

METHODOLOGY

Open Access

Bayesian reference analysis for exponential power regression models

Marco AR Ferreira^{1*} and Esther Salazar²

*Correspondence:

ferreiram@missouri.edu

¹Department of Statistics, University of Missouri, Columbia, USA

Full list of author information is available at the end of the article

Abstract

We develop Bayesian reference analyses for linear regression models when the errors follow an exponential power distribution. Specifically, we obtain explicit expressions for reference priors for all the six possible orderings of the model parameters and show that, associated with these six parameters orderings, there are only two reference priors. Further, we show that both of these reference priors lead to proper posterior distributions. Furthermore, we show that the proposed reference Bayesian analyses compare favorably to an analysis based on a competing noninformative prior. Finally, we illustrate these Bayesian reference analyses for exponential power regression models with applications to two datasets. The first application analyzes per capita spending in public schools in the United States. The second application studies the relationship between sold home videos versus profits at the box office.

MSC: 62F15; 62F35; 62J05

Keywords: Bayesian inference; Exponential power errors; Frequentist properties; Reference prior; Robustness

1 Introduction

A flexible way to deal with outliers in linear regression is to assume that the errors follow an exponential power (EP) distribution. Specifically, assuming an EP distribution decreases the influence of outliers and, as a result, increases the robustness of the analysis (Box and Tiao 1962; Liang et al. 2007; Salazar et al. 2012; West 1984). In addition, the EP distribution includes the Gaussian distribution as a particular case. Further, the EP distribution may have tails either lighter (platykurtic) or heavier (leptokurtic) than Gaussian. Platykurtic distributions may be a result of truncation, whereas leptokurtic distributions provide protection against outliers. Salazar et al. (2012) have developed three types of Jeffreys priors for linear regression models with independent EP errors. Unfortunately, two of those priors lead to useless improper posterior distributions and only one leads to a proper posterior distribution. Here we develop explicit expressions for reference priors for all the six possible orderings of the model parameters.

We show that the six parameters orderings lead to two distinct reference priors. The parameter ordering corresponds to the order of importance of each parameter in the analysis, with the most important parameter appearing first and the least important appearing last (Berger and Bernardo 1992a,b). In addition to the two formally obtained reference priors, we propose an approximate reference prior that shares the same tail behavior

but is much more straightforward to implement in practice. Finally, we show that the two reference priors lead to useful proper posterior distributions.

To make sure that Bayesian reference procedures do not bias the data analysis in an undesirable manner, it is important to study their frequentist properties. To study the frequentist properties of our proposed procedures, we have performed a Monte Carlo study that shows that our proposed Bayesian reference approaches compare favorably to a posterior analysis based on a competing prior in terms of coverage of credible intervals, relative mean squared error, and mean length of credible intervals. While the relative mean squared error and the mean length of credible intervals should be judged in comparison with those yielded by competing priors, the coverage of credible intervals should be as close as possible to the nominal level.

Coverage of credible intervals close to nominal provides a guarantee of level of performance of the procedure when used automatically and independently by many researchers in their problems. In our Monte Carlo study, we have found that the Bayesian reference credible intervals that we have obtained have frequentist coverage close to nominal. These good frequentist properties results agree with previous literature on Bayesian reference analyses for other models such as, for example, Gaussian random fields (Berger et al. 2001), Markov random fields (Ferreira and De Oliveira 2007), multivariate normal models (Sun and Berger 2007), and elapsed times in continuous-time Markov chains (Ferreira and Suchard 2008).

The EP density is given by

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

where $p > 1$, $-\infty < \mu < \infty$ and $\sigma_p > 0$. The EP distribution has three parameters: the location parameter $\mu = E(y)$, the scale parameter $\sigma_p = [E(|y - \mu|^p)]^{1/p}$, and the shape parameter p . The scale parameter σ_p can be seen as a variability index that generalizes the standard deviation. Moreover, σ_p is also known as power deviation of order p (Vianelli 1963). In addition, the kurtosis is $\kappa = \Gamma(1/p)\Gamma(5/p)/(\Gamma(3/p))^2$, implying that the shape parameter p determines the thickness of the tails of the EP density. Specifically, the EP distribution is leptokurtic if $p < 2$ ($\kappa > 3$) and platykurtic if $p > 2$ ($\kappa < 3$). Finally, the EP distribution has several important especial cases such as the Laplace distribution ($p = 1$), the normal distribution ($p = 2$) and, when $p \rightarrow \infty$, the uniform distribution on the interval $(\mu - \sigma_p, \mu + \sigma_p)$ (e.g., see Box and Tiao 1992).

There are just some few Bayesian procedures for the analysis of EP regression models published to date. Moreover, there are no published reference priors for EP regression models. Existing literature has considered the use of EP errors in a number of contexts such as, for example, EP errors to robustify linear models (Box and Tiao 1992; Salazar et al. 2012), and mixtures of regression models with EP errors (Achcar and Pereira 1999). In addition, the EP distribution has been used as a prior for a Gaussian model location parameter (Choy and Smith 1997). To implement simulation-based computation for models with EP errors, one may use 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). As an alternative, Salazar et al. (2012) have developed fast analysis for EP regression models using Laplace approximations and Newton-Cotes integration. Here we use these latter fast computational methods.

The remainder of the paper is organized as follows. Section 2 presents the linear model with exponential power errors and the associated likelihood function. Section 3 derives the two reference priors and shows that both of these priors lead to proper posterior distributions. Section 4.1 presents a simulation study of the frequentist properties of the reference-priors-based Bayesian procedures and those of a competing noninformative prior. Section 4.2 presents applications of Bayesian reference analysis to two datasets. Section 5 concludes with a discussion of major findings and possible future research directions.

2 EP linear model

Let $y = (y_1, \dots, y_n)'$ be the vector of observations and $x = (x_1, \dots, x_n)'$ be the $n \times k$ design matrix of explanatory variables. We consider the linear model

$$y = x\beta + \epsilon, \quad (2)$$

where $\beta = (\beta_1, \dots, \beta_k)' \in \mathbb{R}^k$ is a vector of regression coefficients, and $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ is a vector of errors such that $\epsilon_1, \dots, \epsilon_n$ are independent and identically distributed and follow the exponential power distribution with location parameter equal to zero, scale parameter σ_p , and shape parameter p . We reparameterize the model by defining $\sigma = p^{1/p}\sigma_p\Gamma(1 + 1/p)$. This reparametrization has also been considered by Zhu and Zinde-Walsh (2009) and Salazar et al. (2012). Let us denote the parameter vector by $\theta = (\beta, \sigma, p) \in \mathbb{R}^k \times (0, \infty) \times (1, \infty)$. Then, the log-likelihood function for the model given in Equation (2) is

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

We use the log-likelihood function to develop reference priors for the EP regression model.

3 Methods

In this section, we obtain explicit expressions for reference priors for all the six possible orderings of the parameters of the EP linear model, and show that associated with these six parameters orderings there are only two reference priors. Finally, we show that both of these reference priors lead to proper posterior distributions.

Specifically, we consider here the Bernardo reference priors (Bernardo 1979) that take into account the Kulback-Leibler divergency between the prior distribution and the posterior distribution. In a nutshell, the reference priors proposed by Bernardo maximize the expected value of perfect information about the model parameters (p. 300, Bernardo and Smith 1994). When the parameter space is one-dimensional and asymptotic normality of the posterior distribution holds, the reference prior coincides with Jeffreys prior (Jeffreys 1961). However, when the parameter space is multidimensional Jeffreys prior is known to lead to Bayesian procedures that may have undesirable frequentist properties, such as for example frequentist coverage of credible intervals far away from the desired nominal level.

For the multidimensional parameter case when the parameters may be partitioned in a block of parameters of interest and another block of nuisance parameters, Bernardo

(1979) suggested an approach in three stages. The first stage obtains the conditional distribution of the nuisance parameter conditional on the parameter of interest. The second stage integrates out the nuisance parameter with respect to that conditional distribution to obtain a marginal likelihood. Finally, the third stage applies the reference prior approach to the marginal likelihood to obtain the reference prior for the parameter of interest. This idea can be naturally extended to partitions of the parameter vector with more than two components. The resulting reference prior will then depend on the ordering of the parameter vector components. This multiparameter case has been developed in a series of papers by Berger and Bernardo (1992a,b,c). Here we use the Berger-Bernardo approach to develop reference priors for the parameters of the EP regression model.

