distributions in Bayesian linear regression." *Journal of the Royal Statistical Society, Series B*, 46, 431–439.

— (1987). "On scale mixtures of normal distributions." *Biometrika*, 79, 646–648.

Zhu, D. and Zinde-Walsh, V. (2009). "Properties and estimation of asymmetric exponential power distribution." *Journal of Econometrics*, 148, 86–99.