BAYESIAN MODEL COMPARISON FOR SKEW-ELLIPTICAL LINEAR REGRESSION MODELS

By

Ignacio Vidal García

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
THE PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
STGO. DE CHILE, CHILE
AUGUST 26, 2003

THE PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTMENT OF STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Mathematics for acceptance a thesis entitled "**Bayesian Model Comparison for Skew-Elliptical Linear Regression Models**" by **Ignacio Vidal García** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: August 26, 2003

External Examiners: ______________________
Dr. Manuel Galea-Rojas

____
Dr. Víctor Salinas

Research Supervisor: ______________________
Dr. Pilar Iglesias

Examing Committee: ______________________
Dr. Reinaldo Arellano-Valle

____
Dr. Fernando Quintana

To my son Adrián.

Table of Contents

List of Tables

List of Figures

Abstract

Model comparisons within a Bayesian perspective is probably one of the topics with great impact on statistical literature in recent years. Diverse methods have been developed which are based on several points of view. One of them considers a purely inferential approach, consisting in the deriving of the Bayes factor, which is perhaps the most popular measure to Bayesian models comparing. In this thesis we develop a treatment for Bayesian model comparison for skew-elliptical regression models and our objective is concentrated on the regression linear models together with the distribution of the errors. It also includes the study of the existence of possible measurement errors in the predictor variables.

We start by establishing some results where the Bayes factor and some default Bayes factors do not work. These results are established for elliptical linear models and a class of prior distribution which generalizes the normal-chi-squared family.

Afterwards, we deal with measurement error models (MEM). We provide an expression of the Bayes factor to test the existence of measurement error in the explicative variable and present a method to compute it based on Importance Sampling and Metropolis-Hastings algorithms. Additionally, we construct measures and develop computational methods to evaluate influence of observations for the parameters of the MEM. The measures are based on the perturbation function approach combined with the Bayes factors and other divergence measures. We apply these results to problems with real data.

Finally, we compute Bayes factors to test asymmetry under skew elliptical linear regression models. In the univariate case, we study the problem of the sensitivity of the skewness parameter using the L_1-distance, and provide expressions of the Bayes factor to test skewness under some particular prior distributions. The results are evaluated through simulation problems obtaining expected results. Also, we develop methods to compute the Bayes factor to identify asymmetry in a representable skew elliptical linear regression model and we presented simulation results with multivariate skew-normal distribution for the errors. Application in stock markets are also considered.

Acknowledgements

I would like to thank my advisor Pilar Iglesias for her ideas, constant support during this investigation and for showing me the wonderful world of Bayesian Statistics. Furthermore, I am grateful to Reinaldo Arellano for his valuable suggestions and his willingness to help me at all times.

I must say thanks to the existence of the Pontificia Universidad Católica de Chile, MECESUP and CONICYT. These institutions bring the possibility to some foreign students to apply for scholarships. They will not work without the people that work on them, and that is why I would like to extend my thanks to Pilar Iglesias, Guido del Pino and Fernando Quintana for trusting me and allow me the real possibility to achieve this grade. I also thank all members from the Department of Statistics from the P.U.C. (professors and secretaries) for the good care and warm welcome that they gave me.

I also ought to mention Carlos Mora for helping me contacting Pilar. Again I must mention Pilar, because she was not only my professor, she was the fundamental person in our exit from Cuba, although she did not know us personally. I spoke in plural because I am including my friend for many years Oslando Padilla, without our conversations and mutual support, it would not have been possible to exit. To Pilar and Guido I will always be grateful for your disenfranchised help. They did not know us at that moment, so this gesture by itself speaks of their great human values.

I am grateful to my parents and brother because they were the ones who shaped me and they are responsible for what I am today. I appreciate my mother for her kindness, my father for his scientific mind, and my brother for being so extroverted.

Indeed, I am grateful to my new and small family, Diana and Dianita, for her love, affection, and emotional tranquility that allow me to do this job and carry on with my life.

In general, I am thankful to all the people that consciously helped me in life because their contributions helped me achieve this step in my career.

Santiago de Chile — Ignacio Vidal G.

July 27, 2003

Introduction

In this chapter, we introduce the main concepts that will be used along this thesis. We give the necessary conceptual bases to follow the reading of the next chapters and we describe them. We also reference the recent works in the topics that concern us, together with mentioning some pioneer works.

Bayes factors. Comparison of models within a Bayesian perspective is probably one of the topics with great impact on the statistical literature in recent years. The first steps in Bayesian models comparison was given by Jeffreys (1935, 1961), and ever since, diverse methods have been developed based on several points of view. For further information, please refer to Akaike (1973), Schwarz (1978), Aitkin (1991), Bernardo and Smith (1994), O'Hagan (1995), Kass and Raftery (1995), Berger and Pericchi (1993, 1996a), Pereira and Stern (1999) and Bernardo (1999). An extensive list with more than 100 references is presented in Berger (1998). The problem of model comparison has been approached using different methods. One of them considers a purely inferential approach and consists in the deriving of the Bayes factor (BF), which is perhaps the most popular measure used to comparing models, probably for its simplicity, interpretation and because it is an important quantity in many of the different theories on the Bayesian models comparison.

The existence of a great variety of these methods from a Bayesian point of view reflects the complexity of the problem, in spite of the interest of the Bayesian statistics community in order to find an unifier theory, necessary in any scientific theory. In this sense, the selection of models from the point of view of the Decision Theory seems to better fulfill this goal. More details on this focus are given in Bernardo and Smith (1994, Chapter 6), Bernardo (1999) and Key *et al.* (1999).

The essence of the Bayesian inference resides in extracting information of the observed data through the posterior distribution of some unknown state of the nature and this is of our interest. In the Bayesian methods for hypotheses comparison is similarly so, and in almost all these methods are necessary the calculation of the BF.

Let us suppose that we are comparing two models M_0 and M_1. Then the model M_0 would be more in agreement with the data $\mathbf{x}$ if $p\left(M_0 \left| \mathbf{x}\right.\right) > p\left(M_1 \left| \mathbf{x}\right.\right)$, that is, if

$$1 < \frac{p\left(M_0 \left| \mathbf{x}\right.\right)}{p\left(M_1 \left| \mathbf{x}\right.\right)} = \frac{p\left(\mathbf{x}\right| M_0) \, p\left(M_0\right)}{p\left(\mathbf{x}\right| M_1) \, p\left(M_1\right)},$$

where

$$p\left(\mathbf{x}\right| M_i) = \int p\left(\mathbf{x}\right| \boldsymbol{\theta}_i, M_i) \, \pi\left(\boldsymbol{\theta}_i\right| M_i) \, d\boldsymbol{\theta}_i, \qquad i = 0, 1.$$

The expression on the right involve the BF defined in the follow.

Definition 0.0.1. *Given two hypotheses H_0 and H_1 corresponding to assumptions of two alternative models, M_0 and M_1, for the data $\mathbf{x}$, the Bayes factor in favor of H_0 (and against H_1) is given by the posterior to prior odds ratio:*

$$B_{01} = \frac{p\left(\mathbf{x}\right| M_0)}{p\left(\mathbf{x}\right| M_1)} = \left\{ \frac{p\left(M_0 \left| \mathbf{x}\right.\right)}{p\left(M_1 \left| \mathbf{x}\right.\right)} \right\} \Big/ \left\{ \frac{p\left(M_0\right)}{p\left(M_1\right)} \right\}.$$

Which can also be written as

$$B_{01} = \frac{\int p_0\left(\mathbf{x}\right| \boldsymbol{\theta}_0) \, \pi_0\left(\boldsymbol{\theta}_0\right) d\boldsymbol{\theta}_0}{\int p_1\left(\mathbf{x}\right| \boldsymbol{\theta}_1) \, \pi_1\left(\boldsymbol{\theta}_1\right) d\boldsymbol{\theta}_1},$$

where $p_k(\mathbf{x}|\,\boldsymbol{\theta}_k) = p(\mathbf{x}|\,\boldsymbol{\theta}_k, M_k)$ and $\pi_k(\boldsymbol{\theta}_k) = \pi(\boldsymbol{\theta}_k|\,M_k)$ are the likelihood and a prior density of $\boldsymbol{\theta}_k$ under the model M_k $(k = 0, 1)$ respectively. Kass and Raftery (1995) present a detailed review on the use of the BF in several applied areas, using computation techniques such as asymptotic approximations and Markov Chain Monte Carlo (MCMC) methodology. These authors present also a complete list of publications related to the subject.

The prior distributions $\pi_k(\boldsymbol{\theta}_k)$ $(k = 0, 1)$ are necessary: from a point of view this is an advantage since we could include additional information to that given by the data about the values of the parameter, but it could be very difficult to specify a prior distribution when this information does not exist. On the other hand, these prior densities should be proper since the improper ones depend on an indefinite multiplicative constant, and therefore the BF , in this case, generally, would depend on indefinite constants. More details on this topic can be seen in Berger and Pericchi (1993, 1996a) and O'Hagan (1995).

Other advantages of the BF is that it does not require alternative models with the same parametric space, which is useful to compare any pair of models. The utility of the BF is not only to compare models or to select a model inside a set of possible models. In fact, the BF is involved in the problem of prediction when we adopt the model average approach as we describe below.

If we have a set of models $\mathcal{M} = \{M_i : i \in I\}$ with their respective predictive distributions $p_i(\mathbf{x}) = p(\mathbf{x}|\,M_i)$ and prior probabilities $\pi_i = \pi(M_i)$, $\sum_{i\in I}\pi_i = 1$, then the posterior probability of the model M_j is given by

$$P(M_j|\,\mathbf{x}) = \frac{p_j(\mathbf{x})\,\pi_j}{\sum_{i\in I} p_i(\mathbf{x})\,\pi_i} = \frac{1}{\sum_{i\in I} B_{ij}\frac{\pi_i}{\pi_j}},$$

Therefore, a possible model to choose for the data $\mathbf{y}$, given the previous observations

$\mathbf{x}$, is

$$p\left(\mathbf{y}|\,\mathbf{x}\right)=\sum_{i\in I}p_i\left(\mathbf{y}|\,\mathbf{x}\right)P\left(\,M_i|\,\mathbf{x}\right)$$

where

$$p_i\left(\mathbf{y}|\,\mathbf{x}\right)=\int p_i\left(\mathbf{y}|\,\mathbf{x},\boldsymbol{\theta}_i\right)p_i\left(\boldsymbol{\theta}_i|\,\mathbf{x}\right)d\boldsymbol{\theta}_i=\int p_i\left(\mathbf{y}|\,\boldsymbol{\theta}_i\right)p_i\left(\boldsymbol{\theta}_i|\,\mathbf{x}\right)d\boldsymbol{\theta}_i.$$

The last equality is obtained by assuming $\mathbf{y}\perp\!\!\!\perp\mathbf{x}|\,\boldsymbol{\theta}_i$. From $p\left(\mathbf{y}|\,\mathbf{x}\right)$ follows

$$E\left(\mathbf{y}|\,\mathbf{x}\right)=\sum_{i\in I}E\left(\mathbf{y}|\,\mathbf{x},M_i\right)P\left(\,M_i|\,\mathbf{x}\right).$$

Literature describes also some disadvantages that the BF presents (O'Hagan (1994, Chapter 7), O'Hagan (1995), Kass and Raftery (1995) and Berger and Pericchi (1993, 1996a)). For example, its high sensitivity with respect to the prior distribution even with great sample sizes. Another difficulty is that, in nested hypotheses, using noninformative prior distribution on the parameter of interest, will force the BF to favor the hypothesis H_0. The Bartlett's paradox shows this fact (see Bartlett (1957)). With the purpose of avoiding these difficulties, it has been developed other variants of the BF , such as the Fractional Bayes Factors (FBF) due to O'Hagan (1995) and the Intrinsic Bayes Factors (IBF) by Berger and Pericchi (1996a). Other points of view in the Bayesian models comparison that avoid the great influence of the prior distributions are the Conventional Prior (CP) approached due to Jeffreys (1961, Chapter 5), the Bayesian Information Criterion (BIC) derived by Schwarz (1978), the Bayesian Reference Criterion (BRC) obtained by Bernardo (1999) and the Full Bayesian Significance Test (FBST) introduced by Pereira and Stern (1999). Of course these methods have other difficulties. A wide exposition that includes the bases and motivations of the Bayesian model selection, as well as examples and comparison of some methods of model comparison is in Berger and Pericchi (2001).

Frequently it is difficult to find an analytic solution for the normalization constants of the posterior densities, and therefore, it becomes also difficult the computation of the BF. Due to this has been developed diverse computational methods in order to calculate the marginal probability of the data or, the BF directly. Among these methods are the Laplace approximation (e.g., Tierney and Kadane (1986) and Tierney *et al.* (1989)), Monte Carlo, Importance Sampling and Iterative Quadrature to calculate the predictive (see, for example, Naylor and Smith (1982, 1988), Geweke (1989), McCulloch and Rossi (1991) and Gelfand and Dey (1994)). Another group of calculation procedures is those that use simulations from the posterior distribution, among these are the Sampling Importance Resampling (SIR) referred to Rubin (1988), the Gibbs Sampling referred to Geman and Geman (1984) and the Metropolis-Hastings referred to Metropolis *et al.* (1953) and Hastings (1970). Chib (1995) and Chib and Jeliazkov (2001) use the MCMC draws from a posterior distribution to calculate the marginal likelihood. Different useful numeric methods in the Bayesian Statistic are presented and compared by Chen *et al.* (2000).

Elliptical models. Due to the great idealization of the normal model, unrealistic in many applications, in the last times, the non normal models have gained importance. However, departures from the normal model generally takes an implicit loss of the parsimony and a bigger mathematical complexity. Due to this, and to try not to lose many good properties of the normal model, diverse generalizations of the normal model have been developed, some guided to control the skewness, others to model the weight of the tails, or both at the same time.

A natural generalization from the normal model, with the goal of controlling the weight of the tails, is the class of elliptical models defined in Kelker (1970) and broadly studied in Cambanis *et al.* (1981), Fang *et al.* (1990) and Arellano-Valle (1994). The elliptical models present the advantage of including as particular cases a great variety of important models and they also present very good properties (for example, marginalization and conditionality).

Bayesian inference for normal regression models, including sensitivity analysis, model comparison and error in variables under noninformative and conjugate prior for the parametric model has received considerable attention in the last decades. From a distributional point of view the results can be extended in several directions. One is by considering a wider class of prior distributions for the parameters of the model. Another, is by considering alternative distributions for the error terms.

Usually, the results with non-conjugate priors relies heavily on MCMC methods. On the other hand, many extensions have been obtained by considering the so called dependent elliptical model, which is often used in linear regression analysis to accommodate the kurtosis of the error terms and to accommodate outliers. Bayesian inference with multivariate elliptical models was initially presented in Chu (1973). Posteriorly, Zellner (1976) used the multivariate Student-t distribution, who considered a Bayesian treatment of linear regression models under noninformative prior distributions.

The results of Zellner (1976) were extended to the case where the errors are modelled as scale mixtures of normal distributions by Jammalamadaka *et al.* (1987) and Chib *et al.* (1988) and to the entire class of multivariate elliptical distributions by Osiewalski and Steel (1993). See also, Arellano-Valle *et al.* (2000) and Branco *et al.* (2001) for connections to diagnostic and calibration problems in elliptical regression models.

Arellano-Valle *et al.* (2002a) extended the results of Osiewalski and Steel (1993) for a new class of prior distributions that generalizes the normal-chi-squared family. They showed that, for this class, the posterior analysis for the coefficients is invariant with respect to changes in the generator function under some conditions, and conjugation is achieved for ϕ.

Skew-elliptical models. With the objective of modelling the asymmetry of the data, taking as a base the normal distribution, diverse generalizations have arisen, for example the presented by Fernández and Steel (1998) and the presented by Arellano-Valle *et al.* (2001) which include the epsilon-skew-normal distribution of Mudholkar and Hutson (2000). But the skew density function defined by Azzalini (1985) has several attractive properties such as "strict inclusion" of the normal density, mathematical tractability and a wide range of the indices of skewness and kurtosis. Azzalini (1985) makes notice that if f is a symmetric p.d.f. around zero, and G is a continuous c.d.f. such that G' is symmetric p.d.f. around zero, then

$$\frac{2}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda \frac{x-\mu}{\sigma}\right)$$

is a p.d.f. for any $\lambda \in \mathbb{R}$. Where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and λ is a skewness parameter, when $\lambda = 0$ we recover the symmetric p.d.f., $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Different choices of f and G give us important special cases, e.g., the skew-normal with p.d.f. given by $\frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right)$ and denoted by $SN(\lambda, \mu, \sigma)$.

Generalizations to the multivariate distributions could be seen in Azzalini and Dalla-Valle (1996) and Azzalini and Capitanio (1999). Further extension that can simultaneously account for both skewness and heavy tails are the multivariate skew-elliptical distributions defined and studied in Branco and Dey (2001) and Genton and Loperfido (2001), where an interesting special case is the skew-t distribution with p.d.f. given by $\frac{2}{\sigma}t\left(\frac{x-\mu}{\sigma}\,|0,1,\nu\right)F_T\left(\lambda\frac{x-\mu}{\sigma}\right)$, where $T \sim t\,(0,1,\nu)$, μ is the location parameter, σ is the scale parameter, ν the degree of freedom (control the heaviness of the tail) and λ is the skewness parameter.

An unifier work on asymmetric distributions was made by Arellano-Valle *et al.* (2002b). They defined a general class of skew-distributions that include, as particular cases, the skew-elliptical distributions and also, for this class, they gave two stochastic representations and a general method for computing moments.

Measurement error models. Another generalization of standard regression models are the measurement error models (MEM) (also called errors-in-variables models). Due to practical motivations and simplicity of model, attention has been paid to the linear regression model, where commonly Y_i denote the dependent variable and $x_{i1}, \cdots, x_{ip}$, the explanatory variables that are supposedly known variables. In occasions this assumption is not valid because many real problems exist where is not possible to know the explanatory variables completely, for example, in social sciences and management sciences. In this context the denominated MEM arise.

Adcock (1877, 1878) is usually regarded as the first person specifically to consider such models. It has been written much on the topic, but a detailed and recent analysis about MEM could be found in Fuller (1987) and Cheng and Ness (1999). Carrol *et al.* (1995) concentrated on nonlinear measurement errors models.

Inference problems in MEM typically are approached via classical inference (e.g., Fuller (1987), Carrol *et al.* (1995) and Cheng and Ness (1999)). Literature related to the Bayesian methodology in MEM is less vast than in classical approach. A pioneer work, by Lindley and El-Sayyad (1968), investigates Bayesian inference for normal case. A unification of the results in Lindley and El-Sayyad (1968) with additional considerations on the prior assumptions is considered in the book by Zellner (1971).

Some other results which appeared later include Villegas (1972), Florens *et al.* (1974), Reilly and Patino-Leal (1981) and Bolfarine and Cordani (1993), among others. These more recent works emphasize obtaining posterior distributions for the regression coefficients under different assumptions which include normally distributed errors. The scarcity of such results is probably due to the fact that the analytical treatment of the Bayesian approach for MEM is not simple. This difficulty has been overcame recently by considering MCMC methods. Some very recent works in this directions are considered in Stephens and Dellaportas (1992), Dellaportas and Stephens (1995), Richardson and Gilks (1993), Richardson (1996), Muller and Roeder (1997) and Aoki *et al.* (2003).

However, the development of MEM has been slower than in other areas of statistics. Particularly very few results are considered in connection to model choice and model comparisons in spite of the abundant literature available for ordinary linear models.

Diagnostic. Another important aspect in modelling is the diagnostic of models. Classical and Bayesian diagnostic techniques for normal linear regression models have been extensively studied in the statistical literature. From classical approach, please refer to, for instance, Cook and Weisberg (1982) and Barnett and Lewis (1994). Diagnostic techniques within a Bayesian framework have been studied by Johnson and Geisser (1982, 1983, 1985), Pettit and Smith (1983), Guttman and Peña (1988) and Peng and Dey (1995), among others. These authors developed influence measures based on divergence measures between the joint (and marginal) posterior (predictive) distributions with and without a given subset of observations. Kempthorne (1986), Geisser (1987) and Carlin and Polson (1991) investigate the problem of quantifying the influence of observations in a Bayesian decision framework by examining the changes in Bayes risk under certain specified prior distributions and loss function. Most of the research has been conducted for normal linear models and using noninformative prior distributions. Extensions of the previous results to elliptical (dependent) regression models are considering in Arellano-Valle *et al.* (2000).

In MEM, the problem has been mainly approached from the classical point of view by considering normal distributions and by computing influence measures, see, e.g., Wellman and Gunst (1991), Abdullah (1995) and Kim (2000). We are not aware of any Bayesian literature on the problem of quantifying influence in MEM .

Like in the study of any branch of the science, the development of the linear regression models has gone from simpler suppositions toward other ones more complex. Many authors have dedicated great part of their works to generalize the suppositions that traditionally have been made on the linear regression models. We, in this thesis, will contribute to this development in this same sense.

Main objective. During many years it has been supposed that the errors of the observations in a linear model are normal, however in many situations this assumption is unrealistic, for example, if the errors present some values very far from the rest. In this case it could be better to take an error distribution with heavier tails that those of the normal distribution, in this sense the elliptical models provide a nice alternative. The dependent multivariate elliptical distributions can help to relax the strong supposition of the uncorrelation of errors without complicating too much the analysis and at the same time to maintain a wide class of distributions to control "atypical" observations.

The assumption of symmetrical errors is also a supposition that can be violated in the practice, then it would become necessary to have models that are able to control this characteristic. To control the asymmetry it have been defined different skew models. In particular, we will work with the skew elliptical models defined by Branco and Dey (2001) which they are able to control the skewness, heaviness of the tails and correlation. Another assumption that commonly is made it is to suppose that the predictor variable is measured without error, or rather they are measured with a negligible error, but this is not feasible to suppose in many practical cases, it is here where the models with errors in variables emerge. Different examples and applications of these models are considered in the books of Fuller (1987), Carrol *et al.* (1995) and Cheng and Ness (1999).

The main objective of this research is to select the probability model that best explains the behavior of the observed data with known covariates. The selection of the true model is almost impossible, then, what we will do is to study the topic to provide new tools and knowledge that will allow us to approach this objective. The selection model will be focused inside the class of skew-elliptical linear regression models with measurement error in the predictor variables.

The outline of Figure 1 provides a way to summarize the possible paths with the purpose of determining a specific regression model. To achieve our objective, we will calculate the BF in some of the nodes of the outline of Figure 1, and in the square 1 of the outline. Also, with the model diagnostic objective, in the square 2 will see the importance of the BF when it is used in influence measures to evaluate the sensitivity of posterior estimates when a group of observations is eliminated from the analysis.

Outline of the thesis. The thesis is organized as follows. In Chapter 1, we review

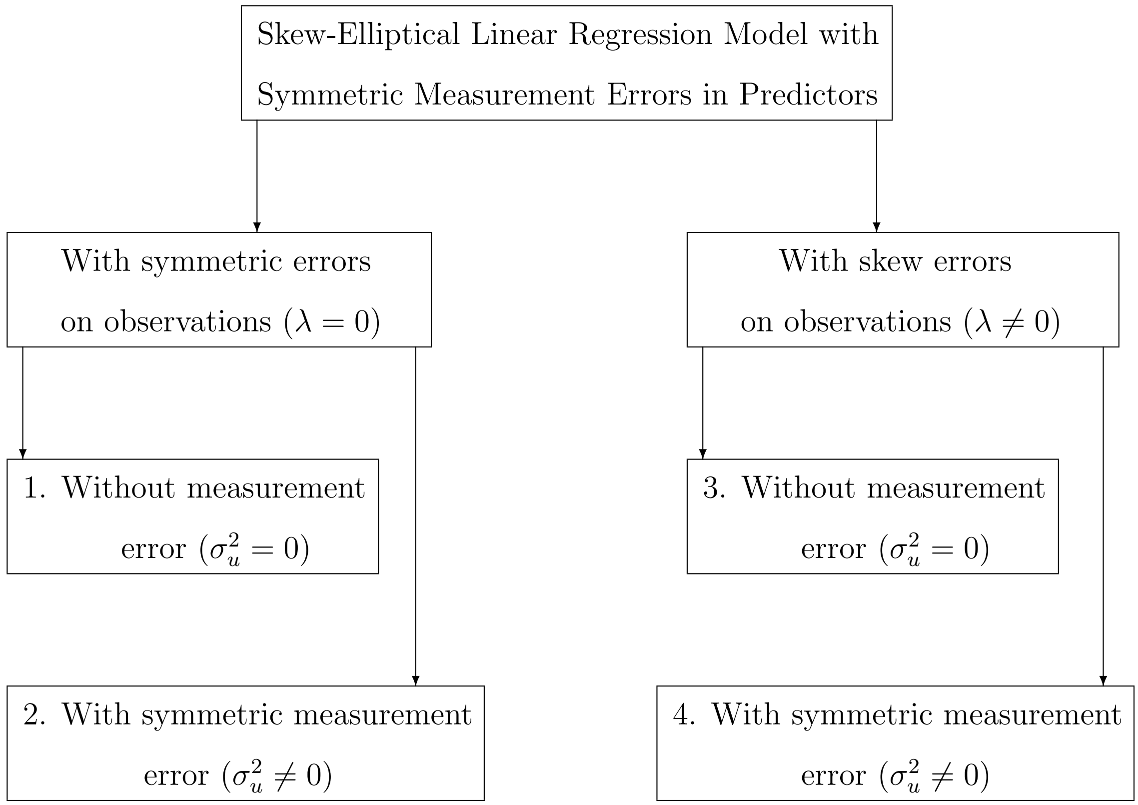

Figure 1: Outline of model selection into skew-elliptical linear regression models

some results related to elliptical distributions, present a Bayesian analysis of elliptical linear models by considering different specifications for the prior distribution and compute default Bayes factors to compare elliptical linear models including a discussion of the performance of these measures as an objective criteria for elliptical model comparison. In Chapter 2, through the Metropolis-Hastings algorithm and the Importance Sampling method, we compute the Bayes factor to test the existence of measurement error, where its behavior is evaluated through simulated and real data. Measures to evaluate influence of observations are studied in Chapter 3. We use the perturbation function to calculate some influence measures on the posterior distribution of the parameters of the MEM , and apply these measures to a problem with real data.

In Chapters 4 and 5, we compute Bayes factors to test asymmetry under skew-elliptical models. In Chapter 4, for the univariate skew-elliptical model, and in Chapter 5, for skew-elliptical linear regression model. In univariate case, we measured the sensitivity of the skewness parameter using the L_1-distance between the symmetric and asymmetric models. We also compute the Bayes factor to test skewness and present simulation results for the skew-normal and skew-t distributions obtaining expected results. Secondly, we compute the Bayes factor to identify asymmetry in a representable skew elliptical linear regression model and present simulation results with multivariate skew-normal distribution for the errors. Application in stock markets are also considered. Conclusions and final remarks are presented in Chapter 6.

Notations. Through this thesis, we employ the usual symbols $\|\|$ to denote Euclidean length of a vector, $\perp\!\!\!\perp$, independence of two random vectors and $\overset{d}{=}$, the equality in distribution. $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, denote n-dimensional Normal distribution with mean vector $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$, and $\phi_n(\mathbf{x}\,|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_n(\mathbf{x}\,|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote its respective p.d.f. and c.d.f. Also, $\mathbf{I}_n$ denote identity matrix, $\mathbb{1}_n$, n -dimensional vector of ones, $Ga(a, b)$, Gamma distribution with expected value equal to $\frac{a}{b}$ and $IGa(a, b)$, Inverted-Gamma such that if $X \sim IGa(a, b)$ then $X^{-1} \sim Ga(a, b)$. $t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, denote n-dimensional Student-t distribution with location vector $\boldsymbol{\mu}$, dispersion matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom. Also, $\mathbf{x} \leq \mathbf{y}$ means that $x_i \leq y_i$, $i = 1, \ldots, n$.

Chapter 1

Comparing Elliptical Linear Regression Models

In this chapter we consider the calculation of Bayes factors between elliptical linear models for a new class of prior distributions that generalizes the normal-chi-squared family. Arellano-Valle *et al.* (2002a) showed that for this class the posterior analysis is simple to perform under some conditions, and conjugation is achieved for ϕ.

The results of Arellano-Valle *et al.* (2002a) detected, in the posterior analysis, a invariance with respect to changes in the generator function. We use this fact to show that the Bayes factors do not depend on the generator function of an elliptical model. We also show that for some noninformative prior distributions belonging to this prior class, some default Bayes factors neither depend on the generator function.

The chapter is organized as follows. In Section 1.1, we introduce the elliptical distribution and review some results related to elliptical distributions. In Section 1.2, we present a Bayesian analysis of elliptical linear models by considering different specifications for the prior distribution. The main result of this section is due to Arellano-Valle *et al.* (2002a), and it bears to that many solutions to inference problems on elliptical regression models are equals with those obtained under normal regression models. Finally, in Section 1.3 we compute default Bayes factors to compare elliptical linear models including a discussion of the performance of these measures as an objective criteria for elliptical model comparison.

1.1 Elliptical Distributions

In this section we summarize the basic properties of elliptical distributions. Roughly speaking, a random real variable Z has a spherical distribution if $Z \stackrel{d}{=} -Z$. In this work we restrict the study to the case when the c.d.f. of Z is absolutely continuous, so that the spherical symmetry implies that Z has density given by $f_Z(z) = h(z^2)I_{\mathbb{R}}(z)$, where

$$\int_0^\infty u^{-\frac{1}{2}}h(u)du = 1. \tag{1.1.1}$$

The function h is called the density generator and we write $Z \sim S_1(h)$. For example, if $h(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u}$ then we obtain the standard normal distribution with $Z \sim N(0,1)$, and if $h(u) = c\{\nu + u\}^{-\frac{\nu+1}{2}}$, for some constant c, then the standard Student-t distribution with ν degrees of freedom, say $Z \sim t(0,1,\nu)$, follows. The class of spherical power exponential distribution can be obtained by setting $h(u) = k\exp(-\frac{1}{2}|u|^s)$, $s > 0$, for some constant k. The class of spherical distributions is a large family and includes the spherical uniform distribution, scale mixture of the spherical normal distribution, among others. Thus, this class of distributions contains symmetric distributions with heavier and lighter tails than those of the normal distribution.

We note that if $Z \sim S_1(h)$ then $T = Z^2$ has density given by

$$f_T(t) = t^{-\frac{1}{2}}h(t)I_{(0,\infty)}(t). \tag{1.1.2}$$

T is usually termed radial random variable and is denoted by $T \sim R^2(h)$. Now, let $Y = \mu + \phi^{\frac{1}{2}} Z$ then

$$f_Y(y) = \phi^{\frac{1}{2}} h((y-\mu)^2\phi) I_{\mathbb{R}}(y),$$

where h satisfy (1.1.1) and $T = \phi(Y-\mu)^2$ has density given by (1.1.2). We say that the random variable Y has elliptical distribution with parameters μ (location) and ϕ (precision), with $\mu \in \mathbb{R}$ and $\phi > 0$, and we write $Y \sim El_1(\mu, \phi^{-1}, h)$.

Multivariate distributions with univariate marginal spherical distributions can be constructed in several ways. The simplest procedure is to consider $\mathbf{Z} = (Z_1, \ldots, Z_n)^t$ a random vector with $Z_i \overset{iid}{\sim} S_1(h)$. In this case we say that $\mathbf{Z}$ has poly-spherical distribution. On the other hand, we can construct a multivariate distribution with constant density function over spheres, that is, $f_{\mathbf{Z}}(\mathbf{z}) = h^{(n)}(\|\mathbf{z}\|^2)$, where

$$\int_0^\infty \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1} h^{(n)}(u) du = 1. \tag{1.1.3}$$

We say that $\mathbf{Z}$ has a multivariate spherical distribution and we write $\mathbf{Z} \sim S_n(h^{(n)})$.

Note that $T = \|\mathbf{Z}\|^2$ has density given by

$$f_T(t) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} h^{(n)}(t) I_{(0,\infty)}(t).$$

We say here that T has radial-squared distribution with n degrees of freedom and density generator function $h^{(n)}$, and we write this as $T \sim R_n^2(h^{(n)})$. Thus, $\mathbf{Z} \sim S_n(h^{(n)})$ if and only if $T = \|\mathbf{Z}\|^2 \sim R_n^2(h^{(n)})$.

Note also that the random variable $S = T^{-1}$ has density function

$$\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} (1/s)^{\frac{n}{2}-1} h^{(n)}(1/s) I_{(0,\infty)}(s),$$

which is refereed to as the inverted radial-squared distribution with n degrees of freedom and density generator $h^{(n)}$, and we write this as $S \sim IR_n^2(h^{(n)})$.

