Nonparametric Bayesian Modelling Using Skewed Dirichlet Processes

Pilar Iglesias Z., Yasna Orellana Z. and Fernando A. Quintana*

July 30, 2008

Abstract

We introduce a new class of discrete random probability measures that extend the definition of Dirichlet Process (DP) by explicitly incorporating skewness. The asymmetry is controlled by a single parameter in such a way that symmetric DPs are obtained as a special case of the general construction. We review the main properties of skewed DPs and develop appropriate Polya urn schemes. We illustrate the modelling in the context of linear regression models of the CAPM type, where assessing symmetry for the error distribution is important to check validity of the model.

Key Words: Bayes Factor, Density Estimation, Dirichlet Process, Linear Regression Model, Polya Sequence, Skewed Distribution.

*Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Casilla 306, Correo 22, Santiago, CHILE. Contact e-mail: quintana@mat.puc.cl. This article is dedicated to the memory of Pilar L. Iglesias, a friend who is still very much missed.

1 Introduction

Over the past years there has been an increased interest in the construction of parametric families of distributions that contain a wide range of alternatives. Such developments have been largely motivated by the need to accurately describe and model data coming from many fields of application. These typically require distributions with heavier or lighter tails than the normal case, asymmetry, multimodality, etc.

Parametric asymmetric families of distributions have been the subject of much work. Azzalini (1985) introduces the skew-normal distribution, which extends the normal distribution, including it as a special case. Alternative ways of introducing asymmetry in a given family of symmetric distributions are discussed, among many others, in Fernández and Steel (1998), Branco and Dey (2001) and Azzalini and Capitanio (2003). See further properties and developments of such families, generically termed "skew-distributions", in Genton (2004). Throughout, we will use the terms "skewness" and "asymmetry" as synonyms.

A typical fact concerning the above families is that the type of "non-normal" behavior that is captured depends on how they are constructed and on the number of parameters involved. Consequently, a price to be paid for flexibility is lack of parsimony. An alternative approach that has become increasingly popular over the last decade is to adopt a semi or nonparametric approach. This implies a modelling process that treats the distribution of interest as itself unknown, placing a prior probability measure on the space of distribution functions. A critical advantage of such procedure is that the inference is no longer constrained by any specific parametric form, which ultimately implies a substantially increased flexibility of the modelling. By far the most popular choice of prior over the space of distribution functions – also called *random probability measures* (RPM) – is the Dirichlet process (DP), introduced by Ferguson (1973). The DP has the property of being closed under i.i.d. sampling, i.e. the posterior distribution is again a DP. Consequently, many properties of the DP have been developed. In particular, Blackwell and MacQueen (1973) and Sethuraman (1994) obtain alternative characterizations for this process, the former as a limit of predictive distributions resulting from Polya Urn Schemes and the latter as an infinite weighted mixture of point masses, with stochastically ordered weights. See further discussion in Müller and Quintana (2004) and references therein.

The DP has the property of being almost surely discrete, which turns out to be inappropriate for many applications. To mitigate this problem, several authors introduced mixtures of Dirichlet processes (MDP) which consist of mixing the DP with a given kernel. If this kernel is continuous, the resulting MDP is a process with support in the set of continuous distributions. See further discussion of such mixture modelling in Escobar and West (1998).

In certain applications it is of interest to define distributions with support given by a subset of the space of continuous distributions. Consider, for instance, the distribution of the error terms in a linear regression model. A potentially interesting model would consider errors with symmetric distributions. It is known, for instance, that a regression model such as the capital asset pricing model (CAPM) is valid provided the underlying distribution is elliptically symmetric. See, for instance, Hamada and Valdez (2008) and references therein. A DP model for these errors would not be adequate to this end, as the RPM is asymmetric with probability 1, which follows immediately from the representation in Sethuraman (1994). More generally, symmetry of errors ϵ_i in linear regression models typically allows for a simple interpretation of coefficients, as $E(|\epsilon_i|) < \infty$ guarantees $E(\epsilon_i) = 0$, i.e., mean regression. Dalal (1979) introduced a class of RPMs called Dirichlet Invariant Processes that may be used to deal with the above problem. In the special case where the sets on which the process is defined are symmetric over the real numbers, the process concentrates its mass over symmetric distributions. We refer to this particular case as the symmetric DP. See also Doss (1984).

A related construction, consisting of a DP-based RPM supported on the set of fixed-median distributions was presented in Doss (1985a,b) and in Burr and Doss (2005). A DP-based RPM that concentrates on the class of symmetric and unimodal densities is proposed in Brunner and Lo (1989) (see also Brunner, 1995). An asymmetric version of this construction is discussed in Kottas and Gelfand (2001), who also consider families of zero-median distributions, including the symmetric and asymmetric unimodal case, with applications to median regression.

In this work we introduce a new class of RPMs that is flexible enough to accommodate symmetry and skewness as special cases. We call them skewed Dirichlet processes (SDPs) and the above properties are specially suitable in the context of flexible modelling of error distributions in linear regression models. Our definition depends on a skewness parameter θ such that the symmetric DP is obtained if and only if $\theta = \theta_0$, where θ_0 is a fixed and known value. Our construction is partially motivated by the problem of assessing symmetry in the underlying distributions. SDP models thus help us to investigate, under a more general framework, validity of models relying on symmetry assumptions, such as the CAPM. To this effect, we will show that SDPs preserve many of the key properties of DPs that are useful from both conceptual and computational viewpoints. In particular, the SDP admits a Polya urn representation (which we call the *skewed Polya urn*) and it is also representable as an infinite mixture of point masses. As in the DP case, SDPs are discrete with probability 1 but are dense on the space of all distributions defined on the support of the baseline measure. Modelling of continuous distributions is approached by means of mixtures with respect to a continuous kernel. We illustrate this with an application to the CAPM, with an explicit symmetry assessment.

The rest of this article is organized as follows. Section 2 gives the definition and some of the main properties of SDPs, with emphasis on a representation via urn schemes. We build the theory to the extent needed by practical application of this new tool.

A small simulation study in the context of density estimation is carried out to learn about practical aspects of the application of SDP models for symmetry assessment and modelling. Section 3 discusses semiparametric SDP linear regression models, including an application to the estimation of systematic risk for a Chilean wine producer company. Some final remarks and further discussion are presented in Section 4. Proofs of the various results stated in the main body of the article are given in the Appendix.

2 Skewed Dirichlet Process: Definition and Main Properties

We introduce here a new random probability measure that extends the definition of the Dirichlet Process (DP) by explicitly controlling skewness, and yielding symmetry as a special case. The basic idea stems from the representation of any random variable X with symmetric distribution as $U \cdot V$ where V has the distribution of $|X|$ (concentrated on $\mathbb{R}^+$) and U is uniform on the discrete set $\{-1, 1\}$. By allowing U to be distributed such that $P(U = 1) = \theta = 1 - P(U = -1)$ with $0 < \theta < 1$ then the distribution of $U \cdot V$ is asymmetric and symmetry holds if and only if $\theta = \frac{1}{2}$.

2.1 Definition of Skewed Dirichlet Processes

We call F a *Skewed Dirichlet Processes* (SDP) on $\mathbb{R}$ with parameters θ, M and F_0 if for any measurable partition $A_1, A_2, \ldots, A_k$ of $\mathbb{R}$ we have

$$(F(A_1), \ldots, F(A_k)) \sim \theta Dir(MF_0(A_1), \ldots, MF_0(A_k)) + (1-\theta) Dir(MF_0(-A_1), \ldots, MF_0(-A_k)), \quad (1)$$

where F_0 is the *centering distribution function* concentrated on $\mathbb{R}^+$, $0 < \theta < 1$ is the *skewness parameter*, $M > 0$ is the *total mass* or *prior weight* parameter, and

$Dir(\alpha_1, ..., \alpha_k)$ represents the Dirichlet Distribution with parameters $(\alpha_1, \dots, \alpha_k)$. We denote this as $F \sim SDP(M, F_0, \theta)$.