As we show below, the reference priors obtained here are of the form

$$\pi(\theta) \propto \frac{\pi(p)}{\sigma^a}, \tag{4}$$

where $a \in \mathbb{R}$ is a hyperparameter and $\pi(p)$ is the ‘marginal’ prior of the shape parameter p . As shown by Salazar et al. (2012), the Jeffreys-rule prior and two independence Jeffreys priors also have the functional form (4). Specifically, using the same notation as in Salazar et al. (2012), the two independence Jeffreys priors have $a = 1$ and their marginal priors for p are respectively given by

$$\pi^{I_1}(p) \propto p^{-1} [(1 + p^{-1}) \Psi'(1 + p^{-1}) - 1]^{1/2}, \tag{5}$$

and

$$\pi^{I_2}(p) \propto p^{-3/2} [(1 + p^{-1}) \Psi'(1 + p^{-1})]^{1/2}. \tag{6}$$

Meanwhile, the Jeffreys-rule prior is such that $a = k + 1$ and its marginal prior for p is

$$\pi^J(p) \propto [\Gamma(p^{-1}) \Gamma(2 - p^{-1})]^{k/2} \pi^{I_1}(p). \tag{7}$$

In what follows we find that the reference priors for the EP regression model are related to the independence Jeffreys priors given in Equations (5) and (6). When developing noninformative priors, it is crucial to study whether the resulting posterior distribution is proper. Salazar et al. (2012) have shown that the Independence Jeffreys prior $\pi^{I_2}(p)$ yields a proper posterior distribution. Unfortunately, both the independence Jeffreys prior $\pi^{I_1}(p)$ and the Jeffreys-rule prior $\pi^J(p)$ yield improper posterior distributions.

The Berger-Bernardo approach to develop reference priors requires the Fisher information matrix. Specifically, for the EP regression model the Fisher information matrix $H(\theta)$, 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 j th element of $\theta = (\beta, \sigma, p)$, is:

$$H(\theta) = \begin{bmatrix} \sigma^{-2} \Gamma(p^{-1}) \Gamma(2 - p^{-1}) \sum_{i=1}^n x_i x_i' & 0 & 0 \\ 0 & n p \sigma^{-2} & -n \sigma^{-1} p^{-1} \\ 0 & -n \sigma^{-1} p^{-1} & n p^{-3} (1 + p^{-1}) \Psi'(1 + p^{-1}) \end{bmatrix}, \tag{8}$$

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

The Fisher information matrix is block diagonal, with one block corresponding to β and another block corresponding to (σ, p) . One of the consequences of this structure is that reference priors that consider β , σ , and p as three separate groups will depend on the

ordering of the groups only with respect to whether σ or p appears first in the ordering. The following theorem provides reference priors for the parameters of the EP regression model.

Theorem 1. *Consider the EP regression model with log-likelihood function given in Equation (3). Then, there are two reference priors for all six possible orderings of the model parameters. Moreover, these two reference priors are of the form (4) with $a = 1$. For the orderings (β, σ, p) , (σ, β, p) , and (σ, p, β) the ‘marginal’ reference prior for p is*

$$\pi^{r_1}(p) \propto p^{-3/2} [(1 + p^{-1}) \Psi'(1 + p^{-1})]^{1/2}, \tag{9}$$

whereas for the orderings (β, p, σ) , (p, β, σ) , and (p, σ, β) the ‘marginal’ reference prior for p is

$$\pi^{r_2}(p) \propto p^{-3/2} [(1 + p^{-1}) \Psi'(1 + p^{-1}) - 1]^{1/2}. \tag{10}$$

Proof. See the Appendix. □

While reference prior π^{r_2} is a new prior that has not appeared before in the literature, there are similarities between the reference priors given in Theorem 1 and the independence Jeffreys priors given in Equations (5) and (6). Reference prior π^{r_1} coincides with the independence Jeffreys prior π^{I_2} given in Equation (6). Moreover, it is important to point out that reference prior π^{r_2} is somewhat similar to the independence Jeffreys prior π^{I_1} given in Equation (5), differing only by a factor of $p^{-1/2}$. However, as we show below this difference between π^{I_1} and π^{r_2} is enough to make π^{I_1} yield a useless improper posterior distribution while the reference prior π^{r_2} yields a useful proper posterior distribution.

Consider a prior of the form (4). Then the integrated likelihood for p is given by

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

Then the prior leads to a proper posterior distribution if and only if

$$\int_1^\infty L^I(p; y) \pi(p) dp < \infty, \tag{11}$$

Thus, in order to determine whether a prior of the form (4) leads to a proper posterior distribution, one needs to investigate the tail behavior of both the marginal prior and the integrated likelihood for p . The tail behavior of the marginal reference priors for p given in Theorem 1 is given in the following lemma.

Lemma 1. *The marginal priors for p given in Theorem 1 are continuous functions in $[1, \infty)$ and are such that $\pi^{r_1}(p) = O(p^{-3/2})$ and $\pi^{r_2}(p) = O(p^{-3/2})$ as $p \rightarrow \infty$.*

Proof. Direct inspection shows that $\pi^{r_1}(p)$ and $\pi^{r_2}(p)$ are continuous functions in $[1, \infty)$. Their tail behavior when $p \rightarrow \infty$ follows from the fact that $\Psi'(1 + p^{-1}) \rightarrow 1.6449$ and $\Gamma(p^{-1}) = O(p)$ as $p \rightarrow \infty$. □

Theorem 1 and Lemma 1 suggest the definition of an approximate reference prior inspired by priors π^{r_1} and π^{r_2} that has the same value for the hyperparameter $a = 1$ and share their tail behavior with respect to p . We define such an approximate reference prior in Definition 1.

Definition 1. We define an approximate reference prior π^{r_3} to be of the form (4) with $a = 1$ and marginal prior for p equal to $\pi^{r_3}(p) \propto p^{-3/2}$.

Computation of prior π^{r_3} is faster and more straightforward than that of priors π^{r_1} and π^{r_2} . In addition, Section 4.1 shows that the frequentist properties of procedures based on π^{r_3} are similar to those based on π^{r_1} and π^{r_2} . As a consequence, the approximate reference prior π^{r_3} may become more widely used than the reference priors π^{r_1} and π^{r_2} . Therefore, henceforth we drop the term “approximate” and simply refer to π^{r_3} as a reference prior.

The following lemma, that was proved by Salazar et al. (2012), provides the tail behavior for the integrated likelihood for p .

Lemma 2 (Salazar et al.2012). *Provided that $n > k + 1 - a$, the integrated likelihood for p under the class of priors (4) is a continuous function in $[1, \infty)$ and is such that $L^I(p; y) = O(1)$ as $p \rightarrow \infty$.*

The following proposition establishes that the two reference priors that we have obtained yield proper posterior distributions.

Proposition 1. *Provided that $n > k + 1 - a$, the two reference priors π^{r_1} and π^{r_2} given in Theorem 1 yield proper posterior distributions.*

Proof. This proposition follows directly from condition (11), and Lemmas 1 and 2. \square

To implement posterior analysis for the parameters of the EP regression model based on the reference priors developed here, we use an approach proposed by Salazar et al. (2012) that combines Laplace approximations and Newton-Cotes integration.

