CHAPTER 4

Bayesian Multivariate Skewed Regression Modelling With an Application to Firm Size

4.1 Introduction

In this chapter, we study the application of Bayesian multivariate linear regression models, where the errors have skewed distributions belonging to one of two parametric classes. In particular we apply the multivariate skewed distributions as defined by Ferreira and Steel (2003), henceforth denoted by FS, and by Sahu et al. (forthcoming), henceforth denoted by SDB. The resulting regression models are employed in a study on firm size. In FS we introduced a novel method for the generation of multivariate skewed distributions. An p-dimensional skewed distribution is defined via an affine linear transformation of independent univariate variables, each with a possibly skewed distribution. As is shown in FS, this method generates a very general class of distributions. For the distribution of the univariate components involved in the transformation we suggested the method defined in Fernández and Steel (1998). We note however that other choices could have been made. The multivariate skewed distributions defined in this fashion share a number of interesting characteristics, including: direct analytical form of the probability density function (pdf), ease of moment calculation, analytical form of Mardia's measure of skewness in most cases, absence of restrictions on mean and covariance structure due to skewness, freedom from conditioning arguments, thus not involving cumulative distribution functions (cdf's) and freedom from the particular choice of coordinate axes. The main disadvantage of this class of distributions is that, in general, it is not closed under marginalisation or conditioning.

The second class of multivariate skewed distributions that will be studied in this chapter is the one developed in SDB. Using a hidden truncation model (Arnold and Beaver 2000) and conditioning on as many unobserved quantities as variables, SDB extend the work of Azzalini and Dalla Valle (1996) and Branco and Dey (2001), and develop a more general class of skewed distribution. This class of distributions is among the most general within the hidden truncation modelling framework. It is closely related to the class of elliptical distributions and it shares some of the latter properties, such as closedness under marginalisation and conditioning. The evaluation of the pdf of these multivariate skewed distributions requires the calculation of an p-dimensional cdf, which can be problematic for high dimensions and/or for certain distributions. Further, it imposes that the skewness of the distribution is introduced along the coordinate axes, consequently restricting the flexibility of the class.

We apply both classes of multivariate skewed distributions in a Bayesian linear regression setup. We compare the methodologies using skewed and fat-tailed distributions, namely the skew-Student distribution as defined by each method. We also analyse the models without fat tails, using skew-Normal distributions. Finally, we compare these alternatives with the symmetric ones: Student and Normal. The prior distribution is always chosen to be proper and, for the common parameters, equal for both classes of distributions. Formal model comparison is carried out using Bayes factors.

We apply the Bayesian regression models to a study of the distribution of firm size. Using data for three hundred publicly traded companies we evaluate the validity of common economic hypotheses, such as the suitability of the law of proportionate effects (Gibrat 1931). For all companies, data is available at two points in time: 1980 and 1990, permitting the study not only of the size distribution of the companies, but also of growth in the 1980's. We also examine the influence of research and development effort and investment on the distribution of the quantities of interest.

The remainder of this chapter is organised into four sections. In Section 4.2 we outline the two classes of multivariate skewed distributions. We introduce the Bayesian multivariate regression models in Section 4.3. Section 4.4 is devoted to the analysis of the firm size application. Finally, we offer a brief discussion in Section 4.5.

4.2 Multivariate skewed distributions

This section provides a brief review of the two classes of multivariate skewed distributions that are going to be studied in this chapter. Further details are available from the respective references. A number of other classes of multivariate skewed distributions is available in the literature and we refer the interested reader to the first part of the current edition. Our particular choice is inspired by the facts that both classes are quite general and have separately been analysed in a Bayesian framework, similar to the one here.

In the sequel, we will apply the notation FS and SDB as prefixes to the skewed distributions, dropping the term "skew" (e.g., the skew-Normal distribution as defined in Ferreira and Steel (2003) will be denoted as FS-Normal).

4.2.1 FS skewed distributions

In FS the authors introduce a general method for the construction of multivariate skewed distributions, based on affine linear transformations of univariate variables with skewed distribution. Let $p \in N_+$ be the dimension of the random variable $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)' \in \Re^p$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)' \in \Re^p_+$. Also, let $\mathbf{f} = (f_1(\cdot), \ldots, f_p(\cdot))'$ denote a vector of p unimodal and symmetric univariate pdf's. The distribution of $\boldsymbol{\epsilon}$ is a multivariate skewed distribution with independent components where, for $j = 1, \ldots, p$, the pdf of ϵ_j is $p(\epsilon_j|\gamma_j, f_j)$, which corresponds to a skewed version of the distribution with pdf $f_j(\cdot)$.
Following an affine transformation, given a vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and a non-singular matrix $A \in R^{p \times p}$, the variable $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)' \in R^p$, defined as

$$\boldsymbol{\eta} = A'\boldsymbol{\epsilon} + \boldsymbol{\mu} \tag{4.1}$$

has a general multivariate skewed distribution, with parameters $\boldsymbol{\mu}, A, \boldsymbol{\gamma}$ and denoted by $FS(\boldsymbol{\mu}, A, \boldsymbol{\gamma}, \mathbf{f})$. The pdf for $\boldsymbol{\eta}$ is then simply given by,