Let φ denote the function defined as $\varphi(x) = -x$ for $x \in \mathbb{R}$. From the definition (1) it follows immediately that for any $F \sim SDP(M, F_0, \theta)$ and $A \in \mathcal{B}(\mathbb{R})$ we have

$$\mathbb{E}(F(A)) = \theta F_0(A) + (1 - \theta) F_0(\varphi(A)), \tag{2}$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel sets on the real line. In other words, the prior expectation for any SDP is the probability distribution that results from reflecting the baseline distribution F_0 about 0, and changing the weights given to $\mathbb{R}^+$ and $\mathbb{R}^-$ according to the value of θ. In the particular case where $\theta = \frac{1}{2}$ and events A are restricted to satisfy $A = \varphi(A)$, then the SDP becomes a Dirichlet invariant process (Dalal, 1979; Tiwari, 1988).

Remark 1 Comparing (1) with the definition of DP given by Ferguson (1973), the SDP can be thought of as a mixture of a regular DP with its reflection about 0, but with the positive sign assigned with probability θ. It follows from (2) that θ represents the expected amount of probability mass assigned to the positive numbers by any RPM $F \sim SDP(M, F_0, \theta)$.

2.2 Skewed Polya Sequences

Practical application of SDPs requires some extra results. Specifically, and in analogy with the DP case, we discuss now the Polya urn representation for the marginal distribution of a sample $X_1, \dots, X_n \mid F \overset{iid}{\sim} F$ and $F \sim SDP(M, F_0, \theta)$. Such representation is particularly useful for posterior simulation schemes. The description mimics the results due to Blackwell and MacQueen (1973).

We start by noting that, by virtue of Remark 1, if $X_1, \dots, X_n \mid F \overset{iid}{\sim} F$ and $F \sim$

$SDP(M, F_0, \theta)$ then for every $B \in \mathcal{B}(\mathbb{R})$

$$P(X_1 \in B) = \mu(B) \tag{3}$$

$$P(X_{l+1} \in B \mid X_1, X_2, \ldots, X_l) = \frac{\mu_l(B)}{\mu_l(\mathbb{R})}, \tag{4}$$

for every $l = 1, \ldots, n-1$, where $\mu(B) = \mathbb{E}(F(B))$ is given by (2), and where for $l \geq 1$

$$\mu_l(B) = M\mu(B) + \theta \sum_{i=1}^{l} \delta_{|X_i|}(B) + (1-\theta) \sum_{i=1}^{l} \delta_{|X_i|}(\varphi(B)), \tag{5}$$

so that $\mu_l(\mathbb{R}) = M + l$. Here $\delta_z(\cdot)$ represents a point mass at z.

Remark 2 Unlike the case of DPs, SDPs are in general not conjugate under independent sampling. But in practice this is not really a limitation (or a weakness of SDPs compared to DPs) as discrete RPMs are typically used in the context of hierarchical models, combined with a continuous kernel. See the illustrations below. A discussion on related issues can be found in Müller and Quintana (2004).

Any sequence of random variables verifying (3) and (4) will be called *Skewed Polya sequence* with parameters (θ, M, F_0). We make an explicit distinction between these and regular Polya sequences for which $\mu(B)$ and $\mu_l(B)$ in (3) and (4) are replaced by $F_0(B)$ and $MF_0(B) + \sum_{i=1}^{l} \delta_{X_i}(B)$, respectively.

Skewed Polya sequences have several properties, extending corresponding results for regular Polya sequences. We summarize some of these next. In the sequel, the notation $A \perp B$ means independence of the random objects A and B and $A \perp B \mid C$ means conditional independence of A and B given C.

Theorem 1 *If $\{X_n,\, n \geq 1\}$ is a Skewed Polya Sequence with parameters (θ, M, F_0) then*

(a) $|X_{n+1}| \perp (X_1, \ldots, X_n) \mid |X_1|, \ldots, |X_n|$.

(b) $\{\mathrm{sgn(X_n), n \geq 1}\}$ *are i.i.d. random variables.*

(c) $\{|X_n|, n \geq 1\}$ *and* $\{-|X_n|, n \geq 1\}$ *are Polya sequences with parameters* (M, F_0) *and* $(M, F_0 \circ \varphi)$*, respectively.*

Theorem 1 gives important clues about the behavior of skewed Polya sequences. Essentially, they can be represented as regular Polya sequences *on the absolute values* (or their negative versions) plus an i.i.d. process for the signs. A modified version of the Chinese restaurant process (CRP) description (Arratia, Barbour, and Tavaré, 1992) can also be given for skewed Polya sequences. The regular CRP involves a customer entering a restaurant and deciding to either join an already started table with probability proportional to the table size or to start a new one with probability proportional to M. Tables are identified with "labels" drawn from the baseline distribution F_0, so that a customer joining an already started table adopts the corresponding label, but when opening a new table the client draws a fresh label from F_0, independently of the previous ones. In the "skewed CRP", tables play a conceptually different role in that they are divided into two sub-tables: in one of them – the "positive table" – the customer spends a proportion θ of his/her time leaving the remaining $(1-\theta)$ proportion of time to the other – the "negative table". If k tables have been thus opened, the next entering costumer "sees" $2k$ paired sub-tables, and each pair of sub-tables accounts for one "skewed" table. The labels are drawn from F_0, and positive and negative tables are identified by the original label and its negative value, respectively.

The next result shows how to construct skewed Polya sequences from regular Polya sequences, giving a converse to (b) and (c) from Theorem 1.

Theorem 2 *Let* $\{|X_n|, n \geq 1\}$ *be a Polya Sequence with parameters* (M, F_0) *and define* $Y_n = U_n \cdot |X_n|$ *where* $\{U_n, n \geq 1\}$ *is an i.i.d. sequence independent of* $\{X_n, n \geq 1\}$ *and with* $P(U_1 = 1) = \theta = 1 - P(U_1 = -1)$*. Then* $\{Y_n, n \geq 1\}$ *is a Skewed Polya Sequence with parameters* (M, F_0, θ)*.*

A basic fact concerning a regular Polya sequence with parameters (M, F_0) is that the predictive distribution $P(X_{n+1} \in B \mid X_1, \ldots, X_n)$ converges with probability 1 to $F^*(B)$ where $F^* \sim DP(M, F_0)$ (Blackwell and MacQueen, 1973). We give now the version of this result that corresponds to skewed Polya sequences.

Theorem 3 *The predictive distribution $P(X_{n+1} \in B \mid X_1, \ldots, X_n)$ defined in (4) converges with probability 1 as $n \to \infty$ to a limiting random probability measure $G^* = \theta F^* + (1-\theta) F^* \circ \varphi$ where $F^* \sim DP(M, F_0)$.*

Our next result extends the infinite mixture representation of DPs by Sethuraman (1994) to the case of SDPs. The proof of the result follows immediately from Theorem 3 and is therefore omitted.

Theorem 4 *Any $F \sim SDP(M, F_0, \theta)$ can be represented as*

$$F(\cdot) = \theta \sum_{h=1}^{\infty} w_h \delta_{V_h}(\cdot) + (1-\theta) \sum_{h=1}^{\infty} w_h \delta_{-V_h}(\cdot), \qquad (6)$$

where $V_1, V_2, \ldots \overset{iid}{\sim} F_0$, and $w_1, w_2, \ldots$ are stochastically ordered weights *that follow a stick breaking process: $w_1 = U_1$ and $w_h = U_h \times \prod_{i=1}^{h-1}(1-U_i)$ for all $h \geq 2$, with $U_1, U_2, \ldots \overset{iid}{\sim} Beta(1, M)$.*

Theorem 4 shows that $F \sim SDP(M, F_0, \theta)$ can be seen as an infinite mixture of point masses with marginal distribution μ in (3). This result also highlights the fact that symmetry is obtained if and only if $\theta = 1/2$. It is also interesting to note that whenever F_0 has density f_0, μ has density $\theta f_0(x) + (1-\theta) f_0(-x)$ for $x \in \mathcal{X}$. We note that this density is discontinuous at $x = 0$ unless $f_0(0) = 0$. In most applications this discontinuity is hardly justifiable. A simple solution to this problem is to consider, for $0 < \theta < 1$

$$F(\cdot) = \theta \sum_{h=1}^{\infty} w_h \delta_{\{\theta V_h\}}(\cdot) + (1-\theta) \sum_{h=1}^{\infty} w_h \delta_{-\{(1-\theta)V_h\}}(\cdot), \qquad (7)$$

with weights $\{w_h\}$ and random variables $\{V_h\}$ defined as before. Under (7) the marginal density of each mixture component is

$$f_0\left(\frac{x}{\theta}\right) I\{x \geq 0\} + f_0\left(-\frac{x}{1-\theta}\right) I\{x < 0\}, \tag{8}$$

which *is* continuous. Expression (8) defines a two-piece density such as in Fernández and Steel (1998) and Arellano-Valle et al. (2005). As discussed in Jones (2006), many of these two-piece distributions are related trough simple reparametrizations. Under the alternative construction that leads to (8), the theory developed earlier is still valid, with the obvious changes. In particular, the skewed CRP works as before, but instead of placing a point mass V_h or $-V_h$ at values sampled from F_0, we label positive and negative tables as θV_h and $-(1-\theta)V_h$, respectively. Note that symmetry is again obtained if and only if $\theta = \frac{1}{2}$.