4 Results and discussion

4.1 Frequentist properties

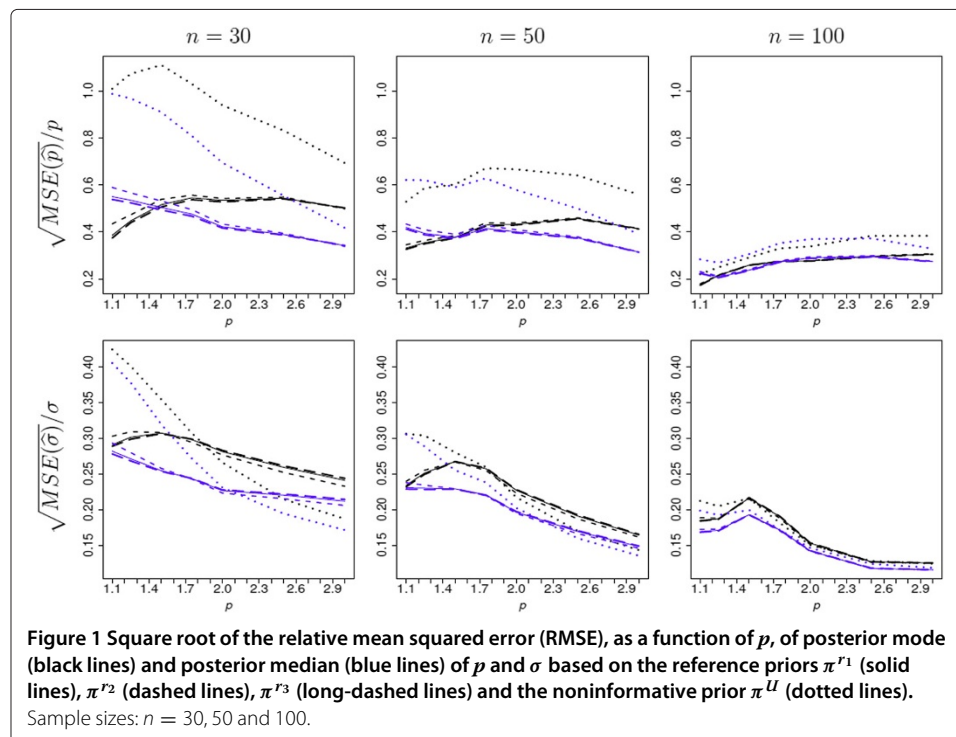
In this section we perform a simulation study to assess the frequentist properties of Bayesian procedures based on the reference priors π^{r_1} , π^{r_2} , and π^{r_3} . In addition, we compare the performance of these reference priors to that of a competing noninformative prior π^U that takes the form (4) with $a = 1$ and $\pi^U(p) \propto 1$ for $1 < p < 10$ and $\pi^U(p) = 0$ otherwise. The joint prior $\pi^U(\theta)$ leads to a proper posterior distribution, however as we see below the uniform prior $\pi^U(p)$ is a naïve way to express lack of information about p . The Bayesian procedures we consider are the posterior modes and posterior medians for point estimation, and the 95% highest posterior density (HPD) credible intervals for interval estimation. Finally, we consider three frequentist measures of quality. For evaluating the quality of point estimation, we consider the square root of the frequentist relative mean squared error. For evaluating the performance of interval estimation, we consider two frequentist measures: the frequentist coverage and the mean length of the credible intervals.

We have considered several combinations of sample sizes and parameters. Specifically, we have considered three sample sizes: $n = 30$, $n = 50$ and $n = 100$. Moreover, we have considered a grid of values for p on the interval from 1 to 3. Further, for each simulated dataset we have used $k = 2$, $x_i = (1, x_{1i})$, $x_{1i} \sim N(2, 1)$, $\beta = (1.5, -3)$, and $\sigma = 1$.

Finally, for each combination of parameter values and sample sizes, we have simulated 1,500 datasets to estimate the frequentist properties of the several procedures.

The square root of the relative mean squared error (RMSE), $\sqrt{MSE(\hat{\theta})}/\theta$, for estimators of p and σ is shown as a function of p in Figure 1. As intuitively expected, for all priors and for both posterior mode and median, as the sample size increases the RMSE decreases. The most substantial differences are between the performances of the posterior mode and posterior median, and between the performances of the reference priors when compared with the π^U prior. First, we compare the performance of the posterior median and the posterior mode. For each prior, for the estimation of p , the posterior median provides smaller RMSE than the posterior mode for most values of p considered except for p close to one. And this advantage of the posterior mode becomes less pronounced as the sample size increases. For each prior, for the estimation of σ , the posterior median provides smaller RMSE than the posterior mode. Therefore, for the reference analysis of the EP regression model we recommend the use of the posterior median.

Second, we compare the RMSE performance of the different priors. For each type of point estimator considered here, in terms of RMSE the reference priors π^{r1} , π^{r2} , and π^{r3} provide qualitatively similar results, with π^{r1} and π^{r3} being slightly better for smaller values of p and π^{r2} being slightly better for larger values of p . In addition, the difference in performance of the three reference priors becomes smaller as the sample size increases. In contrast, the performance of the reference priors differs dramatically from that of the π^U prior. For each class of estimators of p and for all values of p considered, when compared to the π^U prior the reference priors lead to smaller RMSE. For the estimation of σ , the results are mixed; for small sample sizes while the reference priors lead to smaller RMSE when p is small and π^U leads to better results when p is larger. But for larger sample

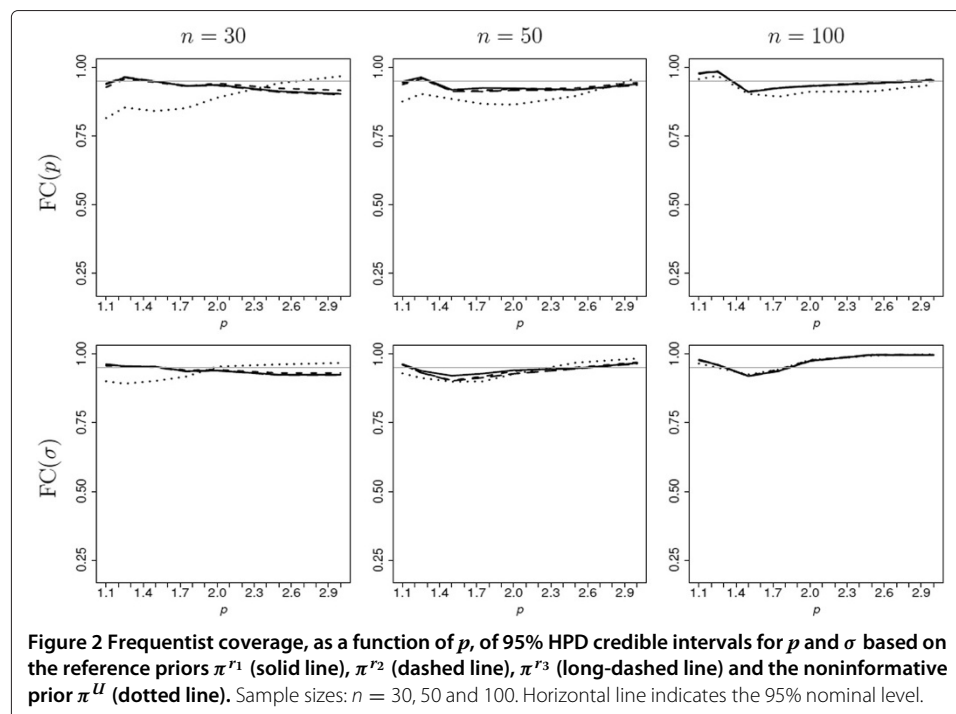


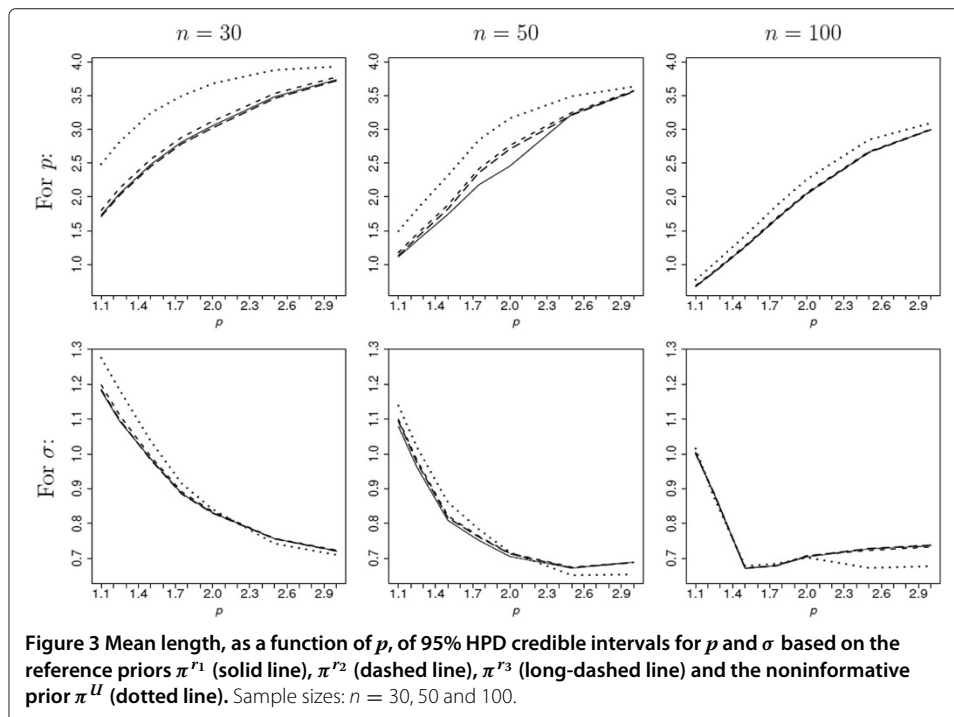
sizes the reference priors-based posterior medians have smaller RMSE for all considered values of p .