$$p(\boldsymbol{\eta}|\boldsymbol{\mu}, A, \boldsymbol{\gamma}, \mathbf{f}) = ||A||^{-1} \prod_{j=1}^{p} p[(\boldsymbol{\eta} - \boldsymbol{\mu})' A^{-1}_{\cdot j} | \gamma_j, f_j], \tag{4.2}$$

where, $A^{-1}_{\cdot j}$ denotes the j-th column of A^{-1}, $||A||$ denotes the absolute value of the determinant of A, and $p(\cdot|\gamma_j, f_j)$ is the pdf corresponding to the univariate skewed distribution.
FS shows that, in contrast to the elliptical distribution case, knowledge of $A'A$ is not sufficient. Further, a decomposition of the nonsingular matrix $A = OU$ is applied, where O is an $p \times p$ orthogonal matrix and U is an $p \times p$ upper triangular matrix with strictly positive diagonal elements. Straightforward manipulation shows that $A'A = U'U$.
The skewed version of the symmetric pdf $f_j(\cdot)$ can be obtained using a number of different methods. In FS, the authors generate univariate skewed distributions using the method proposed by Fernández and Steel (1998). If $f_j(\cdot)$ is a univariate pdf that is symmetric around zero, decreasing in the absolute value of its argument and if $\gamma_j \in (0, \infty)$, then the latter method defines

$$p(\epsilon_j|\gamma_j, f_j) = \frac{2}{\gamma_j + \frac{1}{\gamma_j}} f_j\left(\epsilon_j \gamma_j^{-\mathrm{sign}(\epsilon_j)}\right) \tag{4.3}$$

to be the pdf of a univariate skewed distribution, where sign$(\cdot)$ is the usual sign function.
The multivariate skewed distributions generated by (4.1)-(4.3) have a number of interesting characteristics of which we highlight its validity for any vector of univariate distributions , the dependence of the existence of moments on the existence of moments of the univariate distributions alone,

and the possibility of unrestricted modelling of mean, variance and skewness.

4.2.2 SDB skewed distributions

In SDB the authors introduce a novel method for the introduction of skewness into elliptical distributions.
Let Σ denote a $p \times p$ covariance matrix and $\boldsymbol{\mu} \in \Re^p$. Then, a continuous elliptical distribution of the p-dimensional vector $\boldsymbol{\epsilon}$ can be defined by the pdf

$$f(\boldsymbol{\epsilon}|\boldsymbol{\mu}, \Sigma, g^{(p)}) = |\Sigma|^{-\frac{1}{2}} g^{(k)} \left[(\boldsymbol{\epsilon} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{\epsilon} - \boldsymbol{\mu})\right], \tag{4.4}$$

where $g^{(p)}(\cdot)$ is a function from $\Re_+$ to $\Re_+$ given by

$$g^{(p)}(x) = \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \frac{g(x;p)}{\int_0^\infty r^{\frac{p}{2}-1} g(r;p) dr} \tag{4.5}$$

with $g(x;p)$ a non-increasing, function from $\Re_+$ to $\Re_+$ such that the integral in (4.5) exists. Now let

$$\boldsymbol{\varphi} = (\boldsymbol{\psi}, \boldsymbol{\epsilon})', \ \boldsymbol{\mu} = (\boldsymbol{\mu}^*, \mathbf{0}_p)' \text{ and } \Sigma = \begin{pmatrix} \Sigma^* & 0 \\ 0 & I_p \end{pmatrix},$$

where $\boldsymbol{\psi}, \boldsymbol{\epsilon}$ and $\boldsymbol{\mu}^*$ are in $\Re^p$, $\mathbf{0}_p$ is the p-dimensional zero vector, Σ^* is an $p \times p$ covariance matrix and I_p denotes the p-dimensional identity matrix. Further, let $\boldsymbol{\varphi}$ have pdf $f(\boldsymbol{\varphi}|\boldsymbol{\mu}, \Sigma, g^{(2p)})$ as in (4.4). By defining

$$\boldsymbol{\eta} = D\boldsymbol{\epsilon} + \boldsymbol{\psi},$$

where $D = diag(\boldsymbol{\delta})$, $\boldsymbol{\delta} \in \Re^p$, the random variable $\boldsymbol{\eta}|\boldsymbol{\epsilon} > \mathbf{0}_p$ has a multivariate skewed distribution as defined by SDB, denoted by $SDB(\boldsymbol{\mu}, \Sigma, \boldsymbol{\delta}, g^{(p)})$. The conditional pdf of $\boldsymbol{\eta}|\boldsymbol{\epsilon} > \mathbf{0}_p$ is given by

$$p(\boldsymbol{\eta}|\boldsymbol{\mu}, \Sigma, \boldsymbol{\delta}, g^{(p)}) = 2^p f(\boldsymbol{\eta}|\boldsymbol{\mu}, \Sigma + D^2, g^{(p)}) \times$$
$$P\left[\mathbf{v} > \mathbf{0}_m \middle| D(\Sigma + D^2)^{-1}(\boldsymbol{\eta} - \boldsymbol{\mu}), I_p - D(\Sigma + D^2)^{-1} D, g_\alpha^{(m)}\right], \tag{4.6}$$