All marginal and conditional distributions of a spherical distribution are also spherical (see for example Fang *et al.* (1990)). Any linear combination $W = \mathbf{a}^t\mathbf{Z}$ is spherically distributed too. Moreover, if $\mathbf{Z} = (Z_1, \ldots, Z_n)^t \sim S_n(h^{(n)})$ then $Z_1, \ldots, Z_n$ are independent if and only if $S_n(h^{(n)})$ is the normal spherical distribution (Kelker (1970)).

Thus, except in the normal case, the poly-spherical and multivariate spherical distributions are different classes. Both contain distributions that are long-tailed and short tailed relative to the normal distribution, but the multivariate spherical approach seems to be a more realistic model because the independence assumption is relaxed. For example, in the context of the Student-t model, we have that:

$$\mathbf{Z} \sim \ poly - t(0, 1, \nu) \leftrightarrow f_{\mathbf{Z}}(\mathbf{z}) = k\Pi_{i=1}^n \left\{1 + \frac{z_i^2}{\nu}\right\}^{-\frac{\nu+1}{2}}$$

and

$$\mathbf{Z} \sim t_n\left(\mathbf{0}, \mathbf{I}_n, \nu\right) \leftrightarrow f_{\mathbf{Z}}(\mathbf{z}) = c\left\{1 + \frac{\sum_{i=1}^n z_i^2}{\nu}\right\}^{-\frac{n+\nu}{2}}.$$

Note that the poly-Student-t distribution is not spherically symmetric. This distribution remains invariant only under change of sign. On the other hand, the multivariate Student-t distribution remains invariant under all orthogonal transformations and has Student-t univariate marginal distributions, but the elements of $\mathbf{Z}$ are not independent.

A justification from a predictivistic point of view of the dependence structure in the multivariate Student-t model is given by Arellano-Valle *et al.* (1994), see also Loschi *et al.* (2002).

Figures 1.1 and 1.2 exhibit, respectively, the shape and contours of three types of distributions and Table 1.1 contains different families of spherical generators and their corresponding radial-squared distributions.

Table 1.1: Some subclasses of n–dimensional spherical distributions ($u = \|\mathbf{z}\|^2, \mathbf{z} \in \mathbb{R}^n$).

Distribution	Generator density function	Radial-squared distribution
Normal	$(2\pi)^{-n/2}\exp\{-u/2\}$	χ^2_n
Cauchy	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}}\{1+u\}^{-\frac{(n+1)}{2}}$	$nF_{n,1}$
Student-t	$\frac{\Gamma\left(\frac{n+\nu}{2}\right)\nu^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}\{\nu+u\}^{-\frac{(n+\nu)}{2}}$, $\nu > 0$	$nF_{n,\nu}$
Generalized Student-t	$\frac{\Gamma\left(\frac{n+\nu}{2}\right)\lambda^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}\{\lambda+u\}^{-\frac{(n+\nu)}{2}}$, $\nu, \lambda > 0$	$\frac{n\lambda}{\nu}F_{n,\nu}$
Power Exponential	$\frac{\Gamma\left(\frac{n}{2}\right)\alpha}{\Gamma\left(\frac{n}{2\alpha}\right)\pi^{\frac{n}{2}}2^{\frac{n}{2\alpha}}}\exp\{-u^{\alpha}/2\}$, $\alpha > 0$	$\chi^{\frac{2}{\alpha}}_{\frac{n}{\alpha}}$
Kotz Type	$\frac{\Gamma\left(\frac{n}{2}\right)\alpha r^{\frac{2q+n}{2\alpha}}}{\Gamma\left(\frac{2q+n}{2\alpha}\right)\pi^{\frac{n}{2}}2^{\frac{2q+n}{2\alpha}}}u^q\exp\{-ru^{\alpha}/2\}$, $r, \alpha > 0,\ 2q+n > 0$	$Ga^{\frac{1}{\alpha}}\left(\frac{2q+n}{2\alpha}, \frac{r}{2}\right)$
Pearson Type II	$\frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}(1-u)^{\frac{\nu}{2}-1}$, $\nu > 0$	$Beta\left(\frac{n}{2}, \frac{\nu}{2}\right)$

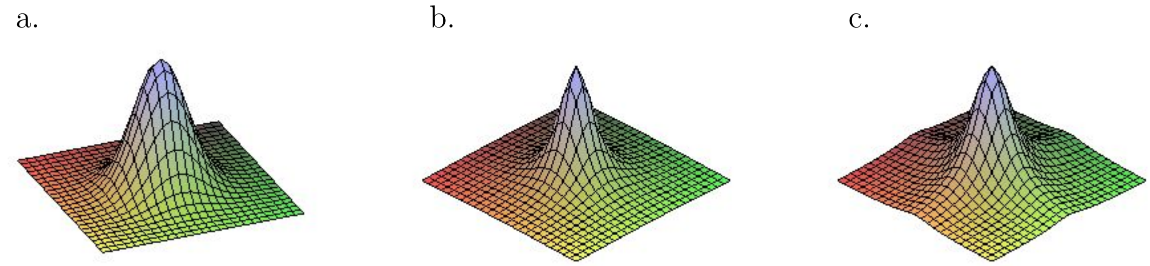

Figure 1.1: Joint densities: a. bivariate standard normal, b. bivariate standard Cauchy and c. independent product of two standard Cauchy.

In Table 1.1, c is an appropriate constant, $Ga^{\frac{1}{s}}(\alpha,\lambda)$ means that $T^s \sim Ga(\alpha,\lambda)$ (the Gamma distribution with parameters α and λ) and $\chi_\nu^{\frac{2}{s}}$ means that $T^s \sim \chi^2_\nu$, where $T = \|\mathbf{Z}\|^2$ is the radial-squared random variable. Moreover, $\phi^n(u) = (2\pi)^{-\frac{n}{2}} e^{-\frac{u}{2}}$ is the normal n-dimensional generator.

In this chapter we deal only with multivariate spherical distributions, more generally with elliptical symmetric distributions.

An $n \times 1$ random vector $\mathbf{Y}$ is said to have an elliptical distribution with parameters $\boldsymbol{\mu}$ (the location vector) and $\boldsymbol{\Sigma}$ (the dispersion matrix) of dimensions $n \times 1$ and $n \times n$, respectively, with $\boldsymbol{\Sigma}$ being positive definite ($\boldsymbol{\Sigma} > 0$), if $\mathbf{Y}$ has density function of the form

$$|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\, h^{(n)}\left[(\mathbf{y}-\boldsymbol{\mu})^t\, \boldsymbol{\Sigma}^{-1}\,(\mathbf{y}-\boldsymbol{\mu})\right],$$

where $h^{(n)}$ satisfies (1.1.3).

In this case, we write $Y \sim El_n\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(n)}\right)$ which is equivalent to $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}\left(\mathbf{Y}-\boldsymbol{\mu}\right) \sim El_n\left(\mathbf{0}, \mathbf{I}_n, h^{(n)}\right) = S_n\left(h^{(n)}\right)$ and therefore to $\left(\mathbf{Y}-\boldsymbol{\mu}\right)^t \boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}-\boldsymbol{\mu}\right) = \|\mathbf{Z}\|^2 = T \sim R_n^2\left(h^{(n)}\right)$.

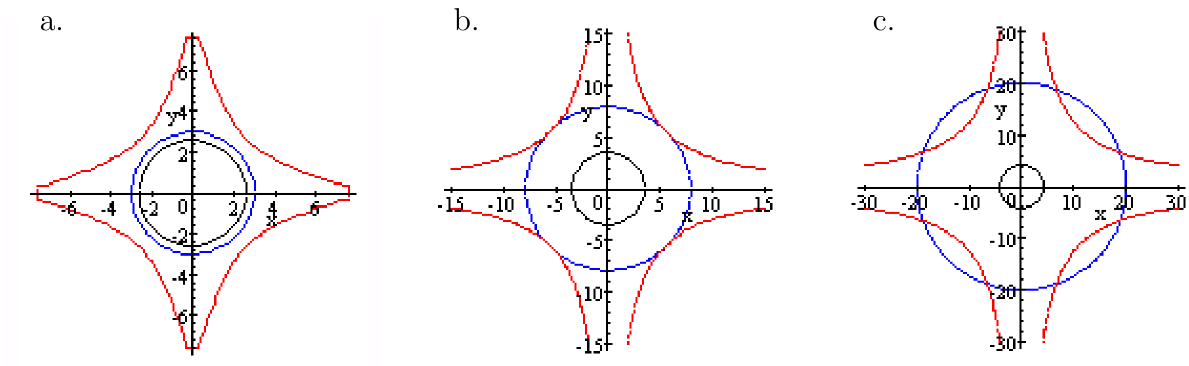

Figure 1.2: Contours of joint densities: bivariate standard normal (black line), bivariate standard Cauchy (blue line) and independent product of standard Cauchy (red line) to a height c. a. $c = (2^6\pi)^{-1}$, b. $c = (2^{10}\pi)^{-1}$ and c. $c = (2^{14}\pi)^{-1}$.

From the above results we can show also that if $\mathbb{E}\left(T\right) < \infty$, then $\mathbb{E}\left(\mathbf{Z}\right) = \mathbf{0}$ and $\mathbb{V}\left(\mathbf{Z}\right) = \alpha_h \mathbf{I}_n$, that is,

$$\mathbb{E}\left(\mathbf{Y}\right) = \boldsymbol{\mu} \quad \text{and} \quad \mathbb{V}\left(\mathbf{Y}\right) = \alpha_h \boldsymbol{\Sigma},$$

where $\alpha_h = \mathbb{E}\left(n^{-1}T\right)$ is the variance parameter associated with the density generator $h^{(n)}$.

Now, let $\mathbf{Z}_k$ be a $k-$dimensional $(1 \leq k < n)$ random sub-vector of $\mathbf{Z} \sim S_n\left(h^{(n)}\right)$. Then, $\mathbf{Z}_k \sim S_k\left(h^{(k)}\right)$ and has density function $h^{(k)}\left(\|\mathbf{z}_k\|^2\right)$, $\mathbf{z}_k \in \mathbb{R}^k$, where

$$h^{(k)}\left(u\right) = \int_0^\infty \frac{\pi^{\frac{n-k}{2}}}{\Gamma\left(\frac{n-k}{2}\right)} v^{\frac{n-k}{2}-1} h^{(n)}(u+v)dv, \tag{1.1.4}$$

so that $T_k = \|\mathbf{Z}_k\|^2 \sim R_k^2\left(h^{(k)}\right)$, see Fang *et al.* (1990). Moreover, provided that the required moments exist, we have that

$$\mathbb{E}\left(k^{-1}T_k\right) = \alpha_h \quad \text{and} \quad \mathbb{V}\left(k^{-1}T_k\right) = \{k^{-1}(k+2)(\kappa_h+1) - 1\}\alpha_h^2,$$

where $\kappa_h = \alpha_h^{-2}\mathbb{E}\left[\{n(n+2)\}^{-1}T^2\right] - 1$ is the kurtosis parameter of the elliptical family with density generator $h^{(n)}$. Similar results hold for the inverse radial-squared random variable $S_k = T_k^{-1}$.

Let us now consider the partition $\mathbf{Z} = \left(\mathbf{Z}_k^t, \mathbf{Z}_{(k)}^t\right)^t \sim S_n\left(h^{(n)}\right)$. Thus, the conditional distribution of $\mathbf{Z}_k$ given $\mathbf{Z}_{(k)} = \mathbf{z}_{(k)}$ is such that

$$\mathbf{Z}_k\left|\mathbf{Z}_k = \mathbf{z}_k \overset{d}{=} \mathbf{Z}_k\right|\|\mathbf{Z}_k\|^2 = t \sim S_k\left(h_t^{(k)}\right), \tag{1.1.5}$$

where $t = \left\|\mathbf{z}_{(k)}\right\|^2$ and, for each $t \geq 0$,

$$h_t^{(k)}(u) = \frac{h^{(n)}(u+t)}{h^{(n-k)}(t)}, \qquad u \geq 0, \tag{1.1.6}$$

is the conditional density generator function. Moreover, by noting that

$$T_k = \|\mathbf{Z}_k\|^2 \sim R_k^2\left(h^{(k)}\right),\ T_{(k)} = \left\|\mathbf{Z}_{(k)}\right\|^2 \sim R_{n-k}^2\left(h^{(n-k)}\right) \quad \text{and} \quad T = \|\mathbf{Z}\|^2 \sim R_n^2\left(h^{(n)}\right),$$

we obtain the following relationship: $R_n^2\left(h^{(n)}\right) \overset{d}{=} R_k^2\left(h^{(k)}\right) + R_{n-k}^2\left(h^{(n-k)}\right)$, since $\|\mathbf{z}\|^2 = \|\mathbf{z}_k\|^2 + \left\|\mathbf{z}_{(k)}\right\|^2$. From (1.1.5), it also follows that

$$T_k|T_{(k)} = t \sim R_k^2\left(h_t^{(k)}\right),$$

so that the variance and kurtosis parameters, $\alpha_{h_t^{(k)}}$ and $\kappa_{h_t^{(k)}}$, respectively, associated with the conditional elliptical model in (1.1.5) are functions of $\mathbf{z}_{(k)}$ through $t = \left\|\mathbf{z}_{(k)}\right\|^2$.

More details about the relationship between elliptical and radial-squared distributions can be found in Arellano-Valle *et al.* (2002a).

1.2 Bayesian Inference for Elliptical Linear Models

In this section we consider the elliptical linear regression model

$$\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \phi, h^{(n)} \sim El_n\left(\mathbf{X}\boldsymbol{\beta}, \phi^{-1}\mathbf{I}_n, h_{a_0\phi}^{(n)}\right), \tag{1.2.7}$$

where, from (1.1.6),

$$h_{a_0\phi}^{(n)}(u) = \frac{h^{(n+d_0)}(u + a_0\phi)}{h^{(d_0)}(a_0\phi)},$$

a_0 is known ($a_0 > 0$), $h^{(n+d_0)}$ is a generator function of a $(n+d_0)$-dimensional elliptical distribution, $\boldsymbol{\beta} \in \mathbb{R}^k$ and $\phi > 0$.

If we adopt the convention that $h^{(0)}(u) = 1$, $h^{(n)}(0) = c$, for some constant c, and $h_0^{(n)} = h^{(n)}$, then, when $d_0 = a_0 = 0$, (1.2.7) yields the standard elliptical model $El_n\left(\mathbf{X}\boldsymbol{\beta}, \phi^{-1}\mathbf{I}_n, h^{(n)}\right)$. In the latter case and under the non-informative prior distribution

$$\pi(\boldsymbol{\beta}, \phi) \propto \phi^{-1}, \tag{1.2.8}$$

Osiewalski and Steel (1993) have shown that the posterior of $\boldsymbol{\beta}$ is the same for all density generators of elliptical distributions $h^{(n)}$, and therefore, for the normal linear model. Similar results hold for the predictive analysis. Only the posterior distribution of ϕ is affected by departures from normality within the class of elliptical distributions. Some results related to posterior moments of ϕ are given in Osiewalski and Steel (1996) by considering the conditional distribution of ϕ given the location parameters $\boldsymbol{\beta}$ and the data $(\mathbf{Y}, \mathbf{X})$. Arellano-Valle *et al.* (2000) provide an alternative proof of this fact, and determine the posterior distribution of ϕ explicitly, obtaining a convenient formula for examining the effects of departures of normality, which are reflected on the posterior of ϕ. Proposition (1.2.1) extends the previous results by considering a more general class of priors distributions for $(\boldsymbol{\beta}, \phi)$.

Specifically, we consider

$$\phi | h^{(n)} \sim a_0^{-1} R_{d_0}^2 \left(h^{(d_0)}\right) \tag{1.2.9}$$

which yields

$$\pi\left(\boldsymbol{\beta},\phi|h^{(n)}\right)=\frac{(a_0\pi)^{\frac{d_0}{2}}}{\Gamma\left(\frac{d_0}{2}\right)}\phi^{\frac{d_0}{2}-1}h^{(d_0)}(a_0\phi)\pi\left(\boldsymbol{\beta}|\phi,h^{(n)}\right), \tag{1.2.10}$$

where $h^{(d_0)}(\cdot)$ is given by (1.1.4).

The dependence on $h^{(n)}$ in (1.2.9) is reasonable, since in the present context the interpretation of the scale parameter changes with the density generator. We will interpret $d_0 = 0$ in (1.2.9) as the non-informative prior $\pi(\phi|h^{(n)}) \propto \phi^{-1}$, so that (1.2.10) is reduced to (1.2.8) when $\pi\left(\boldsymbol{\beta}|\phi,h^{(n)}\right)$ is constant.

Proposition 1.2.1. *Under (1.2.7) and (1.2.10) with $\boldsymbol{\beta}\perp\!\!\!\perp\phi|h^{(n)}$, we have*

$$\pi\left(\boldsymbol{\beta}|\mathbf{X},\mathbf{y},h^{(n)}\right)\propto\left[a_{q(\mathbf{y})}+\left\|\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}}\right\|^2\right]^{-\frac{k+d}{2}}\pi\left(\boldsymbol{\beta}|h^{(n)}\right) \tag{1.2.11}$$

and,

$$\pi\left(\phi|\mathbf{X},\mathbf{y},h^{(n)}\right)\propto\phi^{\frac{k+d}{2}-1}\int_{\mathbb{R}^k}h^{(k+d)}\left(\left[a_{q(\mathbf{y})}+\left\|\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\hat{\boldsymbol{\beta}}\right\|^2\right]\phi\right)\pi\left(\boldsymbol{\beta}|\phi\right)d\boldsymbol{\beta} \tag{1.2.12}$$

where $d = n-k+d_0$ are the remaining degrees of freedom, $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y}$, $a_{q(\mathbf{y})} = a_0 + q(\mathbf{y})$ and $q(\mathbf{y}) = (n-k)S^2$ where $S^2 = \frac{1}{n-k}\left\|\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right\|^2$.

Assuming $h^{(n)}$ as given and $\pi\left(\boldsymbol{\beta}|h^{(n)}\right)$ to be constant, then

$$\boldsymbol{\beta}|\mathbf{X},\mathbf{y}\sim t_k\left(\hat{\boldsymbol{\beta}},a_{q(\mathbf{y})}(\mathbf{X}^t\mathbf{X})^{-1},d\right) \tag{1.2.13}$$

and,

$$\phi|\mathbf{X},\mathbf{y},h^{(n)}\sim\frac{1}{a_{q(\mathbf{y})}}R_d^2\left(h^{(d)}\right). \tag{1.2.14}$$

Proof. Using (1.2.7) and (1.2.9), we have that

$$f\left(\mathbf{y},\phi|\mathbf{X},\boldsymbol{\beta}\right)\propto\phi^{\frac{n+d_0}{2}-1}h^{(n)}_{a_0\phi}\left(\phi\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2\right)\,h^{(d_0)}(a_0\phi).$$

But, from (1.1.6), $h_t^{(n)}(u)h^{(d_0)}(t) = h^{(n+d_0)}(u+t)$, so that

$$f\left(\mathbf{y}, \phi|\mathbf{X}, \boldsymbol{\beta}\right) \propto \phi^{\frac{n+d_0}{2}-1} h^{(n+d_0)} \left(\phi \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right\|^2 + \phi a_0\right). \tag{1.2.15}$$

Using now (1.1.3), we have that

$$f\left(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}\right) \propto \left[a_0 + \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2\right]^{-\frac{n+d_0}{2}},$$

which does not depend on $h^{(n)}$. Thus, the results in (1.2.11) and (1.2.13) follow by using the well known decomposition

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \left\|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right\|^2 + \left\|\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}}\right\|^2 \tag{1.2.16}$$

and from the fact that $\pi\left(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}, h^{(n)}\right) \propto f\left(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}\right)\pi\left(\boldsymbol{\beta}|h^{(n)}\right)$. For the proof of (1.2.12) and therefore (1.2.14), we note that

$$\pi\left(\phi|\mathbf{X}, \mathbf{y}, h^{(n)}\right) \propto \int_{\mathbb{R}^k} f\left(\mathbf{y}, \phi|\mathbf{X}, \boldsymbol{\beta}, h^{(n)}\right)\pi(\boldsymbol{\beta}|h^{(n)})d\boldsymbol{\beta},$$

where $f\left(\mathbf{y}, \phi|\mathbf{X}, \boldsymbol{\beta}, h^{(n)}\right)$ is defined in (1.2.15). Thus, the proof follows by using (1.1.4) jointly with (1.2.16). ■

Note from (1.2.11) that if the prior distribution of $\boldsymbol{\beta}$ does not depend on $h^{(n)}$, then the posterior distribution of $\boldsymbol{\beta}$ is invariant on the class of elliptical models under consideration and can be obtained under the normality assumption. In particular, applying (1.2.13) and (1.2.14) we get $\mathbb{E}\left(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}\right) = \hat{\boldsymbol{\beta}} \quad (n > k+2)$, $\mathbb{V}\left(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}\right) = \frac{a_{q(\mathbf{y})}}{d-2}\left(\mathbf{X}^t\mathbf{X}\right)^{-1} \quad (n > k+4)$ and (provided that they exist) $\mathbb{E}\left(\phi|\mathbf{X}, \mathbf{y}, h^{(n)}\right) = \frac{d\alpha_h}{a_{q(\mathbf{y})}}$, $\mathbb{V}\left(\phi|\mathbf{X}, \mathbf{y}, h^{(n)}\right) = \left[\frac{d+2}{d}(\kappa_h+1) - 1\right]\left[\frac{d\alpha_h}{a_{q(y)}}\right]^2$. For the particular case $d_0 = a_0 = 0$, $d = n-k$ and $a_{q(\mathbf{y})} = (n-k)S^2$: $\boldsymbol{\beta}|\mathbf{X}, \mathbf{y} \sim t_p\left(\hat{\boldsymbol{\beta}}, S^2\left(\mathbf{X}^t\mathbf{X}\right)^{-1}, n-k\right)$ and $\phi|\mathbf{X}, \mathbf{y}, h^{(n)} \sim \frac{1}{(n-k)S^2} R^2_{n-k}\left(h^{(n-k)}\right)$. These results reduce to those in Osiewalski and Steel (1993).

Assuming in Proposition 1.2.1, $h^{(n)}$ as given and $\pi\left(\boldsymbol{\beta}|h^{(n)}\right)$ to be constant, then (1.2.14) implies that the posterior distribution of $\sigma^2 = \phi^{-1}$ satisfies $\sigma^2|\mathbf{X}, \mathbf{y}, h^{(n)} \sim a_{q(\mathbf{y})} IR_d^2\left(h^{(n)}\right)$, that is, $\pi\left(\sigma^2|\mathbf{X}, \mathbf{y}, h^{(n)}\right) = \frac{\pi^{d/2}}{\Gamma(d/2)}\, a_{q(\mathbf{y})}^{d/2} \left(\sigma^{-2}\right)^{\frac{d}{2}+1} h^{(d)}\left(\frac{a_{q(\mathbf{y})}}{\sigma^2}\right)$.

Another posterior inference is the models selection, for this is very common to calculate the Bayes factor between two models, $El_n\left(\mathbf{X}_1\boldsymbol{\beta}_1, \phi^{-1}\mathbf{I}_n, h_{a_0\phi,1}^{(n)}\right)$ and $El_n\left(\mathbf{X}_2\boldsymbol{\beta}_2, \phi^{-1}\mathbf{I}_n, h_{a_0\phi,2}^{(n)}\right)$, that is

$$
\begin{aligned}
BF &= \frac{\int f\left(\mathbf{y}\left|\mathbf{X}_1, \boldsymbol{\beta}_1, \phi, h_1^{(n)}\right.\right)\pi\left(\boldsymbol{\beta}_1, \phi|h_1^{(n)}\right) d\phi d\boldsymbol{\beta}_1}{\int f\left(\mathbf{y}\left|\mathbf{X}_2, \boldsymbol{\beta}_2, \phi, h_2^{(n)}\right.\right)\pi\left(\boldsymbol{\beta}_2, \phi|h_2^{(n)}\right) d\phi d\boldsymbol{\beta}_2} = \frac{\int f\left(\mathbf{y}\left|\mathbf{X}_1, \boldsymbol{\beta}_1\right.\right)\pi\left(\boldsymbol{\beta}_1|h_1^{(n)}\right) d\boldsymbol{\beta}_1}{\int f\left(\mathbf{y}\left|\mathbf{X}_2, \boldsymbol{\beta}_2\right.\right)\pi\left(\boldsymbol{\beta}_2|h_2^{(n)}\right) d\boldsymbol{\beta}_2} \\
&= \frac{\int \left[a_{q_1(\mathbf{y})} + \left\|\mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1\right\|^2\right]^{-\frac{n+d_0}{2}} \pi\left(\boldsymbol{\beta}_1|h_1^{(n)}\right) d\boldsymbol{\beta}_1}{\int \left[a_{q_2(\mathbf{y})} + \left\|\mathbf{X}_2\boldsymbol{\beta}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2\right\|^2\right]^{-\frac{n+d_0}{2}} \pi\left(\boldsymbol{\beta}_2|h_2^{(n)}\right) d\boldsymbol{\beta}_2},
\end{aligned}
$$

where $\hat{\boldsymbol{\beta}}_i = \left(\mathbf{X}_i^t\mathbf{X}_i\right)^{-1}\mathbf{X}_i^t\mathbf{y}$ and $q_i(\mathbf{y}) = \left\|\mathbf{y} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i\right\|^2$ with $i = 1, 2$. An explicit form for the previous Bayes factor is difficult to obtain.

As we could see, invariance of the posterior distribution of $\boldsymbol{\beta}$ facilitates the posterior inferences of $\boldsymbol{\beta}$ since they coincide with the well-known results of the normal model. However, this is a difficulty if our goal is compare different elliptical models under the conditions of Proposition 1.2.1 with $\pi\left(\boldsymbol{\beta}|h^{(n)}\right) \propto 1$, this case is approached in the next section.

1.3 Default Bayes Factors for Elliptical Linear Models

In this section we consider the q alternative standard elliptical linear models

$$
M_j : \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon_j, \quad \epsilon_j \sim El_n\left(\mathbf{0}, \phi^{-1}\mathbf{I}_n; h_j^{(n)}\right), \tag{1.3.17}
$$

$j = 1, \ldots, q$, where the h_j's are n-dimensional generators, $\boldsymbol{\beta} \in \mathbb{R}^k$ $(n > k)$, and the non-informative prior distribution given by (1.2.8).

As is well known, the usual Bayes factors based on non-informative or default improper priors, do not work, because the resulting Bayes factors are undetermined. Several solutions to this difficulty have been proposed and discussed by Berger and Pericchi (2001), they are called objective Bayes model selection methods.

In this section we discuss the model comparison problem under improper prior by using the Intrinsic Bayes Factors (IBF) (Berger and Pericchi (1996a)) and Fractional Bayes Factors (FBF) (O'Hagan (1995)).

Let consider the following partition

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{(1)} \\ \mathbf{y}_{(2)} \end{pmatrix} \text{ and } \mathbf{X} = \begin{pmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} \end{pmatrix}$$

where $\mathbf{y}_{(i)} \in \mathbb{R}^{n_i}$ $(i = 1, 2)$, $n = n_1 + n_2$ and the matrix $\mathbf{X}_{(i)}$ has dimension $n_i \times k$.

Proposition 1.3.1. *If rank*$(\mathbf{X}_{(1)}) = k < n_1$ *then, for each model* M_j *in (1.3.17), the marginal density of the sub-vector* $\mathbf{y}_{(1)}$ *is given by*

$$m_j^N\left(\mathbf{y}_{(1)} \middle| \mathbf{X}_{(1)}\right) = \frac{\Gamma\left(\frac{n_1-k}{2}\right)}{\left(\sqrt{\pi}\right)^{n_1-k}\left(\left|\mathbf{X}_{(1)}^t\mathbf{X}_{(1)}\right|\right)^{\frac{1}{2}}}\left\|\mathbf{y}_{(1)} - \mathbf{X}_{(1)}\hat{\boldsymbol{\beta}}_{(1)}\right\|^{-(n_1-k)},$$

where $\hat{\boldsymbol{\beta}}_{(1)} = \left(\mathbf{X}_{(1)}^t\mathbf{X}_{(1)}\right)^{-1}\mathbf{X}_{(1)}^t\mathbf{y}_{(1)}$.

Proof. Using the same ideas of Proposition 1.2.1 when $a_0 = d_0 = 0$ we can show that

$$f\left(\mathbf{y}, \boldsymbol{\beta} \middle| \mathbf{X}, h\right) = \int_0^\infty \phi^{\frac{n}{2}-1} h^{(n)}\left(\phi\left\|\mathbf{y} - \mathbf{X}\beta\right\|^2\right) d\phi = \frac{\Gamma\left(n/2\right)}{\pi^{n/2}}\left\|\mathbf{y} - \mathbf{X}\beta\right\|^{-n}, \quad (1.3.18)$$

which does not depends on $h^{(n)}$, and can rewrite as

$$f\left(\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \boldsymbol{\beta} \middle| \mathbf{X}\right) = \frac{\Gamma\left(n_1/2\right) t_{n_2}\left(\mathbf{y}_{(2)} \middle| \mathbf{X}_{(2)}\boldsymbol{\beta}, n_1^{-1}\left\|\mathbf{y}_{(1)} - \mathbf{X}_{(1)}\beta\right\|^2 \mathbf{I}_{n_2}, n_1\right)}{\pi^{n_1/2}\left\|\mathbf{y}_{(1)} - \mathbf{X}_{(1)}\beta\right\|^{n_1}},$$

where $t_k\left(\mathbf{y}|\,\boldsymbol{\mu},\boldsymbol{\Sigma},\nu\right)$ is density of $t_k\left(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu\right)$ distribution. If now integrating out $\mathbf{y}_{(2)}$ we obtain

$$f\left(\mathbf{y}_{(1)},\boldsymbol{\beta}|\,\mathbf{X}_{(1)}\right)=\frac{\Gamma\left(n_1/2\right)}{\pi^{n_1/2}}\left\|\mathbf{y}_{(1)}-\mathbf{X}_{(1)}\hat{\beta}_{(1)}\right\|^{-n_1}\left[1+\frac{\left\|\mathbf{X}_{(1)}\boldsymbol{\beta}-\mathbf{X}_{(1)}\hat{\boldsymbol{\beta}}_{(1)}\right\|^2}{S^2_{(1)}\left(n_1-k\right)}\right]^{-n_1/2},$$

where $S^2_{(1)}=\frac{1}{n_1-k}\left\|\mathbf{y}_{(1)}-\mathbf{X}_{(1)}\hat{\boldsymbol{\beta}}_{(1)}\right\|^2$. Since the last factor is the kernel of the Student-t density

$$t_k\left(\boldsymbol{\beta}|\,\hat{\boldsymbol{\beta}}_{(1)},S^2_{(1)}\left(\mathbf{X}^t_{(1)}\mathbf{X}_{(1)}\right)^{-1},n_1-k\right)$$

and $n_1>k$, the result follows. ■

Remark 1.3.2. *From Proposition 1.3.1, the marginal density of any sub-vector $\mathbf{y}_{(1)}$, with $n_1>k$, does not depend on the specific elliptical model under consideration. A similar result is obtained by Berger* et al. *(1998) to compare models of the form (1.3.17), but for a wider class of models. However, their result is not valid for $n_1>k+1$.*

1.3.1 Intrinsic Bayes Factor

The general strategy for computing IBF's begins with the determination of a proper and minimal training sample. It is known that (Berger and Pericchi (1996b)) the minimal training sample for the elliptical models in (1.3.17) is a sub-vector $\mathbf{y}\left(l\right)$ of size $n_1=k+1$ such that the corresponding sub-matrix $\mathbf{X}\left(l\right)$ is of full rank. Computation to compare two elliptical models M_1 and M_2, yields the following Partial Bayes Factor (PBF) for data $\mathbf{y}$,

$$B_{12}\left(l\right)=\frac{m_1^N\left(\mathbf{y}|\,\mathbf{X}\right)}{m_2^N\left(\mathbf{y}|\,\mathbf{X}\right)}\cdot\frac{m_2^N\left(\mathbf{y}\left(l\right)|\,\mathbf{X}\left(l\right)\right)}{m_1^N\left(\mathbf{y}\left(l\right)|\,\mathbf{X}\left(l\right)\right)}=1.$$

Therefore the IBF's would not be useful in order to compare the models in (1.3.17). However, the IBF's are useful and easy to calculate upon comparing elliptical linear models with different design matrices.