The frequentist coverage (FC) of 95% HPD credible intervals for p and σ is shown, as a function of p , in Figure 2. As the sample size increases, the FC of the credible intervals based on the four priors becomes more similar. For both parameters, the π^{r1} -, π^{r2} -, and π^{r3} -based credible intervals have frequentist coverage closer to the nominal level. This superiority of the Bayesian reference analysis is particularly pronounced for sample sizes equal to 30 or 50 and when $p < 2$.

The mean length of the 95% HDP credible intervals for p and σ is shown, as a function of p , in Figure 3. For the credible intervals based on the three reference priors, the mean lengths of the credible intervals are similar with slightly better results for π^{r1} . For interval estimation for σ , the mean lengths of the credible intervals based on the three reference priors are smaller than the mean lengths of the credible intervals based on the π^U when $p < 2$ and are larger when $p > 2$. For interval estimation of p , in the range of values that we consider the π^{r1} -, π^{r2} -, and π^{r3} -based credible intervals are on average shorter than those based on π^U . Therefore, for the interval estimation of p , in the range of values we consider, the credible intervals based on π^{r1} , π^{r2} and π^{r3} provide uniformly superior results.

In summary, the reference priors π^{r1} , π^{r2} , and π^{r3} lead to procedures that have similar frequentist properties. In addition, when compared to the competing noninformative prior π^U , the reference priors π^{r1} , π^{r2} , and π^{r3} lead to overall superior results. Finally, the reference prior π^{r3} has a simpler functional form and is more straightforward to be implemented. Therefore, in cases when there is no prior information for the analysis of EP linear regression models, we recommend the use of the reference prior π^{r3} .





4.2 Applications

This section illustrates the use of the Bayesian reference analysis we propose for exponential power regression models with applications to two real world datasets. The first dataset illustrates leptokurtic errors and the second dataset illustrates platykurtic errors. Because the results based on the reference priors π^{r1} and π^{r3} are extremely similar, we show only the results for priors π^{r1} , π^{r2} , and π^U .

In both applications, we use the same truncation point at $p = 10$ used for $\pi^U(p)$ in Section 4.1 and assume $\pi^U(p) \propto 1$ for $1 < p < 10$ and $\pi^U(p) = 0$ otherwise. We have chosen the truncation point at $p = 10$ because datasets generated with $p = 10$ or with p close to 10 have similar statistical behavior. Hence, to distinguish whether a process follows an EP distribution with $p = 10$ or, say, $p = 10.1$ we would need an extremely large data set. Moreover, the choice of truncation should be made before the analyst looks at the data. For example, for the first application below, after looking at the scatterplot one may think about truncating the prior for values of p that correspond to leptokurtic distributions, that is, $1 < p < 2$. However, doing that would mean to use the data twice in the Bayes Theorem formula: once through the prior, and another time through the likelihood. Usually, such double use of the data leads to underestimation of the uncertainty. Therefore, we prefer to decide the truncation of the prior before looking at the data.

4.2.1 School spending

We analyze the relationship between per capita spending in public schools and per capita income by state in the United States. This dataset has been previously analyzed by Greene (1997), Cribari-Neto et al. (2000), and Fonseca et al. (2008). Specifically, Greene (1997) and Cribari-Neto et al. (2000) proposed analyses based on heuristic approaches to the so-called problem of heterocedasticity-of-unknown-form. In contrast, Fonseca et al. (2008)

have analyzed this dataset in the context of linear regression models with Student-t errors. Fonseca et al. (2008) found that when errors with distributions with heavy tails are assumed, a linear model is superior to a quadratic model. Here, we take a similar approach as that of Fonseca et al. (2008) in that we assume a linear model with errors that may have a heavy tail distribution. However, we assume that the errors follow an exponential power distribution.

Table 1 presents the posterior summaries based on the priors π^U , π^{r_1} , π^{r_2} . For all three priors both the posterior mode and the posterior median for p are smaller than one. In addition, both π^{r_1} - and π^{r_2} -based 95% credible intervals for p are contained in the interval (1, 2) indicating evidence that the errors are leptokurtic. In contrast, the π^U -based 95% credible interval for p is not fully contained in the interval (1, 2). However, from the results in Section 4.1 we know that for small true values of p , the use of the π^U prior leads to on average wider credible intervals for p that have lower coverage than nominal. Thus, this application provides an example when the superiority of the π^{r_1} and π^{r_2} priors matters to the conclusion that in this data set the errors distribution is leptokurtic.

Figure 4(a) shows the scatterplot for the school spending data set along with the fitted EP regression model based on π^{r_1} (solid line), π^{r_2} (dashed line), and π^U (dotted line). Figure 4(a) also shows the fitted Gaussian linear model (dot-dashed line). While the Gaussian model fit is clearly and strongly influenced by the outlier, the use of exponential power errors (with the four priors considered here) automatically makes the analysis robust against outliers. In particular, the model fits using the π^{r_1} and π^{r_2} priors (considering the posterior median) coincide and are equal to $\hat{y} = -88.38 + 600.9x$. Another way to make the analysis robust against outliers is to use Student-t errors. Assuming Student-t errors, a model fitted by Fonseca et al. (2008) was $y = -75.3 + 583.2x$. We can see that both Student-t and exponential power errors fits are robust against outliers. However, the Student-t distribution cannot accommodate platykurtic errors and, therefore, the exponential power distribution provides more flexibility.

Figure 4(b) presents the marginal posterior densities for p based on π^{r_1} (solid line), π^{r_2} (dashed line), π^{r_3} (long-dashed line) and π^U (dotted line). In addition, the vertical lines indicate the limits of the 95% HPD credible intervals. The three reference priors lead to similar posterior densities for p , while the π^U prior leads to a substantially different posterior density for p . Figure 4(b) illustrates why the π^U leads to unnecessarily wider credible intervals. That combined with π^U -based credible intervals having coverage lower than nominal leads us to prefer the data analysis based on the reference priors.

4.2.2 Sold home videos vs. profits at the box office

We analyze a dataset about 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). This dataset has been previously analyzed by Levine et al. (2006) and Salazar et al. (2012) and comprises observations on 30 movies. A scatterplot of the variables of interest is shown in Figure 5(a). Using a linear model with EP errors and the independence Jeffreys prior π^{r_2} given in Equation (6), Salazar et al. (2012) found evidence of a platykurtic distribution for the errors. Here we compare three analyses of this home videos dataset with an EP linear regression model obtained by applying the reference priors π^{r_1} and π^{r_2} , and the noninformative prior π^U .