with $v \in \Re^p$ and $P(\cdot|\boldsymbol{\mu}, \Sigma, g^{(p)})$ the probability function corresponding to the pdf (4.4),

$$g_a^{(p)}(x) = \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}} \frac{g(a + x; 2p)}{\int_0^\infty r^{\frac{p}{2}-1} g(a + r; 2p) dr}, \tag{4.7}$$

and finally $\alpha = (\boldsymbol{\eta} - \boldsymbol{\mu})'(\Sigma + D^2)^{-1}(\boldsymbol{\eta} - \boldsymbol{\mu})$.
The class of distributions described by (4.6) has the property of being closed under marginalisation and conditioning. The main practical problem in using these distributions is that the calculation of the cdf required in (4.6) can be complicated, especially in higher dimensions and for certain distributions. SBD remark that an MCMC sampler does not require evaluating this cdf, but it does necessitate drawings from truncated multivariate distributions which can be computationally difficult.

4.3 Regression models

In the sequel, we assume that n observations from an unknown underlying process are available, each of which is given as a pair $(\mathbf{y}_i, \mathbf{x}_i)$, $i = 1, \dots, n$. For each i, $\mathbf{y}_i \in \Re^p$ represents the variable of interest and $\mathbf{x}_i \in \Re^k$ is a vector of covariates. Throughout, we condition on $\mathbf{x}_i$ without explicit mention in the text. The observations are grouped in $X \in \Re^{n \times k}$ and $Y \in \Re^{n \times p}$, with the ith row corresponding to observation i, $i = 1, \dots, n$.
We assume that the process generating the variable of interest can be described by independent sampling for $i = 1, \dots, n$ from the linear regression model

$$\mathbf{y}_i = \lambda_i^{-1/2} \boldsymbol{\eta}_i + B' \mathbf{x_i}, \tag{4.8}$$

where $\lambda_i \in \Re_+$ has some distribution parameterised by ν, B is a $k \times p$ matrix of real coefficients, and $\boldsymbol{\eta}_i \in \Re^p$ has a distribution with a specific form.
In this chapter we are going to assume that $\boldsymbol{\eta}_i$ follows one of two alternatives: $FS(\mathbf{0}_p, A, \boldsymbol{\gamma}, \mathbf{f})$ or $SDB(\mathbf{0}_p, \Sigma, \boldsymbol{\delta}, g^{(p)})$, where $\mathbf{f}$ is the p-dimensional vector with all components equal to $\phi(\cdot)$, the standard normal pdf and

$$g^{(p)}(u) = \tfrac{\exp\{-u/2\}}{(2\pi)^{p/2}},$$

corresponding to the skew-Normal distributions for both classes.
In a similar manner as when dealing with mixtures of normals, imposing a specific (mixture) distribution on λ_i extends (4.8) to a substantially larger class of distributions. As examples, imposing a Dirac prior on $\lambda_i = 1$ retrieves the skew-Normal distributions, while if λ_i has a Gamma distribution with both precision and shape parameters equal to $\nu/2$, then y_i has a skew-Student distribution with ν degrees of freedom (df). The present chapter will assume one of the two alternatives above. Assuming that $\boldsymbol{\gamma} = \mathbf{1}_p$ or that $\boldsymbol{\delta} = \mathbf{0}_p$ generates the symmetric special cases of the distributions, obviously coinciding for both classes. In summary, in the present chapter we will consider one pair of models that can model both skewness and fat tails (FS-Student and SDB-Student), one pair of models that can model skewness alone (FS-Normal and SDB-Normal), one model than can model only fat tails (Student) and, finally one model that can not model skewness nor fat tails (Normal).

4.3.1 Prior distributions

The definition of the Bayesian models is completed with the specification of the prior distribution of the unknown parameters. In order to compare FS and SDB skewed distributions, we specify common prior distributions whenever possible. For FS we assume that the prior is of the form

$$P_{B,A,\boldsymbol{\gamma},\nu} = P_{B|A} P_A P_{\boldsymbol{\gamma}} P_\nu.$$

In a similar manner we assume that for the SDB models, the prior structure is

$$P_{B,\Sigma,\boldsymbol{\delta},\nu} = P_{B|\Sigma} P_{\Sigma} P_{\boldsymbol{\delta}} P_{\nu}.$$

Using the decomposition of a nonsingular matrix $A = OU$, the decomposition of a covariance matrix $\Sigma = U'U$ and the fact that $A'A = U'U = \Sigma$, we note that O in the FS parameterisation has no counterpart in the SDB model. We impose a prior on U for both models through an inverted Wishart distribution on Σ. The latter has as parameters $q > p - 1$ and Q, an $p \times p$ covariance matrix, and its pdf is given by

$$p(\Sigma|Q,q) \propto |Q|^{\frac{q}{2}} |\Sigma|^{-\frac{p+q+1}{2}} \exp\left(-\tfrac{1}{2} \text{ tr } \Sigma^{-1} Q\right),$$

with tr denoting the trace operation.
As in FS, O has an distribution on $\mathcal{O}^p$ that is invariant to linear orthogonal transformations, where $\mathcal{O}^p$ is a set of orthogonal matrices that ensures identifiability (see Appendix B of FS for details).
The prior on B is set conditional on Σ, and is taken to be a matricvariate Normal with parameters B_0, Σ and M, with B_0 a $k \times p$ matrix of real components and M a $k \times k$ covariance matrix. Then, the pdf of the prior distribution on B is

$$p(B|B_0, \Sigma \otimes M) \propto |M|^{-\frac{p}{2}} |\Sigma|^{-\frac{k}{2}} \exp\left[-\tfrac{1}{2} \text{ tr } \Sigma^{-1}(B - B_0)' M^{-1} (B - B0)\right].$$