Now, we consider the comparison between the elliptical linear models

$$M_j : \mathbf{Y} = \mathbf{X}_j \boldsymbol{\beta}_j + \epsilon_j, \quad \epsilon_j \sim El_n\left(\mathbf{0}, \phi^{-1}\mathbf{I}_n; h_j^{(n)}\right) \tag{1.3.19}$$

$j = 1, \ldots, q$ where the h_j's are the generators, $n > \max_j \{k_j\}$ and $\boldsymbol{\beta}_j \in \mathbb{R}^{k_j}$.

Proposition 1.3.3. *The IBF's in order to compare any two models M_1 and M_2 of type (1.3.19) do not depend on $h_1^{(n)}$ and $h_2^{(n)}$.*

Proof. It suffices to note from Proposition 1.3.1 that the PBF does not depend on the generators. ■

The previous result is very useful since it allow us to calculate IBF's to compare elliptical linear models with different design matrices using the results of Berger and Pericchi (1996b) relating to the IBF's for normal linear models.

1.3.2 Fractional Bayes Factor

Another alternative approach to compare models is the FBF developed in O'Hagan (1995). As mentioned by this author, the FBF has a series of advantages over the IBF (see, for example O'Hagan (1997)), for example it is easier to compute than IBF. It is necessary to note that those models specified by (1.3.19) differ in two aspects, the design matrix $\mathbf{X}_j$ and the generator $h_j^{(n)}$. The next results are related to the comparison of two models $M_1 = \left(\mathbf{X}_1, h_1^{(n)}\right)$ and $M_2 = \left(\mathbf{X}_2, h_2^{(n)}\right)$.

Proposition 1.3.4. *The FBF to compare two models M_1 and M_2 in (1.3.19) with design matrices of full rank is*

$$B_b(\mathbf{y}) = \frac{\Gamma\left(\frac{n-k_1}{2}\right)\Gamma\left(\frac{bn-k_2}{2}\right)\left(\left\|\mathbf{y}-\mathbf{X}_1\hat{\boldsymbol{\beta}}_1\right\|\right)^{n(b-1)}\int_0^\infty u^{\frac{bn}{2}-1}\left(h_2^{(n)}\right)^b(u)\,du}{\Gamma\left(\frac{n-k_2}{2}\right)\Gamma\left(\frac{bn-k_1}{2}\right)\left(\left\|\mathbf{y}-\mathbf{X}_2\hat{\boldsymbol{\beta}}_2\right\|\right)^{n(b-1)}\int_0^\infty u^{\frac{bn}{2}-1}\left(h_1^{(n)}\right)^b(u)\,du},$$

where $0<b<1$, $bn>\max_j\{k_j\}$ *and* $\hat{\boldsymbol{\beta}}_j=\left(\mathbf{X}_j^t\mathbf{X}_j\right)^{-1}\mathbf{X}_j^t\mathbf{y}$, $j=1,2$.

Proof. In this case the FBF is given by $B_b(\mathbf{y})=\frac{q_1(b,\mathbf{y})}{q_2(b,\mathbf{y})}$, where

$$q_j(b,\mathbf{y}) = \frac{\int \pi\left(\boldsymbol{\beta}_j,\phi\right) f\left(\mathbf{y}|\,\mathbf{X}_j,\boldsymbol{\beta}_j,\phi,h_j^{(n)}\right)d\phi d\boldsymbol{\beta}_j}{\int \pi\left(\boldsymbol{\beta}_j,\phi\right) f^b\left(\mathbf{y}|\,\mathbf{X}_j,\boldsymbol{\beta}_j,\phi,h_j^{(n)}\right)d\phi d\boldsymbol{\beta}_j}$$

and $j=1,2$. Using the change of variables $u_j=\phi\left\|\mathbf{y}-\mathbf{X}_j\boldsymbol{\beta}_j\right\|^2$ and integrating out $\boldsymbol{\beta}_j$ we obtain that the denominator of $q_j(b,\mathbf{y})$ is given by

$$\frac{\Gamma\left(\frac{bn-k_j}{2}\right)\pi^{k_j/2}}{\Gamma\left(\frac{bn}{2}\right)\left(\left|\mathbf{X}_j^t\mathbf{X}_j\right|\right)^{\frac{1}{2}}}\left(\left\|\mathbf{y}-\mathbf{X}_j\hat{\boldsymbol{\beta}}_j\right\|\right)^{-(bn-k_j)}\cdot\int u_j^{\frac{bn}{2}-1}\left(h_j^{(n)}\right)^b(u_j)\,du_j.$$

Now, the numerator of $q_j(b,\mathbf{y})$ is just the predictive density of $\mathbf{y}$ under model j. Thus, from Proposition 1.3.1 we have that the numerator is given by

$$\frac{\Gamma\left(\frac{n-k_j}{2}\right)}{\left(\sqrt{\pi}\right)^{n-k_j}\left(\left|\mathbf{X}_j^t\mathbf{X}_j\right|\right)^{\frac{1}{2}}}\left(\left\|\mathbf{y}-\mathbf{X}_j\hat{\boldsymbol{\beta}}_j\right\|\right)^{-(n-k_j)}.$$

Consequently,

$$q_j(b,\mathbf{y}) = \frac{\Gamma\left(\frac{n-k_j}{2}\right)\Gamma\left(\frac{bn}{2}\right)\left(\left\|\mathbf{y}-\mathbf{X}_j\hat{\boldsymbol{\beta}}_j\right\|\right)^{n(b-1)}}{\Gamma\left(\frac{bn-k_j}{2}\right)\left(\sqrt{\pi}\right)^n\int u_j^{\frac{bn}{2}-1}\left(h_j^{(n)}\right)^b(u_j)\,du_j},$$

concluding the proof. ■

The FBF for especial cases are presented in what follows.

Table 1.2: $I\left(b, h^n\right)$ for some subclasses of n-dimensional spherical distributions.

Distribution	Density generator $h^{(n)}(u)$	$I\left(b, h^{(n)}\right)$
Normal	$h_{\phi}^{(n)}(u) = (2\pi)^{-n/2}\exp\{-u/2\}$	$\frac{\Gamma\left(\frac{bn}{2}\right)}{(b\pi)^{bn/2}}$
Contaminated Normal	$(1-\epsilon)\, h_{\phi}(u) + \epsilon\sigma^{-\frac{n}{2}} h_{\phi}\left(\frac{u}{\sigma}\right)$, $0 < \epsilon < 1$, $\sigma > 0$	$\frac{\Gamma\left(\frac{bn}{2}\right)\left(1-\epsilon+\epsilon\sigma^{n(b-1)/2}\right)}{(b\pi)^{bn/2}}$
Student-t	$\frac{\Gamma\left(\frac{n+\nu}{2}\right)\nu^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}\{\nu+u\}^{-\frac{(n+\nu)}{2}}$, $\nu > 0$	$\frac{\Gamma^b\left(\frac{n+\nu}{2}\right)\Gamma\left(\frac{bn}{2}\right)\Gamma\left(\frac{b\nu}{2}\right)}{\Gamma^b\left(\frac{\nu}{2}\right)\Gamma\left(b\frac{n+\nu}{2}\right)\pi^{bn/2}}$
Generalized Student-t	$\frac{\Gamma\left(\frac{n+\nu}{2}\right)\lambda^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}\{\lambda+u\}^{-\frac{(n+\nu)}{2}}$, $\nu, \lambda > 0$	$\frac{\Gamma^b\left(\frac{n+\nu}{2}\right)\Gamma\left(\frac{bn}{2}\right)\Gamma\left(\frac{b\nu}{2}\right)}{\Gamma^b\left(\frac{\nu}{2}\right)\Gamma\left(b\frac{n+\nu}{2}\right)\pi^{bn/2}}$
Pearson Type II	$\frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{n/2}}(1-u)^{\frac{\nu}{2}-1}$, $\nu > 0$	$\frac{\Gamma^b\left(\frac{n+\nu}{2}\right)\Gamma\left(\frac{bn}{2}\right)\Gamma\left(b\frac{\nu-2}{2}+1\right)}{\Gamma^b\left(\frac{\nu}{2}\right)\Gamma\left(b\frac{n+\nu-2}{2}+1\right)\pi^{bn/2}}$
Power Exponential	$\frac{\Gamma\left(\frac{n}{2}\right)s}{\Gamma\left(\frac{n}{2s}\right)\pi^{\frac{n}{2}}2^{\frac{n}{2s}}}\exp\{-u^s/2\}$, $s > 0$	$\frac{\Gamma^b\left(\frac{n}{2}\right)\Gamma\left(\frac{bn}{2s}\right)s^{b-1}}{\Gamma^b\left(\frac{n}{2s}\right)\pi^{\frac{bn}{2}}b^{\frac{bn}{2s}}}$
Kotz Type	$\frac{\Gamma\left(\frac{n}{2}\right)s\rho^{\frac{2q+n}{2s}}}{\Gamma\left(\frac{2q+n}{2s}\right)\pi^{\frac{n}{2}}2^{\frac{2q+n}{2s}}}u^q\exp\{-\rho u^s/2\}$, $\rho, s > 0$, $2q+n > 0$	$\frac{\Gamma^b\left(\frac{n}{2}\right)\Gamma\left(b\frac{2q+n}{2s}\right)s^{b-1}}{\Gamma^b\left(\frac{2q+n}{2s}\right)\pi^{\frac{bn}{2}}b^{b\frac{2q+n}{2s}}}$

Corollary 1.3.5. *The FBF for comparing two models M_1 and M_2 of type (1.3.17) is given by*

$$B_b(\mathbf{y}) = \frac{\int_0^\infty u^{\frac{bn}{2}-1}\left(h_2^{(n)}\right)^b(u)\,du}{\int_0^\infty u^{\frac{bn}{2}-1}\left(h_1^{(n)}\right)^b(u)\,du}.$$

This corollary shows the lack of sensibility of the FBF when distinguishing between two different elliptical linear models, because the FBF depends on the data through the sample size n only. We note also that in order to compare models with different design matrices by using the result of Proposition 1.3.4 it is necessary to know $I\left(b, h_j^{(n)}\right) = \int_0^\infty u^{\frac{bn}{2}-1}\left(h_j^{(n)}\right)^b(u)\,du$. Table 1.2 shows the value of $I\left(b, h_j^{(n)}\right)$ for different generator density functions.

From Table 1.2, note that for the generalized Student-t distribution the value of

$I\left(b, h^{(n)}\right)$ does not depend on the parameter λ, and therefore this value is the same for the Student-t distribution. Similarly, the value of $I\left(b, h^{(n)}\right)$ for the Kotz Type distribution does not depend on the parameter ρ.

The next corollary shows that for comparing two elliptical linear models with different design matrices and common generator function, it is enough to compare two normal linear models with different design matrices, see O'Hagan (1995).

Corollary 1.3.6. *The FBF for comparing two linear models* $M_1 := \left(\mathbf{X}_1, h^{(n)}\right)$ *and* $M_2 := \left(\mathbf{X}_2, h^{(n)}\right)$ *is given by*

$$B_b\left(\mathbf{y}\right) = \frac{\Gamma\left(\frac{n-k_1}{2}\right)\Gamma\left(\frac{bn-k_2}{2}\right)}{\Gamma\left(\frac{n-k_2}{2}\right)\Gamma\left(\frac{bn-k_1}{2}\right)}\left(\frac{\left\|\mathbf{y}-\mathbf{X}_1\hat{\boldsymbol{\beta}}_1\right\|}{\left\|\mathbf{y}-\mathbf{X}_2\hat{\boldsymbol{\beta}}_2\right\|}\right)^{n(b-1)},$$

where $0 < b < 1$*,* $bn > \max_j \{k_j\}$ *and* $\hat{\boldsymbol{\beta}}_j = \left(\mathbf{X}_j^t\mathbf{X}_j\right)^{-1}\mathbf{X}_j^t\mathbf{y}$*,* $j = 1, 2$*.*

Observe that, under conditions of the previous corollary, both: the IBF and FBF, remain invariant for the class of elliptical distributions.

1.3.3 Model Comparison as a Decision Problem

A more general approach is to consider the problem of model comparison within the decision theory framework, as described in Bernardo and Smith (1994). Following the notation used by those authors we will call $\boldsymbol{\omega}$ the unknown state of the nature. In our case, the objective could be inference about $(\boldsymbol{\beta}, \phi)$, $(y_{n+1}, \ldots, y_m)$, etc. Thus, we would like to obtain the conditional distribution of $\boldsymbol{\omega}$ given $\mathbf{y}$ under the true model, by assuming that this model is contained in the class of models that we are comparing. Figure 1.3, taken from Bernardo and Smith (1994), shows the description of our decision problem, where m_i means that, given the data $\mathbf{y}$, we choose model M_i and a_j, $j \in J_i$ is some report of beliefs assuming model M_i.

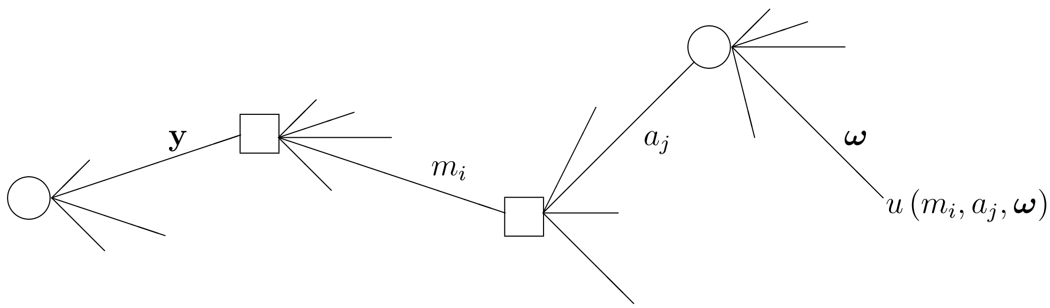

Figure 1.3: Outline of the decision problem

Appropriate utility functions for these cases must be smooth, proper and local score functions; their definitions and more details can be found in Bernardo and Smith (1994). Under these conditions, these authors show that for proper score functions, $u_i(a_j, \boldsymbol{\omega})$, the optimal choice of a_j, $j \in J_i$, is $a_i^* = f(\boldsymbol{\omega}|\, \mathbf{y}, m_i)$ and, therefore the utility function would be

$$u(m_i, a_i^*, \boldsymbol{\omega}) = u_i(f(\boldsymbol{\omega}|\, \mathbf{y}, m_i), \omega), \qquad i = 1, \dots, q.$$

But also, under the assumption of local score function, we have that

$$u(m_i, a_i^*, \boldsymbol{\omega}) = A \log f(\boldsymbol{\omega}|\, \mathbf{y}, m_i) + B(\omega), \qquad i = 1, \dots, q,$$

where $A > 0$ is a constant and $B(\cdot)$ is a function of $\boldsymbol{\omega}$. Therefore, the optimal model is such that maximize the following expected utility function

$$\bar{u}(m_i|\, \mathbf{y}) = \int \{A \log f(\boldsymbol{\omega}|\, \mathbf{y}, m_i) + B(\boldsymbol{\omega})\} f(\boldsymbol{\omega}|\, \mathbf{y})\, d\boldsymbol{\omega} \tag{1.3.20}$$

provided that this exists and where, in our case, if $P(M_i) = 1/q$ for all $i = 1, \dots, q$ then

$$P(M_j|\, \mathbf{y}) = \left(\sum_{i=1}^{q} \frac{m_i^N(\mathbf{y}|\, \mathbf{X}_i)}{m_j^N(\mathbf{y}|\, \mathbf{X}_j)} \right)^{-1}.$$

If additionally all design matrices are equal, then $P\left(M_j|\,\mathbf{y}\right)=1/q$ and

$$f\left(\boldsymbol{\omega}|\,\mathbf{y}\right)=\frac{1}{q}\sum_{j=1}^{q}f\left(\boldsymbol{\omega}|\,\mathbf{y},m_j\right). \tag{1.3.21}$$

That means, except for different design matrices, that there is no posterior preference for any model. In such a case, the expected utility $\bar{u}\left(m_i|\,\mathbf{y}\right)$ depends on the model through $f\left(\boldsymbol{\omega}|\,\mathbf{y},m_j\right)$.

Hereafter, we present results for computing $\bar{u}\left(m_i|\,\mathbf{y}\right)$ to compare models $M_i=(\mathbf{X}_i,h_i)$ and $M_j=(\mathbf{X}_j,h_j)$ for different choices of $\boldsymbol{\omega}$, where $h_j=h_j^{(n)}$. We will also assume that $n>\max_j\{k_j\}$, such that if $\boldsymbol{\omega}=(\boldsymbol{\beta},\phi)$ then $f\left(\boldsymbol{\omega}|\,\mathbf{y},m_j\right)$ and, therefore, $f\left(\boldsymbol{\omega}|\,\mathbf{y}\right)$ are proper, which imply that (1.3.20) exists.

Proposition 1.3.7. *If $\boldsymbol{\omega}=(\boldsymbol{\beta},\phi)$ and $\mathbf{X}_i=\mathbf{X}\ \forall\ i=1,\ldots,q$,*

$$\begin{aligned}\bar{u}\left(m_i|\,\mathbf{y}\right) &= \frac{\pi^{\frac{n}{2}}A}{q\Gamma\left(\frac{n}{2}\right)}\sum_{j=1}^{q}\int\log\left[v^{\frac{n}{2}-1}h_i\left(v\right)\right]v^{\frac{n}{2}-1}h_j\left(v\right)dv\\ &\quad -A\mathbb{E}_{t_k}\left[\log\left(\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^{n-2}\right)\right]-A\log\left[m^N\left(\mathbf{y}|\,\mathbf{X}\right)\right]+\mathbb{E}\left[B\left(\boldsymbol{\beta},\phi\right)|\,\mathbf{y}\right],\end{aligned}$$

where the expected value $\mathbb{E}_{t_k}\left(\cdot\right)$ is calculated with respect to the Student-t distribution $t_k\left(\hat{\boldsymbol{\beta}},S^2\left(\mathbf{X}^t\mathbf{X}\right)^{-1},n-k\right)$.

Proof. Since in this case we are assuming that $\boldsymbol{\omega}=(\boldsymbol{\beta},\phi)$ and $\mathbf{X}_i=\mathbf{X}$, $i=1,\ldots,q$, it follows that $m_i^N\left(\mathbf{y}|\,\mathbf{X}\right)=m^N\left(\mathbf{y}|\,\mathbf{X}\right)$, $i=1,\ldots,q$, so that

$$f\left(\boldsymbol{\beta},\phi|\,\mathbf{y},m_i\right)=\frac{\phi^{\frac{n}{2}-1}h_i\left(\phi\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2\right)}{m^N\left(\mathbf{y}|\,\mathbf{X}\right)}.$$

Then, from (1.3.21) and the previous equation,

$$\begin{aligned}\bar{u}\left(m_i|\,\mathbf{y}\right) &= \frac{A}{qm^N\left(\mathbf{y}|\,\mathbf{X}\right)}\\ &\quad\int\log\left[\phi^{\frac{n}{2}-1}h_i\left(\phi\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2\right)\right]\sum_{j=1}^{q}\phi^{\frac{n}{2}-1}h_j\left(\phi\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2\right)d\left(\boldsymbol{\beta},\phi\right)\\ &\quad -A\log\left[m^N\left(\mathbf{y}|\,\mathbf{X}\right)\right]+\mathbb{E}\left[B\left(\boldsymbol{\beta},\phi\right)|\,\mathbf{y}\right].\end{aligned}$$

In the previous integral, the usual change of variable $v = \phi \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^2$, yields

$$\begin{aligned}
\bar{u}\left(m_i | \mathbf{y}\right) &= \frac{A}{qm^N\left(\mathbf{y}|\mathbf{X}\right)} \int \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^{-n} d\boldsymbol{\beta} \sum_{j=1}^{q} \int \log\left[v^{\frac{n}{2}-1} h_i\left(v\right)\right] v^{\frac{n}{2}-1} h_j\left(v\right) dv \\
&\quad - \frac{A}{qm^N\left(\mathbf{y}|\mathbf{X}\right)} \int \frac{\log\left(\left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^{n-2}\right)}{\left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^{n}} d\boldsymbol{\beta} \sum_{j=1}^{q} \int v^{\frac{n}{2}-1} h_j\left(v\right) dv \\
&\quad - A \log\left[m^N\left(\mathbf{y}|\mathbf{X}\right)\right] + \mathbb{E}\left[B\left(\boldsymbol{\beta}, \phi\right)|\, \mathbf{y}\right].
\end{aligned}$$

Using that

$$\left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^{-n} = \left\| \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right\|^{-n} \left[1 + \frac{\left\| \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\hat{\boldsymbol{\beta}} \right\|^2}{S^2\left(n-k\right)}\right]^{-n/2},$$

and by observing that the last factor is the kernel of a Student-t distribution, we conclude the proof. ■

Remark 1.3.8. *From the above proposition and the previous results it follows that if $\boldsymbol{\omega} = \boldsymbol{\beta}$ and $\mathbf{X}_i = \mathbf{X}$ for all $i = 1, \ldots, q$, then $\bar{u}\left(m_i|\mathbf{y}\right) = A\mathbb{E}_{t_k}\left[\log f\left(\boldsymbol{\beta}|\,\mathbf{y}\right)|\,\mathbf{y}\right] + \mathbb{E}_{t_k}\left[B\left(\boldsymbol{\beta}\right)|\,\mathbf{y}\right]$, which do not depends on h_i.*

Proposition 1.3.9. *If $\omega = \phi$ then*

$$\begin{aligned}
\bar{u}\left(m_i|\mathbf{y}\right) = &\frac{A}{\sum_{r=1}^{q} m_r^N\left(\mathbf{y}|\mathbf{X}_r\right)} \int \log\left[\left(n-k_i\right) S_i^2 v^{\frac{n-k_i}{2}-1} h_i\left(v\right)\right] v^{-1} \\
&\sum_{j=1}^{q} \frac{\left(\pi v\right)^{\frac{n-k_j}{2}}}{\Gamma\left(\frac{n-k_j}{2}\right)} \left(\frac{\left(n-k_j\right) S_j^2}{\left(n-k_i\right) S_i^2}\right)^{\frac{n-k_j}{2}} h_j\left(\frac{\left(n-k_j\right) S_j^2}{\left(n-k_i\right) S_i^2} v\right) m_j^N\left(\mathbf{y}|\mathbf{X}_j\right) dv \\
&+ \mathbb{E}\left[B\left(\phi\right)|\,\mathbf{y}\right] + A \log\left[\frac{\pi^{\frac{n-k_i}{2}}}{\Gamma\left(\frac{n-k_i}{2}\right)}\right].
\end{aligned}$$

Proof. This follows from Proposition 1.2.1, by noting that when $a_0 = d_0 = 0$ then

$$f\left(\phi|\,\mathbf{y}, m_i\right) = \frac{\pi^{\frac{n-k_i}{2}}}{\Gamma\left(\frac{n-k_i}{2}\right)} \left(\left(n-k_i\right) S_i^2 \phi\right)^{\frac{n-k_i}{2}} \phi^{-1} h_i\left(\left(n-k_i\right) S_i^2 \phi\right)$$

where $S_i^2 = \frac{1}{(n-k_i)} \left\| \mathbf{y} - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i \right\|^2$. The change of variable $v = (n - k_i) S_i^2 \phi$ yields the result. ■

Remark 1.3.10. *If in Proposition 1.3.9 we set $\mathbf{X}_i = \mathbf{X}$ for all $i = 1, \ldots, q$, then*

$$\begin{aligned} \bar{u}\left(m_i \middle| \mathbf{y}\right) &= \frac{A\pi^{\frac{n-k}{2}}}{q\Gamma\left(\frac{n-k}{2}\right)} \sum_{j=1}^{q} \int \log\left[v^{\frac{n-k}{2}-1} h_i(v)\right] v^{\frac{n-k}{2}-1} h_j(v)\, dv \\ &\quad + A \log\left[\frac{\pi^{\frac{n-k}{2}}(n-k)s^2}{\Gamma\left(\frac{n-k}{2}\right)}\right] + \mathbb{E}\left[B(\phi)\middle| \mathbf{y}\right], \end{aligned}$$

which depends on h_i.

Let us suppose now that our interest is to select models to make inference about future observations $y_{n+1}, \ldots, y_m$. Thus, we will assume that the vector $\mathbf{y}$, as well as the matrix $\mathbf{X}$, are partitioned as

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_{(n)} \\ \mathbf{y}_{(m-n)} \end{pmatrix} \text{ and } \mathbf{X}_i = \begin{pmatrix} \mathbf{X}_{i(n)} \\ \mathbf{X}_{i(m-n)} \end{pmatrix}$$

with $\mathbf{y}_{(n)} = (y_1, \ldots, y_n)^t$ and $\mathbf{y}_{(m-n)} = (y_{n+1}, \ldots, y_m)^t$ and $\mathbf{X}_{i(n)}$, $\mathbf{X}_{i(m-n)}$ are $n \times k$ and $(m-n) \times k$ dimensional known design matrices. Also, $\mathbf{Y}\left|\mathbf{X}_i, \boldsymbol{\beta}_i, \phi^{-1}, h_i^m\right. \sim El_m\left(\mathbf{X}_i\boldsymbol{\beta}_i, \phi^{-1}\mathbf{I}_m; h_i^m\right)$ and we are comparing the models $M_i = \left(\mathbf{X}_i, h_i^m\right)$ and $M_j = \left(\mathbf{X}_j, h_j^m\right)$.

Proposition 1.3.11. *If $\boldsymbol{\omega} = (y_{n+1}, \ldots, y_m)$,*

$$\begin{aligned} \bar{u}\left(m_i \middle| \mathbf{y}_{(n)}\right) &= \frac{A}{\sum_{r=1}^{q} m_r^N\left(\mathbf{y}_{(n)}\middle| \mathbf{X}_r\right)} \int \log\left[f\left(\mathbf{y}_{(m-n)}\middle| \mathbf{y}_{(n)}, \mathbf{X}_i\right)\right] \\ &\quad \sum_{j=1}^{q} f\left(\mathbf{y}_{(m-n)}\middle| \mathbf{y}_{(n)}, \mathbf{X}_j\right) m_j^N\left(\mathbf{y}_{(n)}\middle| \mathbf{X}_j\right) d\mathbf{y}_{(m-n)} + \mathbb{E}\left[B\left(\mathbf{y}_{(m-n)}\right)\middle| \mathbf{y}_{(n)}\right], \end{aligned}$$

where $\mathbf{y}_{(m-n)}\middle| \mathbf{y}, \mathbf{X}_i \sim t_{m-n}\left(\mathbf{X}_{i(m-n)}\hat{\boldsymbol{\beta}}_i, S_i^2\mathbf{W}_i, n-k_i\right)$ and
$\mathbf{W}_i = \mathbf{X}_{i(m-n)}\left(\mathbf{X}_i^t\mathbf{X}_i\right)^{-1}\mathbf{X}_{i(m-n)}^t + \mathbf{I}_{m-n}$.

Proof. The result follows from Osiewalski and Steel (1993), where it is shown that

$$\mathbf{y}_{(m-n)} \big| \, \mathbf{y}_{(n)}, m_i \sim t_{m-n} \left(\mathbf{X}_{i(m-n)} \hat{\boldsymbol{\beta}}_i, S_i^2 \mathbf{W}_i, n - k_i \right),$$

which depends on m_i through $\mathbf{X}_i$ only. ■

Remark 1.3.12. *In the above proposition, if* $\mathbf{X}_i = \mathbf{X}$ *for all* $i = 1, \ldots, q$, *then*

$$\bar{u} \left(m_i | \, \mathbf{y}_{(n)} \right) = A \mathbb{E}_{t_{m-n}} \left[\log \left[f \left(\mathbf{y}_{(m-n)} \big| \, \mathbf{y}_{(n)}, \mathbf{X} \right) \right] \big| \, \mathbf{y}_{(n)} \right] + \mathbb{E}_{t_{m-n}} \left[B \left(\mathbf{y}_{(m-n)} \right) \big| \, \mathbf{y}_{(n)} \right],$$

where the expected value $\mathbb{E}_{t_{m-n}} (\cdot)$ *is taken with respect to the Student-t distribution*

$$t_{m-n} \left(\mathbf{X}_{(m-n)} \hat{\boldsymbol{\beta}}, S_i^2 \mathbf{W}, n - k \right).$$

We note from propositions 1.3.7 and 1.3.9 that the expected utility function $\bar{u} \left(m_i | \, \mathbf{y} \right)$ depends on the model m_i through $I(h_i) = \int \log \left[v^{\frac{n-k}{2}-1} h_i(v) \right] v^{\frac{n-k}{2}-1} h_j(v) \, dv$, which depends on the data only through the sample size n.

In general, the shape of the density (1.3.21) together with the fact that $f \left(\boldsymbol{\beta} | \, \mathbf{y}, M_i \right)$ and $f \left(y_{n+1}, \ldots, y_m | \, \mathbf{y}, M_i \right)$ do not depend on the elliptical model h_i, and $f \left(\boldsymbol{\beta}, \phi | \, \mathbf{y}, M_i \right) \propto v^{\frac{n}{2}-1} h_i(v)$ and $f \left(\phi | \, \mathbf{y}, M_i \right) \propto u^{\frac{n-k_i}{2}-1} h_i(u)$ with $v = \phi \left\| \mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right\|^2$ and $u = \phi \left\| \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right\|^2$ respectively, is not useful when selecting the most appropriate model after having observed the data $\mathbf{y}$.

The comparison of elliptical models for the errors using the marginal densities, in linear models with prior distribution $\pi \left(\boldsymbol{\beta}, \phi \right) \propto \phi^{-1}$, should be reexamined. On the other hand, if we have chosen an elliptical model, that is to say h_i is fixed, then the comparison is centered in the design matrices, and the comparison could be carried out satisfactorily using the IBF, the FBF or maximizing (1.3.20): in the case of the IBF's, the well-known results of Berger and Pericchi (1996b) could be used, and in the case of the FBF's convenient formulas can be obtained for many models.

Concluding we present a Bayesian analysis of the elliptical linear model under different prior specifications for the parameters. We show that when using squared-radial distributions for ϕ with $\boldsymbol{\beta}\perp\!\!\!\perp\phi|h^{(n)}$ and $\pi(\boldsymbol{\beta}|h^{(n)})$ the posterior of $\boldsymbol{\beta}$ does not depend on $h^{(n)}$. Hence, the inference on $\boldsymbol{\beta}$ is the same as the one obtained under normality. Only the posterior of ϕ depends on $h^{(n)}$, even under improper prior considered here.

Moreover, the IBF to compare two elliptical linear models (with common design matrix) does not work, because the predictive distributions are the same for the models under comparison. On the other hand, even though the FBF depends on $h^{(n)}$, it depends on the data only through the sample size. Similar results are obtained when we adopt the perspective of decision analysis for model comparison. Other alternative methods for nested hypotheses testing that must be explored are presented by Bernardo (1999) and Pereira and Stern (1999), because these procedures involve all parameters in the models being compared.

Thus, many results obtained under the normal model remain valid under dependent elliptical models. In general, the results derived here for the dependent elliptical models do not hold for poly-elliptical models.

In this chapter we specify a conditional distribution for $\mathbf{Y}|\,\mathbf{X}, \boldsymbol{\beta}, \phi, h^{(n)}$ and a prior for $(\boldsymbol{\beta}, \phi)|\,h^{(n)}$, in such a way that $\mathbf{Y}\perp\!\!\!\perp h^{(n)}$, considering $h^{(n)}$ as random (i.e., $h^{(n)}$ is marginally ancillary). Thus, any procedure for model comparisons that is based on the predictive distributions would be not useful to discriminating among different density generators. Even if we introduce a prior for $h^{(n)}$, this would not be updated under the hypotheses imposed in this chapter. On the other hand, it becomes clear that Bayesian model comparisons should include not only the predictive distribution, but also all the model components.

Chapter 2

On the Existence of Measurement Error

The problem of estimating parameters in the regression of a response variable $\mathbf{Y}$ on an explanatory variable $\boldsymbol{\xi}$ from observations on $(\mathbf{Y}, \mathbf{X})$, where $\mathbf{X}$ is a measurement of $\boldsymbol{\xi}$, is a special case of what has historically been called errors-in-variables problem.

In many practical situations, measurement processes are subject to measurement errors, in particular, while collecting information related to a phenomenon that could be described through a regression model. In this case, the predictor variables would become unknown parameters causing a decrease of the parsimony and a greater complexity of the model.

The standard measurement error regression model with one explanatory variable can be expressed by

$$Y_i = \alpha + \beta \xi_i + \epsilon_i, \qquad (2.0.1)$$

$i = 1, \ldots, n$, where the ξ_i are unobservable, however they are related with observable variables X_i by the equation

$$X_i = \xi_i + u_i, \tag{2.0.2}$$

$i = 1, \ldots, n$.