2.3 Computational Issues

An important advantage of semiparametric modelling based on the DP is that the computational complexity is, in principle, independent of the dimension of the space where this is defined. MCMC schemes for DP-related models have been extensively studied. See, e.g., MacEachern and Müller (1998), Neal (2000) and references therein. A key factor in these approaches is the regular Polya urn representation discussed earlier and the underlying exchangeability.

Fortunately, MCMC schemes for SDPs enjoy the same good properties as their DP cousins. For the sake of presentation, consider the simple model $y_i \mid \mu_i \overset{ind}{\sim} p(y_i \mid \mu_i)$, where $\mu_1, \ldots, \mu_n \mid F \overset{iid}{\sim} F$ and $F \sim SDP(M, F_0, \theta)$, where $F_0(\cdot)$ has density $f_0(\cdot)$. By Theorem 1, the parameter vector $\mu = (\mu_1, \ldots, \mu_n)$ can be represented as $(|\mu|, \nu)$, where $|\mu| = (|\mu_1|, \ldots, |\mu_n|)$ is the vector of absolute values of μ and $\nu_i = \text{sgn}(\mu_\text{i})$. Furthermore, by the cluster structure implied by the discreteness of SDPs, $|\mu|$ can be represented by means of the set of unique values $|\mu_1|^*, \ldots, |\mu_k|^*$ among the components of $|\mu|$ (i.e., the cluster *locations*) and the cluster *memberships* $s = (s_1, \ldots, s_n)$ defined as $s_i = j$ if and only if the ith subject belongs to the jth cluster. Of course, this implies $|\mu_i| = |\mu_{s_i}|^*$.

MCMC schemes for posterior inference can be then implemented as in the regular DP case. Indeed, we can apply any of the available algorithms to $|\mu|$, which has been already shown to be a regular DP in Theorem 1, and then update the signs ν from their corresponding complete conditionals. Moreover, if for any $S \subset \{1, \ldots, n\}$ the integrals $\int_0^\infty \prod_{i \in S} p(y_i \mid \mu) f_0(\mu)\, d\mu$ and $\int_0^\infty \prod_{i \in S} p(y_i \mid -\mu) f_0(\mu)\, d\mu$ are available in closed form, which will typically happen if the likelihood $p(y|\mu)$ and $f_0(\mu)$ are of conjugate form, then the collapsed algorithm described in, e.g., MacEachern (1998) can be adopted. This amounts to integrating out the locations $|\mu|^*$ and then updating the memberships s from their complete conditionals. In the non-conjugate case, it is possible to adapt the "no-gaps" algorithm of MacEachern and Müller (1998) to the present case.

2.4 A Simulation Study

To learn about practical issues and potential limitations in the applications of SDP-models we consider here a small simulation study focused on symmetry assessments in the context of density estimation. Specifically, we adapt the mixture model of Escobar and West (1995) to our current context. In doing so we take full advantage of the flexibility provided by the semiparametric density estimation, with the extra advantage of an explicit procedure to handle symmetry. We note that the model in Escobar and West (1995) is bivariate, that is, it postulates an infinite mixture for the mean *and* variance parameters. Since the SDP is meant for distributions supported on the whole real line, we consider a combined model that assigns point masses as in the usual DP for the variance parameter (a non-negative quantity) and our construction for the mean parameter, including an extra location term that plays the role of symmetry point.

Specifically, we assume the following hierarchical model:

$$\begin{aligned} y_i \mid \mu_i, V_i &\sim N(\mu_i, V_i) \\ (\mu_1, V_1), \ldots, (\mu_n, V_n) \mid F &\sim F \end{aligned}$$

and

$$F(\cdot) = \theta \sum_{h=1}^{\infty} w_h \delta_{(m+\theta A_h, B_h)}(\cdot) + (1-\theta) \sum_{h=1}^{\infty} w_h \delta_{(m-(1-\theta)A_h, B_h)}(\cdot),$$

with weights $\{w_h\}$ defined by means of the usual stick-breaking process with parameter M, and where $\{(A_h, B_h)\}$ are sampled from $F_0(A, B)$ such that

$$A \mid B \sim |N(0, \tau B)|, \qquad B^{-1} \sim \Gamma(a_1, b_1),$$

i.e., $A \mid B$ has the same distribution as the absolute value of a random variable $\mathcal{A} \sim N(0, \tau B)$. Assume also

$$\tau^{-1} \sim \Gamma(\lambda_0, \lambda_1), \quad m \sim N(m_0, S), \quad M \sim \Gamma(a, b),$$

for known hyperparameters $(a_1, b_1, a, b, \lambda_0, \lambda_1, m_0, S)$. For the skewness parameter we assume the following prior:

$$\theta \sim (1-\pi) B(a_0, b_0) + \pi \delta_{1/2}(\theta), \tag{9}$$

for known values of (π, a_0, b_0), which corresponds to a mixture between a point mass at 1/2 and a Beta distribution, with weights defined by π. This assumption allows for an explicit assessment of symmetry, as the corresponding posterior $p(\theta \mid y_1, \ldots, y_n)$ will be again a mixture between a continuous distribution and a point mass at 1/2. Consequently, we can compute the Bayes factor in favor of symmetry as

$$\frac{P(\theta = \frac{1}{2} \mid \text{data})(1-\pi)}{P(\theta \neq \frac{1}{2} \mid \text{data})\pi}, \tag{10}$$

which can be easily evaluated from any posterior simulation output.

As a more pragmatic estimation alternative, i.e., where the symmetry assessment is not the main focus, we may assume $\pi = 0$ in (9) above, that is,

$$\theta \sim B(a_0, b_0). \tag{11}$$

This assumption still induces a model that includes symmetry as a special case, but does not specifically favor it by forcing the posterior distribution to have a corresponding atom. Thus, this may be a choice that is more appropriate for the sake of general density estimation.

The above model sets up a prior RPM for the unknown distribution F that has three parts: an overall center m, individual locations for all mixture components, coming from the SDP representation (7) and expressed as $(m + \theta A_h, m - (1-\theta)A_h)$, and the corresponding variances B_h, $h \geq 1$. The resulting marginal model resembles traditional kernel density estimation procedures, in the spirit described by Escobar and West (1995), but based on the SDPs.

We simulated data samples of size $n = 100$ from the following distributions: (i) $N(0, 4)$; (ii) double exponential with density $f(x) = \exp(-|x|/3)/6$; and the two- and three-component mixtures (iii) $0.5N(2, 3) + 0.5N(-2, 3)$; (iv) $0.7N(3, 2) + 0.3N(-3, 1)$; and (v) $0.3N(-10, 3) + 0.5N(0, 1) + 0.2N(4, 0.5)$. In all cases, and following Escobar and West (1995), we fitted the above model using $m_0 = 0$, $S = 10^5$, $\lambda_0 = 1/2$, $\lambda_1 = 50$, $a = 2$, $b = 4$, $a_1 = 2$, $b_1 = 1$, $a_0 = b_0 = 1$ and $\pi = 0.5$, giving substantial weight to prior symmetry. Examples (i) through (iii) are clearly symmetric and uni- or bi-modal. Example (iv) is bimodal and asymmetric, and (v) is multimodal and also asymmetric.

The Bayes factors in favor of symmetry (10) for the five simulated datasets were as follows: 2.216 for (i); 3.788 for (ii); 3.593 for (iii); 0.0160 for (iv); and 0.0001 for (v). All these Bayes factors support the right hypotheses. This is corroborated by the histograms and posterior predictive densities shown in Figure 1, for examples (iv) and (v) only, using priors (9) and (11). For ease of comparison the true simulation density is included as a dashed line. For example (iii) the prediction looks very close to the histogram but slightly more so for prior (11). The same happens for cases (i) and (ii) which are not shown here. For case (iv) the prediction is essentially indistinguishable for both priors, and also very close to the histogram, but not too close to the true simulation density. In this case, the rather limited sample size is certainly determinant. We repeated experiment (v) using a sample of size $1,000$ and both predictions were much closer to the true generating density (not shown). As a final experiment, we tried the same model on the galaxy data described in Escobar and West (1995), with similar results to case (v) above (data not shown). In fact, these results are very similar to what Escobar and West (1995) reported using regular DPs.