For the models with heavy tails, the parameter ν controls the df. An exponential prior on ν, with hyperparameter d and restricted to $(3, \infty)$ is imposed. This prior does not allow for extremely heavy tails, as it imposes the existence of the first three moments of the distributions. This was not seen to be too restrictive for the applications in Section 4.4.
The priors on $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ used here are the ones suggested by FS and SDB. We assume that $\boldsymbol{\gamma} \in \Re_+^p$ has prior distribution with pdf

$$p(\boldsymbol{\gamma}) = \prod_{j=1}^{p} (2\pi s)^{-1/2} \Big\{ \gamma_j^{-2} \exp\Big[-\tfrac{(1-\gamma_j^{-1})^2}{2s}\Big] I_{(0,1)}(\gamma_j) + \\ + \exp\Big[-\tfrac{(\gamma_j - 1)^2}{2s}\Big] I_{[1,\infty)}(\gamma_j) \Big\},$$

imposing that for any two constants such that $1 < \gamma_a < \gamma_b$, we have that $P[\gamma_j \in (\gamma_a, \gamma_b)] = P\left[\gamma_j \in \left(\frac{1}{\gamma_b}, \frac{1}{\gamma_a}\right)\right]$, which is inspired by symmetry considerations in the FS model. The vector $\boldsymbol{\delta}$ is assumed to have a Normal distribution with zero mean and covariance Γ.

4.3.2 Numerical implementation

In order to conduct inference numerical methods have to be adopted. In particular, we construct Markov chain Monte Carlo methods (MCMC). For both classes of models, we used hybrid samplers composed of Metropolis-Hastings steps for all parameters except ν for both classes and for λ_i, $i = 1, \ldots, n$ for the FS models, where Gibbs steps are easy to implement. For

the FS models, the sampler is close to the one in FS. For the SDB models, we use a sampler similar to the one suggested in SDB.
Details of the MCMC samplers are omitted here on grounds of brevity but, a description of the samplers as well as our Matlab implementation of the Bayesian models can be obtained from the authors upon request.