Letting $\mathbf{Y} = (Y_1, \ldots, Y_n)^t$, $\mathbf{X} = (X_1, \ldots, X_n)^t$, $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)^t$, $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^t$ and $\mathbf{u} = (u_1, \cdots, u_n)^t$, model (2.0.1) and (2.0.2) can be represented in matricial form as follows

$$\begin{aligned} \mathbf{Y} &= \alpha + \beta\boldsymbol{\xi} + \boldsymbol{\epsilon} \\ \mathbf{X} &= \boldsymbol{\xi} + \mathbf{u}. \end{aligned} \tag{2.0.3}$$

As usual, it is assumed that

$$\begin{pmatrix} \epsilon_i \\ u_i \end{pmatrix} \overset{iid}{\sim} N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right], \qquad i = 1, \ldots, n.$$

Thus, for this model we have two groups of parameters: $(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2)$ called structural parameters and $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)^t$ called latent (or incidental) parameters. The model (2.0.3) with fixed ξ_i is called functional model, and if we assume that due to a sampling process the ξ_i are random, as for example $\xi_i \overset{iid}{\sim} N(\mu, \tau^2)$, this model is called structural. A complete and unified treatment which also discusses possible applications is presented in Fuller (1987) and more recently in Cheng and Ness (1999).

The statistical treatment of this model is not easy, since the dimension of the parametric space increases with the sample size. Notice that if in the functional MEM, $\sigma_u^2 = 0$, then it becomes a normal simple linear regression model (NSLRM), while in the structural MEM, it becomes a model with random effects (normal simple linear regression model with random explanatory variables). Because simplifications are obtained when $\sigma_u^2 = 0$, we would want to build procedures that allow to compare between the simplest model or the most complex, or equivalently to compare the hypothesis $H_0 : \sigma_u^2 = 0$ versus $H_1 : \sigma_u^2 > 0$.

It is well known that the functional model as well as the structural model are unidentifiable, therefore, from the classic point of view, unless additional assumptions are made on the parameters, it is impossible to solve the proposed problem above. This situation is widely discussed in the literature, please refer to e.g. Fuller (1987) and Cheng and Ness (1999). From a Bayesian perspective, the problem of nuisance parameters is solved marginalizing in the posterior distribution, and the exploratory model comparison can be done by computing Bayes factors. To the best of our knowledge, this problem has not been treated in the Bayesian literature.

The chapter is organized as follows. In Sections 2.1 we compute the BF to compare the hypothesis $H_0 : \sigma_u^2 = 0$ versus $H_1 : \sigma_u^2 > 0$ under functional model and in Section 2.2, for structural model. Section 2.3 is devote to define computational strategies. A version of Importance Sampling method is used to compute the BF presented in Sections 2.1 and 2.2. In Section 2.4 the behavior of the method is evaluated through simulations. Finally, we illustrate the obtained results with an application to real data in the field of Agriculture.

2.1 Bayes Factor in the Functional MEM

In this section we consider the problem of comparing the hypotheses

$$\begin{cases} H_0 : \sigma_u^2 = 0, \alpha, \beta, \sigma_\epsilon^2 \\ H_1 : \sigma_u^2 > 0, \alpha, \beta, \sigma_\epsilon^2, \boldsymbol{\xi} \end{cases}, \tag{2.1.4}$$

in the functional MEM, which is equivalent to compare the models M_0 : NSLRM and M_1 : functional MEM, that is to compare,

$$\begin{cases} M_0 : Y_i \overset{indep.}{\sim} N\left(\alpha + \beta x_i, \sigma_\epsilon^2\right), \ i = 1, \ldots, n \\ \text{versus} \\ M_1 : \begin{pmatrix} Y_i \\ X_i \end{pmatrix} \overset{indep.}{\sim} N_2\left(\begin{pmatrix} \alpha + \beta\xi_i \\ \xi_i \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right), \ i = 1, \ldots, n, \end{cases}$$

where x_i denote the observed value of X_i. Note that, under the M_0 model, we obtain $x_i = \xi_i$.

Denote the likelihood functions by $p\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \boldsymbol{\xi}\right) = N_n\left(\mathbf{y} \middle| \alpha, \beta, \sigma_\epsilon^2\right)$ for M_0 model and, $p\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2, \boldsymbol{\xi}\right) = N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2, \boldsymbol{\xi}\right)$ for M_1 model, then the BF is given by

$$B_{01}^{FM} = \frac{\int N_n\left(\mathbf{y} \middle| \alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) d\alpha d\beta d\sigma_\epsilon^2}{\int N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2, \boldsymbol{\xi}\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2, \boldsymbol{\xi}\right) d\alpha d\beta d\sigma_\epsilon^2 d\sigma_u^2 d\boldsymbol{\xi}}.$$

Observe that the dimension of the integral of the denominator is very high due to the presence of the vector of latent parameters $\boldsymbol{\xi}$. However, if we assume a prior specification such that $\boldsymbol{\xi} \perp\!\!\!\perp (\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2)$ and $\xi_i \overset{iid}{\sim} N\left(\mu, \tau^2\right)$, with μ and τ^2 known, we could integrate on the space of the parameters $\boldsymbol{\xi}$ to obtain a simpler shape of the BF. With these assumptions, we obtain $\left.\binom{Y_i}{X_i}\right| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2$ are conditionally *i.i.d.* normal distributed, with mean and variance given by

$$\begin{aligned} \mathbb{E}\left\{\left.\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\right| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right\} &= \mathbb{E}\left\{\mathbb{E}\left[\left.\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\right| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2, \boldsymbol{\xi}\right]\right\} = \mathbb{E}\left\{\begin{pmatrix} \alpha + \beta\xi_i \\ \xi_i \end{pmatrix}\right\} \\ &= \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}\left\{\left.\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\right| \alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right\} &= \mathbb{V}\left\{\mathbb{E}\left[\left.\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\right| \alpha,\beta,\sigma_\epsilon^2,\sigma_u^2,\boldsymbol{\xi}\right]\right\} \\
&\quad +\mathbb{E}\left\{\mathbb{V}\left[\left.\begin{pmatrix} Y_i \\ X_i \end{pmatrix}\right| \alpha,\beta,\sigma_\epsilon^2,\sigma_u^2,\boldsymbol{\xi}\right]\right\} \\
&= \mathbb{V}\left\{\begin{pmatrix} \alpha+\beta\xi_i \\ \xi_i \end{pmatrix}\right\} + \mathbb{E}\left\{\begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}\right\} \\
&= \begin{pmatrix} \beta^2\tau^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 \end{pmatrix} + \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \\
&= \begin{pmatrix} \beta^2\tau^2+\sigma_\epsilon^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2+\sigma_u^2 \end{pmatrix}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p\left(\mathbf{y},\mathbf{x}\right|\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2) &= \int p\left(\mathbf{y},\mathbf{x}\right|\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2,\boldsymbol{\xi})\,\pi\left(\boldsymbol{\xi}\right)d\boldsymbol{\xi} \\
&= N_{2n}\left[\left.\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}\right| \mathbb{1}_n \otimes \begin{pmatrix} \alpha+\beta\mu \\ \mu \end{pmatrix}, \mathbf{I}_n \otimes \begin{pmatrix} \beta^2\tau^2+\sigma_\epsilon^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2+\sigma_u^2 \end{pmatrix}\right],
\end{aligned}$$

where $\otimes$ denote the Kronecker product of two matrixes. The p.d.f. of previous distribution we will denote by $N_{2n}\left(\mathbf{y},\mathbf{x}\right|\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2)$. Then, we obtain

$$B_{01}^{FM} = \frac{\int N_n\left(\mathbf{y}\right|\alpha,\beta,\sigma_\epsilon^2)\,\pi\left(\alpha,\beta,\sigma_\epsilon^2\right)d\alpha d\beta d\sigma_\epsilon^2}{\int N_{2n}\left(\mathbf{y},\mathbf{x}\right|\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2)\,\pi\left(\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right)d\alpha d\beta d\sigma_\epsilon^2 d\sigma_u^2}. \tag{2.1.5}$$

As we can see, the formula (2.1.5) is friendlier since the dimension of the integral of the denominator does not depend of n. These integrals are not easy to calculate even if we consider standard prior distributions as Normal, Inverted-Gamma, etc. The computational implementation will be discussed in Section 2.3. However, if we assume the prior distribution

$$\pi\left(\alpha,\beta,\sigma_\epsilon^2\right) = \pi\left(\alpha,\beta\left|\sigma_\epsilon^2\right.\right)\pi\left(\sigma_\epsilon^2\right) = N_2\left[\left.\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right|\begin{pmatrix} a_0 \\ b_0 \end{pmatrix}, \sigma_\epsilon^2\mathbf{B}\right] \times IGa\left(\sigma_\epsilon^2\left|a_\epsilon, b_\epsilon\right.\right),$$

where $\mathbf{B}$ is a known 2×2 matrix, the integral of the numerator can be solved in a closed form, since it is well known that

$$f\left(\boldsymbol{\beta}, \sigma_\epsilon^2 \left|\mathbf{y}, \mathbf{X}\right.\right) \propto \frac{b_\epsilon^{a_\epsilon} \left(\sigma_\epsilon^{-2}\right)^{\frac{n+2}{2}+a_\epsilon+1}}{\Gamma\left(a_\epsilon\right)(2\pi)^{\frac{n+2}{2}} \left|\mathbf{B}\right|^{1/2}} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t \mathbf{V}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left[\mathbf{y}^t\mathbf{y} + \boldsymbol{\beta}_0^t\mathbf{B}^{-1}\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}^t\mathbf{V}\hat{\boldsymbol{\beta}} + 2b_\epsilon\right]\right\},$$

where $\boldsymbol{\beta} = (\alpha, \beta)^t$, $\mathbf{V} = \mathbf{X}^t\mathbf{X} + \mathbf{B}^{-1}$, $\boldsymbol{\beta}_0 = (a_0, b_0)^t$, $\hat{\boldsymbol{\beta}} = \mathbf{V}^{-1}\left(\mathbf{X}^t\mathbf{y} + \mathbf{B}^{-1}\boldsymbol{\beta}_0\right)$ and

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

In this case,

$$f(\mathbf{y}) = \int N_n\left(\mathbf{y}\right|\alpha, \beta, \sigma_\epsilon^2\left.\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) d\alpha d\beta d\sigma_\epsilon^2 = \frac{(2b_\epsilon)^{a_\epsilon}\Gamma\left(\frac{n}{2}+a_\epsilon\right)}{\Gamma\left(a_\epsilon\right)\pi^{n/2} b^{n/2+a_\epsilon} \left|\mathbf{B}\right|^{1/2} \left|\mathbf{V}\right|^{1/2}},$$

where $b = \mathbf{y}^t\mathbf{y} + \boldsymbol{\beta}_0^t\mathbf{B}^{-1}\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}^t\mathbf{V}\hat{\boldsymbol{\beta}} + 2b_\epsilon$. Also, it can be show that the predictive distribution is

$$f(\mathbf{y}) = t_n\left(\mathbf{y}\left|\hat{\mathbf{y}}, \frac{r}{2a_\epsilon}\boldsymbol{\Sigma}^{-1}, 2a_\epsilon\right.\right),$$

where $\hat{\mathbf{y}} = \boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{V}^{-1}\mathbf{B}^{-1}\boldsymbol{\beta}_0$, $\boldsymbol{\Sigma} = \mathbf{I}_n - \mathbf{X}\mathbf{V}^{-1}\mathbf{X}^t$ and $r = \boldsymbol{\beta}_0^t\left[\mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{V}^{-1}\mathbf{B}^{-1}\right]\boldsymbol{\beta}_0 - \hat{\mathbf{y}}^t\boldsymbol{\Sigma}\hat{\mathbf{y}} + 2b_\epsilon$.

The denominator of equation (2.1.5) can be estimated using the procedure described by Chib and Jeliazkov (2001).

2.2 Bayes Factor in the Structural MEM

In this section we consider the structural MEM, that is to say we assume that $(\epsilon_i, u_i, \xi_i)^t \overset{iid}{\sim} N_3\left[(0,0,\mu)^t, \operatorname{diag}\left(\sigma_\epsilon^2, \sigma_u^2, \tau^2\right)\right]$, where $i = 1, \ldots, n$ and $\operatorname{diag}\left(\sigma_\epsilon^2, \sigma_u^2, \tau^2\right)$

is a diagonal matrix with the given elements on the diagonal. Here, μ and τ will be also considered known.

Observe that under assumed prior distributions in both cases, functional and structural, if $\sigma_u^2 > 0$ then the predictive distribution of observations are equals, making the functional and structural MEM equivalent. However, if $\sigma_u^2 = 0$, the functional MEM becomes a NSLRM, and the structural, in a NSLRM with random predictor variables.

It is easy to see that when $\sigma_u^2 > 0$ the likelihood function for structural MEM is defined by

$$M_1 : \begin{pmatrix} Y_i \\ X_i \end{pmatrix} \overset{iid}{\sim} N_2 \left(\begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta^2\tau^2 + \sigma_\epsilon^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma_u^2 \end{pmatrix} \right), i = 1, \ldots, n$$

and when $\sigma_u^2 = 0$, it is obtained the model with random effects, $M_0 : Y_i = \alpha + \beta X_i + \epsilon_i$, where $X_i \overset{iid}{\sim} N(\mu, \tau^2)$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$ and $i = 1, \ldots, n$. Then, the problem to compare these two models is equivalent to compare (2.1.4).

In a similar way to the previous case, it can be proved that

$$B_{01}^{SM} = \frac{\int N_{2n} \left(\mathbf{y}, \mathbf{x} \right| \alpha, \beta, \sigma_\epsilon^2) \, \pi(\alpha, \beta, \sigma_\epsilon^2) \, d\alpha d\beta d\sigma_\epsilon^2}{\int N_{2n} \left(\mathbf{y}, \mathbf{x} \right| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2) \, \pi(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2) \, d\alpha d\beta d\sigma_\epsilon^2 d\sigma_u^2}, \tag{2.2.6}$$

where

$$N_{2n} \left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2 \right) = N_{2n} \left[\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \middle| \mathbb{1}_n \otimes \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \mathbf{I}_n \otimes \begin{pmatrix} \beta^2\tau^2 + \sigma_\epsilon^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 \end{pmatrix} \right].$$

Notice that the denominator of (2.2.6) is equal to one of (2.1.5) due to the similarity, among the functional and structural models, that we mentioned at the beginning. In the next section we discuss implementation issues.

Also, since

$$
\begin{aligned}
N_{2n}\left(\mathbf{y},\mathbf{x}\middle|\,\alpha,\beta,\sigma_\epsilon^2,\mu,\tau^2\right) &= N_n\left(\mathbf{y}\middle|\,\alpha,\beta,\sigma_\epsilon^2,\mu,\tau^2,\mathbf{x}\right)\cdot N_n\left(\mathbf{x}\middle|\,\alpha,\beta,\sigma_\epsilon^2,\mu,\tau^2\right)\\
&= N_n\left(\mathbf{y}\middle|\,\alpha,\beta,\sigma_\epsilon^2,\mu,\tau^2\right)\cdot N_n\left(\mathbf{x}\middle|\,\mu,\tau^2\right),
\end{aligned}
$$

then

$$
B_{01}^{SM} = N_n\left(\mathbf{x}\middle|\,\mu,\tau^2\right) B_{01}^{FM}, \tag{2.2.7}
$$

where $N_n\left(\mathbf{x}|\,\mu,\tau^2\right)$ is the p.d.f. of $N_n\left(\mu\mathbb{1}_n,\tau^2\mathbf{I}_n\right)$.

2.3 Computational Strategy

The integration methods Monte Carlo and Importance Sampling are less precise and are more computational demanding than the quadrature methods, although in this case are feasible to use because the complexity of the models hinders to use other integration methods. Generally, in this type of computational methods is required of evaluating of the likelihood function or the posterior density. Even, the dimensionality of the integrals in the expressions (2.1.5) and (2.2.6) is high, computational evaluation of the integrand is not complex because the covariance matrix has a friendly structure. In fact, note that

$$
\begin{aligned}
N_{2n}\left(\mathbf{y},\mathbf{x}\middle|\,\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right) &= N_{2n}\left[\left(\begin{array}{c}\mathbf{y}\\ \mathbf{x}\end{array}\right)\middle|\,\mathbb{1}_n\otimes\left(\begin{array}{c}\alpha+\beta\mu\\ \mu\end{array}\right),\mathbf{I}_n\otimes\left(\begin{array}{cc}\beta^2\tau^2+\sigma_\epsilon^2 & \beta\tau^2\\ \beta\tau^2 & \tau^2+\sigma_u^2\end{array}\right)\right]\\
&= \prod_{i=1}^n N_2\left[\left(\begin{array}{c}y_i\\ x_i\end{array}\right)\middle|\left(\begin{array}{c}\alpha+\beta\mu\\ \mu\end{array}\right),\left(\begin{array}{cc}\beta^2\tau^2+\sigma_\epsilon^2 & \beta\tau^2\\ \beta\tau^2 & \tau^2+\sigma_u^2\end{array}\right)\right]\\
&= \prod_{i=1}^n N_2\left(y_i,x_i\middle|\,\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
N_{2n}\left(\mathbf{y},\mathbf{x}\right|\alpha,\beta,\sigma_\epsilon^2\left.\right) &= \prod_{i=1}^{n} N_2\left(y_i,x_i\right|\alpha,\beta,\sigma_\epsilon^2\left.\right) = \prod_{i=1}^{n} N\left(x_i\right|\mu,\tau^2\left.\right) \times \prod_{i=1}^{n} N\left(y_i\right|\alpha,\beta,\sigma_\epsilon^2\left.\right) \\
&= N_n\left(\mathbf{x}\left|\mu\mathbb{1}_n,\tau^2\mathbf{I}_n\right.\right) N_n\left(\mathbf{y}\right|\alpha,\beta,\sigma_\epsilon^2\left.\right).
\end{aligned}
$$

Then, the three likelihoods involved in the comparison problem, can be written as a product of likelihoods which are easy to evaluate numerically.

We note that from expressions (2.1.5) and (2.2.6), the dimension of the integral of the numerator is one minus than that of the denominator. In these cases there are specific variants to calculate the Bayes factor. A summary of some of these methods can be found in Chen *et al.* (2000). We will use one of these methods which is a variant of the Importance Sampling method. We choose this method because it requires only the generation of one posterior distribution, the corresponding to the most complex model.

2.3.1 Importance Sampling Extended to Hypotheses with Different Dimensions

Now, we will explain the method that we will use to find estimators of the Bayes factors of interest. For details about the method, see Chen *et al.* (2000).

Suppose that we want to test the hypotheses $H_0 : \omega = \omega_0$ versus $H_1 : \omega \neq \omega_0$ where $\boldsymbol{\theta} = (\omega, \boldsymbol{\psi}) \in \Theta = \Omega \times \Psi$, $\pi_0(\boldsymbol{\psi})$ is the prior p.d.f. for $\boldsymbol{\psi}$ under H_0 and $\pi(\omega, \boldsymbol{\psi})$ is the prior density to $\boldsymbol{\theta}$ under H_1. Observe that the Bayes factor to compare both hypotheses is,

$$
B_{01} = \frac{\int p_0\left(\mathbf{D}\right|\boldsymbol{\psi}\left.\right)\pi_0\left(\boldsymbol{\psi}\right)d\boldsymbol{\psi}}{\int p_1\left(\mathbf{D}\right|\omega,\boldsymbol{\psi}\left.\right)\pi\left(\omega,\boldsymbol{\psi}\right)d\omega d\boldsymbol{\psi}},
$$

where $\mathbf{D}$ denotes the observed data. Now, if we find a p.d.f. $g\left(\omega\right|\boldsymbol{\psi}\left.\right)$, since $\int g\left(\omega\right|\boldsymbol{\psi}\left.\right)d\omega =$

1, then the previous Bayes factor can be expressed as

$$\begin{aligned}
B_{01} &= \frac{\int p_0\left(\mathbf{D}|\,\boldsymbol{\psi}\right)\pi_0\left(\boldsymbol{\psi}\right)g\left(\omega|\,\boldsymbol{\psi}\right)d\omega d\boldsymbol{\psi}}{\int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)d\omega d\boldsymbol{\psi}} = \int\left[\frac{p_0\left(\mathbf{D}|\,\boldsymbol{\psi}\right)\pi_0\left(\boldsymbol{\psi}\right)g\left(\omega|\,\boldsymbol{\psi}\right)}{\int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)d\omega d\boldsymbol{\psi}}\right]d\omega d\boldsymbol{\psi} \\
&= \int\frac{p_0\left(\mathbf{D}|\,\boldsymbol{\psi}\right)\pi_0\left(\boldsymbol{\psi}\right)g\left(\omega|\,\boldsymbol{\psi}\right)}{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)}\frac{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)}{\int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)d\omega d\boldsymbol{\psi}}d\omega d\boldsymbol{\psi} \\
&= \int\frac{p_0\left(\mathbf{D}|\,\boldsymbol{\psi}\right)\pi_0\left(\boldsymbol{\psi}\right)g\left(\omega|\,\boldsymbol{\psi}\right)}{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)}p_1\left(\omega,\boldsymbol{\psi}|\,\mathbf{D}\right)d\omega d\boldsymbol{\psi} \\
&= \mathbb{E}_1\left\{\frac{p_0\left(\mathbf{D}|\,\boldsymbol{\psi}\right)\pi_0\left(\boldsymbol{\psi}\right)g\left(\omega|\,\boldsymbol{\psi}\right)}{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)}\right\},
\end{aligned}$$

where $\mathbb{E}_1\left(\cdot\right)$ is the expected value with respect to $p_1\left(\omega,\boldsymbol{\psi}|\,\mathbf{D}\right)$.

Chen *et al.* (2000) show that an optimal selection of $g\left(\omega|\,\boldsymbol{\psi}\right)$, in the sense of minimizes the asymptotic relative mean-square error, is by taking

$$\begin{aligned}
g\left(\omega|\,\boldsymbol{\psi}\right) &= p_1\left(\omega|\,\boldsymbol{\psi},\mathbf{D}\right) = \frac{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega|\,\boldsymbol{\psi}\right)}{\int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega|\,\boldsymbol{\psi}\right)d\omega} \\
&= \frac{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\frac{\pi(\omega,\boldsymbol{\psi})}{\pi(\boldsymbol{\psi})}}{\int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\frac{\pi(\omega,\boldsymbol{\psi})}{\pi(\boldsymbol{\psi})}d\omega} = \frac{p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)}{\int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)d\omega}. \qquad (2.3.8)
\end{aligned}$$

In such case, the asymptotic relative mean-square error is defined by

$$ARE^2\left(\hat{B}_{01}\right) = \lim_{n\to\infty} n\frac{\mathbb{E}\left(\hat{B}_{01}-B_{01}\right)^2}{B_{01}^2}. \qquad (2.3.9)$$

If we choose g as in (2.3.8), then

$$B_{01} = \mathbb{E}_1\left\{\frac{p_0\left(\mathbf{D}|\,\boldsymbol{\psi}\right)\pi_0\left(\boldsymbol{\psi}\right)}{c\left(\boldsymbol{\psi}\right)}\right\}, \qquad (2.3.10)$$

where $c\left(\boldsymbol{\psi}\right) = \int p_1\left(\mathbf{D}|\,\omega,\boldsymbol{\psi}\right)\pi\left(\omega,\boldsymbol{\psi}\right)d\omega$ and therefore, a Monte Carlo estimator is

$$\hat{B}_{01opt} = \frac{1}{m}\sum_{i=1}^{m}\frac{p_0\left(\mathbf{D}|\,\boldsymbol{\psi}^{(i)}\right)\pi_0\left(\boldsymbol{\psi}^{(i)}\right)}{c\left(\boldsymbol{\psi}^{(i)}\right)},$$

where $\left\{\left(\omega^{(i)}, \boldsymbol{\psi}^{(i)}\right), i = 1, \cdots, m\right\}$ is a random sample from $p_1\left(\omega, \boldsymbol{\psi} \middle| \mathbf{D}\right)$. This method is, in general, computational demanding, since for each term of the sum, it is necessary to estimate $c\left(\boldsymbol{\psi}^{(i)}\right)$.

However, for our case, taken into account that (2.1.5) and (2.2.6) have the same denominator, and if we assume $\sigma_u^2 \perp\!\!\!\perp (\alpha, \beta, \sigma_\epsilon^2)$ then, we obtain in both cases that

$$
\begin{aligned}
p_1\left(\omega, \boldsymbol{\psi} \middle| \mathbf{D}\right) &= p_1\left(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2 \middle| \mathbf{y}, \mathbf{x}\right) \\
&= \frac{p_1\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right)}{\int p_1\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) d\alpha d\beta d\sigma_\epsilon^2 d\sigma_u^2} \\
&= \frac{N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\sigma_u^2\right)}{\int N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\sigma_u^2\right) d\alpha d\beta d\sigma_\epsilon^2 d\sigma_u^2},
\end{aligned}
\tag{2.3.11}
$$

$$
\begin{aligned}
g\left(\omega \middle| \boldsymbol{\psi}\right) &= p_1\left(\omega \middle| \boldsymbol{\psi}, \mathbf{D}\right) = \frac{p_1\left(\mathbf{D} \middle| \omega, \boldsymbol{\psi}\right) \pi\left(\omega, \boldsymbol{\psi}\right)}{\int p_1\left(\mathbf{D} \middle| \omega, \boldsymbol{\psi}\right) \pi\left(\omega, \boldsymbol{\psi}\right) d\omega} \\
&= \frac{p_1\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\sigma_u^2\right)}{\int p_1\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\sigma_u^2\right) d\sigma_u^2} \\
&= \frac{N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\sigma_u^2\right)}{\int N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\sigma_u^2\right) d\sigma_u^2}
\end{aligned}
$$

and

$$
\begin{aligned}
c\left(\boldsymbol{\psi}\right) &= \int p_1\left(\mathbf{D} \middle| \omega, \boldsymbol{\psi}\right) \pi\left(\omega, \boldsymbol{\psi}\right) d\omega \\
&= \int N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\sigma_u^2\right) d\sigma_u^2.
\end{aligned}
$$

To calculate (2.1.5) we have

$$
p_0\left(\mathbf{D} \middle| \boldsymbol{\psi}\right) \pi_0\left(\boldsymbol{\psi}\right) = N_n\left(\mathbf{y} \middle| \alpha, \beta, \sigma_\epsilon^2\right) \pi\left(\alpha, \beta, \sigma_\epsilon^2\right)
$$

and from (2.3.10),

$$
B_{01}^{FM} = \mathbb{E}_1\left\{\frac{N_n\left(\mathbf{y} \middle| \alpha, \beta, \sigma_\epsilon^2\right)}{\int N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) \pi\left(\sigma_u^2\right) d\sigma_u^2}\right\}.
$$

Therefore, the corresponding Monte Carlo estimator is:

$$\hat{B}_{01opt}^{FM} = \frac{1}{m}\sum_{i=1}^{m}\frac{N_n\left(\mathbf{y}|\,\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)}\right)}{c\left(\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)}\right)}, \tag{2.3.12}$$

where

$$c\left(\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)}\right) = \frac{1}{rm}\sum_{j=1}^{rm} N_{2n}\left(\mathbf{y},\mathbf{x}|\,\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)},\sigma_u^{2(j)}\right),$$

$\left\{\left(\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)}\right), i=1,\cdots,m\right\}$ is a random sample draws from the posterior distribution (2.3.11) and $\left\{\sigma_u^{2(j)}, j=1,\cdots,rm\right\}$, from the prior distribution $\pi\left(\sigma_u^2\right)$.

From (2.2.7), the Monte Carlo estimator for (2.2.6) is given by

$$\hat{B}_{01opt}^{SM} = N_n\left(\mathbf{x}|\,\mu,\tau^2\right)\hat{B}_{01opt}^{FM} = \frac{1}{m}\sum_{i=1}^{m}\frac{N_{2n}\left(\mathbf{y},\mathbf{x}|\,\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)}\right)}{c\left(\alpha^{(i)},\beta^{(i)},\sigma_\epsilon^{2(i)}\right)}. \tag{2.3.13}$$

These last results can be summarized in the following two propositions.

Proposition 2.3.1. *An optimal estimator for the Bayes factor to compare a NSLRM against a MEM, under prior conditions* $(\alpha,\beta,\sigma_\epsilon^2)\perp\!\!\!\perp\sigma_u^2\perp\!\!\!\perp\boldsymbol{\xi}$ *and* $\boldsymbol{\xi}\sim N_n\left(\mu\mathbb{1}_n,\tau^2\mathbf{I}_n\right)$ *is given by (2.3.12).*

Proposition 2.3.2. *An optimal estimator for the Bayes factor to compare a NSLRM with normal random predictors against a MEM, under prior conditions* $(\alpha,\beta,\sigma_\epsilon^2)\perp\!\!\!\perp\sigma_u^2\perp\!\!\!\perp\boldsymbol{\xi}$ *and* $\boldsymbol{\xi}\sim N_n\left(\mu\mathbb{1}_n,\tau^2\mathbf{I}_n\right)$ *is given by (2.3.13).*

Here, the optimality is focused in the minimization of the asymptotic relative mean-square error, $ARE^2\left(\hat{B}_{01}\right)$, given by (2.3.9).

2.4 Simulation Results

The behavior of estimators established in (2.3.12) and (2.3.13) will be illustrated by using generated data from different MEM with prior distributions $\left.\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right|\sigma_\epsilon^2\sim$

$N_2\left[\left(\begin{array}{c} a_0 \\ b_0 \end{array}\right), \sigma_\epsilon^2 \mathbf{B}\right]$, $\sigma_\epsilon^2 \sim IGa\left(a_\epsilon, b_\epsilon\right)$, $\sigma_u^2 \sim IGa\left(a_u, b_u\right)$ and $\boldsymbol{\xi} \sim N_n\left(\mu \mathbb{1}_n, \tau^2 \mathbf{I}_n\right)$.

For functional and structural models, the sample $\left\{\left(\alpha^{(i)}, \beta^{(i)}, \sigma_\epsilon^{2(i)}, \sigma_u^{2(i)}\right)\right\}$, $i = 1, \cdots, m$, were drawn from the posterior distribution using the Metropolis-Hastings (M-H) algorithm (Metropolis *et al.* (1953) and Hastings (1970)) with initial values

$$\beta^{(0)} = \frac{\sum_{j=1}^n y_j x_j - n\bar{y}\bar{x}}{\sum_{j=1}^n \left(x_j - \bar{x}\right)^2},$$

$$\alpha^{(0)} = \bar{Y} - \beta^{(0)} \bar{x},$$

$$\sigma_\epsilon^{2(0)} = \frac{b_\epsilon + \frac{1}{2}\sum_{j=1}^n \left(y_j - \alpha^{(0)} - \beta^{(0)} x_j\right)^2}{a_\epsilon + \frac{n}{2}},$$

$$\sigma_u^{2(0)} = \frac{b_u + \frac{1}{2}\sum_{j=1}^n \left(x_j - \bar{x}\right)^2}{a_u + \frac{n}{2}}$$

and transition probability functions

$$\left(\begin{array}{c} \alpha^{(i+1)} \\ \beta^{(i+1)} \end{array}\right) \sim N_2\left[\left(\begin{array}{c} \alpha^{(i)} \\ \beta^{(i)} \end{array}\right), R\left(\alpha^{(i)}, \beta^{(i)}, \mathbf{y}, \mathbf{x}\right)\left(\begin{array}{cc} \bar{x}^2 + n^{-1}\sum_{j=1}^n \left(x_j - \bar{x}\right)^2 & -\bar{x} \\ -\bar{x} & 1 \end{array}\right)\right],$$

$$\sigma_\epsilon^{-2(i+1)} \sim Ga\left(1, \sigma_\epsilon^{2(i)}\right)$$

and

$$\sigma_u^{-2(i+1)} \sim Ga\left(1, \sigma_u^{2(i)}\right),$$

where $R\left(\alpha^{(i)}, \beta^{(i)}, \mathbf{y}, \mathbf{x}\right) = \frac{\sum_{j=1}^n \left(y_j - \alpha^{(i)} - \beta^{(i)} x_j\right)^2}{(n-2)\sum_{j=1}^n \left(x_j - \bar{x}\right)^2}$. To estimate the posterior distribution of the parameters, 2200 samples were generated. The initial 200 iterations were discarded to assure stationarity and a lag of 10 was selected to avoid autocorrelation. That means that a net sample size of 200 was used. The M-H algorithm was programmed in MATLAB package, version 6.0.0.88.

2.4.1 The Functional MEM

The data (y_j, x_j) , $j = 1, \ldots, 50$ were drawn from the following models

$$Y_j = 2.0 + 1.0\xi_j + \epsilon_j$$

where the values ξ_j were 5 replicates of the values (-4, -3, -2, -1, 0, 1, 2, 3, 4, 5), σ_u assumed values from 0.0 to 3.0 with step equal to 0.2 and $\sigma_\epsilon = \frac{\sigma_u}{10}$, except when $\sigma_u = 0$ where σ_ϵ was equal to 0.1. Since $\sigma_\epsilon < \sigma_u$, there are measurement error, then we hope that $B_{01} < 1$.