Table 1 School spending data set: Posterior summaries based on the noninformative prior π^U and the reference priors π^{r1} and π^{r2}

	π^U			π^{r1}			π^{r2}		
	Mode	Median	95% C.I.	Mode	Median	95% C.I.	Mode	Median	95% C.I.
ρ	1.18	1.33	(1.02, 2.03)	1.06	1.26	(1.00, 1.91)	1.08	1.27	(1.00, 1.92)
σ	52.73	54.75	(38.59, 74.44)	51.21	53.23	(38.08, 73.43)	51.72	53.23	(38.08, 73.43)
β_1	-89.37	-92.51	(-131.79, -37.85)	-91.51	-88.38	(-131.81, -36.86)	-91.51	-88.38	(-131.81, -36.86)
β_2	616.07	603.95	(525.16, 667.59)	609.99	600.90	(525.15, 667.57)	609.99	600.90	(525.15, 667.57)

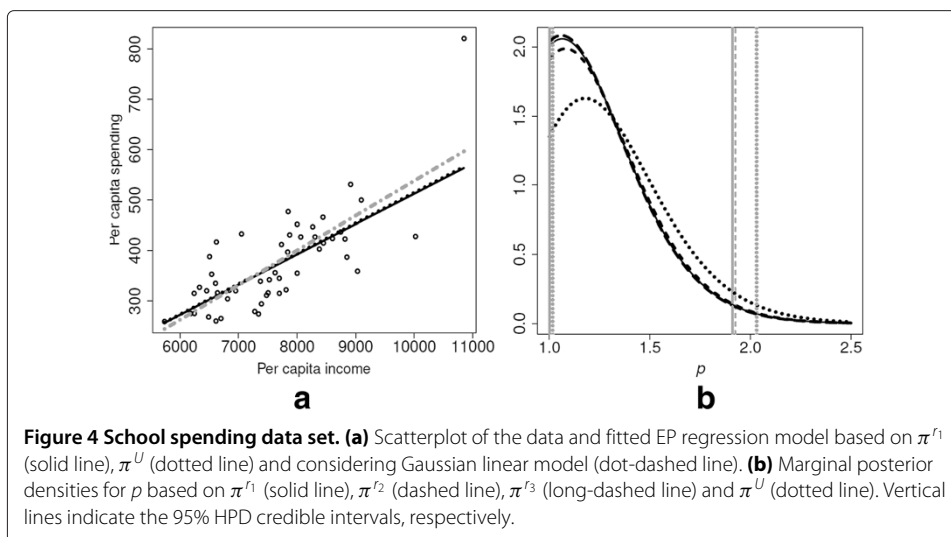


Figure 5(a) shows the model fit for each of the priors we consider. The fits based on the reference priors visually coincide, whereas the fit based on π^U is slightly different. This is confirmed by Table 2, that shows that the slopes for the three fits are similar and around 4.33, whereas the intercept for the π^U -based fit is about 4.5% larger than the intercept for the π^{r1} - and π^{r2} -based fits. Even more striking are the differences between the reference analyses and the π^U -based analysis for σ and p . For σ , both posterior medians based on π^{r1} and π^{r2} are very similar and equal to 67.37 and 68.38 respectively, while the posterior median based on π^U is 77.50. Moreover, the 95% credible intervals for σ based on π^{r1} and π^{r2} are very similar and equal to (38.08, 93.64) and (38.10, 93.64) respectively, while the interval based on π^U is substantially different and equal to (47.17, 98.69).

The reference analyses for p are also strikingly distinct from the π^U -based analysis for p . First, the posterior medians for p based on π^{r1} and π^{r2} coincide and are equal to 2.64 while the π^U -based posterior median differs tremendously and is equal to 4.36. Second, the 95% credible intervals for p based on π^{r1} and π^{r2} are similar and equal to (1.00, 7.01) and (1.00, 7.18) respectively, while the π^U -based interval for p differs tremendously from

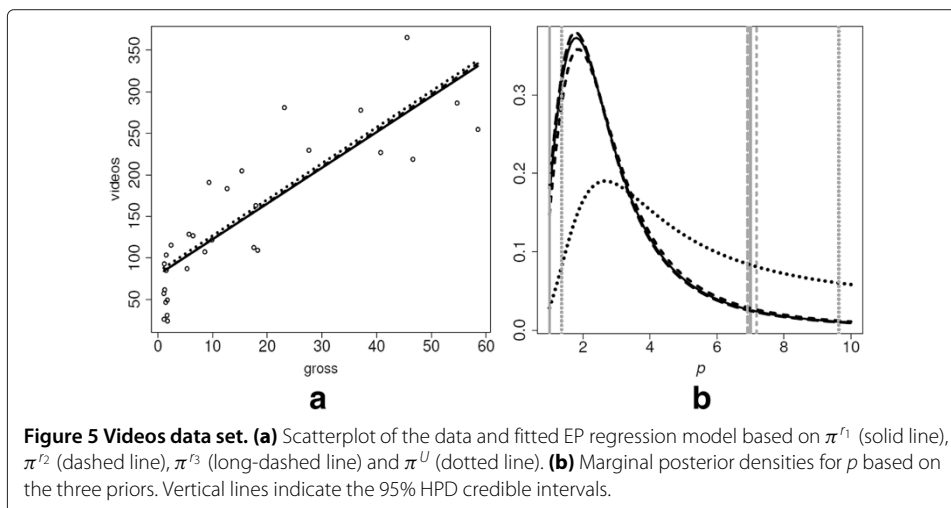


Table 2 Videos data set: Posterior summaries based on the noninformative prior π^U and the reference priors π^{r_1} and π^{r_2}

	π^U			π^{r_1}			π^{r_2}		
	Mode	Median	95% C.I.	Mode	Median	95% C.I.	Mode	Median	95% C.I.
ρ	2.64	4.36	(1.36, 9.64)	1.82	2.64	(1.00, 7.01)	1.83	2.64	(1.00, 7.18)
σ	80.51	77.50	(47.17, 98.69)	67.37	67.37	(38.08, 93.64)	68.38	68.38	(38.10, 93.64)
β_1	83.11	83.11	(54.92, 107.65)	79.39	79.40	(53.03, 105.76)	79.53	79.53	(53.16, 104.98)
β_2	4.31	4.35	(3.42, 5.24)	4.32	4.32	(3.31, 5.33)	4.33	4.33	(3.32, 5.34)

the reference CIs and is equal to (1.36, 9.64). Hence, the π^U -based CI is more than 30% wider than the reference CIs. This undesirable feature of π^U -based CIs coincides with the results from the simulation study presented in Section 4.1.

Finally, Figure 5(b) presents the marginal posterior densities for p based on π^{r1} , π^{r2} , and π^U . This figure sheds light on the reason for the striking difference between the π^{r1} - and π^{r2} -based CIs and the π^U -based CI. The problem with the π^U -based analysis is that the right tail of the marginal posterior density for p decays too slowly. As a result, for the home video dataset the π^U -based CI depends dramatically on the right side truncation of the prior, which in this manuscript has been fixed at 10. Figure 5(b) makes it really clear that a larger truncation point would have a huge impact in the resulting π^U -based CI for p . This dataset clearly illustrates the superiority of the Bayesian reference analyses.

5 Conclusions

We have developed Bayesian reference analysis for linear models with exponential power errors. Specifically, we have developed three reference priors that lead to useful proper posterior distributions. In addition, we have shown through a simulation study that both priors yield procedures that have better frequentist properties than procedures resulting from a competing noninformative prior. Finally, we have illustrated our Bayesian reference analysis methodology with two real world applications that highlight the flexibility of the exponential power distribution to accommodate both cases when there are outliers in the dataset and also cases when the errors follow a platykurtic distribution.

The fact that the reference priors we have obtained for the EP regression model lead to proper posterior distributions is of substantial theoretical interest. The propriety of these reference posterior distributions contrasts with the impropriety of the posterior distribution associated with the Jeffreys-rule prior found by Salazar et al. (2012). Moreover, Salazar et al. (2012) found two independence Jeffreys priors, one of which leads to an improper posterior distribution whereas the other leads to a proper posterior distribution. We have found that the independence Jeffreys prior that yields a proper posterior distribution coincides with our reference prior π^{r1} . Further, the independence Jeffreys prior that yields a useless improper posterior distribution differs only by a factor of $p^{-1/2}$ from the reference prior π^{r2} . However, this difference is enough to make our reference prior π^{r2} yield a useful proper posterior distribution.

Our results motivate many possible directions for future research. First, an open question is whether there exist general conditions under which reference priors yield proper posterior distributions. In addition, the existence of general conditions for posterior propriety may be investigated for Jeffreys-rule and independence Jeffreys priors. The search of general conditions for posterior propriety may benefit from our present work on EP regression and previous literature on examples of impropriety of posterior distributions for distinct objective Bayes priors (Berger et al. 2001; Ferreira and De Oliveira 2007; Salazar et al. 2012; Wasserman 2000).

We have considered the frequentist properties of the proposed Bayesian approaches via a simulation study. In particular, we have shown that credible intervals based on π^{r1} , π^{r2} , and π^{r3} have similar frequentist properties with coverage close to nominal for p and σ . This is a reflection of the fact that for any prior satisfying some regularity conditions the frequentist coverage of credible intervals and the nominal level agree up to $O(n^{-1/2})$ (for a discussion and conditions, see Ghosh et al. 2006). A prior that leads to a more

stringent agreement of order $O(n^{-1})$ is called a first-order probability matching prior. Such priors have to be derived with a specific parameter of interest in mind, and their derivation is far from trivial. Therefore, promising directions for future research for the EP regression model would be the derivation of priors that lead to Bayesian predictions that have approximate frequentist validity (Datta et al. 2000b) and the derivation of first-order probability matching priors (Datta and Ghosh 1995; Datta et al. 2000a).