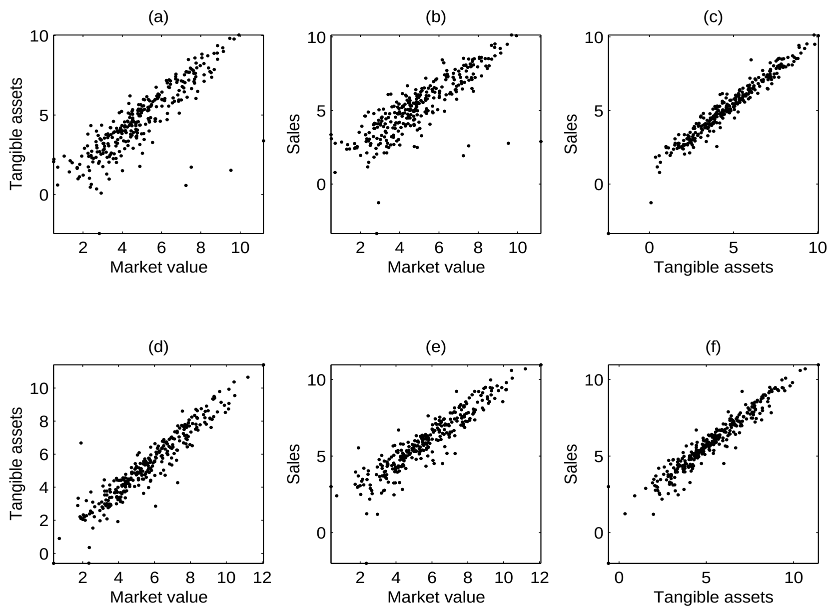
4.4 Application to firm size

The relative sizes of firms, their dynamics, and their relation to the firms' particular characteristics is an important problem in economics and the focus of substantial research effort. A large review of studies on firm size can be found in Ahn (2001). In this section we perform a study of the size distribution of a cohort of three hundred companies, in the manufacturing sector, in the years 1980 and 1990.
The set of firms that we study here originates from a larger cohort, created and maintained by Bronwyn H. Hall, containing information on about 3000 publicly traded companies in the U.S. manufacturing sector. The original dataset contains the records of an unbalanced panel from 1951-1991. Further information about the dataset and panel can be found in Hall (1993a,b).
From the complete dataset we randomly selected a cohort of three hundred firms for which data is available for both 1980 and 1990. We are interested in studying the overall distribution of three measures of firm size (market value, tangible assets and sales), under the influence of two cofactors (research and development (R&D) effort, and investment). The quantifiers of firm size are expressed as the logarithm of the original values, measured in millions of dollars. The covariates R&D and Investment are measured as the ratio between quantity spent and total assets, both standardised to have mean zero and unit variance. In addition to our predictors we also included a constant term.
We stress that the cohort of firms that we study is not a random sample from the set of all of U.S. manufacturing firms. All firms in this study were publicly traded during the period 1980-1990, implying that small firms, less likely to be quoted in the stock market, are underrepresented in the cohort. Further, we impose that the firms have survived the period of study, implying that our cohort of firms does not contain failing firms. See Geroski *et al.* (2003, p. 51) for a discussion of how this could affect the results.
In this chapter we analyse the joint distribution of the firm size variables for 1980 and 1990 and their growth between 1980 and 1990. We are especially interested in testing for the presence of skewness and fat tails in the distributions. The existence of skew distributions for measures of firm size is not a novel hypothesis in economic theory. In fact, it has been suggested in many previous studies, a good review of which is provided by Sutton (1997). Gibrat (1931) introduced the law of proportionate effect,

also known as Gibrat's law, where the firm size variables are assumed to have a Lognormal distribution.

In order to investigate the presence of skewness, Figure 4.1 presents pairwise scatterplots for the firm size measures at 1980 and 1990. Particularly from the plots on the two left-most columns of the figure, some skewness is apparent. The presence of skewness is also suggested by marginal skewness measures. For the 1980 data the sample skewness is 0.34, 0.10 and -0.16 for Market value, Tangible assets and Sales, respectively; for the 1990 data the sample skewness is, in the same order, 0.24, 0.17 and -0.03.

Figure 4.1 *Pairwise scatterplots of the firm size measures at 1980 (top row) and 1990 (bottom row).*



The full definition of the Bayesian models introduced in Subsection 4.3 requires the setting of the hyperparameters. In what follows, B_0 and M, in the matricvariate prior for B are set to the $k \times p$ zero matrix and $100 I_k$, respectively. For the prior on Σ, Q is set to I_p and q is set to $p+2$ (ensuring the existence of a prior mean). The remaining hyperparameters are set as in FS and SDB. In particular, s is set to unity, $\Gamma = 100 I_p$ and $d = 0.1$. These settings generate a rather vague prior on the model parameters.

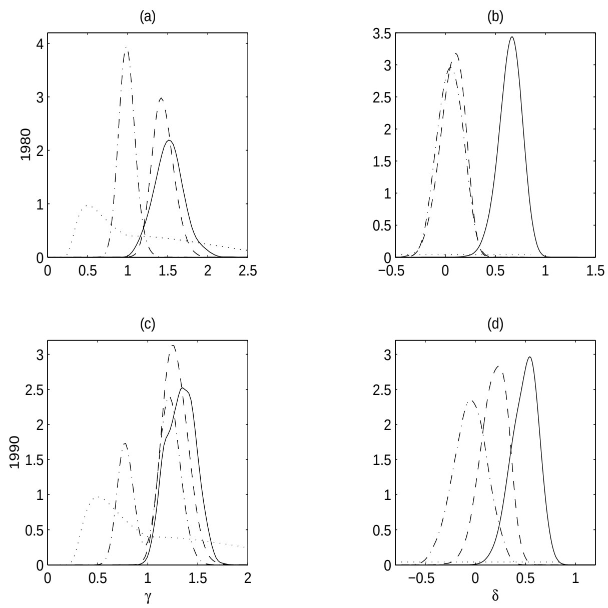
Inference for the models was always conducted using every tenth realisation from a chain of 50,000 iterations. A burn-in period of 10,000 samples preceded the collection phase. For all models, convergence was achieved early in the burn-in period.

Formal model comparison will be conducted by comparing marginal likelihoods. The p_4 estimator in Newton and Raftery (1994) is used to provide estimates of marginal likelihoods, with their δ set to 0.1.
We now provide posterior and predictive inference divided into two sections, the first for the cross-sectional studies on the distribution of firm size at 1980 and 1990 and the second for the analysis of firm growth for that period.