3 Application to Linear Regression Models

We now consider linear regression, with a model for errors based on SDPs.

3.1 Some Preliminaries

There is now a substantial literature on parametric and semiparametric Bayesian linear regression models with asymmetric errors. From a parametric viewpoint, the pioneering work by Fernández and Steel (1998) introduced a method for constructing univariate asymmetric distributions from a symmetric kernel, with a skewness parameter. Sahu, Dey, and Branco (2003) provide a multivariate construction of skew distributions, again with a skewness parameter. Some recent discussion and applications are presented in Liseo (2004), Ferreira and Steel (2004) and references therein. A general discussion of univariate skewed distributions can be found in Ferreira and Steel (2006). Alternatively one may consider median, or more generally, quantile regression. Yu and Moyeed (2001) consider asymmetric Laplace distributions for quantile regression, fixing the pth quantile to be 0. Related discussion and applications can be found in Yu, Van Kerm, and Zhang (2005) and references therein, and a substitution likelihood approach can be found in Dunson and Taylor (2005).

Semiparametric approaches to linear regression have also been the subject of much work. Gelfand (1999) provides a general discussion. West, Müller, and Escobar (1994) consider linear regression with a DP-based mixture on regression coefficients and variances. Brunner and Lo (1989) and Brunner (1995) consider regression with error terms having symmetric unimodal densities. Lavine and Mockus (1995) use the Dirichlet process to model error distributions in the context of isotonic regression. Quantile and median regression have also been studied under a semiparametric standpoint. Kottas and Gelfand (2001) and Gelfand and Kottas (2003) study median regression modelling, describing a family of zero-median asymmetric distributions using a particular scale-mixture of DPs (a fixed median and interquartile range DP had been previously described in Newton, Czado, and Chappell, 1996). They also describe a fully nonparametric approach zero-median family of distributions, based on mixing two DPs, but with less stable results (Kottas and Gelfand, 2001). More recently, Kottas and Krnjajic (2005) develop a nonparametric methodology for quantile regression using dependent DPs (MacEachern, 1999), letting the shape of the resulting error distributions be driven by covariate values. Finally, Hanson and Johnson (2002) consider a mixture of absolutely continuous zero-median Polya trees for the error distribution.

3.2 The Model

We consider an alternative proposal for the distribution of errors in linear regression, using mixtures of SDPs on the location rather on the scale. Specifically, let $(y_i, \boldsymbol{x}_i)$ denote the response and associated vector of p-dimensional covariates for the ith subject,

$i = 1, \ldots, n$. We assume the following hierarchical model for the data:

$$\begin{aligned} y_i \mid \mu_i, \boldsymbol{\beta}, \sigma^2 &\sim N(\mu_i + \boldsymbol{\beta}^T \boldsymbol{x}_i, \sigma^2) \\ \mu_1, \ldots, \mu_n \mid F &\sim F \\ F \mid \theta, M, \tau &\sim SDP(\theta, M, F_0(\tau)) \end{aligned} \tag{12}$$

where we use specification (7) for F, $F_0(\tau)$ is the distribution of $|Z|$ with $Z \sim N(0, \tau)$ and with the following prior definition: $\sigma^{-2} \sim \Gamma(\nu_0, \nu_1)$, $\tau^{-1} \sim \Gamma(\lambda_0, \lambda_1)$, and $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{S})$, for known hyperparameters $(\nu_0, \nu_1, \lambda_0, \lambda_1, \boldsymbol{S})$. The model specification is completed by assuming the prior distribution for θ to be the mixture defined in (9). As earlier, this assumption allows for an explicit assessment of symmetry, as the corresponding posterior $p(\theta \mid y_1, \ldots, y_n)$ will be again a mixture between a continuous distribution and a point mass at $1/2$.

The hierarchical structure adopted in (12) has several motivations. Firstly, we consider that asymmetry may be introduced by the presence of outliers or groups of points with similar deviation from the regression curve, which is taken to be a linear combination of the available covariates. The flexibility of the SDP can capture such abnormalities, through the intercept parameters μ_i, similarly to the method discussed by Quintana and Iglesias (2003) in the context of product partition models. Secondly, it will be shown that the marginal distribution of responses, after integrating out the SDP, is characterized by the regression coefficients and the skewness parameter θ, and that symmetry holds if and only if $\theta = 1/2$. This will result in an explicit symmetry assessment procedure that may be useful when dealing with CAPM models, as mentioned in the introduction. Of course, symmetry of the error distribution is not enough to guarantee validity of the theoretical derivation of the CAPM, i.e. ellipticity. Nevertheless, this is still useful as a procedure to check whether the data show any features that contradict the basic symmetry assumption. Specifically, the Bayes factor (10) can be used to this effect. See the discussion below. Thirdly, the hierarchical model (12) has the advantage of interpretability. The nonparametric model is placed on the intercept parameters μ_i. Therefore, given the μ_i's and the RPM F, the regression coefficients $\boldsymbol{\beta}$ and variance σ^2 have the usual interpretation. The prior specification can thus be done conditionally on the intercepts.

Note that the likelihood component in (12) can be represented as $y_i = \mu_i + \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with $N(0, \sigma^2)$ distribution and independent of all other terms in the model. This can be rewritten as $y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + r_i$, $i = 1, \ldots, n$, where $r_i = \mu_i + \epsilon_i$ represents the error term in the likelihood model. We now formally show that r_i has marginal symmetric distribution if and only if $\theta = 1/2$.

Theorem 5 *Under model (12), the distribution of errors has marginal density given by* $f(x) = \int N(x;\, \mu, \sigma^2)\, dF(\mu)$*, for which symmetry holds if and only if* $\theta = \frac{1}{2}$.

We remark here that the conclusion of Theorem 5 holds under either (6) or (7). Therefore, when assuming the prior distribution (9) for θ we can still compute a Bayes factor in favor of symmetry as in (10).

3.3 Illustration: Stock Market Returns

We consider now data on a monthly series of stock returns from Concha y Toro, Chilean wine producers. It is of interest to study the relationship between these stock returns and a variable describing the behavior of the market. We use here the corresponding values of IPSA, the Chilean version of the Dow-Jones index. Capital asset pricing models (CAPM) are routinely used in such cases (see, e.g., Elton and Gruber, 1995). Specifically, we consider the regression model $y_i \sim N(\mu + \beta x_i, \sigma^2)$, where y_i represents the return on the asset in excess of the risk-free rate during the i-th period, and x_i is the excess return on the market portfolio of assets in the i-th period. In most cases, estimating the regression slope β or *systematic risk* is the main target of the analysis, as it describes how the stock returns vary relative to the market.

Figure 2 shows a scatter plot of the data and a histogram of the regression residuals after fitting a regression line using ordinary least squares. The histogram contains the kernel density estimator provided by the function density available in the **R** system. The plot suggests the need to use asymmetric distributions. In fact, the kernel estimator suggests the presence of a few minor modes in the distribution of errors.

It is well known that atypical observations (outliers, gross errors, etc.), such as those found in Figure 2 can have a severe impact on the estimation of regression coefficients. To deal with such problems, Quintana and Iglesias (2003) proposed using product partition models with random effects in the form of subject-specific intercepts. We now reanalyze these data using model (12) with Concha y Toro stock returns as response, and IPSA values as the single predictor. Our aim here is three-fold: (i) we want a robust estimate of the systematic risk β, which is accomplished by the flexible distribution of errors; (ii) we want to capture the heterogeneity which provokes the presence of several outliers; and (iii) we want to formally test for symmetry.