Appendix

Proof of Theorem 1. To prove Theorem 1, we follow the methodology to obtain reference priors proposed by Berger and Bernardo (1992a). In particular, we assume that the reader is familiar with both the notation and the methodology of Berger and Bernardo (1992a). This proof is divided in two parts. In the first part, we obtain the reference prior for the orderings (β, σ, p) , (σ, β, p) , and (σ, p, β) . Because the proofs are analogous for each of these three orderings, in the first part we obtain the reference prior for the ordering (σ, β, p) . In the second part, we obtain the reference prior for the orderings (β, p, σ) , (p, β, σ) , and (p, σ, β) . Because the proofs are analogous for each of these three orderings, in the second part we obtain the reference prior for the ordering (p, β, σ) .

Part 1. Consider the ordering $\theta = (\sigma, \beta, p)$.

After rearranging the Fisher information matrix $H(\theta)$ given in Equation (8) to conform to this ordering, the inverse of the Fisher information matrix becomes

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

Thus,

$$S_1 = \frac{\sigma^2}{np} \frac{(1+p^{-1})\Psi'(1+p^{-1})}{(1+p^{-1})\Psi'(1+p^{-1})-1},$$

$$S_2 = \begin{bmatrix} \frac{\sigma^2}{np} \frac{(1+p^{-1})\Psi'(1+p^{-1})}{(1+p^{-1})\Psi'(1+p^{-1})-1} & 0 \\ 0 & \sigma^2 \left\{ \Gamma(p^{-1}) \Gamma(2-p^{-1}) \sum_{i=1}^n x_i x'_i \right\}^{-1} \end{bmatrix},$$

and $S_3 = S(\theta)$. Moreover, let $H_j = S_j^{-1}$. Thus,

$$H_1 = \frac{np}{\sigma^2} \frac{(1+p^{-1})\Psi'(1+p^{-1})-1}{(1+p^{-1})\Psi'(1+p^{-1})},$$

$$H_2 = \begin{bmatrix} \frac{np}{\sigma^2} \frac{(1+p^{-1})\Psi'(1+p^{-1})-1}{(1+p^{-1})\Psi'(1+p^{-1})} & 0 \\ 0 & \sigma^{-2} \Gamma(p^{-1}) \Gamma(2-p^{-1}) \sum_{i=1}^n x_i x'_i \end{bmatrix},$$

and $H_3 = H(\theta)$.

Let h_j be the $n_j \times n_j$ lower right corner of H_j . Thus,

$$h_1 = \frac{np}{\sigma^2} \frac{(1+p^{-1})\Psi'(1+p^{-1})-1}{(1+p^{-1})\Psi'(1+p^{-1})},$$

$$h_2 = \sigma^{-2} \Gamma(p^{-1}) \Gamma(2-p^{-1}) \sum_{i=1}^n x_i x'_i, \quad \text{and}$$

$$h_3 = np^{-3} (1+p^{-1}) \Psi'(1+p^{-1}).$$

Let $\theta_{(1)} = \sigma$, $\theta_{(2)} = \beta$, and $\theta_{(3)} = p$. In addition, let $\theta_{[1]} = \theta_{(1)} = \sigma$, $\theta_{[2]} = (\theta_{(1)}, \theta_{(2)}) = (\sigma, \beta)$, and $\theta_{[3]} = (\theta_{(1)}, \theta_{(2)}, \theta_{(3)}) = (\sigma, \beta, p)$. Moreover, let $\theta_{[\sim 1]} = (\theta_{(2)}, \theta_{(3)}) = (\beta, p)$ and $\theta_{[\sim 2]} = (\theta_{(3)}) = p$. Further, consider the following compact sets: for σ , $\Theta_{(1)}^l = [l^{-1}, l]$; for β , $\Theta_{(2)}^l = [-l, l]^k$; for p , $\Theta_{(3)}^l = [1, l]$.

Then,

$$\begin{aligned} \pi_3^l(p \mid \sigma, \beta) &= \pi_3^l(\theta_{[\sim 2]} \mid \theta_{[2]}) \\ &= \frac{|h_3(\theta)|^{1/2} \mathbf{1}_{\Theta_{(3)}^l}(\theta_{(3)})}{\int_{\Theta_{(3)}^l} |h_3(\theta)|^{1/2} d\theta_{(3)}} \\ &= \frac{\{np^{-3}(1+p^{-1})\Psi'(1+p^{-1})\}^{1/2} \mathbf{1}_{[1,l]}(p)}{\int_1^l \{np^{-3}(1+p^{-1})\Psi'(1+p^{-1})\}^{1/2} dp} \\ &= \{c_1(l)\}^{-1} p^{-3/2} (1+p^{-1})^{1/2} \{\Psi'(1+p^{-1})\}^{1/2} \mathbf{1}_{[1,l]}(p), \end{aligned}$$

where $c_1(l) = \int_1^l p^{-3/2} (1+p^{-1})^{1/2} \{\Psi'(1+p^{-1})\}^{1/2} dp$.

Now,

$$\begin{aligned} \pi_2^l(\beta, p \mid \sigma) &= \pi_2^l(\theta_{[\sim 1]} \mid \theta_{[1]}) \\ &= \frac{\pi_3^l(\theta_{[\sim 2]} \mid \theta_{[2]}) \exp\{0.5E_2^l[\log|h_2(\theta)||\theta_{[2]}]\} \mathbf{1}_{\Theta_{(2)}^l}(\theta_{(2)})}{\int_{\Theta_{(2)}^l} \exp\{0.5E_2^l[\log|h_2(\theta)||\theta_{[2]}]\} d\theta_{(2)}}, \end{aligned}$$

where

$$\begin{aligned} E_2^l[\log|h_2(\theta)||\theta_{[2]}] &= \int_{\Theta_{(3)}^l} \log|h_2(\theta)| \pi_3^l(\theta_{[\sim 2]} \mid \theta_{[2]}) d\theta_{[\sim 2]} \\ &= \int_1^l \left\{ -2k \log \sigma + k \log \Gamma(p^{-1}) + k \log \Gamma(2-p^{-1}) + \log \left| \sum_{i=1}^n x_i x_i' \right| \right\} \\ &\quad \{c_1(l)\}^{-1} p^{-3/2} (1+p^{-1})^{1/2} \{\Psi'(1+p^{-1})\}^{1/2} dp \\ &= -2k \log \sigma + c_2(l), \end{aligned}$$

with

$$\begin{aligned} c_2(l) &= \{c_1(l)\}^{-1} \int_1^l \left\{ k \log \Gamma(p^{-1}) + k \log \Gamma(2-p^{-1}) + \log \left| \sum_{i=1}^n x_i x_i' \right| \right\} \\ &\quad p^{-3/2} (1+p^{-1})^{1/2} \{\Psi'(1+p^{-1})\}^{1/2} dp. \end{aligned}$$

Hence,

$$\begin{aligned} \pi_2^l(\beta, p \mid \sigma) &= \frac{\pi_3^l(p \mid \sigma, \beta) \exp\{0.5[-2k \log \sigma + c_2(l)]\} \mathbf{1}_{[-l,l]^k}(\beta)}{\int_{[-l,l]^k} \exp\{0.5[-2k \log \sigma + c_2(l)]\} d\beta} \\ &= \pi_3^l(p \mid \sigma, \beta) (2l)^{-k} \mathbf{1}_{[-l,l]^k}(\beta). \end{aligned}$$

Finally,

$$\begin{aligned} \pi_1^l(\sigma, \beta, p) &= \pi_1^l(\theta_{[\sim 0]} \mid \theta_{[0]}) \\ &= \frac{\pi_2^l(\theta_{[\sim 1]} \mid \theta_{[1]}) \exp\{0.5E_1^l[\log|h_1(\theta)||\theta_{[1]}]\} \mathbf{1}_{\Theta_{(1)}^l}(\theta_{(1)})}{\int_{\Theta_{(1)}^l} \exp\{0.5E_1^l[\log|h_1(\theta)||\theta_{[1]}]\} d\theta_{(1)}}, \end{aligned}$$

with

$$E_1^l [\log |h_1(\theta)| | \theta_{[1]}] = \int_{[-l, l]^k} \int_1^l \log \left\{ \frac{np (1+p^{-1}) \Psi'(1+p^{-1}) - 1}{\sigma^2 (1+p^{-1}) \Psi'(1+p^{-1})} \right\} \pi_2^l(\beta, p | \sigma) dp d\beta$$

$$= -2 \log \sigma + c_3(l),$$

where

$$c_3(l) = \int_{[-l, l]^k} \int_1^l \log \left\{ np \frac{(1+p^{-1}) \Psi'(1+p^{-1}) - 1}{(1+p^{-1}) \Psi'(1+p^{-1})} \right\} \pi_2^l(\beta, p | \sigma) dp d\beta$$

does not depend on $\theta = (\sigma, \beta, p)$.