4.4.1 Distribution of firm size

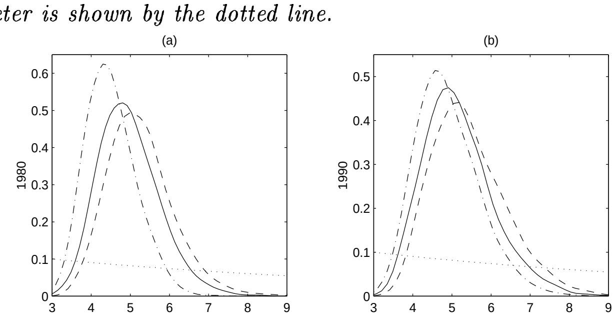
The presence of skewness in the distribution of firm size can easily be determined by examining the marginal posterior distributions of $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. For the FS models, $\boldsymbol{\gamma}$ different from the p-dimensional vector of ones indicates skewness. For the SDB models, the same holds if $\boldsymbol{\delta}$ differs from the p-dimensional zero vector. Figure 4.2 presents estimates of the posterior density of the components of $\boldsymbol{\gamma}$ (left column) and of $\boldsymbol{\delta}$ (right column), for the data relating to 1980 (upper row) and 1990 (lower row) for the skewed Student models. In each plot, the prior distribution for the parameter is also presented. For both points in time, for the FS models evidence on the presence of skewness is rather strong. The $\boldsymbol{\gamma}$ parameter of the FS-Student, estimated for the 1980 data has two components with distributions markedly centred away from unity. For 1990 the same model reveals that all components are different from unity. The presence of skewness in the SDB models is less evident, with only one component of $\boldsymbol{\delta}$ substantially different from zero.
We now examine the effect of R&D and Investment in the distribution of firm size. Table 4.1 presents summaries of the marginal posterior distributions of the regression coefficients, estimated using the FS-Student model. The magnitude of the coefficients provides evidence that the influence of covariates is restricted. Nevertheless, some conclusions can be drawn. Research and development seem to have a mostly negative effect on the firm size variables in 1980. A possible explanation is the fact that if a firm assigns a substantial amount to R & D, then its immediate sales and turnover are reduced and, as a consequence, so is its value. In contrast, Investment is seen to have a mostly positive effect on Market value and Tangible assets. Both covariates appear to have been more influential in 1980 than in 1990. The SDB model leads to similar inference on the regression coefficients.
So far, we have noted the presence of skewness in the distributions. Now we study the tails of the distributions. Figure 4.3 presents the posterior density for ν for the heavy tailed models, for 1980 (left) and 1990 (right). As is evident, heavy tails are strongly supported by the data. All models have similar estimates for the pdf of ν, with the symmetric Student model focusing on slightly lower values. This may be a consequence of modelling with an elliptical distribution which does not capture the skewness present

Figure 4.2 *Marginal posterior pdf of the components of* $\boldsymbol{\gamma}$ *(left column) and* $\boldsymbol{\delta}$ *(right column), for the data relating to 1980 (upper row) and 1990 (lower row), for the skewed Student models. The prior distribution of the parameter is shown by the dotted line.*



in the data. Heavier tails are then required to account for the inadequacy of the elliptical model.

Figure 4.3 *Marginal posterior pdf of* ν *for the FS-Student (solid line), the SDB-Student (dashed line) and the Student (dot-dashed line) models. Panel (a) plots the estimates for 1980 and (b) the estimates for 1990. The prior distribution of the parameter is shown by the dotted line.*



We conclude the analysis of the distribution of firm size with a formal

Table 4.1 *Summaries of the marginal posterior distributions of the regression coefficients, estimated using the FS-Student model*

Year	Covariate	Size var.	Mean	Std	5% quant.	95% quant.
1980	R&D	M. value	0.02	0.13	-0.18	0.24
		T. assets	-0.36	0.13	-0.57	-0.14
		Sales	-0.30	0.12	-0.50	-0.10
	Investment	M. value	0.32	0.10	0.15	0.50
		T. assets	0.13	0.11	-0.05	0.31
		Sales	0.04	0.10	-0.12	0.22
1990	R&D	M. value	-0.03	0.12	-0.22	0.16
		T. assets	-0.05	0.11	-0.22	0.14
		Sales	0.01	0.10	-0.16	0.18
	Investment	M. value	0.28	0.27	-0.18	0.73
		T. assets	0.15	0.25	-0.26	0.58
		Sales	0.19	0.21	-0.14	0.57

comparison of the different models. Table 4.2 shows the difference in log marginal likelihood between the FS-Student model and all other models. A negative value denotes an advantage for the FS-Student model. If prior model probabilities are assumed equal for all models, the exponentials of the values in Table 4.2 are posterior odds versus the FS-Student model. For both years, models that allow for heavier tails are shown to be more adequate. Also, skewed models are preferred, with the FS models getting more support than their SDB counterparts. In particular, the FS-Normal models dominate the SDB-Normal ones. In summary, the results from Table 4.2 suggest strong data support for skewness and fat tails. Thus, the symmetry and Normal tail behaviour assumptions (the law of proportionate effects) do not hold for the set of companies that we analyse in this chapter.

4.4.2 Analysis of firm growth

The second part of our application is devoted to the study of firm growth. We analyse the growth of Market value and of Tangible assets between 1980 and 1990. We define $\mathbf{y}_i = (y_i^1, y_i^2)'$, $i = 1, \ldots, n$, where for observation i, y_i^1 represents the difference in the logarithm of Market value between 1990 and 1980 and, equivalently, y_i^2 denotes the difference in the logarithm of Tangible assets in the same period. We also want to assess the impact of the

Table 4.2 *Distribution of firm size - difference in log marginal likelihood between the FS-Student model and all other models. Negative values denote advantage for the FS-Student model.*

	Student			Normal		
	FS	SDB	Elliptical	FS	SDB	Elliptical
1980	0	-6.2	-12.7	-56.0	-95.2	-117.3
1990	0	-3.3	-11.0	-46.5	-49.7	-58.4

1980 effort in R&D and Investment on the growth of the firms. Therefore, we use these measures as covariates. We also include a constant term in our analysis.