To fit model (12) to these data we chose the following hyperparameter values: $\pi = 0.2$, $\nu_0 = \lambda_0 = 2.01$, $\nu_1 = \lambda_1 = 1.01$, $S = 10^4$, and $a_0 = b_0 = 1$. Some posterior summaries are given in Table 1. Figure 3 shows the marginal posterior densities for β, σ^2 and τ. It also shows the conditional posterior of θ given $\theta \neq 1/2$. With the choices stated earlier, we found the posterior probability that $\theta = 1/2$ to be given by 0.2168, from which the Bayes factor (10) turns out be equal to 1.107. A similar value was obtained when fitting again this model but now increasing π to 0.5. This is certainly inconclusive in terms of choosing between symmetry and asymmetry, and therefore, it is not to be taken as strong evidence against the validity of CAPM in this problem. It is interesting to observe that the conditional marginal posterior of θ given that $\theta \neq 1/2$, i.e., the continuous component in the mixture that corresponds to $p(\theta \mid \boldsymbol{y})$, is more concentrated on $(0.5, 1)$ than on $(0, 0.5)$. This is consistent with what is reported in the right panel of Figure 2, which suggests some possible skewness to the right, mainly due to the presence of a few outliers. For a better understanding of how the model fits such outlying points we considered the 95% HPDs for regression residuals shown in Figure 4. Most of the HPDs do include 0, but some of them do not. This may be interpreted as an indication of atypical points.

To further understand how residuals are handled by the model, we sorted the absolute value of the posterior means for all residuals. Next we considered the three smallest ones, the three in the middle and the three largest ones. The corresponding posterior densities are shown in Figure 5, from top to bottom. Not surprisingly, the posteriors in the first row are tightly concentrated, but less so in the middle. The final row exhibits bimodal shapes, revealing that a standard parametric model would be inappropriate to fit the error distributions.

4 Discussion

A basic motivation of this work is the development of probability models that incorporate skewness while at the same time providing increased flexibility with respect to alternative parametric forms. When incorporating skewness, it is important to have symmetry as a special case, as comparing both scenarios is usually quite important. The SDP introduced here provides both the flexibility of nonparametric modelling and the possibility of assessing symmetry. Some useful probabilistic representations of SDPs were derived, namely, Polya urn schemes and stick-breaking mixtures. The discreteness of the construction can be easily handled by considering mixtures with a continuous kernel. The skewness parameter θ is such that the process supports symmetric distributions with probability one if and only if $\theta = 1/2$. A Bayes factor for such hypothesis can be easily obtained from posterior simulation output. One can also examine the posterior distribution of θ, in particular, positive values of $2P(\theta > 1/2 \mid \text{data}) - 1$ may be used as a rough indication of skewness to the right.

We applied the model to the problem of estimating systematic risk for a monthly series of stock returns. The specific context is a linear regression model under the presence of skewness *and* outliers, where symmetry lies at the heart of the model justification. Our modelling allows to handle both problems simultaneously. These results, together with what we learned in the simulation study allow us to conclude that SDPs can be used as a general nonparametric modelling tool, with flexibility comparable to the usual DP-based models. The specific symmetry assessment though may be more effective in the context of regression analysis while a general SDP model seems more appropriate for density estimation.

We are currently studying the performance of SDP modelling in other problems such as random effects models and models with errors in variables. Multivariate extensions of this construction are also of interest, but we anticipate a number of practical and computational difficulties when dealing with many dimensions. These and other issues are the subject of ongoing research.

Acknowledgments

Fernando Quintana was partially funded by grants FONDECYT 1020712, 1060729 and Laboratorio de Análisis Estocástico PBCT-ACT13. We thank the Associate Editor and two referees for their valuable comments and suggestions.

Appendix: Proof of Theorems

Proof of Theorem 1

Note that the predictive probabilities (4) depend on previous values in the sequence only through their absolute values. This immediately implies

$$P(X_{n+1} \in B \mid X_1, \ldots, X_n) = P(X_{n+1} \in B \mid |X_1|, \ldots, |X_n|), \tag{13}$$

from which (a) follows easily. Also,

$$P(\mathrm{sgn(X_{n+1})} = 1 \mid \mathrm{sgn(X_1)}, \ldots, \mathrm{sgn(X_n)}) = \mathrm{E(I\{X_{n+1} > 0\} \mid sgn(X_1), \ldots, sgn(X_n))}.$$

The right-hand side of this last equation can be expressed as

$$\begin{aligned} &E(E(I\{X_{n+1} > 0\} \mid \mathrm{sgn(X_1)}, \ldots, \mathrm{sgn(X_n)}, |\mathrm{X_1}|, \ldots, |\mathrm{X_n}|) \mid \mathrm{sgn(X_1)}, \ldots, \mathrm{sgn(X_n)}) \\ &\qquad = E(E(I\{X_{n+1} > 0\} \mid X_1, \ldots, X_n)|\mathrm{sgn(X_1)}, \ldots, \mathrm{sgn(X_n)}) \\ &\qquad\qquad = E(P(X_{n+1} > 0 \mid X_1, \ldots, X_n) \mid \mathrm{sgn(X_1)}, \ldots, \mathrm{sgn(X_n)}). \end{aligned}$$

Using (4), we find that $P(X_{n+1} > 0 \mid X_1, \ldots, X_n) = \theta$. Thus $P(\mathrm{sgn(X_{n+1})} = 1 \mid \mathrm{sgn(X_1)}, \ldots, \mathrm{sgn(X_n)}) = \theta$, from which (b) follows. Now, for all $A \in \mathcal{B}(\mathbb{R}^+)$, (13) implies

$$\begin{aligned} &P(|X_{n+1}| \in A \mid |X_1|, \ldots, |X_n|) \\ &\qquad = P(X_{n+1} \in A \mid |X_1|, \ldots, |X_n|) + P(X_{n+1} \in -A \mid |X_1|, \ldots, |X_n|). \end{aligned}$$

But $P(X_{n+1} \in A \mid |X_1|, \ldots, |X_n|) = \theta(M\mu(A) + \sum_{i=1}^n \delta_{|X_i|}(A))/(M\mu(\mathbb{R}^+) + n)$ and also, $P(X_{n+1} \in -A \mid |X_1|, \ldots, |X_n|) = (1-\theta)(M\mu(\varphi(A)) + \sum_{i=1}^n \delta_{|X_i|}(\varphi(A)))/(M\mu(\mathbb{R}^+) + n)$, where $\mu(\mathbb{R}^+) = \theta$ and $\mu(A) = \theta F_0(A) + (1-\theta)F_0(\varphi(A)) = \theta F_0(A)$. Consequently,

$$P(|X_{n+1}| \in A \mid |X_1|, \ldots, |X_n|) = \frac{M\mu(A) + \sum_{i=1}^n \delta_{|X_i|}(A)}{M\mu(\mathbb{R}^+) + n},$$

and the corresponding marginal probability for X_1 is given by

$$P(|X_1| \in A) = \frac{\mu(A)}{\mu(\mathbb{R}^+)} = \frac{\theta F_0(A)}{\theta} = F_0(A),$$

which shows the first part of (c). The other one is analogous. ■

Proof of Theorem 2

By definition we have $Y_i = |X_i| I\{U_i = 1\} - |X_i| I\{U_i = -1\}$. Hence we must necessarily have that $|X_{n+1}| \perp U_1, \ldots, U_n \mid |X_1|, \ldots, |X_n|$, which implies

$$P(Y_{n+1} \in B | Y_1, \ldots, Y_n) = \begin{cases} P(|X_{n+1}| \in B | |X_1|, \ldots, |X_n|) & \text{if } U_{n+1} = 1 \\ P(-|X_{n+1}| \in B | -|X_1|, \ldots, -|X_n|) & \text{if } U_{n+1} = -1, \end{cases}$$

for any borel set B in the appropriate space. Therefore,

$$\begin{aligned} P(Y_{n+1} \in B | Y_1, \ldots, Y_n) = \theta P(|X_{n+1}| \in B | |X_1|, \ldots, |X_n|) \\ + (1-\theta) P(-|X_{n+1}| \in B | -|X_1|, \ldots, -|X_n|). \end{aligned}$$

But $|Y_i| = |X_i|$ from which we get

$$P(Y_{n+1} \in B | Y_1, \ldots, Y_n) = \theta(F_0(B) + \sum_{i=1}^{n} \delta_{|Y_i|}(B)) + (1-\theta)(F_0(\varphi(B)) + \sum_{i=1}^{n} \delta_{-|Y_i|}(B)),$$

which shows the result. ■

Proof of Theorem 3

Denote $m_n(B) = P(|X_{n+1}| \in B \mid |X_1|, \ldots, |X_n|)$ for $B \in \mathcal{B}(\mathbb{R}^+)$. Then, by Theorem 1

$$\frac{\mu_n(B)}{\mu_n(\mathbb{R})} = \theta m_n(B) + (1-\theta) m_n(\varphi(B)).$$