Hence,

$$\pi_1^l(\sigma, \beta, p) = \frac{\pi_2^l(\beta, p | \sigma) \exp\{0.5[2 \log \sigma + c_3(l)]\} \mathbf{1}_{(l^{-1}, l)}(\sigma)}{\int_{l^{-1}}^l \exp\{0.5[2 \log \sigma + c_3(l)]\} d\sigma}$$

$$= \frac{\pi_2^l(\beta, p | \sigma) \sigma^{-1} \mathbf{1}_{(l^{-1}, l)}(\sigma)}{2 \log l}$$

Thus,

$$\pi_1^l(\sigma, \beta, p) = \sigma^{-1} p^{-3/2} (1+p^{-1})^{1/2} \{\Psi'(1+p^{-1})\}^{1/2}$$

$$\times \{c_1(l)\}^{-1} (2l)^{-k} (2 \log l)^{-1} \mathbf{1}_{[l^{-1}, l]}(\sigma) \mathbf{1}_{[-l, l]^k}(\beta) \mathbf{1}_{[1, l]}(p).$$

Now take any point $\theta^* = (\sigma^*, \beta^*, p^*) \in [l^{-1}, l] \times [-l, l]^k \times [1, l]$. Then, the reference prior for the ordering (σ, β, p) is

$$\pi(\sigma, \beta, p) \propto \lim_{l \rightarrow \infty} \frac{\pi_1^l(\sigma, \beta, p)}{\pi_1^l(\sigma^*, \beta^*, p^*)}$$

$$= \sigma^{-1} p^{-3/2} (1+p^{-1})^{1/2} \{\Psi'(1+p^{-1})\}^{1/2},$$

which is of the form (4).

Part 2. Consider the ordering $\theta = (p, \beta, \sigma)$.

After rearranging the Fisher information matrix $H(\theta)$ given in Equation (8) to conform to this ordering, the inverse of the Fisher information matrix becomes

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

Thus,

$$S_1 = \frac{p^3}{n} \frac{1}{(1+p^{-1}) \Psi'(1+p^{-1}) - 1},$$

$$S_2 = \begin{bmatrix} \frac{p^3}{n} \frac{1}{(1+p^{-1})\Psi'(1+p^{-1})-1} & 0 \\ 0 & \sigma^2 \{\Gamma(p^{-1}) \Gamma(2-p^{-1}) \sum_{i=1}^n x_i x_i'\}^{-1} \end{bmatrix},$$

and $S_3 = S(\theta)$. Moreover, let $H_j = S_j^{-1}$. Thus,

$$H_1 = np^{-3} \{(1+p^{-1}) \Psi'(1+p^{-1}) - 1\},$$

$$H_2 = \begin{bmatrix} np^{-3} \{(1+p^{-1}) \Psi'(1+p^{-1}) - 1\} & 0 \\ 0 & \sigma^{-2} \Gamma(p^{-1}) \Gamma(2-p^{-1}) \sum_{i=1}^n x_i x_i' \end{bmatrix},$$

and $H_3 = H(\theta)$.

Let h_j be the $n_j \times n_j$ lower right corner of H_j . Thus,

$$h_1 = np^{-3}\{(1 + p^{-1})\Psi'(1 + p^{-1}) - 1\}, h_2 = \sigma^{-2}\Gamma(p^{-1})\Gamma(2 - p^{-1})\sum_{i=1}^n x_i x'_i, \text{ and } h_3 = np\sigma^{-2}.$$

Let $\theta_{(1)} = p$, $\theta_{(2)} = \beta$, and $\theta_{(3)} = \sigma$. In addition, let $\theta_{[1]} = \theta_{(1)} = p$, $\theta_{[2]} = (\theta_{(1)}, \theta_{(2)}) = (p, \beta)$, and $\theta_{[3]} = (\theta_{(1)}, \theta_{(2)}, \theta_{(3)}) = (p, \beta, \sigma)$. Moreover, let $\theta_{[\sim 1]} = (\theta_{(2)}, \theta_{(3)}) = (\beta, \sigma)$ and $\theta_{[\sim 2]} = (\theta_{(3)}) = \sigma$. Further, consider the following compact sets: for p , $\Theta^l_{(1)} = [1, l]$; for β , $\Theta^l_{(2)} = [-l, l]^k$; for σ , $\Theta^l_{(3)} = [l^{-1}, l]$.

Then,

$$\begin{aligned} \pi_3^l(\sigma | p, \beta) &= \pi_3^l(\theta_{[\sim 2]} | \theta_{[2]}) \\ &= \frac{|h_3(\theta)|^{1/2} \mathbf{1}_{\Theta^l_{(3)}}(\theta_{(3)})}{\int_{\Theta^l_{(3)}} |h_3(\theta)|^{1/2} d\theta_{(3)}} \\ &= \frac{\{np\sigma^{-2}\}^{1/2} \mathbf{1}_{[l^{-1}, l]}(\sigma)}{\int_{l^{-1}}^l \{np\sigma^{-2}\}^{1/2} d\sigma} \\ &= \sigma^{-1} (2 \log l)^{-1} \mathbf{1}_{[l^{-1}, l]}(\sigma). \end{aligned}$$

Moreover,

$$\begin{aligned} \pi_2^l(\beta, p | \sigma) &= \pi_2^l(\theta_{[\sim 1]} | \theta_{[1]}) \\ &= \frac{\pi_3^l(\theta_{[\sim 2]} | \theta_{[2]}) \exp\{0.5E_2^l[\log |h_2(\theta)| | \theta_{[2]}]\} \mathbf{1}_{\Theta^l_{(2)}}(\theta_{(2)})}{\int_{\Theta^l_{(2)}} \exp\{0.5E_2^l[\log |h_2(\theta)| | \theta_{[2]}]\} d\theta_{(2)}}, \end{aligned}$$

where

$$\begin{aligned} E_2^l[\log |h_2(\theta)| | \theta_{[2]}] &= \int_{\Theta^l_{(3)}} \log |h_2(\theta)| \pi_3^l(\theta_{[\sim 2]} | \theta_{[2]}) d\theta_{[\sim 2]} \\ &= \int_{l^{-1}}^l \left\{ -2k \log \sigma + k \log \Gamma(p^{-1}) + k \log \Gamma(2 - p^{-1}) + \log \left| \sum_{i=1}^n x_i x'_i \right| \right\} \\ &\quad \sigma^{-1} (2 \log l)^{-1} d\sigma \\ &= c_1(l, p), \text{ which does not depend on } \beta. \end{aligned}$$

Hence,

$$\begin{aligned} \pi_2^l(\beta, \sigma | p) &= \frac{\pi_3^l(\sigma | p, \beta) \exp\{0.5c_1(l, p)\} \mathbf{1}_{[-l, l]^k}(\beta)}{\int_{[-l, l]^k} \exp\{0.5c_1(l, p)\} d\beta} \\ &= \pi_3^l(\sigma | p, \beta) (2l)^{-k} \mathbf{1}_{[-l, l]^k}(\beta). \end{aligned}$$