Table 4.3 presents the difference in log marginal likelihood between the different models and the one for the FS-Student model with all covariates. We also include the models with only the constant term. As in the previous subsection, the skewed models are supported by the data. Also, heavy-tailed models are strongly favoured. The skewed Student models provide equivalent alternatives for the distributions. However, as in the previous subsection, there is a marked difference in terms of the skewed Normal models in favor to the FS ones, especially when covariate information is used. This ranking of the models is not affected by the inclusion of the R&D and Investment cofactors, even though the covariates are assessed to be quite important. For all models, the covariates prove to be quite important.

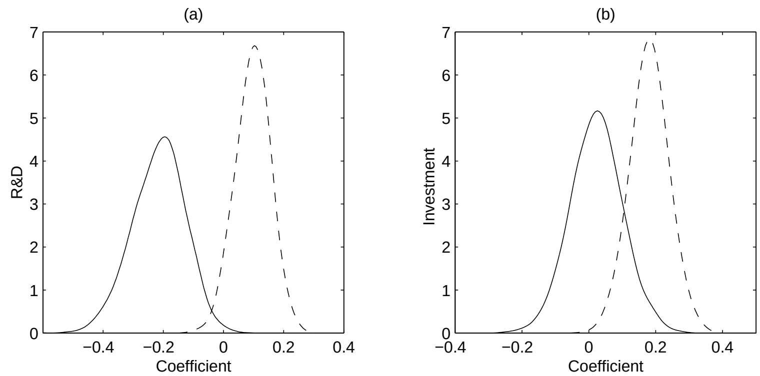
Table 4.3 *Distribution of firm growth - difference in log marginal likelihood between the FS-Student model with all covariates and all other models. Negative values denote advantage for the FS-Student model with covariates.*

	Student			Normal		
	FS	SDB	Elliptical	FS	SDB	Elliptical
With cofactors	0	0.4	-25.6	-49.3	-68.4	-87.8
Without	-32.2	-33.0	-60.6	-75.6	-77.2	-111.5

The influence of the covariates on firm growth can be assessed by the plots in Figure 4.4, where posterior pdf's for the coefficients of the covariates are presented. These estimates were obtained from the FS-Student models. Figure 4.4 (a) shows the distinct effect of R&D on the two measures of firm growth. R&D effort in 1980 has a positive effect on Tangible asset growth and a negative effect on the growth of Market value. From Figure

4.4 (b) we realize that Investment has almost no effect on Market value and has a strong positive effect on Tangible assets. In summary, growth in Tangible assets is positively affected by both R&D and Investment, while growth in Market value is negatively affected by R&D and is not affected by Investment.

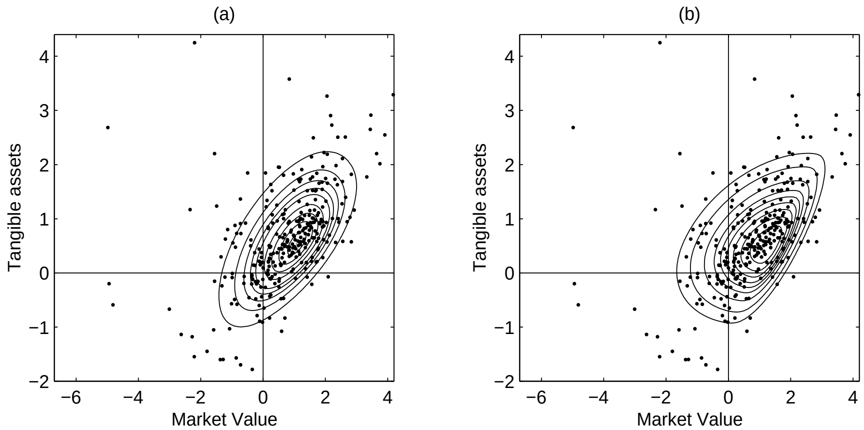
Figure 4.4 *FS-Student estimate of the marginal posterior densities of the coefficients of cofactors R&D (a) and Investment (b). The solid lines represents the estimates of the coefficient for Market value growth and the dashed lines the ones for Tangible asset growth.*



As evidenced in Table 4.3 heavy tails are supported by the data. The posterior median value of ν for the FS-Student model is 3.80 when covariates are used and 3.70 when only the constant term is used. The median values for the remaining heavy-tailed models are similar. The presence of heavy tails does not support the common assumption in economics that firm growth has a Normal distribution.

Until now, we have not graphically presented any skewed distribution. The application in Section 4.4.1, with its three covariates, is not directly suitable for visualization. However, here we analyse a bi-dimensional problem and we can provide some contour plots of the posterior predictive pdf's. Figure 4.5 presents contour plots of the posterior predictive densities, for the FS-Student model and FS-Normal with only the constant term, overplotted with the true data represented by the dots. The presence of skewness in the distribution is obvious, especially for the FS-Normal model. By fixing tail behaviour to be the one of the Normal distribution, the distribution has to be substantially more skewed than if heavier tails are allowed. This can explain the reason for the near equivalence of the FS and SDB models for the Student alternatives, and for the advantage of the FS models when only Normal tails are allowed. The majority of firms experienced growth during the ten-years period up to 1990. However, a considerable number of firms exhibited a decrease, sometimes considerable, in one or both measures. In particular, after inspecting the contours of the pdf's, we can conclude that a firm having its Market value diminishing while increasing Tangible assets is more likely than the opposite. This can be generated by a poor opinion of the firm by the market, even when Tangible assets are increasing. As these are manufacturing firms, this could be linked with the general decline in the manufacturing sector in that period.