We consider the following prior specifications $\left.\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right| \sigma_\epsilon^2 \sim N_2\left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, 2\sigma_\epsilon^2 \mathbf{I}_2\right]$, $\sigma_\epsilon^2 \sim IGa\,(2, 0.1)$, $\sigma_u^2 \sim IGa\,(2, 1)$ and $\xi_j \overset{iid}{\sim} N\,(\mu, \tau^2)$, where $\mu = 0, 3$ and $\tau^2 = 1, 3, 5$. Table A.1 of Appendix A exhibits the Bayes factors computed for different combinations of μ and τ^2.

Figure 2.1 (the ordinates axis is multiplied by 10^{78}) shows the obtained results for $\mu = 0$ and $\tau^2 = 3$, in this case high values of BF are appreciated for $\sigma_u \leq 1.4$ and values approximately equal to zero when $\sigma_u \geq 1.6$, then for high values of σ_u the BF favors to MEM model as we expected. Similarly, it happened when $\mu = 0$ and $\tau^2 = 5$, see Table A.1. For the other cases, where prior information is not in agreement with the true values of ξ_j, the BF does not work well, see Table A.1. Since the true values of ξ_j are replicates of the values (-4, -3, -2, -1, 0, 1, 2, 3, 4, 5). A nice prior distribution for ξ_j would be an uniform.

These results show the high sensitivity of the BF with respect to prior distribution of ξ_j. This could be stimulating to explore non-subjective Bayesian methods for model comparison.

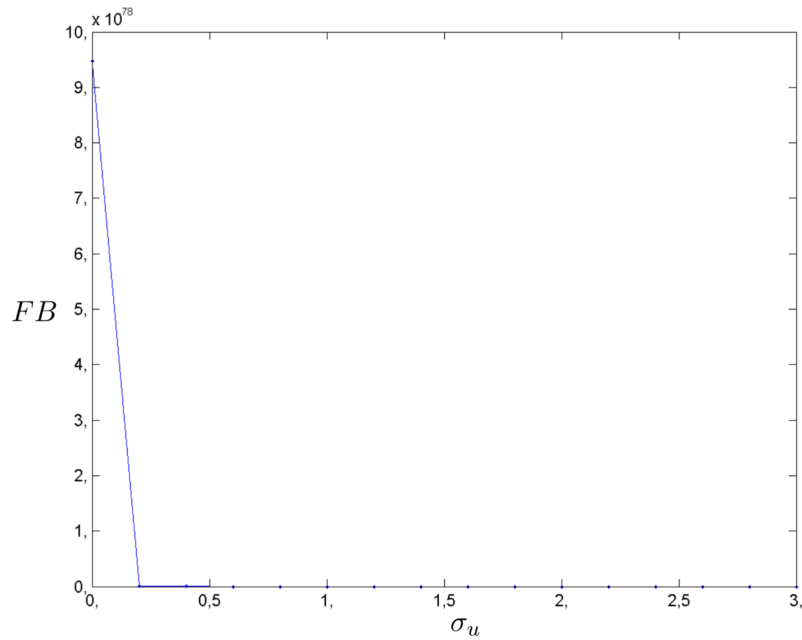

Figure 2.1: Optimal Bayes factors for functional MEM with $\mu = 0$ and $\tau^2 = 3$.

2.4.2 The Structural MEM

The data (y_j, x_j), $j = 1, \ldots, 50$ were drawn from the following models

$$Y_j = 2.0 + 1.0\xi_j + \epsilon_j$$

where the values ξ_j were drawn from a normal distribution $N(0, 9)$, σ_u assumed the values from 0.0 to 3.0 with step equal to 0.2 and σ_ϵ was selected according to the ratio $\delta = \frac{\sigma_\epsilon}{\sigma_u}$, where δ took the values 0.1, 1 and 5; when $\sigma_u = 0$ then σ_ϵ was equal to 0.1, 1 and 5, respectively.

We considered the prior distributions $\left(\begin{array}{c} \alpha \\ \beta \end{array} \right) \Big| \sigma_\epsilon^2 \sim N_2 \left[\left(\begin{array}{c} 2 \\ 1 \end{array} \right), 2\sigma_\epsilon^2 \mathbf{I}_2 \right]$, $\sigma_\epsilon^2 \sim IGa(2, 0.1)$ and $\sigma_u^2 \sim IGa(2, 1)$. Table A.2 of Appendix A shows the obtained Bayes factors for different combinations of δ.

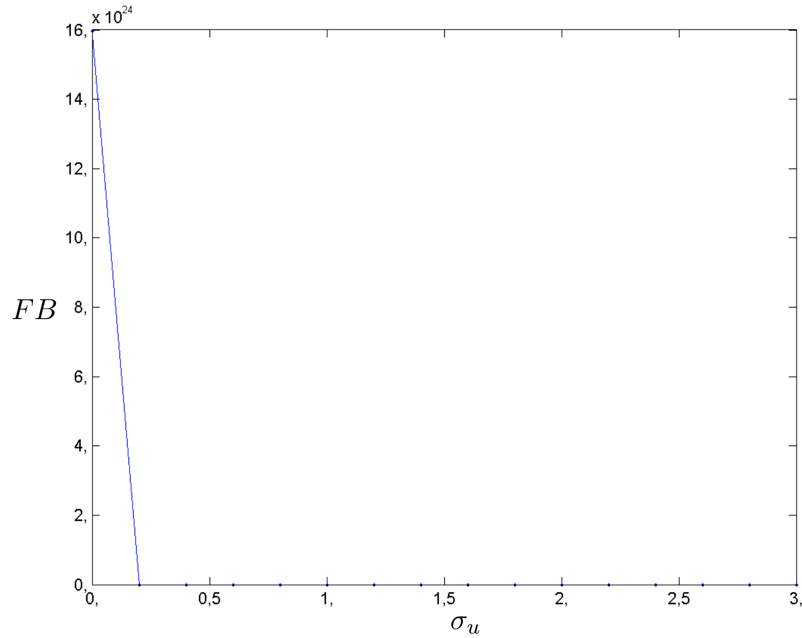

Figure 2.2: Optimal Bayes factors for structural MEM with $\delta = 0.1$.

Figures 2.2 to 2.4 show better the optimal Bayes factors values for each $\delta = 0.1, 1, 5$, respectively. From these results we appreciate the importance of ratio δ for testing the existence of measurement error, the results are only good when $\delta < 1$, that is to say, when the measurement error is more evident, $\sigma_\epsilon < \sigma_u$.

2.5 An Application

Now, we use the estimator (2.3.13) with real data to test the existence of measurement errors in variables. The data were taken from Fuller (1987, Chapter 1) and consist on areas under corn crop. Two different methods were digitized: aerial photography and personal interview with the farm operator. We denoted the hectares of corn

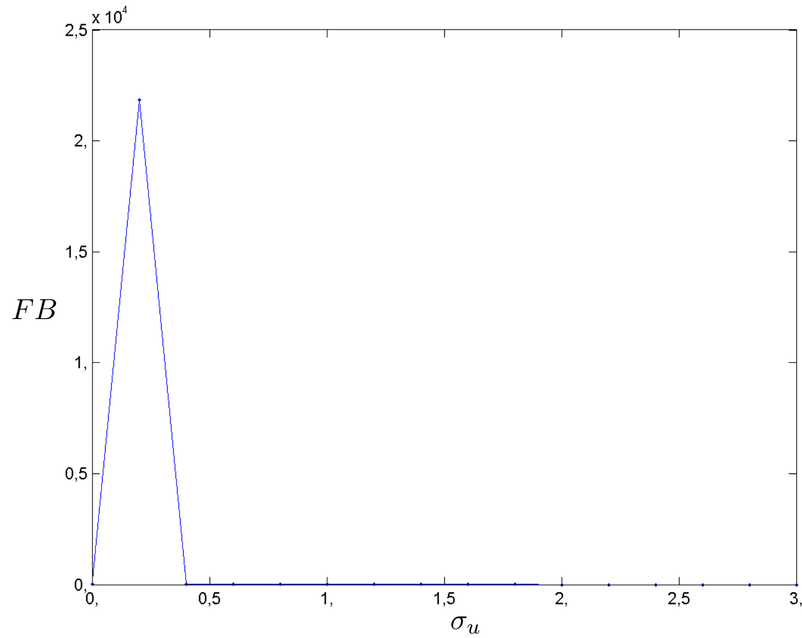

Figure 2.3: Optimal Bayes factors for structural MEM with $\delta = 1$.

determined for an area segment by Y_i for aerial photography and X_i, for personal interview. An area segment is an area of the earth's surface of approximately 250 hectares. Observations for a sample of 37 area segments are given in Table A.3 of Appendix A.

For the data description we should hope Bayes factor indicates existence of measurement error, and also, the model should be described by the equations,

$$\mathbf{Y} = \alpha + \beta\boldsymbol{\xi} + \boldsymbol{\epsilon}$$

$$\mathbf{X} = \boldsymbol{\xi} + \mathbf{u}.$$

Since two different methods were used to measure the same object, we decide to take a $N_2\left[\begin{pmatrix} 0 \\ 1 \end{pmatrix}, 3\sigma_\epsilon^2\mathbf{I}_2\right]$ as prior distribution for $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\Bigg| \sigma_\epsilon^2$ and as prior distribution

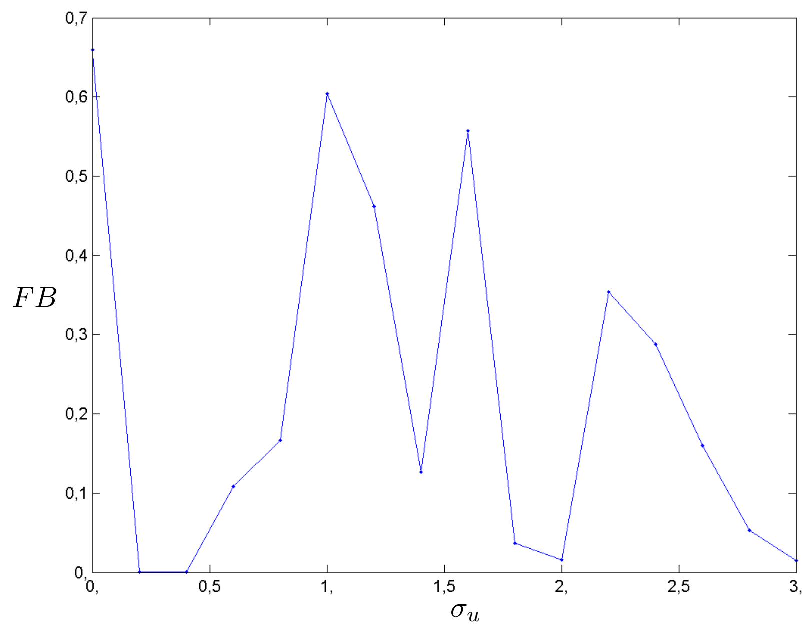

Figure 2.4: Optimal Bayes factors for structural MEM with $\delta = 5$.

for the planted hectare sizes in each area segment we consider a normal distribution centered in $\frac{250hect.}{2}$ and with variance $(32)^2$, that is to say $\boldsymbol{\xi} \sim N_{37}\left(125\mathbb{1}_{37}, 1024\mathbf{I}_{37}\right)$. The computations for the optimal Bayes factor were made under different Inverted-Gamma prior distributions for σ_ϵ^2 and σ_u^2 given in Table 2.1.

The results are shown in Table 2.2, where Column 1 displays the three sets of prior distributions considered in Table 2.1, Column 2 gives the optimal BF estimates from (2.3.13) and Columns 3 to 6 show the posterior expected values of α, β, σ_ϵ^2 and σ_u^2, respectively.

To make the computation of the optimal Bayes factor, 301,000 samples from the posterior distribution were generated with the M-H algorithm described in the previous section. The initial 1000 iterations were discarded to assure stationarity and

Table 2.1: Prior distributions for σ_ϵ^2 and σ_u^2 in the corn hectares example.

Set priors	Prior distributions	prior mean $\pm$ prior s.d.
I	$\sigma_\epsilon^2 \sim IGa\,(2.25, 12.5)$	10 ± 20
	$\sigma_u^2 \sim IGa\,(3, 40)$	20 ± 20
II	$\sigma_\epsilon^2 \sim IGa\,(2.0001, 10.001)$	10 ± 1000
	$\sigma_u^2 \sim IGa\,(2.0004, 20.008)$	20 ± 1000
III	$\sigma_\epsilon^2 \sim IGa\,(2, 10)$	$10 \pm \infty$
	$\sigma_u^2 \sim IGa\,(2, 20)$	$20 \pm \infty$

Table 2.2: Results for corn hectares data in structural MEM.

Set priors	$\hat{B}_{01opt}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_u^2$
I	0.1903	-0.9980	1.0346	13.3654	20.7529
II	0.2904	-0.9331	1.0339	14.9646	19.3656
III	0.3521	-0.9639	1.0342	14.8065	19.4108

a lag of 10 was selected to avoid autocorrelation. That means that a net sample size of 30,000 was used.

As we expected, the optimal Bayes factor, for these data and prior distributions considered, always favored the presence of measurement errors. Also, the parameters estimates are quite near to those obtained by Fuller (1987, Chapter 1).

2.6 Other Approaches

Because the BF is highly sensible with respect to the prior specification, several authors have proposed objective Bayesian methods for model comparison, as for example the Bayesian Information Criterion (BIC) by Schwarz (1978), the Fractional Bayes Factors (FBF) by O'Hagan (1995), the Intrinsic Bayes Factors (IBF) by Berger and Pericchi (1996a), the Bayesian Reference Criterion (BRC) by Bernardo (1999), among others.

Suppose that we are comparing two models $M_0 : f_0\,(\mathbf{y}\,|\boldsymbol{\theta}_0)$ and $M_1 : f_1\,(\mathbf{y}\,|\boldsymbol{\theta}_1)$

with non informative prior distribution $\pi_j^N(\boldsymbol{\theta}_j)$, $j = 0, 1$.

The BIC has the "advantages" of simplicity and freedom from prior distributions. It is given by

$$B_{01}^{IC} = \frac{f_0\left(\mathbf{y}\left|\hat{\boldsymbol{\theta}}_0\right.\right)}{f_1\left(\mathbf{y}\left|\hat{\boldsymbol{\theta}}_1\right.\right)} n^{(d_1-d_0)/2}, \tag{2.6.14}$$

where the $\hat{\boldsymbol{\theta}}_j$ is the maximum likelihood estimator (mle) of $\boldsymbol{\theta}_j$ and $d_j = \dim(\boldsymbol{\theta}_j)$. But, obtaining of an unique mle under MEM requires of additional assumptions over the model. One of these assumptions which will make the normal structural model identifiable is to assume the ratio of the error variances $\delta = \frac{\sigma_\epsilon}{\sigma_u}$ known. In this case the mle's are given by

$$\hat{\beta} = \frac{s_{yy} - \delta^2 s_{xx} + \left[(s_{yy} - \delta^2 s_{xx})^2 + 4\delta^2 s_{xy}^2\right]^{\frac{1}{2}}}{2s_{xy}},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x},$$

$$\hat{\sigma}_u^2 = \frac{s_{yy} - 2\hat{\beta}s_{xy} + \hat{\beta}^2 s_{xx}}{\delta^2 + \hat{\beta}^2},$$

$$\hat{\sigma}_\epsilon^2 = \delta^2 \hat{\sigma}_u^2,$$

$$\hat{\tau}^2 = \frac{s_{xy}}{\hat{\beta}}$$

and

$$\hat{\mu} = \bar{x},$$

where

$$s_{xx} = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

$$s_{yy} = n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

and

$$s_{xy} = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}),$$

see, for example, Zellner (1971, Chapter 5), Fuller (1987, Chapter 1) and Cheng and Ness (1999, Chapter 1).

On the other hand, the mle's under M_0 model, that is NSLRM with normal explanatory variables, are $\hat{\mu}_0 = \bar{x}$, $\hat{\tau}_0^2 = s_{xx}$, $\hat{\beta}_0 = \frac{s_{xy}}{s_{xx}}$, $\hat{\alpha}_0 = \bar{y} - \hat{\beta}_0 \bar{x}$ and $\hat{\sigma}^2_{\epsilon,0} = \frac{s_{yy}s_{xx} - s_{xy}^2}{s_{xx}}$. The following proposition shows that in this case the BIC do not discriminate between the models under comparison.

Proposition 2.6.1. *The BIC to compare a NSLRM with normal explanatory variables versus a structural MEM with ratio of measurement variances known is equal to one.*

Proof. In this case the BIC (2.6.14) is given by

$$B^{IC}_{01,\delta}(\mathbf{y}, \mathbf{x}) = \frac{N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \hat{\alpha}_0, \hat{\beta}_0, \hat{\sigma}^2_{\epsilon,0}, \hat{\mu}_0, \hat{\tau}_0^2\right)}{N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2_{\epsilon}, \hat{\sigma}^2_{u}, \hat{\mu}, \hat{\tau}^2\right)}.$$

Thus, we have to prove

$$\ln\left(B^{IC}_{01,\delta}(\mathbf{y}, \mathbf{x})\right) = l\left(\hat{\alpha}_0, \hat{\beta}_0, \hat{\sigma}^2_{\epsilon,0}, \hat{\mu}_0, \hat{\tau}_0^2\right) - l\left(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2_{\epsilon}, \hat{\sigma}^2_{u}, \hat{\mu}, \hat{\tau}^2\right) = 0,$$

where $l\left(\hat{\alpha}_0, \hat{\beta}_0, \hat{\sigma}^2_{\epsilon,0}, \hat{\mu}_0, \hat{\tau}_0^2\right) = \ln\left[N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \hat{\alpha}_0, \hat{\beta}_0, \hat{\sigma}^2_{\epsilon,0}, \hat{\mu}_0, \hat{\tau}_0^2\right)\right]$ and $l\left(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2_{\epsilon}, \hat{\sigma}^2_{u}, \hat{\mu}, \hat{\tau}^2\right) = \ln\left[N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \hat{\alpha}, \hat{\beta}, \hat{\sigma}^2_{\epsilon}, \hat{\sigma}^2_{u}, \hat{\mu}, \hat{\tau}^2\right)\right]$.

For M_0 model, the log-likelihood function evaluated in the corresponding mle's is given by

$$\begin{aligned}
l\left(\hat{\alpha}_0,\hat{\beta}_0,\hat{\sigma}^2_{\epsilon,0},\hat{\mu}_0,\hat{\tau}^2_0\right) &= -n\ln(2\pi)-\frac{n}{2}\ln\left(\hat{\tau}^2_0\hat{\sigma}^2_{\epsilon,0}\right)-\frac{1}{2\hat{\sigma}^2_{\epsilon,0}}\left\|\mathbf{y}-\left(\hat{\alpha}_0+\hat{\beta}_0\hat{\mu}_0\right)\mathbb{1}_n\right\|^2\\
&\quad+\frac{\hat{\beta}_0}{\hat{\sigma}^2_{\epsilon,0}}\left(\mathbf{y}-\left(\hat{\alpha}_0+\hat{\beta}_0\hat{\mu}_0\right)\mathbb{1}_n\right)^t(\mathbf{x}-\hat{\mu}_0\mathbb{1}_n)\\
&\quad-\frac{\hat{\beta}_0^2\hat{\tau}_0^2+\hat{\sigma}^2_{\epsilon,0}}{2\hat{\tau}^2_0\hat{\sigma}^2_{\epsilon,0}}\|\mathbf{x}-\hat{\mu}_0\mathbb{1}_n\|^2\\
&= -n\ln(2\pi)-\frac{n}{2}\ln\left(\hat{\tau}^2_0\hat{\sigma}^2_{\epsilon,0}\right)-\frac{ns_{yy}}{2\hat{\sigma}^2_{\epsilon,0}}+\frac{n\hat{\beta}_0 s_{xy}}{\hat{\sigma}^2_{\epsilon,0}}\\
&\quad-n\frac{\left(\hat{\beta}_0^2\hat{\sigma}^{-2}_{\epsilon,0}+\hat{\tau}_0^{-2}\right)s_{xx}}{2}\\
&= -n\left[1+\ln(2\pi)+\frac{1}{2}\ln\left(s_{yy}s_{xx}-s^2_{xy}\right)\right].
\end{aligned}$$

For M_1 model, we are assuming $\delta=\frac{\sigma_\epsilon}{\sigma_u}$ known. Thus, the log-likelihood function evaluated in the corresponding mle's is given by

$$\begin{aligned}
l\left(\hat{\alpha},\hat{\beta},\hat{\sigma}^2_u,\hat{\mu},\hat{\tau}^2\right) &= -n\ln(2\pi)-\frac{n}{2}\ln\left(\hat{D}\right)-\frac{\left(\hat{\tau}^2+\hat{\sigma}^2_u\right)}{2\hat{D}}\left\|\mathbf{y}-\left(\hat{\alpha}+\hat{\beta}\hat{\mu}\right)\mathbb{1}_n\right\|^2\\
&\quad+\frac{\hat{\beta}\hat{\tau}^2}{\hat{D}}\left(\mathbf{y}-\left(\hat{\alpha}+\hat{\beta}\hat{\mu}\right)\mathbb{1}_n\right)^t(\mathbf{x}-\hat{\mu}\mathbb{1}_n)\\
&\quad-\frac{\hat{\beta}^2\hat{\tau}^2+\delta^2\hat{\sigma}^2_u}{2\hat{D}}\|\mathbf{x}-\hat{\mu}\mathbb{1}_n\|^2\\
&= -n\ln(2\pi)-\frac{n}{2}\ln\left(\hat{D}\right)-\frac{n\left(\hat{\tau}^2+\hat{\sigma}^2_u\right)s_{yy}}{2\hat{D}}+\frac{n\hat{\beta}\hat{\tau}^2 s_{xy}}{\hat{D}}\\
&\quad-\frac{n\left(\hat{\beta}^2\hat{\tau}^2+\delta^2\hat{\sigma}^2_u\right)s_{xx}}{2\hat{D}},
\end{aligned}$$

where $\hat{D}=\hat{\sigma}^2_u\left(\hat{\tau}^2+\delta^2\hat{\tau}^2+\delta^2\hat{\sigma}^2_u\right)$. Now, using the well known relationships

$$s_{xx}=\hat{\tau}^2+\hat{\sigma}^2_u,$$

$$s_{yy}=\hat{\beta}^2\hat{\tau}^2+\delta^2\hat{\sigma}^2_u$$

and

$$s_{xy} = \hat{\beta}\hat{\tau}^2,$$

(see, for example, Zellner (1971, Chapter 5), Fuller (1987, Chapter 1) and Cheng and Ness (1999, Chapter 1)), we obtain

$$\begin{aligned} l\left(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_u^2, \hat{\mu}, \hat{\tau}^2\right) &= -n \ln(2\pi) - \frac{n}{2} \ln\left(s_{yy}s_{xx} - s_{xy}^2\right) - \frac{s_{xx}s_{yy}}{2\left(s_{yy}s_{xx} - s_{xy}^2\right)} \\ &\quad + \frac{s_{xy}^2}{s_{yy}s_{xx} - s_{xy}^2} - \frac{s_{xx}s_{yy}}{2\left(s_{yy}s_{xx} - s_{xy}^2\right)} \\ &= -n\left[1 + \ln(2\pi) + \frac{1}{2}\ln\left(s_{yy}s_{xx} - s_{xy}^2\right)\right], \end{aligned}$$

and therefore, we conclude the proof. ■

The general strategy for computing IBF's begins with the determination of a proper and minimal training sample, which is a subset of the entire data $\mathbf{y}$. Due to there are a variety of training samples, we index them by l.

A training sample, $\mathbf{y}(l)$, is called proper if $0 < m_j(\mathbf{y}(l)) < \infty$ for $j = 0, 1$, and minimal if it is proper and no subset is proper, where $m_j(\mathbf{y})$ is the corresponding marginal or predictive p.d.f.,

$$m_j(\mathbf{y}) = \int f_j(\mathbf{y}\,|\boldsymbol{\theta}_j)\, \pi_j^N(\boldsymbol{\theta}_j)\, d\boldsymbol{\theta}_j.$$

Since in training sample $0 < m_j(\mathbf{y}(l)) < \infty$, then $\pi_j(\boldsymbol{\theta}_j|\,\mathbf{y}(l)) \propto f_j(\mathbf{y}(l)\,|\boldsymbol{\theta}_j)\,\pi_j^N(\boldsymbol{\theta}_j)$ is proper. Now taking these posterior distributions as prior distributions we can use the remaining data, $\mathbf{y}(-l)$, for model comparison and to compute a Bayes factor,

$$B_{01}^P(l) = \frac{\int f_0(\mathbf{y}(-l)\,|\boldsymbol{\theta}_0, \mathbf{y}(l))\,\pi_0(\boldsymbol{\theta}_0|\,\mathbf{y}(l))\,d\boldsymbol{\theta}_0}{\int f_1(\mathbf{y}(-l)\,|\boldsymbol{\theta}_1, \mathbf{y}(l))\,\pi_1(\boldsymbol{\theta}_1|\,\mathbf{y}(l))\,d\boldsymbol{\theta}_1}.$$

$B_{01}^P(l)$ is called Partial Bayes Factor (PBF). It is easy to prove that

$$B_{01}^P(l) = B_{01}(\mathbf{y}) \cdot B_{10}(\mathbf{y}(l)),$$

where $B_{10}(\mathbf{y}(l)) = \frac{m_1(\mathbf{y}(l))}{m_0(\mathbf{y}(l))}$. Due to PBF depends on the arbitrary choice of the training sample, to eliminate this dependence and to increase stability, Berger and Pericchi (1996a) define the IBF's averaging the $B_{01}^{P}(l)$ over all possible training samples $\mathbf{y}(l)$, $l = 1, \ldots, L$. Then, for each type of average there is an IBF, for example, the arithmetic IBF is defined as $B_{01}^{AI}(\mathbf{y}) = \frac{B_{01}(\mathbf{y})}{L} \sum_{l=1}^{L} B_{10}(\mathbf{y}(l))$, the geometric IBF, as $B_{01}^{GI}(\mathbf{y}) = B_{01}(\mathbf{y}) \exp\left\{L^{-1} \sum_{l=1}^{L} \ln B_{10}(\mathbf{y}(l))\right\}$, and the median IBF, as $B_{01}^{MI}(\mathbf{y}) = B_{01}(\mathbf{y}) \,\mathrm{med}\left[B_{10}(\mathbf{y}(l))\right]$, where med denotes the median of a data set. Berger and Pericchi (1996a) defined others alternative strategies for model comparison, such as the expected IBF, intrinsic prior distributions and encompassing IBF.

In our case, measurement error model, the computation of B_{01} was computationally demanding, so to compute the IBF's could be worse. The use of intrinsic prior distributions requires of a nice asymptotic behavior of the mle's, but the mle's of the MEM they do not satisfy those limiting properties, unless we make additional suppositions on the parameters of the MEM. For example, if we assume $\delta = \frac{\sigma_\epsilon}{\sigma_u}$ known we obtain consistent mle's, but it yields to that the models we are comparing are no longer nested. In this case, the encompassing IBF could be used, but this would bring other additional complications, see for example O'Hagan (1997).

The FBF, with training fraction $b \in (0, 1)$, is defined by

$$B_{01}^{F}(\mathbf{y}) = \frac{q_0(\mathbf{y}, b)}{q_1(\mathbf{y}, b)},$$

where

$$q_j(\mathbf{y}, b) = \frac{\int f_j(\mathbf{y} \,|\boldsymbol{\theta}_j)\, \pi_j^N(\boldsymbol{\theta}_j)\, d\boldsymbol{\theta}_j}{\int \left[f_j(\mathbf{y} \,|\boldsymbol{\theta}_j)\right]^b \pi_j^N(\boldsymbol{\theta}_j)\, d\boldsymbol{\theta}_j}.$$

O'Hagan (1995) proposed different choices for b, for example, $b = m_0/n$, $b = n^{-1} \max\{m_0, \sqrt{n}\}$ and $b = n^{-1} \max\{m_0, \ln(n)\}$, where m_0 is the minimal training sample size. Also, it is easy to show that the B_{01}^F is given by

$$B_{01}^F(\mathbf{y}) = B_{01}(\mathbf{y}) \frac{\int [f_1(\mathbf{y}\,|\boldsymbol{\theta}_1)]^b \pi_1^N(\boldsymbol{\theta}_1)\, d\boldsymbol{\theta}_1}{\int [f_0(\mathbf{y}\,|\boldsymbol{\theta}_0)]^b \pi_0^N(\boldsymbol{\theta}_0)\, d\boldsymbol{\theta}_0},$$

and B_{01} can be computed for MEM from Propositions 2.3.1 and 2.3.2. Here, difficulty resides in calculating the quotient of the previous expression, although this seems to be simpler than to calculate the IBF's.

The BRC was developed by Bernardo (1999). He combines decision theory, Kullback-Leibler information and reference analysis to propose a non-subjective Bayesian approach to nested hypotheses testing. BRC is a very nice models selection tool, however, it can be very difficult to carry out, for example in MEM it could be complicated to find the reference priors.

Other nested model comparison procedure is the following. For a posterior density, $f_1(\boldsymbol{\theta}\,|\mathbf{y})$ and for some $0 < p < 1$, a highest posterior density (HPD) credible set for $\boldsymbol{\theta}$ is defined to be the event $R_p(\mathbf{y})$, which is the smallest region such that

$$\int_{R_p(\mathbf{y})} f_1(\boldsymbol{\theta}\,|\mathbf{y})\, d\boldsymbol{\theta} = p.$$

Intuitively, for large p, $R_p(\mathbf{y})$ contains those values of $\boldsymbol{\theta}$ which are most plausible given the model M_1 and the data $\mathbf{y}$. Then, given a specified p and derived $R_p(\mathbf{y})$, we are going to assert that the true value of $\boldsymbol{\theta}$ lies in $R_p(\mathbf{y})$.

Defining the decision problem of the choice of p in $[0, 1]$, with the state of the world defined to be the true $\boldsymbol{\theta}$, we have to choose a value of p. An appropriate utility function may be

$$u(p, \boldsymbol{\theta}) = h(p)\, I_{\{R_p(\mathbf{y})\}}(\boldsymbol{\theta}) + g(1-p)\, I_{\left\{R_p^c(\mathbf{y})\right\}}(\boldsymbol{\theta}),$$

where $R_p^c(\mathbf{y}) = \Theta \backslash R_p(\mathbf{y})$, and h and g are decreasing functions defined on $[0, 1]$.

Then the expected utility of choosing p is given by

$$\overline{u}\left(p\right)=ph\left(p\right)+\left(1-p\right)g\left(1-p\right),$$

from which the optimal p may be derived for any specific choices of h and g. Here, the problem resides in being able to calculate the posterior distributions or obtain samples of them. So, this method is not too difficult.

In the next chapter, more specifically, in Section 3.3 we present another approach. Essentially, it consists on choosing the model with smaller posterior variance with respect to some parameter of interest.

Chapter 3

Influential Observations in Functional Measurement Error Model

In this chapter we propose measures to determine the influence of a given subset of observations on the posterior distribution of the structural parameters in a functional MEM. This topic was treated by Wellman and Gunst (1991) and Abdullah (1995) from classical point of view and by Arellano-Valle *et al.* (2000) for elliptical linear regression models. The model that we analyze in this chapter is the normal simple linear regression model with additive measurement error in variables. Thus, we consider the model given by

$$\begin{aligned} Y_i &= \alpha + \beta\xi_i + \epsilon_i, \qquad (3.0.1) \\ X_i &= \xi_i + u_i, \end{aligned}$$

$i = 1, \ldots, n$, where as in Chapter 2, (Y_i, X_i) are observed quantities, ξ_i are unobserved quantities, $\epsilon_i \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$, $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$ and $\boldsymbol{\epsilon} \perp\!\!\!\perp \mathbf{u}$, where $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^t$ and

$\mathbf{u} = (u_1, \cdots, u_n)^t$.