Because $m_n(\cdot)$ is the predictive distribution of a regular Polya sequence we have that $m_n(B)$ and $m_n(\varphi(B))$ converge with probability 1 to $F^*(B)$ and $F^*(\varphi(B))$, respectively, where $F^* \sim DP(M, F_0)$. The result follows. ■

Proof of Theorem 5

We have that $\mu_1|\boldsymbol{\beta}, \sigma^2$ has density given by $\theta f_0(\mu_1) I\{\mu_1 > 0\} + (1-\theta) f_0(-\mu_1) I\{\mu_1 < 0\}$ which is symmetric if and only if $\theta = 1/2$. By exchangeability of $\mu_1, \ldots, \mu_n$, the same conclusion applies to all the μ_i's. Since ϵ_i is independent of μ_i and has symmetric distribution, r_i has symmetric distribution if and only if $\theta = 1/2$. ■

References

Arellano-Valle, R., H. Gómez, and F. Quintana (2005). Statistical Inference for A General Class of Asymmetric Univariate Distributions. *Journal of Statistical Planning and Inference 128*(2), 427–443.

Arratia, R., A. D. Barbour, and S. Tavaré (1992). Poisson process approximations for the Ewens sampling formula. *Annals of Applied Probability 2*, 519–535.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*, 171–178.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B. Statistical Methodology 65*, 367–389.

Blackwell, D. and J. B. MacQueen (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics 1*, 353–355.

Branco, M. D. and D. K. Dey (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis 79*(1), 99–113.

Brunner, L. J. (1995). Bayesian linear regression with error terms that have symmetric unimodal densities. *Journal of Nonparametric Statistics 4*, 335–348.

Brunner, L. J. and A. Y. Lo (1989). Bayes methods for a symmetric unimodal density and its mode. *The Annals of Statistics 17*, 1550–1566.

Burr, D. and H. Doss (2005). A Bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association 100*(469), 242–251.

Dalal, S. R. (1979). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Processes and their Applications 9*, 99–107.

Doss, H. (1984). Bayesian estimation in the symmetric location problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 68*(2), 127–147.

Doss, H. (1985a). Bayesian nonparametric estimation of the median. I. Computation of the estimates. *The Annals of Statistics 13*, 1432–1444.

Doss, H. (1985b). Bayesian nonparametric estimation of the median. II. Asymptotic properties of the estimates. *The Annals of Statistics 13*, 1445–1464.

Dunson, D. B. and J. A. Taylor (2005). Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics 17*(3), 385–400.

Elton, E. J. and M. J. Gruber (1995). *Modern Portfolio Theory and Investment Analysis, 5th Edition.* New York: Wiley.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*, 577–588.

Escobar, M. D. and M. West (1998). Computing nonparametric hierarchical models. In D. Dey, P. Müller, and D. Sinha (Eds.), *Practical nonparametric and semiparametric Bayesian statistics*, Volume 133 of *Lecture Notes in Statist.*, pp. 1–22. New York: Springer.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*, 209–230.

Fernández, C. and M. F. J. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association 93*, 359–371.

Ferreira, J. T. A. S. and M. F. J. Steel (2004). Bayesian multivariate skewed regression modeling with an application to firm size. In M. Genton (Ed.), *Skew-elliptical distributions and their applications*, pp. 175–190. Chapman & Hall/CRC, Boca Raton, FL.

Ferreira, J. T. A. S. and M. F. J. Steel (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association 101*, 823–829.

Gelfand, A. E. (1999). Approaches for semiparametric Bayesian regression. In S. Ghosh (Ed.), *Asymptotics, nonparametrics, and time series*, Volume 158 of *Statist. Textbooks Monogr.*, pp. 615–638. New York: Dekker.

Gelfand, A. E. and A. Kottas (2003). Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics. Theory and Applications 30*(4), 651–665.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications*. Chapman & Hall/CRC, Boca Raton, FL.

Hamada, M. and E. A. Valdez (2008). CAPM and Option Pricing with Elliptically Contoured Distributions. *Journal of Risk and Insurance 75*(2), 387–409.

Hanson, T. and W. Johnson (2002). Modeling Regression Error with a Mixture of Polya Trees. *Journal of the American Statistical Association 97*, 1020–1033.

Jones, M. C. (2006). A note on rescalings, reparametrizations and classes of distributions. *Journal of Statistical Planning and Inference 136*(10), 3730–3733.

Kottas, A. and A. E. Gelfand (2001). Bayesian Semiparametric Median Regression Modeling. *Journal of the American Statistical Association 96*, 1458–1468.

Kottas, A. and M. Krnjajic (2005). Bayesian Nonparametric Modeling in Quantile Regression. Technical report, University of California, Santa Cruz, Department of Applied Mathematics and Statistics.

Lavine, M. and A. Mockus (1995). A Nonparametric Bayes Method for Isotonic Regression. *Journal of Statistical Planning and Inference 46*, 235–248.

Liseo, B. (2004). Skew-elliptical distributions in Bayesian inference. In M. Genton (Ed.), *Skew-elliptical distributions and their applications*, pp. 153–171. Chapman & Hall/CRC, Boca Raton, FL.

MacEachern, S. N. (1998). Computational methods for mixture of Dirichlet process models. In *Practical nonparametric and semiparametric Bayesian statistics*, Volume 133 of *Lecture Notes in Statist.*, pp. 23–43. New York: Springer.

MacEachern, S. N. (1999). Dependent nonparametric process. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association, Alexandria, VA.

MacEachern, S. N. and P. Müller (1998). Estimating Mixture of Dirichlet Process Models. *Journal of Computational and Graphical Statistics 7*(2), 223–338.

Müller, P. and F. Quintana (2004). Nonparametric Bayesian Data Analysis. *Statistical Science 19*, 95–110.

Neal, R. M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics 9*, 249–265.

Newton, M. A., C. Czado, and R. Chappell (1996). Bayesian inference for semiparametric binary regression. *Journal of the American Statistical Association 91*(433), 142–153.

Quintana, F. and P. Iglesias (2003). Bayesian Clustering and Product Partition Models. *Journal of the Royal Statistical Society Series B 65*, 557–574.

Sahu, S. K., D. K. Dey, and M. D. Branco (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics. La Revue Canadienne de Statistique 31*(2), 129–150.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica 4*(2), 639–650.

Tiwari, R. C. (1988). Convergence of Dirichlet invariant measures and the limits of Bayes estimates. *Communications in Statistics. Theory and Methods 17*(2), 375–393.

West, M., P. Müller, and M. D. Escobar (1994). Hierarchical priors and mixture models, with application in regression and density estimation. In A. F. M. Smith and P. R. Freeman (Eds.), *Aspects of uncertainty*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pp. 363–386. Chichester: Wiley.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*(4), 437–447.

Yu, K., P. Van Kerm, and J. Zhang (2005). Bayesian Quantile Regression: An application to the Wage distribution in 1990s Britain. *Sankhya: Special Issue on Quantile regression and related methods 67*, 359–377.