Figure 4.5 *Data (dots) and contours of the posterior predictive density of the FS-Student (a) and FS-Normal (b) models with only the constant term as regressor*



Contour plots of the posterior predictive pdf's for the SDB models are not included due to the fact that their computation is much more demanding. Due to the pdf form in (4.6)-(4.7) the computation of the posterior predictive pdf contours would require substantial computing effort.

4.5 Discussion

In this chapter we review and compare two different alternatives for skewed distributions. The first methodology, introduced in Ferreira and Steel (2003), generates multivariate skewed distributions by using linear transformations of univariate skewed distributions. The second, proposed by Sahu et al. (forthcoming), uses a hidden truncation framework, where one unobserved component is required for each dimension.

We provide general Bayesian linear regression models that can allow for both skewness and heavy tails. FS have derived sufficient conditions for inference under general improper prior. However, as similar conditions are not available for the SDB models, we conduct inference here under a proper prior, when possible common to both models. We also analyse restricted versions of the most general models, by excluding either skewness, heavy tails or both.

The regression models are used in two econometric applications: cross-sectional studies of the distribution of a cohort of manufacturing firms, in two time periods, and the analysis of the growth of the same firms between the two periods. We also assess the relevance of R&D and Investment on the distributions. Fat tails and skewness found support in both applications.
The FS models were seen to be more suited to the applications at hand. The preference was substantially stronger for the Normal models, indicating the flexibility of the FS models. One methodological reason for this flexibility is that the FS class of distributions can model mean, covariance and skewness separately. The same is not true for the SDB class, where fixing Σ and increasing $\boldsymbol{\delta}$ in absolute value leads to a decrease in the correlation between the variables. Thus, in the SDB class of distributions, it is not possible to model simultaneously highly correlated and heavily skewed data. We feel that this is an advantage of the FS models, and according to the model comparison we performed in this article, one that can have strong practical relevance.
In addition, the models proposed in FS are much more computationally efficient, both in terms of model fitting and model comparison. The numerical methods necessary to conduct inference under the SDB models, rely on a data augmentation procedure requiring one truncated multivariate variable for each observation. The updating of these variables can be quite demanding. Model comparison based on marginal posterior probabilities is also an expensive computational issue. Estimation of these marginal probabilities can be done, at least, in two different ways. The first involves calculating multivariate cdf values, which is difficult for most common distributions, especially in high dimensions. The second is to apply Monte Carlo integration using the data augmentation procedure used in model fitting, with the augmented variables sampled from the prior. None of these procedures is required for the FS models, where the pdf of the skewed distribution has a much more explicit form which is straightforward to evaluate for most cases.

Acknowledgements The work of the first author was supported by grant SFRH BD 1399 2000 from Fundação para a Ciência e Tecnologia, Ministério para a Ciência e Tecnologia, Portugal.

References

Ahn, S. (2001). Firm dynamics and productivity growth: a review of micro evidence from OECD countries, *Technical Report ECO/WKP(2001)23*, OECD, Paris.

Arnold, B. C. and Beaver, R. J. (2000). Hidden truncation models, *Sankhya A* **62**: 23–35.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika* **83**: 715–726.

Branco, M. and Dey, D. K. (2001). A general class of multivariate skew elliptical distributions, *J. Multivariate Anal.* **79**: 99–113.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness, *J. Amer. Statist. Assoc.* **93**: 359–371.

Ferreira, J. T. A. S. and Steel, M. F. J. (2003). Bayesian multivariate regression analysis with a new class of skewed distributions, *Technical report*, IMS, University of Kent.

Geroski, P. A., S, L., Urga, G. and Walters, C. F. (2003). Are differences in firm size transitory or permanent?, *J. Appl. Econ.* **18**: 47–59.

Gibrat, R. (1931). *Les inégalités économiques; applications: aux inégalités des richesses, à la concentration des enterprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel*, Librairie du Recueil Sirey, Paris.

Hall, B. H. (1993a). Industrial research during the 1980's: Did the rate of return fall?, *Brookings Pap. Econ. Act.* **2**: 289–344.

Hall, B. H. (1993b). The stock market's valuation of r&d investment during the 1980's, *Amer. Econ. Rev.* **83**: 259–263.

Newton, M. A. and Raftery, A. E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap (with discussion), *J. Roy. Statist. Soc. B* **56**: 3–48.

Sahu, S., Dey, D. K. and Branco, D. (forthcoming). A new class of multivariate skew distributions with applications to Bayesian regression models, *Can. J. Statist.*

Sutton, J. (1997). Gibrat's legacy, *J. Econ. Lit.* **35**: 40–59.