For this model there are two kinds of parameters: $(\alpha, \beta, \sigma_\epsilon^2, \sigma_u^2)$ called structural parameters and $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_n)^t$ called latent (nuisance or incidental) parameters. Assuming as prior distribution $\boldsymbol{\xi} \sim N_n(\mu\ \mathbb{1}_n, \tau^2\mathbf{I}_n)$ with μ and τ known, and $\boldsymbol{\xi}$ independent of $\boldsymbol{\epsilon}$ and $\mathbf{u}$, then, integrating out $\boldsymbol{\xi}$, the likelihood of this model can be specified by

$$\begin{aligned} N_{2n}\left(\mathbf{y}, \mathbf{x} \middle| \alpha, \beta, \sigma_\epsilon^2, \sigma_u^2\right) &= N_{2n}\left[\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \middle| \mathbb{1}_n \otimes \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \mathbf{I}_n \otimes \begin{pmatrix} \beta^2\tau^2 + \sigma_\epsilon^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma_u^2 \end{pmatrix}\right] \\ &= \prod_{i=1}^{n} N_2\left[\begin{pmatrix} y_i \\ x_i \end{pmatrix} \middle| \begin{pmatrix} \alpha + \beta\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \beta^2\tau^2 + \sigma_\epsilon^2 & \beta\tau^2 \\ \beta\tau^2 & \tau^2 + \sigma_u^2 \end{pmatrix}\right]. \end{aligned}$$

Let I be any subset with k elements of the set $\{1, \ldots, n\}$, and as usual, when a subset I has been deleted from the data, then $(\mathbf{y}_I, \mathbf{x}_I)$ and $\left(\mathbf{y}_{(I)}, \mathbf{x}_{(I)}\right)$ are the corresponding eliminated and remaining data.

In Section 3.1 we present the perturbation function and highlight its utility. In Section 3.2 we use the perturbation function together with the BF to calculate influence measures based on q-divergence. The perturbation function also is used in Section 3.3, where we show a proposition that allow us to compute some influence measures based on posterior Bayes risk under quadratic loss function using only a sample from the unperturbed posterior distribution. Finally, we apply these results to data from concrete compressive strengths.

3.1 The Perturbation Function

The perturbation functions were introduced by Kass *et al.* (1989), and Weiss (1996), generalizes the problem of assessment of the influence of model assumptions on a posterior distribution $f(\boldsymbol{\theta}\,|\mathbf{y}, M_0)$ in a general context, using a perturbation function. The perturbation function is defined by

$$h(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\theta}\,|\mathbf{y}, M_1)}{f(\boldsymbol{\theta}\,|\mathbf{y}, M_0)},$$

where $f(\boldsymbol{\theta}\,|\mathbf{y}, M_1)$ is the perturbed posterior distribution with respect to the posterior distribution $f(\boldsymbol{\theta}\,|\mathbf{y}, M_0)$.

We note that,

$$h(\boldsymbol{\theta}) = \frac{f(\mathbf{y}, \boldsymbol{\theta}\,|M_1)}{f(\mathbf{y}\,|M_1)} \cdot \frac{f(\mathbf{y}\,|M_0)}{f(\mathbf{y}, \boldsymbol{\theta}\,|M_0)} = h^*(\boldsymbol{\theta})\,B_{01},$$

where B_{01} is the Bayes factor to compare the models M_0 and M_1, and $h^*(\boldsymbol{\theta}) = f(\mathbf{y}, \boldsymbol{\theta}\,|M_1)\,/f(\mathbf{y}, \boldsymbol{\theta}\,|M_0)$. The next example shows that a suitable choice of $h^*(\boldsymbol{\theta})$ can be used to assess the influence of model assumptions.

Example 3.1.1. *Let $\mathbf{y} = (y_1, \ldots, y_n)$ be conditionally independent observations from a regression model with p.d.f.*

$$f(\mathbf{y}\,|\boldsymbol{\theta}, \mathbf{X}) = \prod_{i=1}^{n} f(y_i\,|\boldsymbol{\theta}, \mathbf{x}_i),$$

where $\mathbf{x}_i$ is the corresponding vector of predictors for the observation y_i and $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^t$ is the design matrix. The perturbation function corresponding to deletion case is such that

$$h(\boldsymbol{\theta}) \propto h^*(\boldsymbol{\theta}) = \frac{f\left(\mathbf{y}_{(I)}\,|\boldsymbol{\theta}, \mathbf{X}_{(I)}\right)\pi(\boldsymbol{\theta})}{f(\mathbf{y}\,|\boldsymbol{\theta}, \mathbf{X})\,\pi(\boldsymbol{\theta})} = \left[f(\mathbf{y}_I\,|\boldsymbol{\theta}, \mathbf{X}_I)\right]^{-1},$$

where $\mathbf{X}_I$ is the formed matrix by the vectors of predictors corresponding to the excluded observations $\mathbf{y}_I$. Similarly, for prior perturbations we obtain $h(\boldsymbol{\theta}) \propto \pi_1(\boldsymbol{\theta})\,/\pi(\boldsymbol{\theta})$, where $\pi_1(\boldsymbol{\theta})$ is an alternative prior distribution; and for likelihood perturbations, $h(\boldsymbol{\theta}) \propto f_1(\mathbf{y}\,|\boldsymbol{\theta}, \mathbf{X})\,/f(\mathbf{y}\,|\boldsymbol{\theta}, \mathbf{X})$, where $f_1(\mathbf{y}\,|\boldsymbol{\theta}, \mathbf{X})$ is an alternative likelihood function for $\boldsymbol{\theta}$.

On the other hand, we notice that

$$\begin{aligned}
B_{01} &= \frac{f(\mathbf{y}\,|M_0)}{f(\mathbf{y}\,|M_1)} = \frac{f(\mathbf{y}\,|M_0)}{\int f(\mathbf{y},\boldsymbol{\theta}\,|\mathbf{X},M_1)\,d\boldsymbol{\theta}} = \frac{f(\mathbf{y}\,|M_0)}{\int h^*(\boldsymbol{\theta})\, f(\mathbf{y},\boldsymbol{\theta}\,|\mathbf{X},M_0)\,d\boldsymbol{\theta}} \\
&= \left[\int h^*(\boldsymbol{\theta})\,\frac{f(\mathbf{y}\,|\boldsymbol{\theta},\mathbf{X},M_0)\,\pi(\boldsymbol{\theta}\,|M_0)}{f(\mathbf{y}\,|M_0)}\,d\boldsymbol{\theta}\right]^{-1} \\
&= \mathbb{E}^{-1}\left[h^*(\boldsymbol{\theta})\,|\mathbf{y}\right],
\end{aligned} \tag{3.1.2}$$

where the expected value is with respect to the posterior distribution obtained from the unperturbed model (see, Weiss (1996)).

Equation (3.1.2) expresses the BF in function of the perturbation $h^*(\boldsymbol{\theta})$, and also implies that $\mathbb{E}\left[h(\boldsymbol{\theta})\,|\mathbf{y}\right] = 1$. Moreover, a valuable advantage of formula (3.1.2) is that it requires only a sample from the unperturbed posterior distribution that can be obtained through MCMC methods, and then to apply Monte Carlo approximation. Different methods to detect outliers, based on (3.1.2), can be found in Pettit (1992) and Weiss (1996).

For our case, that is under the model (3.0.1), the perturbation function corresponding to deletion case is given by

$$\begin{aligned}
h\left(\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right) &= B_{01} h^*\left(\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right) \\
&= \frac{\left[\prod_{i\in I} N_2\left(y_i,x_i|\,\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right)\right]^{-1}}{\mathbb{E}_{\pi^*}\left[\left[\prod_{i\in I} N_2\left(y_i,x_i|\,\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right)\right]^{-1}\Big|\,\mathbf{y},\mathbf{x}\right]},
\end{aligned} \tag{3.1.3}$$

where π^* is given by equation (2.3.11).

Remark 3.1.2. *If $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, the perturbation function to obtain the influence on* *posterior distribution of* $\boldsymbol{\theta}_1$*is given by*

$$\begin{aligned} h(\boldsymbol{\theta}_1) &= \frac{f_1(\boldsymbol{\theta}_1|\,\mathbf{y})}{f(\boldsymbol{\theta}_1|\,\mathbf{y})} = \frac{\int f_1(\boldsymbol{\theta}|\,\mathbf{y})\,d\boldsymbol{\theta}_2}{\int f(\boldsymbol{\theta}|\,\mathbf{y})\,d\boldsymbol{\theta}_2} = \frac{\int h(\boldsymbol{\theta})\,f(\boldsymbol{\theta}|\,\mathbf{y})\,d\boldsymbol{\theta}_2}{\int f(\boldsymbol{\theta}|\,\mathbf{y})\,d\boldsymbol{\theta}_2} \\ &= \int h(\boldsymbol{\theta})\frac{f(\boldsymbol{\theta}|\,\mathbf{y})}{\int f(\boldsymbol{\theta}|\,\mathbf{y})\,d\boldsymbol{\theta}_2}d\boldsymbol{\theta}_2 = \int h(\boldsymbol{\theta})\,f(\boldsymbol{\theta}_2|\,\boldsymbol{\theta}_1,\mathbf{y})\,d\boldsymbol{\theta}_2 \\ &= B_{01}\int h^*(\boldsymbol{\theta})\,f(\boldsymbol{\theta}_2|\,\boldsymbol{\theta}_1,\mathbf{y})\,d\boldsymbol{\theta}_2, \end{aligned}$$

but in the MEM given by (3.0.1), is hard to calculate $f(\boldsymbol{\theta}_2|\,\boldsymbol{\theta}_1,\mathbf{y},\mathbf{x})$ *and therefore,* $h(\boldsymbol{\theta}_1)$ *too. Chib and Jeliazkov (2001) present a method for estimating* $f(\boldsymbol{\theta}_2^*|\,\boldsymbol{\theta}_1^*,\mathbf{y},\mathbf{x})$ *for an arbitrary* $\boldsymbol{\theta}^*$ *(from MCMC chains produced by the M-H algorithm), but for estimation efficiency, the point* $\boldsymbol{\theta}^*$ *has to be taken with a high density under the posterior distribution. For estimating* $h(\boldsymbol{\theta}_1)$*, we need many points like this, and in this way, we would only achieve a poor estimate of* $h(\boldsymbol{\theta}_1)$*.*

3.2 Influence Measures Based on q-divergence

Other appealing ways of quantifying influence is by computing divergence measures between posteriors computed with and without a given subset of the data. That is, measures that take into account the full distributions involved. The problem of quantifying the effect of subsets of data using divergence measures has been considered by several authors, following the approach proposed by Johnson and Geisser (1982). Weiss and Cook (1992) provide a unified treatment, based on divergence measures, to examine influence of model perturbations. These author define the q-divergence measure between two densities π_1 and π_2 on $\boldsymbol{\theta}$ by

$$d_q(\pi_1,\pi_2) = \int q\left[\frac{\pi_1(\boldsymbol{\theta})}{\pi_2(\boldsymbol{\theta})}\right]\pi_2(\boldsymbol{\theta})\,d\boldsymbol{\theta}, \tag{3.2.4}$$

where q is a convex function such that $q(1) = 0$. From (3.2.4) a wide class of different divergence measures is obtained. For example, when $q(z) = -\log(z)$ the Kullback-Leibler divergence follows, when $q(z) = (z-1)\log(z)$, the J-distance (or the symmetric version of Kullback-Leibler divergence), when $q(z) = \frac{1}{2}|z-1|$, the L_1-divergence and χ^2-divergence follows of take $q(z) = (z-1)^2$.

Thus, taking $\pi_1(\boldsymbol{\theta}) = f\left(\boldsymbol{\theta}\left|\mathbf{y}_{(I)}, \mathbf{x}_{(I)}\right.\right)$ and $\pi_2(\boldsymbol{\theta}) = f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})$, in (3.2.4) we have that $d_q(I) = d_q(\pi_1, \pi_2)$ can be interpreted as the q-influence of the data $(\mathbf{y}_I, \mathbf{x}_I)$ on posterior distribution of $\boldsymbol{\theta}$, which can be written as

$$d_q(I) = \int q\left[\frac{f\left(\boldsymbol{\theta}\left|\mathbf{y}_{(I)}, \mathbf{x}_{(I)}\right.\right)}{f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})}\right] f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})\, d\boldsymbol{\theta} = \mathbb{E}\left\{q\left[h(\boldsymbol{\theta})\right]|\,\mathbf{y}, \mathbf{x}\right\}, \tag{3.2.5}$$

where the expected value is taken with respect to the unperturbed posterior distribution $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})$.

Notice that these measures provide an order relation on the set of all partitions of subsets of $\{1, \ldots, n\}$, according to their relative influence. In fact, by adopting the approach developed by Girón *et al.* (1992), we can define a subset I_1 as being more q-influential than a subset I_2 for the parameter $\boldsymbol{\theta}$ if $d_q(I_1) \geq d_q(I_2)$. A similar ordering can be introduced using the measures M_E, M_I and M_R considered in the next section.

The most commonly used q-influence measures are the J-influence and the L_1-influence measures. The later measure and χ^2-influence has been recommended by several authors (see, for example, Peng and Dey (1995) and Weiss (1996)) because these are easier to interpret. However, it is difficult to obtain explicit expressions for these influence measures, even in simple cases, but in order to estimate $d_q(I)$ we can use the formula given by Weiss (1996).

For the model (3.0.1), through MCMC methods, we can obtain a sample $\boldsymbol{\theta}^{(j)}$ from the unperturbed posterior distribution $f(\boldsymbol{\theta}\,|\mathbf{y},\mathbf{x})$, where $\boldsymbol{\theta}=(\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2)$ and thus, (3.2.5) can be estimated by the formula given by

$$\hat{d}_q(I)=m^{-1}\sum_{j=1}^{m}q\left[\hat{B}_{01}h^*\left(\boldsymbol{\theta}^{(j)}\right)\right],$$

where $h^*(\boldsymbol{\theta})=\left[\prod_{i\in I}N_2\left(y_i,x_i|\,\alpha,\beta,\sigma_\epsilon^2,\sigma_u^2\right)\right]^{-1}$ and

$$\hat{B}_{01}=m^{-1}\sum_{j=1}^{m}h^*\left(\boldsymbol{\theta}^{(j)}\right).$$

Notice that $\hat{d}_q(I)$ requires only a sample from the unperturbed posterior distribution $f(\boldsymbol{\theta}\,|\mathbf{y},\mathbf{x})$.

3.3 Measures Based on the Posterior Bayes Risk

Kempthorne (1986) defined different influence measures in a Bayesian decision theory framework. In this context, the influence of a subset I of observations on the decision problem is defined as its impact on the posterior Bayes risk. Consequently, if $\mathcal{A}$ denotes the space of all actions and Θ the unknown states of the world, then preferences among actions are determined by their risk:

$$r(\pi^*,\mathbf{a})=\mathbb{E}_{\pi^*}\left[L(\boldsymbol{\theta},\mathbf{a})\right],$$

that is, the expected loss with respect to the posterior distribution π^* when we choose the action $\mathbf{a}$ and $\boldsymbol{\theta}$ is the true state of the world. In the context of parametric inference, it may be of interest estimation, prediction, hypotheses testing, models selection, etc., as considered by Berger (1985), O'Hagan (1994) and Bernardo and Smith (1994). In this section we consider the estimation problem following Kempthorne (1986). Thus, we consider the loss function given by

$$L(\boldsymbol{\theta},\mathbf{a})=(\boldsymbol{\theta}-\mathbf{a})^t\,\mathbf{W}\,(\boldsymbol{\theta}-\mathbf{a})=\|\boldsymbol{\theta}-\mathbf{a}\|_{\mathbf{W}}^2\,, \tag{3.3.6}$$

with $\mathbf{W}$ being a known symmetric, positive semi-definite matrix. The optimal action is the Bayes action $\mathbf{a}^* = \mathbb{E}_{\pi^*}(\boldsymbol{\theta})$ which gets the smallest posterior Bayes risk.

The influence measures, of a subset I of observations, on $\boldsymbol{\theta}$, can be measured in three different ways,

$$M_{E,\boldsymbol{\theta}}(I) = r\left(\pi^*, \mathbf{a}^*_{(I)}\right) - r\left(\pi^*, \mathbf{a}^*\right),$$

$$M_{I,\boldsymbol{\theta}}(I) = r\left(\pi^*_{(I)}, \mathbf{a}^*\right) - r\left(\pi^*_{(I)}, \mathbf{a}^*_{(I)}\right)$$

and

$$M_{R,\boldsymbol{\theta}}(I) = r\left(\pi^*_{(I)}, \mathbf{a}^*_{(I)}\right) - r\left(\pi^*, \mathbf{a}^*\right),$$

where $\pi^*_{(I)}$ denote the posterior distribution on $\boldsymbol{\theta}$ when the subset I of observations is excluded from the analysis and $\mathbf{a}^*_{(I)}$ is the corresponding Bayes action.

The influence measure M_E is the cost of *excluding* the subset I of observations from the analysis in terms of the posterior Bayes risk. In this case it is considered that all the data follow the same model since the risk it is taken with respect to the same posterior distribution π^* which is best characterization of the belief on $\boldsymbol{\theta}$ given all the data. Therefore, a subset of observations does not have influence if its exclusion of the data does not increase the posterior Bayes risk. Also, note that because $\mathbf{a}^*$ is the Bayes action under π^*, then $r\left(\pi^*, \mathbf{a}^*_{(I)}\right) \geq r\left(\pi^*, \mathbf{a}^*\right)$ and the measure M_E is always non-negative. Assuming that all data follow the same model, to exclude a subset of data from the analysis does not reduce the posterior Bayes risk.

The influence measure M_I assumes that all data, except the subset I, follow the same model. Thus, if we do not know anything about the true model of the observations I, then the analysis of the decision problem is valid if the observations I are excluded. Therefore, the posterior distribution $\pi^*_{(I)}$, is the to best characterizes the belief on $\boldsymbol{\theta}$. According to these suppositions, M_I measures the increment of the posterior Bayes risk when the subset I is incorrectly *included* in the data. Similar to the previous measure, M_I is non-negative since incorrectly including a the data I in the analysis never reduces the posterior Bayes risk.

In the third influence measure we assume that all data follow the same model, then π^* and $\pi^*_{(I)}$ are valid and M_R is the *reduction* of the posterior Bayes risk when we increase the set of data adding the cases I. Including a subset of observations I which are very different to the rest of the data of the analysis may reduce the precision of the posterior distribution of $\boldsymbol{\theta}$ and therefore, it produces an increment in the posterior Bayes risk. Then, contrary to the two previous influence measures, M_R is not restricted to be non negative.

The next lemma gives general expressions for the three measures M_E, M_I and M_R that can be used for the numeric computation of these measures using samples from the posterior distribution generated by some MCMC method.

Lemma 3.3.1. *Under the quadratic loss function (3.3.6),*

$$M_{E,\boldsymbol{\theta}}\left(I\right) = M_{I,\boldsymbol{\theta}}\left(I\right) = \left\|\mathbf{a}^* - \mathbf{a}^*_{(I)}\right\|^2_{\mathbf{W}}$$

and

$$M_{R,\boldsymbol{\theta}}\left(I\right) = tr\left[\mathbf{W}\left(\mathbf{V}^*_{(I)} - \mathbf{V}^*\right)\right],$$

where $\mathbf{V}^* = \mathbb{V}_{\pi^*}\left(\boldsymbol{\theta}\right)$ *and* $\mathbf{V}^*_{(I)} = \mathbb{V}_{\pi^*_{(I)}}\left(\boldsymbol{\theta}\right)$.

Proof. It is an immediate consequence of the well-known expression,
$\mathbb{E}_{\pi^*}\left[\left(\boldsymbol{\theta} - \mathbf{a}\right)^t \mathbf{W}\left(\boldsymbol{\theta} - \mathbf{a}\right)\right] = \mathrm{tr}\left(\mathbf{W}\mathbf{V}^*\right) + \left\|\mathbf{a}^* - \mathbf{a}\right\|^2_{\mathbf{W}}$. ■

Remark 3.3.2. *Let us notice that, if we are interested in selecting the model with better* $\boldsymbol{\theta}$ *estimate, then we should choose the model with the smaller Bayes risk,* $r\left(\pi^*, \mathbf{a}\right)$.

That is to say, if we are comparing two models M_0 and M_1 with posterior distributions π_0^* and π_1^*, respectively, then we should choose the model M_1 if $r(\pi_0^*, \mathbf{a}_0^*) - r(\pi_1^*, \mathbf{a}_1^*) > 0$, where $\mathbf{a}_j^* = \mathbb{E}_{\pi_j^*}(\boldsymbol{\theta})$, $j = 0, 1$. Thus, under the quadratic loss function (3.3.6) and from the previous lemma, we select the model M_1 if $tr[\mathbf{W}(\mathbf{V}_0^* - \mathbf{V}_1^*)] > 0$. Likewise, if we are only interested in the estimate of one parameter θ_1, then we select the model M_1 if $\mathbb{V}_{\pi_0^*}(\theta_1) > \mathbb{V}_{\pi_1^*}(\theta_1)$.

Now we see these influence measures with respect to the posterior distribution of the parameters of the model (3.0.1). The next proposition provides an expression to M_E and M_I that only involves the expected value of the unperturbed posterior distribution and the standardized perturbation function h.

Proposition 3.3.3. *Under the quadratic loss function (3.3.6),*

$$M_{E,\boldsymbol{\theta}}(I) = M_{I,\boldsymbol{\theta}}(I) = \left\| \mathbb{E}_{\pi^*}\left[(1 - h(\boldsymbol{\theta}))\, \boldsymbol{\theta} \,|\mathbf{y}, \mathbf{x}\right] \right\|_{\mathbf{W}}^2,$$

where $h(\boldsymbol{\theta})$ is the perturbation function of $\pi^*(\boldsymbol{\theta}\,|\mathbf{y}, \mathbf{x})$ to $\pi_{(I)}^*\left(\boldsymbol{\theta}\,\middle|\mathbf{y}_{(I)}, \mathbf{x}_{(I)}\right)$.

Proof. From

$$\begin{aligned}
\mathbf{a}^* - \mathbf{a}_{(I)}^* &= \mathbb{E}_{\pi^*}(\boldsymbol{\theta}\,|\mathbf{y}, \mathbf{x}) - \mathbb{E}_{\pi_{(I)}^*}\left(\boldsymbol{\theta}\,\middle|\mathbf{y}_{(I)}, \mathbf{x}_{(I)}\right) = \mathbb{E}_{\pi^*}(\boldsymbol{\theta}\,|\mathbf{y}, \mathbf{x}) - \mathbb{E}_{h\pi^*}(\boldsymbol{\theta}\,|\mathbf{y}, \mathbf{x}) \\
&= \mathbb{E}_{\pi^*}(\boldsymbol{\theta} - \boldsymbol{\theta} h(\boldsymbol{\theta})\,|\mathbf{y}, \mathbf{x})
\end{aligned}$$

and the previous lemma, the result is obtained. ■

For our case, that is under the model (3.0.1), the function $h(\boldsymbol{\theta})$ is given by (3.1.3), and π^* by (2.3.11).

3.4 An Application

The results established in the previous sections will now be illustrated by using measured compressive strength of concrete data. The data were taken from Wellman and Gunst (1991) and consist of 41 pairs of observed (y_i, x_i) values (see Figure 3.1, and Table A.4), where the y_i and x_i represent the measured compressive strength of concrete taken after 28 days and 2 days of pouring, respectively. The measured strengths of concrete differ from their respective true underlying values due to various sources of measurement errors. Thus an appropriate model for the data is given by (3.0.1). Wellman and Gunst (1991) and Abdullah (1995) used these data to evaluate the performance of various diagnostic techniques in linear regression with errors in variables, but from classical point of view.

The calculations were made numerically and using the M-H algorithm described in Section 2.4 with prior distributions $(\alpha, \beta)^t \,|\sigma_\epsilon^2 \sim N_2\left[(2000, 1)^t, (200^2, 4)\,\sigma_\epsilon^2\mathbf{I}_2\right]$, $\sigma_\epsilon^2 \sim IGa\,(3, 4\times 10^5)$, $\sigma_u^2 \sim IGa\,(3, 4\times 10^5)$ and $\boldsymbol{\xi} \sim N_{41}\,(3000\mathbb{1}_{41}, 500^2\mathbf{I}_{41})$. 301,000 samples from the posterior distribution were generated, where the initial 1000 iterations were discarded to assure stationarity and a lag of 10 was selected to avoid autocorrelation. That means that a net sample size of 30,000 was used.

Figures A.1 to A.4 show the influence of one observation based on the influence measures described in the previous sections. Looking at Figures A.1 to A.4, we see that the most influential observations are 17, 21, 22 and 37 coinciding with the observations analyzed by Wellman and Gunst (1991) and Abdullah (1995). The observations 26 and 34 have a moderated influence.

Table 3.1 gives the influence measures for these observations, Columns 2 and 3 should be multiplied by 10^{-6} and 10^{10}, respectively. The effect of each of these six

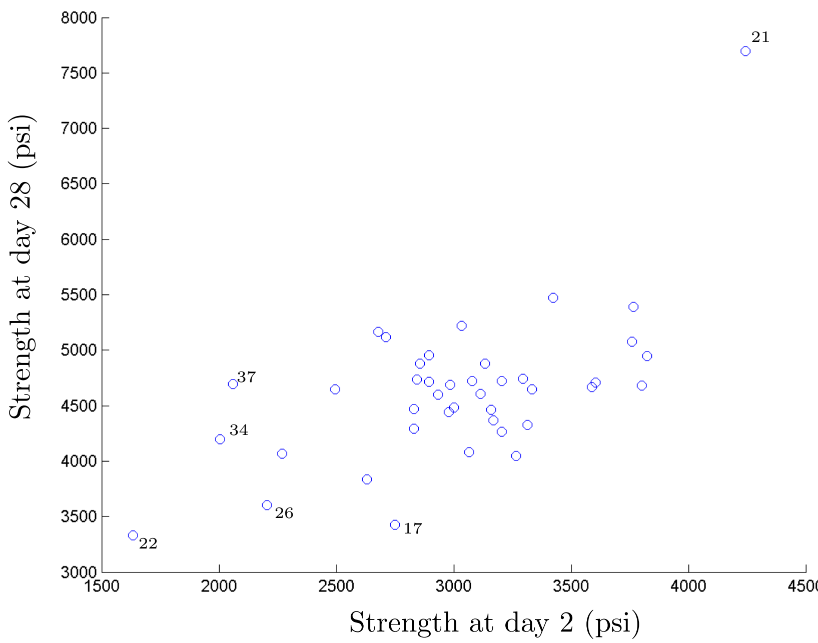

Figure 3.1: Concrete compressive strengths, in pounds per square inch, at 2 and 28 days.

observations over all parameters of the MEM given by (3.0.1) is show in Columns 2 to 5. Due to Remark 3.1.2, the influence measures given in Sections 3.2 and 3.3, over some particular parameter are hard to estimate using perturbation function. However, it is not difficult to calculate $\mathbb{E}_{\pi^*}\left(\boldsymbol{\theta}\,|\mathbf{y},\mathbf{x}\right) - \mathbb{E}_{\pi^*_{(I)}}\left(\boldsymbol{\theta}\,\middle|\mathbf{y}_{(I)},\mathbf{x}_{(I)}\right)$, so that, Columns 6 to 9 give the differences between Bayes estimators for each parameter of the MEM.

From Table 3.1 and Figures A.1 to A.4, we see that observation 21 have an strong influence on model parameters. This is followed by observations 17 and 37. Columns 8 and 9 indicate that the effect of the sample 37 is over the estimation of σ_ϵ and σ_u. The effect of the observation 22 is bigger on σ_ϵ. Summarizing, the observation

Table 3.1: Influence measures for concrete compressive strength data under MEM in deletion case.

i	$\hat{B}_{01}$	$M_{E,\boldsymbol{\theta}}(i)$	$\hat{d}_{L_1}(i)$	$\hat{d}_{\chi^2}(i)$	$\hat{\alpha}-\hat{\alpha}_{(i)}$	$\hat{\beta}-\hat{\beta}_{(i)}$	$\hat{\sigma}_\epsilon-\hat{\sigma}_{\epsilon(i)}$	$\hat{\sigma}_u-\hat{\sigma}_{u(i)}$
17	0.0673	0.0585	0.1589	0.1908	-166.6826	0.0472	7.4454	-0.2708
21	0	0.6962	0.7753	155.3351	-1385.4	0.4796	40.9677	26.0708
22	0.0268	0.0016	0.071	0.0327	-121.9643	0.0394	-16.5134	11.4325
26	0.1093	0.0017	0.0684	0.0320	-99.6000	0.0302	-8.7644	-0.6540
34	0.1205	0.0138	0.0589	0.0214	129.9191	-0.0395	0.5022	8.3000
37	0.0893	0.01	0.0897	0.0555	232.7790	-0.0691	15.2408	14.3426

21 is the one that really has a great influence on the model. From this analysis, the separate effect of these observations is clearly established.

This influence measures can be used for examining of posterior distributions from different models. Then, we can use this influence measures for model comparison. It is well know that for the linear regression model, $\mathbf{y}\,|\boldsymbol{\beta},\sigma_\epsilon^2,\mathbf{X} \sim N_n\left(\mathbf{X}\boldsymbol{\beta},\sigma_\epsilon^2\mathbf{I}_n\right)$, with prior distributions $\boldsymbol{\beta}\,|\sigma_\epsilon^2 \sim N_k\left[\boldsymbol{\beta}_0,\sigma_\epsilon^2\mathbf{B}\right]$ and $\sigma_\epsilon^2 \sim IGa\left(a_\epsilon,b_\epsilon\right)$, the posterior distribution is

$$f\left(\boldsymbol{\beta},\sigma_\epsilon^2\,|\mathbf{y},\mathbf{X}\right) \propto \frac{b_\epsilon^{a_\epsilon}\left(\sigma_\epsilon^{-2}\right)^{\frac{n+k}{2}+a_\epsilon+1}}{\Gamma\left(a_\epsilon\right)\left(2\pi\right)^{\frac{n+k}{2}}\left|\mathbf{B}\right|^{1/2}}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\mathbf{V}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}$$
$$\times\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left[\mathbf{y}^t\mathbf{y}+\boldsymbol{\beta}_0^t\mathbf{B}^{-1}\,\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}^t\mathbf{V}\hat{\boldsymbol{\beta}}+2b_\epsilon\right]\right\}, \quad (3.4.7)$$

where $\mathbf{V}=\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}$ and $\hat{\boldsymbol{\beta}}=\mathbf{V}^{-1}\left(\mathbf{X}^t\mathbf{y}+\mathbf{B}^{-1}\boldsymbol{\beta}_0\right)$. From this follows that

$$\boldsymbol{\beta}\,|\mathbf{y},\mathbf{X}\sim t_k\left(\hat{\boldsymbol{\beta}},\frac{b}{n+2a_\epsilon}\mathbf{V}^{-1},n+2a_\epsilon\right) \quad (3.4.8)$$

and

$$\sigma_\epsilon^2\,|\mathbf{y},\mathbf{X}\sim IGa\left(\frac{n}{2}+a_\epsilon,\frac{b}{2}\right), \quad (3.4.9)$$

where $b=\mathbf{y}^t\mathbf{y}+\boldsymbol{\beta}_0^t\mathbf{B}^{-1}\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}^t\mathbf{V}\hat{\boldsymbol{\beta}}+2b_\epsilon$.

Table 3.2 displays fitted models using the complete data set and the 40 samples excluding sample 21. To estimate (α,β) and σ_ϵ^2 we used, respectively, expected value

Table 3.2: Comparison of linear regression model and MEM fits.

	Complete data set		Sample 21 deleted	
	Linear model	MEM	Linear model	MEM
$\hat{\alpha}$	2250.7	873.7	3016.4	2259.1
$\hat{\beta}$	0.7894	1.2482	0.5159	0.7686
$\hat{\sigma}_\epsilon$	515.1397	397.5288	397.6253	356.5611
$\hat{\sigma}_u$	–	354.7970	–	328.7262
sd (α)	463.6781	739.8793	380.8376	598.4807
sd (β)	0.1512	0.2436	0.1257	0.1984
sd (σ_ϵ^2)	5.9893×10^4	7.3983×10^4	3.6145×10^4	4.6123×10^4
sd (σ_u^2)	–	4.8429×10^4	–	4.6115×10^4

of $\boldsymbol{\beta}\,|\mathbf{y}, \mathbf{X}$ and $\sigma_\epsilon^2\,|\mathbf{y}, \mathbf{X}$ under each models. sd (θ) denotes the posterior standard deviation of θ .

From Table 3.2 we can see the strong effect that the estimates have with each model. In both models we notice the great influence of the observation 21, but for the MEM fit is greater than for the linear regression fit. The σ_u estimate suggests the presence of measurement error. However, in both cases the complete data set and the data set without sample 21, the standard deviation for each parameter in the linear model was smaller than using MEM. Thus, from Remark 3.3.2, we should choose the linear regression model to fit the data. For the complete data set, the Bayes factor given by Proposition 2.3.2 was equal to 0.3817, while for the data set with sample 21 deleted was equal to 2.3112. However, these values of $\hat{B}_{01opt}$ are not reliable because the simulation results of Section 2.4.2 were only good when $\sigma_\epsilon < \sigma_u$, and due to $\hat{\sigma}_\epsilon$ and $\hat{\sigma}_u$ values, we can assume that $\sigma_\epsilon \approx \sigma_u$. Also, the compressive strengths of concrete taken after 28 days and 2 days of pouring were taken with the same measurement instrument, thus, this also makes us think that $\sigma_\epsilon \approx \sigma_u$. We note that the sample 21 has a great influence on $\hat{B}_{01opt}$.