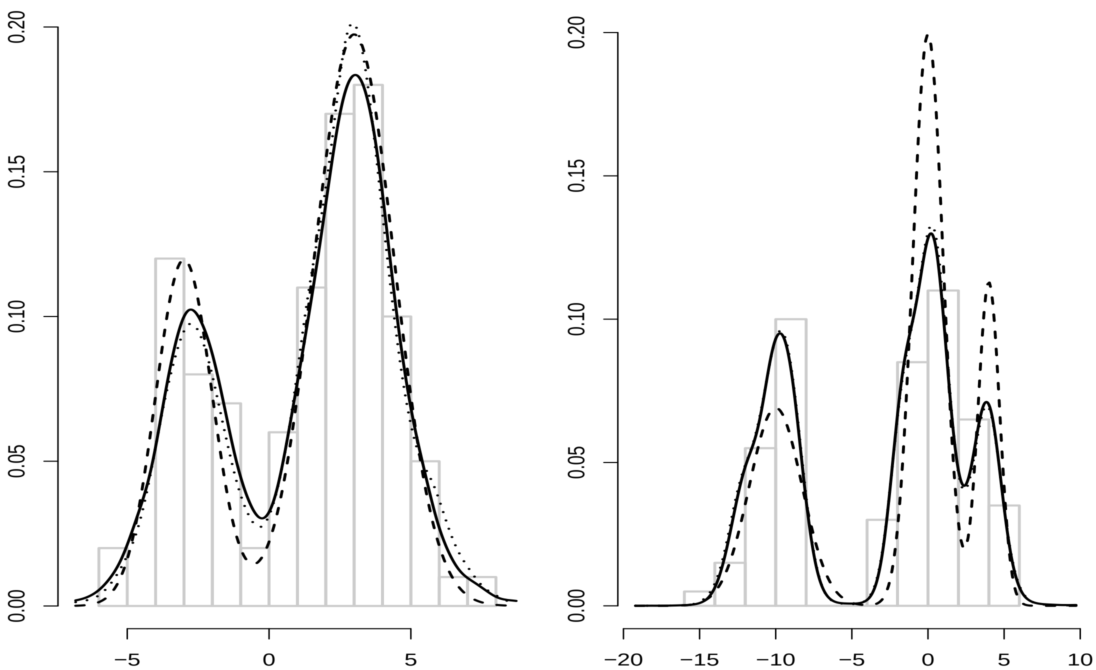

Figure 1: *Histograms and posterior predictive densities for some simulation examples. The left panel corresponds to case (iv) and the right panel depicts case (v). True simulation densities are shown as dashed lines. Prediction using prior (9) are represented as solid lines and prior (11) as dotted lines. See the text for an explanation.*

Table 1: Some posterior summaries for the regression example. Monte Carlo standard errors are indicated within parentheses.

Parameter	Posterior Mean	95% HPD
M	0.929 (0.0268)	$(0.209, 1.830)$
β	0.611 (0.0056)	$(0.174, 0.990)$
σ^2	0.030 (0.0002)	$(0.020, 0.040)$
τ	0.652 (0.0222)	$(0.102, 1.654)$
θ	0.574 (0.0162)	$(0.164, 0.999)$

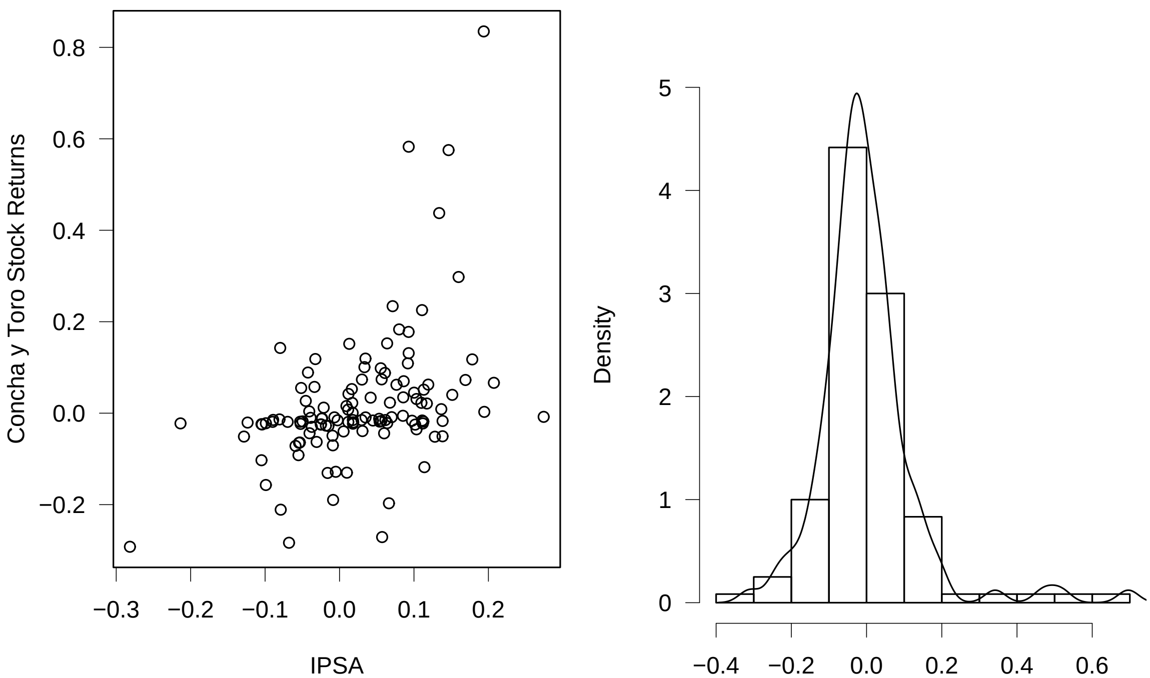

Figure 2: *Left panel shows the scatter plot of the Concha y Toro versus the IPSA index. Right panel shows the regression residuals for a standard least squares fit. Overlaid on top is a kernel density estimator.*

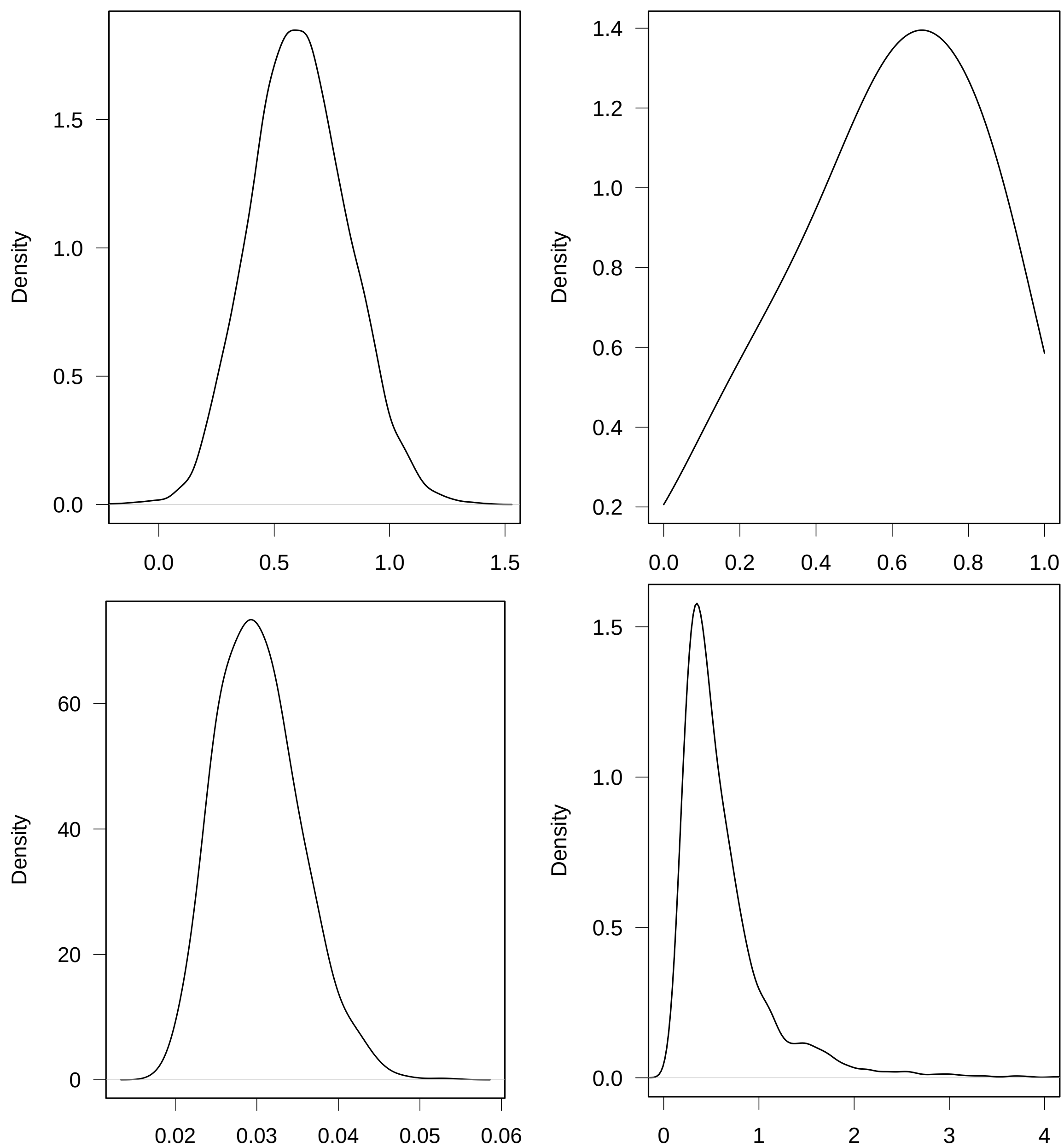

Figure 3: *For the regression dataset, the left upper panel shows the marginal posterior density of* β*. The right upper panel shows the conditional marginal posterior* $p(\theta|\theta \neq 1/2, \boldsymbol{y})$*. The bottom panels show, from left to right, the marginal posterior densities of* σ^2 *and* τ*.*