Further,

$$\begin{aligned} \pi_1^l(p, \beta, \sigma) &= \pi_1^l(\theta_{[\sim 0]} | \theta_{[0]}) \\ &= \frac{\pi_2^l(\theta_{[\sim 1]} | \theta_{[1]}) \exp\{0.5E_1^l[\log |h_1(\theta)| | \theta_{[1]}]\} \mathbf{1}_{\Theta^l_{(1)}}(\theta_{(1)})}{\int_{\Theta^l_{(1)}} \exp\{0.5E_1^l[\log |h_1(\theta)| | \theta_{[1]}]\} d\theta_{(1)}}, \end{aligned}$$

with

$$\begin{aligned} E_1^l[\log |h_1(\theta)| | \theta_{[1]}] &= \int_{[-l, l]^k} \int_{l^{-1}}^l \log [np^{-3}\{(1 + p^{-1})\Psi'(1 + p^{-1}) - 1\}] \pi_2^l(\beta, \sigma | p) d\sigma d\beta \\ &= \log [np^{-3}\{(1 + p^{-1})\Psi'(1 + p^{-1}) - 1\}]. \end{aligned}$$

Hence,

$$\begin{aligned}\pi_1^l(\sigma, \beta, p) &= \frac{\pi_2^l(\beta, \sigma | p) \exp\{0.5 \log [np^{-3}\{(1+p^{-1})\Psi'(1+p^{-1})-1\}]\} \mathbf{1}_{(1,l)}(p)}{\int_1^l \exp\{0.5 \log [np^{-3}\{(1+p^{-1})\Psi'(1+p^{-1})-1\}]\} dp} \\ &= \pi_2^l(\beta, p | \sigma) p^{-3/2} \{(1+p^{-1})\Psi'(1+p^{-1})-1\}^{1/2} c_2(l) \mathbf{1}_{[1,l]}(p),\end{aligned}$$

where

$$\{c_2(l)\}^{-1} = \int_1^l \exp\{0.5 \log [np^{-3}\{(1+p^{-1})\Psi'(1+p^{-1})-1\}]\} dp.$$

Thus,

$$\pi_1^l(p, \beta, \sigma) = \sigma^{-1} p^{-3/2} \{(1+p^{-1})\Psi'(1+p^{-1})-1\}^{1/2} c_2(l) (2l)^{-k} (2 \log l)^{-1} \mathbf{1}_{[1,l]}(p) \mathbf{1}_{[-l,l]^k}(\beta) \mathbf{1}_{[l^{-1},l]}(\sigma).$$

Now take any point $\theta^* = (p^*, \beta^*, \sigma^*) \in [1, l] \times [-l, l]^k \times [l^{-1}, l]$. Then, the reference prior for the ordering (p, β, σ) is

$$\begin{aligned}\pi(p, \beta, \sigma) &\propto \lim_{l \rightarrow \infty} \frac{\pi_1^l(p, \beta, \sigma)}{\pi_1^l(p^*, \beta^*, \sigma^*)} \\ &= \sigma^{-1} p^{-3/2} \{(1+p^{-1})\Psi'(1+p^{-1})-1\}^{1/2},\end{aligned}$$

which is of the form (4).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MARF proved Theorem 1, Lemma 1, and Proposition 1, and wrote the manuscript. ES performed the computations for the simulation study and for the application, and wrote the manuscript. Both authors read and approved the final manuscript.

Acknowledgement

The work of Ferreira was supported in part by National Science Foundation Grant DMS-0907064. The authors gratefully acknowledge the constructive comments and suggestions made by three anonymous referees that led to a substantially improved article.

Author details

¹Department of Statistics, University of Missouri, Columbia, USA. ²Department of Electrical and Computer Engineering, Duke University, Durham, USA.

Received: 17 October 2013 Accepted: 11 April 2014

Published: 17 June 2014

References

- Achcar, JA, Pereira, GA: Use of exponential power distributions for mixture models in the presence of covariates. *J. Appl. Stat.* **26**(6), 669–679 (1999)
- Berger, JO, Bernardo, JM: On the development of the reference prior method. In: Bernardo, JM, Berger, JO, Dawid, AP, Smith, AFM (eds.) *Bayesian Statistics 4*, pp. 35–60. Oxford University Press, London, (1992a)
- Berger, JO, Bernardo, JM: Ordered group reference priors with applications to a multinomial problem. *Biometrika* **79**, 25–37 (1992b)
- Berger, JO, Bernardo, JM: Reference priors in a variance components problem. In: Goel, PK, Iyengar, NS (eds.) *Bayesian Analysis in Statistics and Econometrics*, pp. 323–340. Springer, Berlin, (1992c)
- Berger, JO, de Oliveira, V, Sansó, B: Objective Bayesian analysis of spatially correlated data. *J. Am. Stat. Assoc.* **96**(456), 1361–1374 (2001)
- Bernardo, JM: Reference posterior distribution for Bayes inference. *J. Roy. Stat. Soc. B.* **41**, 113–147 (1979)
- Bernardo, JM, Smith, AFM: *Bayesian Theory*. Wiley, New York (1994)
- Box, GEP, Tiao, GC: A further look at robustness via Bayes's theorem. *Biometrika* **49**, 419–432 (1962)
- Box, GEP, Tiao, GC: *Bayesian Inference in Statistical Analysis*. Wiley-Interscience, Hoboken (1992)
- Choy, STB, Smith, AFM: On robust analysis of a normal location parameter. *J. Roy. Stat. Soc. B.* **59**(2), 463–474 (1997)
- Cribari-Neto, F, Ferrari, SLP, Cordeiro, GM: Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika* **87**, 907–918 (2000)
- Datta, GS, Ghosh, JK: Noninformative priors for maximal invariant parameter in group models. *Test.* **4**, 95–114 (1995)
- Datta, GS, Ghosh, M, Mukerjee, R: Some new results on probability matching priors. *Bull. Calcutta Stat. Assoc.* **50**(199–200), 179–192 (2000a)
- Datta, GS, Mukerjee, R, Ghosh, M, Sweeting, TJ: Bayesian prediction with approximate frequentist validity. *Ann. Stat.* **28**, 1414–1426 (2000b)

- Ferreira, MAR, De Oliveira, V: Bayesian reference analysis for Gaussian Markov Random Fields. *J. Multivariate Anal.* **98**, 789–812 (2007)
- Ferreira, MAR, Suchard, MA: Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36**, 355–368 (2008)
- Fonseca, TCO, Ferreira, MAR, Migon, HS: Objective Bayesian analysis for the Student-*t* regression model. *Biometrika.* **95**(2), 325–333 (2008)
- Greene, WH: *Econometric Analysis*. Prentice-Hall, Upper Saddle River (1997)
- Ghosh, JK, Delampady, M, Samanta, T: *An Introduction to Bayesian Statistics – Theory and Methods*. Springer, New York (2006)
- Jeffreys, H: *Theory of Probability*. Oxford University Press, Oxford (1961)
- Levine, DM, Krehbiel, TC, Berenson, ML: *Business Statistics: A First Course*. Pearson Prentice Hall, Upper Saddle River (2006)
- Liang, F, Liu, C, Wang, N: A robust sequential Bayesian method for identification of differentially expressed genes. *Statistica Sinica.* **17**, 571–597 (2007)
- Salazar, E, Ferreira, MAR, Migon, HS: Objective Bayesian analysis for exponential power regression models. *Sankhya - Series B.* **74**, 107–125 (2012)
- Sun, D, Berger, JO: Objective Bayesian analysis for the multivariate normal model. In: Bernardo, JM, Bayarri, MJ, Berger, JO, Dawid, AP, Heckerman, D, Smith, AFM, West, M (eds.) *Bayesian Statistics 8*, pp. 525–547. Oxford University Press, Oxford, (2007)
- Vianelli, S: La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza. *Statistica.* **23**, 447–474 (1963)
- Walker, SG, Gutiérrez-Peña, E: Robustifying Bayesian procedures. In: *Bayesian Statistics 6*. Oxford University Press, New York, (1999)
- Wasserman, L: Asymptotic inference for mixture models using data-dependent priors. *J. Roy. Stat. Soc. B.* **62**, 159–180 (2000)
- West, M: Outlier models and prior distributions in Bayesian linear regression. *J. Roy. Stat. Soc. B.* **46**, 431–439 (1984)
- West, M: On scale mixtures of normal distributions. *Biometrika.* **79**, 646–648 (1987)
- Zhu, D, Zinde-Walsh, V: Properties and estimation of asymmetric exponential power distribution. *J. Econometrics.* **148**, 86–99 (2009)

doi:10.1186/2195-5832-1-12

Cite this article as: Ferreira and Salazar: Bayesian reference analysis for exponential power regression models. *Journal of Statistical Distributions and Applications* 2014 **1**:12.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com