In short, from these data we can infer two important pieces of evidence: first, due to the problem description (and not due to $\hat{B}_{01opt}$), there is measurement error, and second, due to the influence measures, the sample 21 is an outlier. Therefore, we would choose the estimates of Column 5 of Table 3.2. Although, we think that the best fit is reached by the estimates of Column 4.

Chapter 4

Testing of Asymmetry in Univariate Skew Elliptical Model

In many applications the assumption of the normal distribution is not appropriate and more realistic models are needed. However, these more flexible models increase the mathematical complexity. Computational technics can solve partially the problem, even so, some mathematical calculation needs to be done if we want to obtain accurate results. New models have been developed with the goal to preserve good properties of the normal model and also to be more flexible to control the skewness and kurtosis of the distribution. These general models include the normal case as a special one.

The model proposed by Azzalini (1985) has the above qualities. If f and g are symmetric p.d.f.'s around zero and G is a continuous c.d.f. associate with g, then

$$\frac{2}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda \frac{x-\mu}{\sigma}\right) \tag{4.0.1}$$

is a p.d.f. for any $\lambda \in \mathbb{R}$. Where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and λ is a skewness parameter. When $\lambda = 0$ we obtain the symmetric p.d.f., $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Different choice of the functions f and G give us important special cases, for example, the skew-normal with p.d.f. given by $\frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right)$ and denoted by $SN(\lambda,\mu,\sigma)$.

The elliptical models given by Kelker (1970) are other well known generalizations of the normal model. Theses models have been studied, for example, Cambanis *et al.* (1981), Fang *et al.* (1990) and Arellano-Valle (1994). The elliptical models include a vast variety of important distributions (the Student-t distribution, double exponential, Pearson type II) and also have good properties (marginalization, conditionally among others). These models are symmetric as the normal model, with different kurtosis coefficient.

Extension for the normal model using the two above mentioned ideas, skewness and heavy tails, have been studied by Branco and Dey (2001) and Genton and Loperfido (2001). An interesting special case is the skew-t distribution with p.d.f. given by $\frac{2}{\sigma}t\left(\frac{x-\mu}{\sigma}|0,1,\nu\right)F_T\left(\lambda\frac{x-\mu}{\sigma}\right)$, where $T\sim t\,(0,1,\nu)$, μ is the location parameter, σ is a scale parameter, ν the degree of freedom (control the heaviness of the tails) and λ is the skewness parameter.

In this chapter we approach the problem of model comparison within skew-elliptical families. In Section 4.1, we measure the sensitivity of the skewness parameter using the L_1-distance between the symmetric and asymmetric models. Computation of the Bayes factor to examine asymmetry is presented in Section 4.2. Also, in Section 4.3 we present simulation results for the skew-normal and skew-t distributions obtaining expected results. Application in stock markets are also considered.

4.1 Sensitivity Analysis for the Skewness Parameter

In this section, we study the L_1 distance between a symmetric and an asymmetric models.

The model comparison here to seek by evidences from the data set to decide about one of the models below

$$\begin{aligned} M_0 &: \sigma^{-1} f\left(\frac{x-\mu}{\sigma}\right) \\ M_1 &: 2\sigma^{-1} f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda \frac{x-\mu}{\sigma}\right). \end{aligned} \tag{4.1.2}$$

Interesting questions are, how much different are M_0 and M_1? Is it possible to obtain an expression as function of λ? In Figure 4.1, we plot the skew normal p.d.f. for three different values of λ.

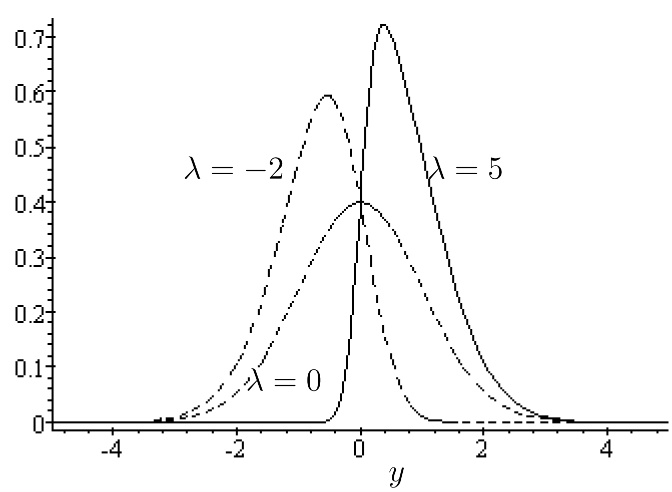

Figure 4.1: Skew-normal densities for $\lambda = -2$, $\lambda = 0$ and $\lambda = 5$.

There are many ways to measure the distance between models. An interesting measurement is the L_1 distance, because it has a easy and nice interpretation (see, for example, Peng and Dey (1995), Weiss (1996) and Arellano-Valle *et al.* (2000)). The L_1 distance between two densities f_1 and f_2 is

$$L_1(f_1, f_2) = \frac{1}{2}\int |f_1(x) - f_2(x)|\,dx = \sup_{A\in\mathcal{B}} |\mathbb{P}(A\,|f_1) - \mathbb{P}(A\,|f_2)|\,,$$

where $\mathcal{B}$ are the Borel's sets. The L_1 distance is bounded and takes values in $[0,1]$, where $L_1(f_1, f_2) = 0$ indicates that $f_1(x) = f_2(x)$ for all x value. Also, $L_1(f_1, f_2)$ is an upper bound on the differences $|\mathbb{P}(A\,|f_1) - \mathbb{P}(A\,|f_2)|$ for any set A, where $\mathbb{P}(\cdot\,|f)$ denote the probability measure defined by f. Generally, it is difficult to obtain explicit expressions for the L_1 distance, even in simple cases. However, for our case, the following proposition provide an useful expression to compute and understand the distance.

Proposition 4.1.1. *For any μ and σ fixed, the L_1 distance between M_0 and M_1, specified in (4.1.2), is*

$$L_1(M_0, M_1) = \mathbb{E}_{f^*}\left[G\left(|\lambda|\,Z\right)\right] - \frac{1}{2}, \tag{4.1.3}$$

where $f^*(z) = 2f(z)\,I_{[0,+\infty)}(z)$ is the p.d.f. f truncated on zero.

Proof. By letting $z = \frac{x-\mu}{\sigma}$,

$$\begin{aligned}
L_1(M_0, M_1) &= \frac{1}{2\sigma}\int_{\mathbb{R}} \left| f\left(\frac{x-\mu}{\sigma}\right) - 2f\left(\frac{x-\mu}{\sigma}\right) G\left(\lambda\frac{x-\mu}{\sigma}\right)\right| dx \\
&= \sigma^{-1}\int_{\mathbb{R}} \left|\frac{1}{2} - G\left(\lambda\frac{x-\mu}{\sigma}\right)\right| f\left(\frac{x-\mu}{\sigma}\right) dx = \int_{\mathbb{R}} \left|\frac{1}{2} - G(\lambda z)\right| f(z)\,dz \\
&= \int_{-\infty}^{0} \left|\frac{1}{2} - G(\lambda z)\right| f(z)\,dz + \int_{0}^{\infty} \left|\frac{1}{2} - G(\lambda z)\right| f(z)\,dz.
\end{aligned}$$

Now, by symmetry of g, we notice that if $z > 0$ and $\lambda > 0$ or if $z < 0$ and $\lambda < 0$, then

$$\frac{1}{2} - G(\lambda z) = -\int_0^{\lambda z} g(u)\,du$$

and

$$\left|\frac{1}{2} - G(\lambda z)\right| = \int_0^{\lambda z} g(u)\,du.$$

On the other hand, if $z < 0$ and $\lambda > 0$ or if $z > 0$ and $\lambda < 0$, then

$$\frac{1}{2} - G(\lambda z) = \int_{\lambda z}^{0} g(u)\,du = \int_0^{-\lambda z} g(u)\,du > 0.$$

But since, f and g are symmetric,

$$\int_{-\infty}^{0}\int_0^{-\lambda z} g(u)\,du f(z)\,dz = \int_0^{\infty}\int_0^{\lambda z} g(u)\,du f(z)\,dz$$

and

$$\int_{-\infty}^{0}\int_0^{\lambda z} g(u)\,du f(z)\,dz = \int_0^{\infty}\int_0^{-\lambda z} g(u)\,du f(z)\,dz,$$

then, if $\lambda > 0$ one gets

$$\begin{aligned} L_1(M_0, M_1) &= \int_{-\infty}^{0}\int_0^{-\lambda z} g(u)\,du f(z)\,dz + \int_0^{\infty}\int_0^{\lambda z} g(u)\,du f(z)\,dz \\ &= 2\int_0^{\infty}\int_0^{\lambda z} g(u)\,du f(z)\,dz = \mathbb{E}_{f^*}\left[G(\lambda Z)\right] - \frac{1}{2}, \end{aligned}$$

and if $\lambda < 0$,

$$\begin{aligned} L_1(M_0, M_1) &= \int_{-\infty}^{0}\int_0^{\lambda z} g(u)\,du f(z)\,dz + \int_0^{\infty}\int_0^{-\lambda z} g(u)\,du f(z)\,dz \\ &= 2\int_0^{\infty}\int_0^{-\lambda z} g(u)\,du f(z)\,dz = \mathbb{E}_{f^*}\left[\frac{1}{2} - G(\lambda Z)\right] \\ &= \mathbb{E}_{f^*}\left[G(|\lambda|\,Z)\right] - \frac{1}{2}. \end{aligned}$$

■

Note that $L_1(M_0, M_1) = L_1(\lambda)$ does not depend on μ and σ.

Corollary 4.1.2. *If L_1 is as in (4.1.3), then*

$$\max_{\lambda} L_1(\lambda) = \frac{1}{2}.$$

Proof. Since G is a c.d.f., follows that $G(|\lambda| z) \to 1$ when $|\lambda| \to \infty$. On the other side, $\mathbb{E}_{f^*}[G(|\lambda| Z)]$ exist because $0 < G(x) < 1$ for all $x \in \mathbb{R}$. Therefore

$$\lim_{|\lambda|\to\infty} \mathbb{E}_{f^*}[G(|\lambda| Z)] = \mathbb{E}_{f^*}\left[\lim_{|\lambda|\to\infty} G(|\lambda| Z)\right] = 1.$$

■

The next examples will show us some special cases, where the L_1 distance can be obtained in a close way. For these examples we will assume, without loss of generality, $\mu = 0$ and $\sigma = 1$.

Example 4.1.3. ***(Uniform)***

Let $f(x) = \frac{1}{2} I_{[-1,1]}(x)$ and $G(x) = \frac{x+1}{2} I_{[-1,1]}(x) + I_{(1,+\infty)}(x)$, the p.d.f. and c.d.f. of the uniform distribution $U_{[-1,1]}$, respectively. Then, $f^*(x) = I_{[0,1]}(x)$, and for $\lambda > 0$,

$$\begin{aligned} G(\lambda x) &= \frac{\lambda x + 1}{2} I_{[-1,1]}(\lambda x) + I_{(1,+\infty)}(\lambda x) \\ &= \frac{\lambda x + 1}{2} I_{\left[-\frac{1}{\lambda}, \frac{1}{\lambda}\right]}(x) + I_{\left(\frac{1}{\lambda}, +\infty\right)}(x). \end{aligned}$$

Now, if $\lambda > 1$, then

$$L_1(\lambda) = \mathbb{E}_{f^*}[G(\lambda X)] - \frac{1}{2} = \int_0^{\frac{1}{\lambda}} \frac{\lambda x + 1}{2} dx + \int_{\frac{1}{\lambda}}^1 dx - \frac{1}{2} = \frac{1}{2} - \frac{1}{4\lambda},$$

and, If $0 < \lambda < 1$, then

$$L_1(\lambda) = \mathbb{E}_{f^*}[G(\lambda X)] - \frac{1}{2} = \int_0^1 \frac{\lambda x + 1}{2} dx - \frac{1}{2} = \frac{\lambda}{4}.$$

Therefore

$$L_1(\lambda) = \begin{cases} \frac{1}{2} - \frac{1}{4|\lambda|} \ \mathit{if} \ |\lambda| > 1 \\ \frac{|\lambda|}{4} \ \mathit{if} \ |\lambda| \leq 1 \end{cases} .$$

Solid line in Figure 4.2 shows this function.

Example 4.1.4. ***(Double Exponential)***

Let $f(x) = \frac{1}{2}e^{-|x|}I_{\mathbb{R}}(x)$ *and* $G(x) = \frac{e^x}{2}I_{(-\infty,0)}(x) + \left(1 - \frac{e^{-x}}{2}\right)I_{[0,+\infty)}(x)$, *the p.d.f. and c.d.f. of the double exponential, respectively. Then,* $f^*(x) = e^{-x}I_{[0,+\infty)}(x)$ *and, for* $\lambda > 0$ *and* $x > 0$ *it follows that*

$$G(\lambda x) = 1 - \frac{1}{2}e^{-\lambda x},$$

so that

$$L_1(\lambda) = \mathbb{E}_{f^*}[G(|\lambda|X)] - \frac{1}{2} = \int_0^\infty \left(1 - \frac{1}{2}e^{-|\lambda|x}\right)e^{-x}dx - \frac{1}{2} = \frac{|\lambda|}{2(|\lambda|+1)}.$$

Dashed line in Figure 4.2 shows this function.

Example 4.1.5. ***(Normal)***

The L_1 *distance for the normal and skew-normal distributions is*

$$\begin{aligned} L_1(\lambda) &= \mathbb{E}_{f^*}[\Phi(|\lambda|X)] - \frac{1}{2} = \frac{1}{\pi}\int_0^\infty \int_{-\infty}^{|\lambda|x} e^{-\frac{1}{2}\left(t^2+x^2\right)}dtdx - \frac{1}{2} \\ &= \frac{1}{\pi}\int_0^\infty \int_{-\infty}^0 e^{-\frac{1}{2}\left[(y+|\lambda|x)^2+x^2\right]}dydx - \frac{1}{2} \\ &= \frac{1}{\pi}\int_{-\infty}^0 \int_{-\infty}^0 \exp\left\{-\frac{1}{2}\left[\left(1+|\lambda|^2\right)x^2 - 2|\lambda|xy + y^2\right]\right\}dydx - \frac{1}{2} \\ &= 2F_{\mathbf{U}}(0,0) - \frac{1}{2} = 2\left[\frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right)\right] - \frac{1}{2} \\ &= \frac{1}{2} - \frac{1}{\pi}\arccos\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right), \end{aligned}$$

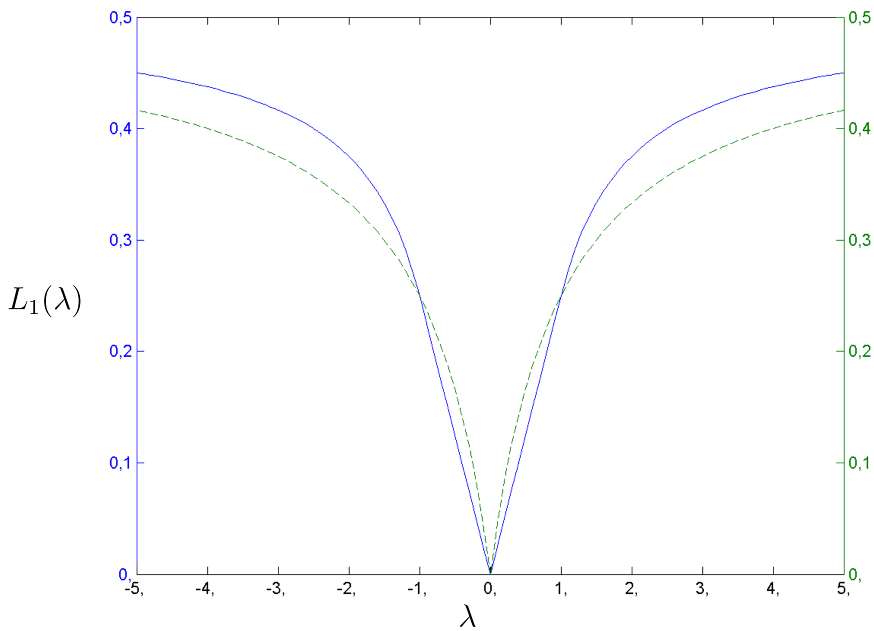

Figure 4.2: L_1-distance: solid line for the uniform and skew-uniform densities, and dashed line for double exponential and skew-double-exponential densities.

where $\mathbf{U} \sim N_2(\mathbf{0}, \mathbf{\Sigma})$ *with* $\mathbf{\Sigma} = \begin{pmatrix} 1 & |\lambda| \\ |\lambda| & 1+\lambda^2 \end{pmatrix}$. *Thus,*

$$L_1(\lambda) = \frac{1}{2} - \frac{1}{\pi} \arccos\left(\frac{|\lambda|}{\sqrt{1+\lambda^2}}\right).$$

However, in more general cases, numerical methods are necessary to calculate the L_1 distance, as an example, the skew-t model.

Figure 4.3 present values of $L_1(\lambda)$ for two groups of models: the solid line shows the distance between $N(0,1)$ and $SN(\lambda, 0, 1)$ densities. The circles line, the distance between the Student-t distribution and the skew-t model. In this case, we calculated $L_1(\lambda)$ using the S-PLUS integration function, this function which implements adaptive 15-point Gauss-Kronrod quadrature based on the Fortran function *dqage* and *dqagie* from QUADPACK (Piessens *et al.* (1983)) in NETLIB (Dongarra and Grosse (1987)).

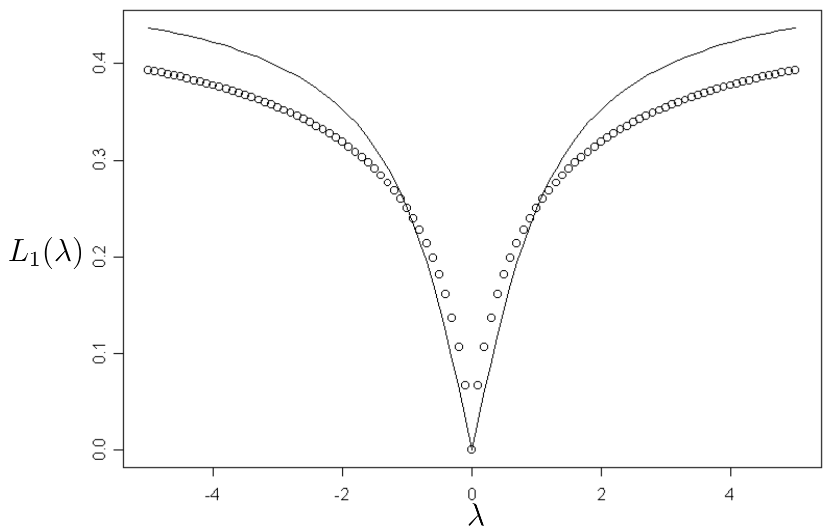

Figure 4.3: L_1-distance: solid line for the normal and skew-normal densities, and circles line for Cauchy and skew-Cauchy densities.

An important way to measure the sensitivity of the λ parameter is considering the effect on the posterior distribution for the parameters (μ, σ). With this objective we obtained the following results to compute the posterior distributions.

Proposition 4.1.6. *Under the* $SN(\lambda, \mu, \sigma)$ *and the prior assumptions* $\lambda \perp\!\!\!\perp (\mu, \sigma)$, $\mu|\,\sigma \sim N\left(m, \frac{\sigma^2}{v}\right)$ *and* $\sigma^{-2} \sim Ga\,(a, b)$, *then:*

1.

$$\begin{aligned} \pi\left(\mu, \sigma|\,\lambda, \mathbf{x}\right) &= \frac{r^{n+2a}\sqrt{v+n}}{2^{\frac{n-1}{2}+a}\Gamma\left(\frac{n}{2}+a\right)\sqrt{\pi}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)} \\ & \quad \frac{1}{(\sigma^2)^{\frac{n}{2}+a+1}}\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[(n+v)\left(\mu-\hat{\mu}\right)^2+r^2\right]\right\}, \end{aligned}$$

where $\hat{\mu} = \frac{n\bar{x}+mv}{n+v}$, $r^2 = ns^2 + \frac{nv}{n+v}(m-\bar{x})^2 + 2b$, $s^2 = n^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ *and* $F_{\mathbf{T}}$ *is the c.d.f. of the* $t_n(\mathbf{0}, \mathbf{\Sigma}, n+2a)$ *with* $\mathbf{\Sigma} = \mathbf{I}_n + \frac{\lambda^2}{v+n}\mathbb{1}_n\mathbb{1}_n^t$.

2.

$$\pi(\mu|\lambda,\mathbf{x}) = \frac{r^{n+2a}\sqrt{v+n}\,\Gamma\left(\frac{n+1}{2}+a\right)}{\Gamma\left(\frac{n}{2}+a\right)\sqrt{\pi}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)}\frac{F_{\mathbf{Y}}\left(\frac{\lambda\sqrt{n+2a+1}}{\sqrt{r^2+(n+v)(\mu-\hat{\mu})^2}}(\mathbf{x}-\mu\mathbb{1}_n)\right)}{\left[r^2+(n+v)(\mu-\hat{\mu})^2\right]^{\left(\frac{n+1}{2}+a\right)}}$$

and

3.

$$\pi(\sigma|\lambda,\mathbf{x}) = \frac{r^{n+2a}\sigma^{-n-2a-1}}{2^{\frac{n}{2}+a-1}\Gamma\left(\frac{n}{2}+a\right)F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)}\exp\left\{-\frac{r^2}{2\sigma^2}\right\}F_{\mathbf{U}}\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma}\right),$$

where $F_{\mathbf{Y}}$ *is the c.d.f. of the* $t_n(\mathbf{0},\mathbf{I}_n, n+2a+1)$ *and* $F_{\mathbf{U}}$ *is the c.d.f. of the* $N_n(\mathbf{0},\mathbf{\Sigma})$.

Proof. See Appendix C. ∎

From this proposition and (3.4.7)-(3.4.9) we have

$$\pi(\mu,\sigma|\lambda,\mathbf{x}) = \frac{\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)}{F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)} \times \pi(\mu,\sigma|\lambda=0,\mathbf{x}),$$

$$\pi(\mu|\lambda,\mathbf{x}) = \frac{F_{\mathbf{Y}}\left(\frac{\lambda\sqrt{n+2a+1}}{\sqrt{r^2+(n+v)(\mu-\hat{\mu})^2}}(\mathbf{x}-\mu\mathbb{1}_n)\right)}{F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)} \times \pi(\mu|\lambda=0,\mathbf{x})$$

and

$$\pi(\sigma|\lambda,\mathbf{x}) = \frac{F_{\mathbf{U}}\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma}\right)}{F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)} \times \pi(\sigma|\lambda=0,\mathbf{x}).$$

The first term in the three previous equations can be seen as a factor of sensitivity.

Also, notice that for $\lambda = 0$ (normal model), $\mu|\lambda=0,\mathbf{x} \sim t_1\left(\hat{\mu}, \frac{r^2}{(n+v)(n+2a)}, n+2a\right)$ and $\sigma^{-2}|\lambda=0,\mathbf{x} \sim Ga\left(\frac{n}{2}+a, \frac{r^2}{2}\right)$. From the previous proposition it is easy to calculate $\pi(\mu|\sigma,\lambda,\mathbf{x})$ and $\pi(\sigma|\mu,\lambda,\mathbf{x})$, and these two conditional posterior distributions are necessary in Gibbs Sampling algorithm.

On the other side, also we can calculate the L_1 distance considering these posterior distributions, for example, between $\pi\left(\sigma|\, \lambda = 0, \mathbf{x}\right)$ and $\pi\left(\sigma|\, \lambda, \mathbf{x}\right)$,

$$\begin{aligned} L_1(\lambda) &= \frac{1}{2}\int_0^\infty \left|\pi\left(\sigma|\, \lambda = 0, \mathbf{x}\right) - \pi\left(\sigma|\, \lambda, \mathbf{x}\right)\right| d\sigma \\ &= \frac{r^{n+2a}}{2^{\frac{n}{2}+a}\Gamma\left(\frac{n}{2}+a\right)}\int_0^\infty \left|1 - \frac{F_{\mathbf{U}}\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{\sigma}\right)}{F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)}\right| \frac{\exp\left\{-\frac{r^2}{2\sigma^2}\right\}}{\sigma^{n+2a+1}} d\sigma \\ &= \frac{1}{2}\mathbb{E}\left[\left|1 - \frac{F_{\mathbf{U}}\left(\lambda\sqrt{S}\left(\mathbf{x}-\hat{\mu}\mathbb{1}_n\right)\right)}{F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)}\right|\right], \end{aligned}$$

where $S \sim Ga\left(\frac{n}{2}+a, \frac{r^2}{2}\right)$. Also, the L_1 distance between $\pi\left(\mu|\, \lambda = 0, \mathbf{x}\right)$ and $\pi\left(\mu|\, \lambda, \mathbf{x}\right)$ is given by

$$\begin{aligned} L_1(\lambda) &= \frac{1}{2}\int_{-\infty}^\infty \left|\pi\left(\mu|\, \lambda = 0, \mathbf{x}\right) - \pi\left(\mu|\, \lambda, \mathbf{x}\right)\right| d\mu \\ &= \frac{\Gamma\left(\frac{n+1}{2}+a\right) r^{n+2a}\sqrt{v+n}}{2\sqrt{\pi}\Gamma\left(\frac{n}{2}+a\right)} \\ &\quad \int_{-\infty}^\infty \left|1 - \frac{F_{\mathbf{Y}}\left(\frac{\lambda\sqrt{n+2a+1}}{\sqrt{r^2+(n+v)(\mu-\hat{\mu})^2}}\left(\mathbf{x}-\mu\mathbb{1}_n\right)\right)}{F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)}\right| \left[r^2 + (n+v)\left(\mu-\hat{\mu}\right)^2\right]^{-\frac{n+2a+1}{2}} d\mu \\ &= \frac{1}{2}\mathbb{E}\left[\left|1 - \frac{F_{\mathbf{Y}}\left(\frac{\lambda\sqrt{n+2a+1}}{\sqrt{r^2+(n+v)(M-\hat{\mu})^2}}\left(\mathbf{x}-M\mathbb{1}_n\right)\right)}{F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}_n}{r}\right)}\right|\right], \end{aligned}$$

where $M \sim t_1\left(\hat{\mu}, \frac{r^2}{(n+v)(n+2a)}, n+2a\right)$.

These last results depend on the factor of sensitivity and can be used to study the influence of the skewness parameter over the posterior distribution of μ and σ. The influence measures given in Chapter 3 and their calculus methods can be used.

4.2 Bayes Factor

In this section we use the Bayes factor to do model comparison (see, for example, Kass and Raftery (1995), Lavine and Schervish (1999) and Berger and Pericchi (2001)). Liseo and Loperfido (2002) used the Bayes factor with prior reference to compare normal versus skew normal models. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the data set, it is comes from *i.i.d.* random samples. Then, the Bayes factor to test the hypotheses given by (4.1.2), in favor of the M_0 (the symmetric model) is given as

$$BF = \frac{\int \sigma^{-n} \left[\prod_{i=1}^{n} f\left(\frac{x_i-\mu}{\sigma}\right)\right] \pi\left(\mu, \sigma\right) d\mu d\sigma}{2^n \int \sigma^{-n} \left[\prod_{i=1}^{n} f\left(\frac{x_i-\mu}{\sigma}\right) G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right] \pi\left(\mu, \sigma, \lambda\right) d\mu d\sigma d\lambda}, \tag{4.2.4}$$

where $\pi\left(\cdot\right)$ is the prior distribution for the respective parameter. As we can note by expression (4.2.4), a closed form to Bayes factor is not possible to obtain in a general way. Also the numeric calculation is complex. However, when μ and σ are known, the Bayes factor have a better expression

$$BF_{\mu,\sigma} = \frac{1}{2^n \int \left[\prod_{i=1}^{n} G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right] \pi\left(\lambda\right) d\lambda}. \tag{4.2.5}$$

The next proposition gives another expression for (4.2.5).

Proposition 4.2.1. *The Bayes factor (4.2.5) is given by*

$$BF_{\mu,\sigma} = \left[2^n \mathbb{P}\left(\mathbf{Z} - \lambda\frac{\mathbf{x} - \mu\mathbb{1}_n}{\sigma} \leq \mathbf{0}\right)\right]^{-1},$$

where $Z_1, \ldots, Z_n \overset{iid}{\sim} G$, $\lambda \sim \pi\left(\lambda\right)$ *and* $Z_i \perp\!\!\!\perp \lambda$ *for all* $i = 1, \ldots, n$.

Proof. By noting that

$$\begin{aligned}
\int \left[\prod_{i=1}^{n} G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right] \pi\left(\lambda\right) d\lambda &= \int \mathbb{P}\left(\mathbf{Z} \leq \lambda\frac{\mathbf{x} - \mu\mathbb{1}_n}{\sigma}\right) \pi\left(\lambda\right) d\lambda \\
&= \int \mathbb{P}\left(\mathbf{Z} \leq \lambda\frac{\mathbf{x} - \mu\mathbb{1}_n}{\sigma} \middle| \lambda\right) \pi\left(\lambda\right) d\lambda \\
&= \mathbb{P}\left(\mathbf{Z} \leq \lambda\frac{\mathbf{x} - \mu\mathbb{1}_n}{\sigma}\right).
\end{aligned}$$

■

Corollary 4.2.2. *If G is the standard normal c.d.f. and $\lambda \sim N\left(m, v^2\right)$, then*

$$BF_{\mu,\sigma} = \left[2^n \Phi_n\left(\mathbf{0} \left| m\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}, \mathbf{I}_n + (v/\sigma)^2 \left(\mathbf{x}-\mu\mathbb{1}_n\right)\left(\mathbf{x}-\mu\mathbb{1}_n\right)^t\right.\right)\right]^{-1}.$$

Proof. It is enough to note that,

$$\mathbf{Z} - \lambda\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma} \sim N_n\left[m\frac{\mathbf{x}-\mu\mathbb{1}_n}{\sigma}, \mathbf{I}_n + (v/\sigma)^2 \left(\mathbf{x}-\mu\mathbb{1}_n\right)\left(\mathbf{x}-\mu\mathbb{1}_n\right)^t\right].$$

■

On practical point of view it is natural to consider the sign of the λ parameter known, i.e., we know the direction of the skewness. In this case, the comparison to be consider is an unilateral test for the parameter λ. In the next two propositions, we obtain expressions for the Bayes factor that can be helpful for numerical implementation.

Proposition 4.2.3. *Let $\mathbf{x} = (x_1, \ldots, x_n)$ a random sample from (4.0.1), where G is the standard normal c.d.f. , $\lambda \perp\!\!\!\perp (\mu, \sigma)$, $\lambda^2 \sim Ga\,(a, b)$ and $\mathbb{P}\left(\lambda > 0\right) = 1$. Then the Bayes factor (4.2.4) to compare the models specified by the hypotheses $H_0 : \lambda = 0$ and $H_1 : \lambda > 0$ is given by*

$$BF = \frac{\int_0^\infty \int_{\mathbb{R}} \sigma^{-n}\left[\prod_{i=1}^n f\left(\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\mu,\sigma\right)d\mu d\sigma}{2^n \int_0^\infty \int_{\mathbb{R}} \sigma^{-n}\left[\prod_{i=1}^n f\left(\frac{x_i-\mu}{\sigma}\right)\right] F_{\mathbf{T}}\left(\frac{x_1-\mu}{\sigma}\sqrt{\frac{a}{b}}, \ldots, \frac{x_n-\mu}{\sigma}\sqrt{\frac{a}{b}}\right)\pi\left(\mu,\sigma\right)d\mu d\sigma},$$

where $F_{\mathbf{T}}$ is the c.d.f. of the $t_n\left(\mathbf{0}, \mathbf{I}_n, 2a\right)$.