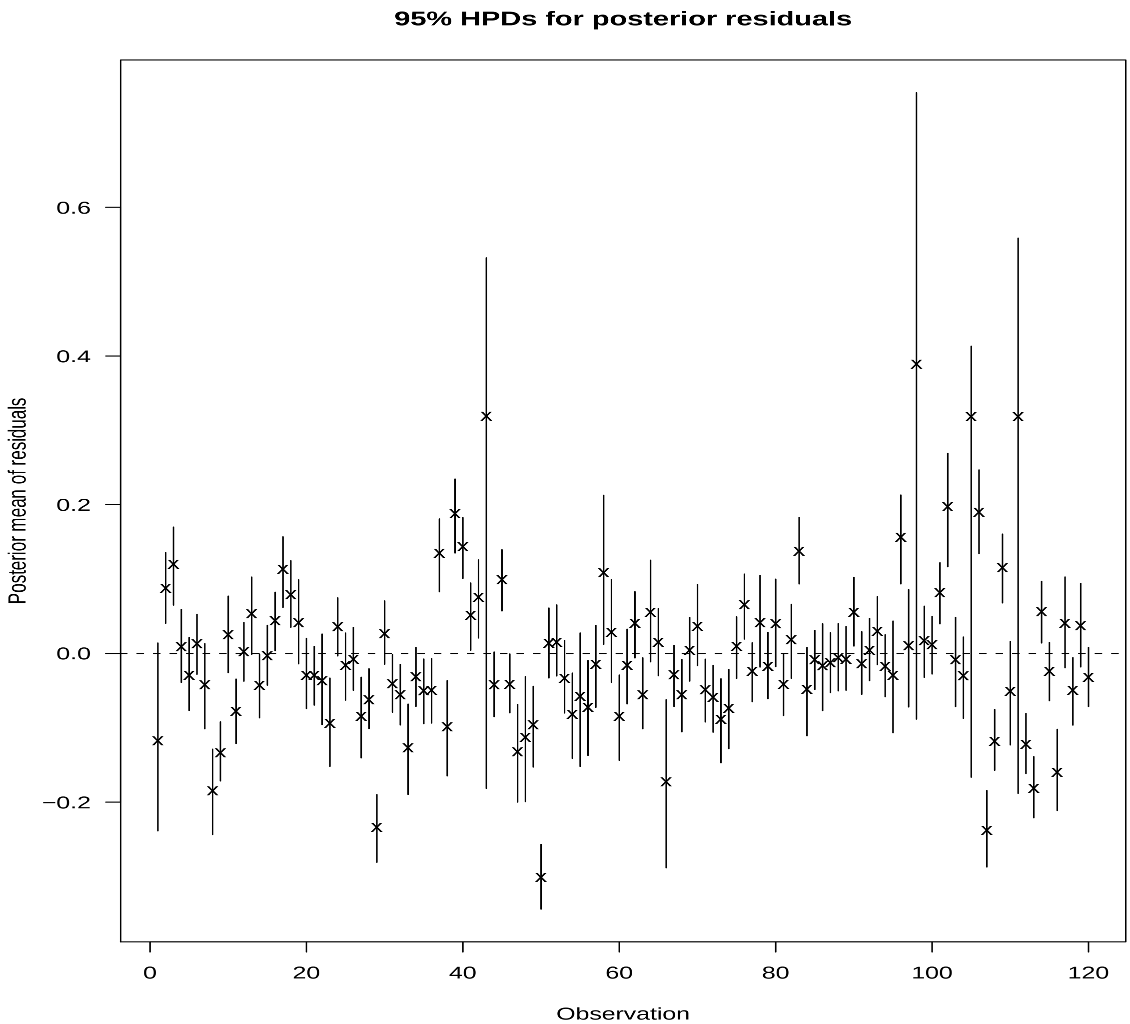

Figure 4: *95% posterior HPDs for regression residuals corresponding to all* 120 *data points.*

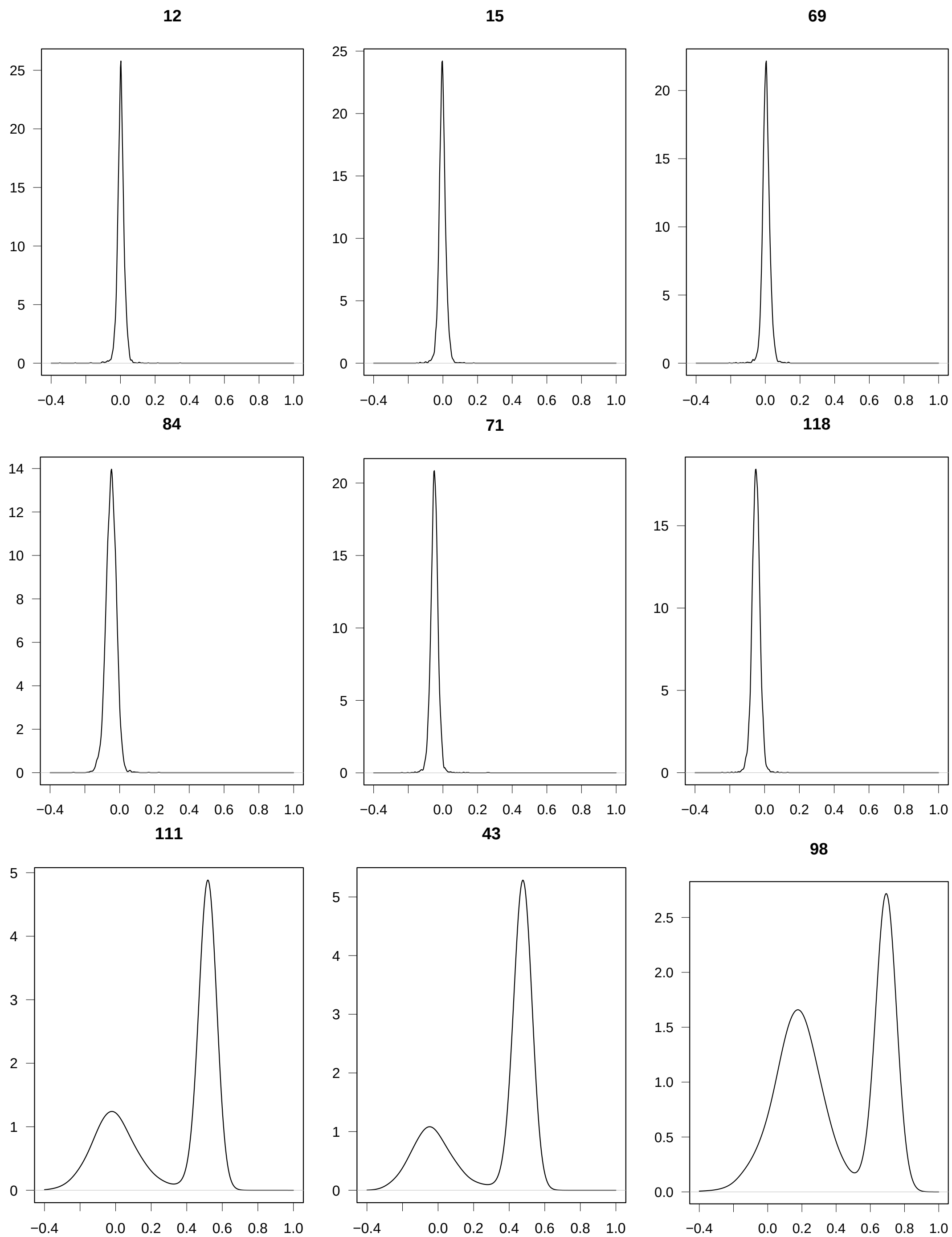

Figure 5: *Posterior distributions for some selected regression residuals.*