Proof. Notice that if $\lambda^2 \sim Ga\,(a, b)$, then λ has probability density function $f\left(\lambda\,|a, b\right) =$

$\frac{2b^a}{\Gamma(a)}\left(\lambda^2\right)^{a-\frac{1}{2}}\exp\left(-b\lambda^2\right)$. Thus,

$$\begin{aligned}\int\left[\prod_{i=1}^{n}G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\lambda\right)d\lambda &= \frac{2b^a}{(2\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int_0^\infty\left[\prod_{i=1}^{n}\int_{-\infty}^{\lambda\frac{x_i-\mu}{\sigma}}\exp\left(-\frac{t_i^2}{2}\right)dt_i\right]\\ &\quad \lambda^{2a-1}\exp\left(-b\lambda^2\right)d\lambda\\ &= \frac{2b^a}{(2\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int_0^\infty\int\cdots\int_{-\infty}^{\lambda\frac{x_i-\mu}{\sigma}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}t_i^2\right)dt_i\\ &\quad \lambda^{2a-1}\exp\left(-b\lambda^2\right)d\lambda.\end{aligned}$$

Making the change of variables $t_i=\lambda y_i$, for all $i=1,\ldots,n$, and exchanging the integration order, we obtain

$$\begin{aligned}\int\left[\prod_{i=1}^{n}G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\lambda\right)d\lambda &= \frac{2b^a}{(2\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int\cdots\int_{-\infty}^{\frac{x_i-\mu}{\sigma}}\\ &\quad \int_0^\infty\lambda^{2a+n-1}\exp\left[-\left(b+\frac{1}{2}\sum_{i=1}^{n}y_i^2\right)\lambda^2\right]d\lambda dy_i.\end{aligned}$$

Then, with the change of variables $l=\lambda^2$, one obtains

$$\begin{aligned}\int\left[\prod_{i=1}^{n}G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\lambda\right)d\lambda &= \frac{b^a}{(2\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int\cdots\int_{-\infty}^{\frac{x_i-\mu}{\sigma}}\\ &\quad \int_0^\infty l^{a+\frac{n}{2}-1}\exp\left[-\left(b+\frac{\mathbf{y}^t\mathbf{y}}{2}\right)l\right]dl dy_i\\ &= \frac{b^a\Gamma\left(a+\frac{n}{2}\right)2^{a+\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int\cdots\int_{-\infty}^{\frac{x_i-\mu}{\sigma}}\left[2b+\mathbf{y}^t\mathbf{y}\right]^{-\frac{n+2a}{2}}dy_i\\ &= \frac{b^{-\frac{n}{2}}\Gamma\left(a+\frac{n}{2}\right)}{(2\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int\cdots\int_{-\infty}^{\frac{x_i-\mu}{\sigma}}\left[1+\frac{\mathbf{y}^t\mathbf{y}}{2b}\right]^{-\frac{n+2a}{2}}dy_i.\end{aligned}$$

But making the change of variables $y_i=t_i\sqrt{\frac{b}{a}}$, for all $i=1,\ldots,n$, we obtain

$$\begin{aligned}\int\left[\prod_{i=1}^{n}G\left(\lambda\frac{x_i-\mu}{\sigma}\right)\right]\pi\left(\lambda\right)d\lambda &= \frac{\Gamma\left(a+\frac{n}{2}\right)}{(2a\pi)^{\frac{n}{2}}\Gamma\left(a\right)}\int\cdots\int_{-\infty}^{\frac{x_i-\mu}{\sigma}\sqrt{\frac{a}{b}}}\left[1+\frac{\mathbf{t}^t\mathbf{t}}{2a}\right]^{-\frac{n+2a}{2}}dt_i\\ &= F_{\mathbf{T}}\left(\frac{x_1-\mu}{\sigma}\sqrt{\frac{a}{b}},\ldots,\frac{x_n-\mu}{\sigma}\sqrt{\frac{a}{b}}\right).\end{aligned}$$

Then, the proposition result is immediate. ■

From Proposition 4.2.3, for μ and σ known,

$$BF = \frac{1}{2^n F_{\mathbf{T}}\left(\frac{x_1-\mu}{\sigma}\sqrt{\frac{a}{b}}, \ldots, \frac{x_n-\mu}{\sigma}\sqrt{\frac{a}{b}}\right)}. \tag{4.2.6}$$

Remark that to use the expression (4.2.5) and (4.2.6) is not necessary to determine the p.d.f. $\sigma^{-1} f\left(\frac{x_i-\mu}{\sigma}\right)$. The next result is obtained using similar argument.

Proposition 4.2.4. *Let $\mathbf{x} = (x_1, \ldots, x_n)$ a random sample from the p.d.f. (4.0.1), where G is the standard normal c.d.f. , $\lambda \perp\!\!\!\perp (\mu, \sigma)$, $\lambda^2 \sim Ga\,(a, b)$ and $\mathbb{P}\,(\lambda < 0) = 1$. Then the Bayes factor (4.2.4) to compare the models defined by the hypotheses H_0 : $\lambda = 0$ and $H_1 : \lambda < 0$ is given by*

$$BF = \frac{\int_0^\infty \int_{\mathbb{R}} \sigma^{-n}\left[\prod_{i=1}^n f\left(\frac{x_i-\mu}{\sigma}\right)\right] \pi\,(\mu, \sigma)\, d\mu d\sigma}{2^n \int_0^\infty \int_{\mathbb{R}} \sigma^{-n}\left[\prod_{i=1}^n f\left(\frac{x_i-\mu}{\sigma}\right)\right] F_{\mathbf{T}}\left(\frac{\mu-x_1}{\sigma}\sqrt{\frac{a}{b}}, \ldots, \frac{\mu-x_n}{\sigma}\sqrt{\frac{a}{b}}\right) \pi\,(\mu, \sigma)\, d\mu d\sigma},$$

where $F_{\mathbf{T}}$ is the c.d.f. of the $t_n\,(\mathbf{0}, \mathbf{I}_n, 2a)$.

Proof. Notice that if $\lambda^2 \sim Ga\,(a, b)$, then $\lambda < 0$ has probability density function $f\,(\lambda\,|a, b) = \frac{2b^a}{\Gamma(a)}\,(\lambda^2)^{a-\frac{1}{2}} \exp\left(-b\lambda^2\right)$. So that, making the change of variable $l = -\lambda$, we obtain

$$\begin{aligned}
\int \left[\prod_{i=1}^n G\left(\lambda \frac{x_i-\mu}{\sigma}\right)\right] \pi\,(\lambda)\, d\lambda &= -\int_\infty^0 \left[\prod_{i=1}^n \Phi\left(-l \frac{x_i-\mu}{\sigma}\right)\right] f\,(-l\,|a, b)\, dl \\
&= \int_0^\infty \left[\prod_{i=1}^n \Phi\left(l \frac{\mu-x_i}{\sigma}\right)\right] f\,(l\,|a, b)\, dl.
\end{aligned}$$

Then, it is enough to continue the proof of Proposition 4.2.3. ■

A general way to calculate (4.2.5), is using the Monte Carlo method. Note that if we can generate a sample $\lambda_1, \ldots, \lambda_m$ from $\pi\,(\lambda)$, then the Monte Carlo estimator for (4.2.5) is

$$\hat{BF}_{\mu,\sigma} = \frac{m}{\sum_{j=1}^m \left[\prod_{i=1}^n 2G\left(\lambda_j \frac{x_i-\mu}{\sigma}\right)\right]},$$

where the variance of the estimator $\hat{BF}_{\mu,\sigma}$ depends on the prior variance of λ: prior distributions less informative for λ give estimation with greater variance. However, we can obtain a superior limit for the variance of $\left(\hat{BF}_{\mu,\sigma}\right)^{-1}$. It is well known if X is a random variable in $[0,1]$ then $\mathbb{V}(X) \leq \frac{1}{4}$. Therefore,

$$\mathbb{V}\left[\left(\hat{BF}_{\mu,\sigma}\right)^{-1}\right] = \frac{2^{2n}}{m^2}\sum_{j=1}^{m}\mathbb{V}\left[\prod_{i=1}^{n} G\left(\lambda_j \frac{x_i-\mu}{\sigma}\right)\right] \leq \frac{2^{2n}}{m^2}\frac{m}{4} = \frac{2^{2(n-1)}}{m}.$$

The superior limit above gives us an idea how big need to be m in order to control the variability of the Monte Carlo estimation. However, this limit increases quickly with n, that is to say, it gives us a very large limit value for moderate sample size.

4.2.1 Bayes Factor for Representable Skew Distributions

In the latter section we discussed the difficulties to obtain a general form for the Bayes factor. Therefore, it is important to consider restrictions in the functions f and G to make the calculations of (4.2.4) simpler and also to keep the class of the asymmetric distributions general. Next, we consider a subclass of the elliptical distribution more simple to work with.

Definition 4.2.5. *$Z|\,\lambda,\mu,\sigma$ has a skew representable distribution under the c.d.f. H if the p.d.f. can be written by*

$$f_{Z|\lambda,\mu,\sigma}(z) = \int_0^\infty \frac{2}{\sigma\sqrt{\omega}}\phi\left(\frac{z-\mu}{\sigma\sqrt{\omega}}\right)\Phi\left(\lambda\frac{z-\mu}{\sigma\sqrt{\omega}}\right)dH(\omega),$$

where H is the c.d.f. of a random variable ω, which is non-negative and such that $\omega \perp\!\!\!\perp (\lambda,\mu,\sigma)$.

An equivalent definition is given by: $Z|\,\lambda,\mu,\sigma$ is a skew representable if and only if there is $\omega \sim H$ and $\omega \perp\!\!\!\perp (\lambda,\mu,\sigma)$ such that $Z|\,\lambda,\mu,\sigma,\omega \sim SN\left(\lambda,\mu,\sigma\sqrt{\omega}\right)$. Properties and examples can be found in Branco and Dey (2001).

Proposition 4.2.6. *Let* $\mathbf{x} = (x_1, \ldots, x_n)$ *a random sample from a skew representable distribution. If a priori* $\mu|\,\sigma \sim N\left(m, \frac{\sigma^2}{v}\right)$ *and* $\sigma^{-2} \sim Ga\,(a, b)$*, then the Bayes factor (4.2.4) is given by*

$$BF\,(\mathbf{x}) = \frac{\int \cdots \int r^{-n-2a} \left[(\eta + v) \prod_{i=1}^{n} \omega_i\right]^{-\frac{1}{2}} dH\,(\omega_1) \cdots dH\,(\omega_n)}{2^{\frac{n}{2}} \int \cdots \int r^{-n-2a} \left[(\eta + v) \prod_{i=1}^{n} \omega_i\right]^{-\frac{1}{2}} g\,(\boldsymbol{\omega})\, dH\,(\omega_1) \cdots dH\,(\omega_n)}. \qquad (4.2.7)$$

where

$$g\,(\boldsymbol{\omega}) = \int F_{\mathbf{T}} \left(\lambda \sqrt{n + 2a} \frac{\mathbf{x} - \hat{\mu}\mathbb{1}_n}{r} \left[\mathbf{D}\,(\boldsymbol{\omega})\right]^{-1}\right) \pi\,(\lambda)\, d\lambda,$$

$F_{\mathbf{T}}$ *is the c.d.f. of a* $t_n\,(\mathbf{0}, \boldsymbol{\Sigma}, n + 2a)$*,* $\hat{\mu} = \frac{\eta}{\eta+v}\left(\sum_{i=1}^{n} \nu_i x_i + vm\right)$*,* $\eta = \sum_{i=1}^{n} \omega_i^{-1}$*,* $\boldsymbol{\Sigma} = \left[\mathbf{D}\,(\boldsymbol{\omega})\right]^{-1} + \frac{\lambda^2}{v+\eta} \left[\mathbf{D}\,(\boldsymbol{\omega})\right]^{-1} \mathbb{1}_n \mathbb{1}_n^t \left[\mathbf{D}\,(\boldsymbol{\omega})\right]^{-1}$*,* $\mathbf{D}\,(\boldsymbol{\omega}) = diag\,(\omega_1, \ldots, \omega_n)$*,* $r^2 = \eta S_\omega^2 + \frac{\eta v}{\eta+v}\left(m - \sum_{i=1}^{n} \nu_i x_i\right)^2 + 2b$*,* $S_\omega^2 = \sum_{i=1}^{n} \nu_i \left(x_i - \sum_{j=1}^{n} \nu_j x_j\right)^2$ *and* $\nu_i = \frac{\omega_i}{\eta}$ *for each* $i = 1, \ldots, n$*.*

Proof. See Appendix D. ■

In the special case where $\omega_1 = \omega_2 = \cdots = \omega_n = 1$, the result of Proposition 4.2.6 agrees with Liseo and Loperfido (2002) result.

4.3 Simulation Results

In this section we perform a simulation study to describe the behavior of the Bayes factor given by (4.2.5). In Subsection 4.3.1 we make use of the fact that the Student-t distribution can be written as a mixture of normal distributions. In this case (4.2.6) is given by

$$\left\{\int_0^\infty \prod_{i=1}^{n} \left[2\Phi\left(\frac{x_i - \mu}{\sigma}\sqrt{\frac{\omega a}{b}}\right)\right] Ga\,(\omega\,|a, a\,)\, d\omega\right\}^{-1}.$$

We calculate the previous integral using the MATLAB integration function (*quad*) based on the recursive adaptive Simpson quadrature method.

Table 4.1: Mean and variance for prior distributions in simulation study.

b	$\mathbb{E}(\lambda)$	$\mathbb{V}(\lambda)$	$\frac{a}{b}$
0.1	2.8	2.15	10
1	0.89	0.21	1
5	0.4	0.04	0.2

4.3.1 Normal Versus Skew-Normal

For each value of λ, $\lambda = 0, 0.1, \ldots, 0.5$, we generate 1000 independent data sets, $y_1, \ldots, y_n$, from p.d.f. $2\phi(y)\,\Phi(\lambda y)$ with $n = 10, 50, 100$. Then, for each data set we calculate (4.2.6) considering $\mu = 0$, $\sigma = 1$ and $\lambda^2 \sim Ga(1, b)$, where $b = 0.1, 1, 5$.

From $\lambda^2 \sim Ga(a, b)$ the prior p.d.f. for λ is given by

$$f(\lambda\,|a, b) = \frac{2b^a}{\Gamma(a)}\left(\lambda^2\right)^{a-\frac{1}{2}}\exp\left(-b\lambda^2\right),$$

and the variance and mean are given, respectively, by

$$\begin{aligned}\mathbb{V}(\lambda) &= \frac{a}{b} - \frac{\Gamma^2\left(a+\frac{1}{2}\right)}{b\Gamma^2(a)},\\ \mathbb{E}(\lambda) &= \frac{\Gamma\left(a+\frac{1}{2}\right)}{\sqrt{b}\Gamma(a)} \text{ if } \lambda > 0\end{aligned}$$

and

$$\mathbb{E}(\lambda) = -\frac{\Gamma\left(a+\frac{1}{2}\right)}{\sqrt{b}\Gamma(a)} \text{ if } \lambda < 0.$$

Therefore, for the prior distributions that we use in the simulations, we obtain the Table 4.1. Please note that the prior variance of λ is always smaller than $\frac{a}{b}$. Thus, if we want to have a big prior variance, then we have to consider a much bigger than b. Therefore, in this case the prior mean will be big also. Similar results are obtained for negatives values of λ using Proposition 4.2.4.

For each sample size (n) and λ value, the Bayes factor was calculated using 1000 simulated samples. We considered some strong pieces of evidence in favor of the

Table 4.2: Simulation results with $\lambda^2 \sim Ga\,(1, 0.1)$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
10	0	4.4	89.4	8.4193	28.4309	66.0642
	0.1	5.1	86.9	6.9357	21.2874	52.7914
	0.2	9.1	82.5	3.5738	13.9176	41.8535
	0.3	14.0	72.3	1.6913	7.3060	24.1999
	0.4	17.4	66.4	0.9266	5.4030	18.6577
	0.5	24.3	58.9	0.5541	3.5752	12.0782
50	0	0.5	98.4	60.5811	158.1004	319.9009
	0.1	1.7	95.2	21.4082	67.7103	169.2791
	0.2	4.9	87.7	6.2691	26.0408	77.3927
	0.3	12.4	72.4	1.6805	8.4769	38.5235
	0.4	27.4	55.3	0.3907	2.8809	12.9466
	0.5	47.1	33.3	0.0636	0.6080	3.7100
100	0	0.2	99.3	121.7138	310.8615	643.1082
	0.1	2.3	95.2	27.9092	103.7148	261.5884
	0.2	10.6	79.7	2.9946	17.4966	74.9616
	0.3	25.8	56.9	0.4623	3.3920	18.7598
	0.4	55.5	26.8	0.0282	0.3287	2.3751
	0.5	81.9	9.1	0.0016	0.0240	0.2455

asymmetric model if $\hat{BF} < 0.5$, and in favor of the symmetric model if $\hat{BF} > 2$. Tables 4.2 to 4.4 display, in Column 3 and 4, the percentage of the samples that presented evidence in favor of the asymmetric and symmetric models, respectively; from Column 5 to 7, the 25^{th} percentile, median and 75^{th} percentile of the Bayes factor values, respectively.

Tables 4.2 to 4.4 show very good results. Note that for each sample size the $\hat{BF}$ decreases when the value of λ increases, and for $\lambda = 0$ tend to be quite big. This desired behavior is appreciated better for higher sample sizes. In general, the calculated Bayes factors show correct evidence when $\lambda = 0$ and when $\lambda \geq 0.4$, and while the sample size increases, this evidence improves. For $\lambda \geq 1$, the Bayes factor

Table 4.3: Simulation results with $\lambda^2 \sim Ga\,(1,1)$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
10	0	10.4	68.5	1.4078	3.7729	7.8112
	0.1	12.4	63.3	1.2134	2.9230	6.3771
	0.2	17.9	50.6	0.7471	2.0380	5.1615
	0.3	28.1	37.5	0.4165	1.2357	3.2828
	0.4	34.9	31.5	0.2951	0.9964	2.6781
	0.5	42.4	23.8	0.2186	0.7602	1.8649
50	0	2.1	91.5	6.9877	17.8707	36.4650
	0.1	8.7	76.6	2.2403	7.7218	19.0447
	0.2	20.1	57.3	0.7349	2.8510	8.3711
	0.3	38.5	36.8	0.2347	1.0252	3.6563
	0.4	59.6	20.1	0.0514	0.2602	1.3642
	0.5	73.6	8.6	0.0130	0.1075	0.5436
100	0	1.1	96.5	13.2021	34.0111	76.7385
	0.1	7.2	81.6	3.1714	9.8879	25.5317
	0.2	24.9	54.3	0.5055	2.5016	9.1792
	0.3	53.5	27.1	0.0539	0.3897	2.3266
	0.4	80.4	7.6	0.0045	0.0465	0.3220
	0.5	93.8	2	0.0002	0.0035	0.0342

Table 4.4: Simulation results with $\lambda^2 \sim Ga\,(1,5)$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
10	0	11.8	38.0	0.8493	1.5772	2.5897
	0.1	13.8	30.0	0.7558	1.3369	2.2663
	0.2	20.5	23.7	0.5797	1.0459	1.9641
	0.3	33.2	12.5	0.4136	0.7660	1.4244
	0.4	38.4	10.0	0.3463	0.6743	1.2524
	0.5	42.9	5.1	0.2952	0.5855	1.0312
50	0	7.2	73.0	1.8322	4.1703	7.7046
	0.1	16.7	50.3	0.7710	2.0095	4.3708
	0.2	35.1	27.7	0.2951	0.9229	2.2482
	0.3	56.5	14.9	0.1107	0.3736	1.2628
	0.4	73.9	6.7	0.0359	0.1637	0.5405
	0.5	87.7	1.5	0.0110	0.0523	0.2118
100	0	4.3	83.0	3.0349	7.2055	14.2496
	0.1	17.7	57.4	0.8089	2.6087	6.1743
	0.2	48.6	24.6	0.1147	0.5354	1.9529
	0.3	72.4	7.0	0.0232	0.1234	0.5619
	0.4	92.0	0.9	0.0023	0.0176	0.0937
	0.5	98.4	0.0	0.0002	0.0019	0.0136

Table 4.5: Simulation results with $\nu = 1$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
	0	11.7	71.7	1.5312	6.0873	24.8469
	0.1	26.0	48.0	0.4909	1.7623	5.1188
10	0.2	36.1	31.2	0.2949	0.8545	2.6339
	0.3	42.7	23.2	0.2176	0.6198	1.7899
	0.4	54.1	14.9	0.1393	0.4391	1.0920
	0.5	60.8	14.5	0.1032	0.3198	0.9775
	0	1.4	95.6	29.7884	133.4624	594.1999
	0.1	29.4	49.9	0.3438	1.9686	8.9924
50	0.2	59.6	21.4	0.0316	0.2662	1.4956
	0.3	79.7	8.4	0.0049	0.0427	0.3417
	0.4	89.3	4.0	0.0018	0.0140	0.1033
	0.5	93.4	2.4	0.0004	0.0041	0.0370
	0	0.3	98.7	148.4	608.8	2737.9
	0.1	41.9	38.2	0.1	0.8	6.2
100	0.2	84.1	7.6	0.0	0.0	0.2
	0.3	94.9	2.8	0.0	0.0	0.0
	0.4	99.0	0.3	0.0	0.0	0.0
	0.5	99.6	0.0	0.0	0.0	0.0

values are almost zero. We can see the strong dependence of the values of Bayes factor from prior specification.

4.3.2 Student Versus Skew-Student

For each $\lambda = 0, 0.1, \ldots, 0.5$, we generate 1000 independent data sets, $y_1, \ldots, y_n$, with $n = 10, 50, 100$, from the skew-t distribution, where $f(y) = t(y\,|0, 1, \nu)$. We consider $\nu = 1, 3, 10, 20$ and calculate (4.2.5) considering $\lambda^2 \sim Ga(1, 1)$.

Tables 4.5 to 4.8 exhibit the results of the performed simulations for each different values of ν. Similar results to the normal case were obtained, highlighting those better results obtained for small values of ν. We note that, in this case as well as in the normal case, the BF has a nice behavior for small values of λ.

Table 4.6: Simulation results with $\nu = 3$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
10	0	9.6	65.9	1.3456	3.9168	10.7615
	0.1	18.9	55.3	0.7665	2.4418	6.2970
	0.2	26.1	44.6	0.4711	1.5983	4.7274
	0.3	34.1	32.6	0.3219	0.9850	2.9351
	0.4	46.8	21.1	0.1937	0.5647	1.6797
	0.5	48.1	19.4	0.1583	0.5458	1.4894
50	0	2.8	92.1	8.8131	26.0701	64.3461
	0.1	10.4	74.5	1.9732	8.1181	22.9338
	0.2	29.7	48.0	0.3317	1.8392	6.8588
	0.3	49.9	26.0	0.0701	0.5011	2.1223
	0.4	73.2	12.5	0.0146	0.1110	0.5871
	0.5	84.1	4.3	0.0034	0.0283	0.1942
100	0	1.0	96.6	20.2344	64.4118	142.0406
	0.1	10.8	76.8	2.1960	9.6536	32.6166
	0.2	43.3	37.7	0.1266	0.8049	4.5909
	0.3	75.6	10.9	0.0074	0.0679	0.4631
	0.4	91.8	2.5	0.0005	0.0071	0.0676
	0.5	98.3	0.6	0.0000	0.0005	0.0061

Table 4.7: Simulation results with $\nu = 10$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
10	0	10.3	69.3	1.5578	4.3809	9.2094
	0.1	14.7	56.4	0.9371	2.5278	6.4705
	0.2	20.6	49.0	0.6525	1.9159	4.8182
	0.3	29.2	38.7	0.4043	1.2866	3.3779
	0.4	37.3	27.6	0.2820	0.8020	2.2204
	0.5	46.7	21.1	0.1758	0.5682	1.6956
50	0	2.6	91.2	7.0075	19.2956	43.3332
	0.1	25.5	45.6	0.4844	1.6349	4.6176
	0.2	22.1	58.7	0.6381	3.0221	9.1743
	0.3	41.5	33.0	0.1552	0.7937	3.2732
	0.4	61.8	17.4	0.0372	0.2468	1.1372
	0.5	78.3	6.5	0.0078	0.0593	0.3789
100	0	2.1	94.7	15.2573	38.3028	87.9117
	0.1	7.5	81.7	3.4485	11.4648	32.0435
	0.2	27.0	51.5	0.3982	2.1520	7.8946
	0.3	58.3	19.0	0.0294	0.2585	1.3151
	0.4	86.3	5.5	0.0018	0.0179	0.1602
	0.5	95.6	1.1	0.0002	0.0022	0.0241

Table 4.8: Simulation results with $\nu = 20$ and for different values of n and λ.

n	λ	$\hat{BF} < 0.5$	$\hat{BF} > 2$	25^{th} p.	Median	75^{th} p.
10	0	9.2	67.2	1.3782	3.6613	8.6506
	0.1	15.0	58.0	0.8423	2.6160	6.5973
	0.2	22.8	48.2	0.6223	1.9014	4.5525
	0.3	27.4	36.1	0.4471	1.2501	3.1800
	0.4	34.4	30.9	0.2813	0.9360	2.4928
	0.5	45.2	23.2	0.1925	0.6397	1.8215
50	0	2.3	92.1	6.7759	17.6408	37.4567
	0.1	8.9	77.3	2.2818	7.9469	19.0925
	0.2	20.4	56.4	0.7003	2.6450	8.7393
	0.3	39.3	36.8	0.1860	0.9700	3.3768
	0.4	61.4	19.1	0.0406	0.2736	1.2038
	0.5	79.2	7.6	0.0079	0.0624	0.3695
100	0	0.8	95.3	14.8925	35.9162	76.7650
	0.1	6.9	79.6	2.6278	10.5830	28.1940
	0.2	27.4	51.5	0.4086	2.2116	8.2338
	0.3	53.9	24.1	0.0466	0.3766	1.9179
	0.4	81.9	6.3	0.0037	0.0360	0.2745
	0.5	95.0	1.3	0.0002	0.0027	0.0332

Table 4.9: Bayes factors by company.

	$\lambda^2 \sim Ga\,(1,1)$	$\lambda^2 \sim Ga\,(1,0.001)$
Company	$\hat{BF}_{\lambda>0}$	$\hat{BF}_{\lambda>0}$
Cementos	0.6297	137.2086
Cervezas	0.3033	16.6370
Chilquinta	0.1936	10.7234
Copec	0.5601	51.4174
Iansa	1.3841	408.8198

4.4 An Application

In this section we use the results from proposition 4.2.3 and 4.2.4 in a real data set. The data set comes from the "Bolsa de Comercio de Santiago de Chile" and it consists in the monthly rentability of five Chilean companies measured between March, 1990 and April, 1999. The sample size is $n = 110$ for each company. In Appendix B, we present some descriptive statistics, including the skewness and kurtosis.

For each company, Table 4.9 presents in Columns 2 and 3 the Bayes factor under two prior distributions $\lambda^2 \sim Ga\,(1,1)$ and $\lambda^2 \sim Ga\,(1,0.001)$, respectively. We consider the λ sign known, $\lambda > 0$. It can be justified because in this period the Chilean companies presented an affluent economy.

From Table 4.9, we can see that using a more informative prior, the only data set in favor of the symmetry assumption is the Iansa company. On the other side, with the less informative prior, all Bayes factor values are extremely large. This confirms the high sensitivity of the Bayes factor to prior chosen and its wrong behavior when we used vague prior distributions. From the practical point of view, we consider the results given in the second column of Table 4.9. Therefore, the positive asymmetry is more evident in the company Chilquinta following by company Cervezas.

Chapter 5

Testing of Asymmetry in Linear Regression Model

In this chapter we present some results related with the Bayes factors with the purpose of detecting asymmetry on the errors distribution in a linear regression model. With this aim, in the first section, we define the multivariate skew-elliptical distribution which we use throughout this chapter. In the second section, we estimate the Bayes factor for some particular cases and in the third, we perform a simulation study in order to discern the behavior of the Bayes factor, given in Subsection 5.2.1, with respect to different prior distributions. Finally, we apply the results to real data.

5.1 Multivariate Skew-Elliptical Distributions

In this section we define the multivariate skew-elliptical distribution which we employ through this chapter. We use the definition given by Branco and Dey (2001), this definition includes interesting particular cases discussed in Kelker (1970), Fang and Zhang (1990), Fang *et al.* (1990), Azzalini and Dalla-Valle (1996), and Azzalini and

Capitanio (1999).

Consider that the random vector $\mathbf{X}^* = (X_0, X_1, \ldots, X_n)^t$ has $El_{n+1}\left(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, h^{(n+1)}\right)$ distribution, where the location parameter is the vector $\boldsymbol{\mu}^* = (0, \boldsymbol{\mu}^t)^t$ and the scale parameter is given by the matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \boldsymbol{\delta}^t \\ \boldsymbol{\delta} & \boldsymbol{\Omega} \end{pmatrix},$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^t$ and $\boldsymbol{\Omega}$ is the scale matrix associated to the vector $\mathbf{X} = (X_1, \ldots, X_n)^t$.

Under these assumptions, we will say that the random vector $\mathbf{Y} \stackrel{d}{=} \mathbf{X}|\, X_0 > 0$ has skew-elliptical distribution. The next results obtained by Branco and Dey (2001) will be used in the following section.

Proposition 5.1.1. *If the p.d.f. of the random vector $\mathbf{X}^*$ exists and is continuous, then the p.d.f. of $\mathbf{Y}$ is given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = 2 f_{h^{(n)}}(\mathbf{y})\, F_{h_{q(\mathbf{y})}}\left[\boldsymbol{\lambda}^t \boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})\right],$$

where $f_{h^{(n)}}(\mathbf{y})$ is the p.d.f. associated with the distribution $El_n\left(\boldsymbol{\mu}, \boldsymbol{\Omega}, h^{(n)}\right)$ and $F_{h_{q(\mathbf{y})}}$ is the c.d.f. of $El\left(0, 1, h_{q(\mathbf{y})}\right)$, where

$$\boldsymbol{\lambda}^t = \frac{\boldsymbol{\delta}^t \boldsymbol{\Omega}^{-\frac{1}{2}}}{\left(1 - \boldsymbol{\delta}^t \boldsymbol{\Omega}^{-1} \boldsymbol{\delta}\right)^{\frac{1}{2}}},$$

$$h^{(n)}(u) = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty h^{(n+1)}\left(r^2 + u\right) r^{n-1} dr, \qquad u \geq 0,$$

$$h_{q(\mathbf{y})}(u) = \frac{h^{(n+1)}[u + q(\mathbf{y})]}{h^{(n)}[q(\mathbf{y})]},$$

and $q(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu})^t \boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

Proof. From the Bayes theorem we obtain $f_{\mathbf{Y}}(\mathbf{y}) = 2\mathbb{P}(X_0 > 0|\, \mathbf{y})\, f_{\mathbf{X}}(\mathbf{y})$. On the other hand, $\mathbf{X}^* \sim El_{n+1}\left(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, h^{(n+1)}\right)$ and by the marginalization and conditionally properties of the elliptical distributions (see Fang and Zhang (1990)), we get $\mathbf{X} \sim El_n\left(\boldsymbol{\mu}, \boldsymbol{\Omega}, h^{(n)}\right)$. So that $f_{\mathbf{X}}(\mathbf{y}) = f_{h^{(n)}}(\mathbf{y})$ and $X_0|\,\mathbf{y} \sim El\left(\boldsymbol{\delta}^t\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu}), 1-\boldsymbol{\delta}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}, h_{q(\mathbf{y})}\right)$. Now, considering $Z_0 = \frac{X_0 - \boldsymbol{\delta}^t\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\left(1-\boldsymbol{\delta}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}\right)^{\frac{1}{2}}}$ and the symmetric property of the elliptical distribution, it follows that

$$\mathbb{P}\left(X_0 > 0|\,\mathbf{y}\right) = \mathbb{P}\left(Z_0 > -\frac{\boldsymbol{\delta}^t\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\left(1-\boldsymbol{\delta}^t\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}\right)^{\frac{1}{2}}}\middle|\,\mathbf{y}\right) = F_{h_{q(\mathbf{y})}}\left[\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})\right].$$

■

Hereafter, to denote that the random vector $\mathbf{Y}$ has a skew-elliptical distribution, we will note down $\mathbf{Y} \sim SE_n\left(\boldsymbol{\mu}, \boldsymbol{\Omega}, h^{(n)}, \boldsymbol{\lambda}\right)$. Particular cases and properties of this class of distributions can be found in Branco and Dey (2001).

Note that if $\boldsymbol{\lambda} = \mathbf{0}$, we get the symmetric model,

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{h^{(n)}}(\mathbf{y}) = |\boldsymbol{\Omega}|^{-\frac{1}{2}}\, h^{(n)}\left[(\mathbf{y}-\boldsymbol{\mu})^t\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right].$$

The next corollary gives an alternative expression to the p.d.f. of a skew-elliptical distribution that could be convenient in many occasions.

Corollary 5.1.2. *An alternative expression to the p.d.f.* $SE_n\left(\boldsymbol{\mu}, \boldsymbol{\Omega}, h^{(n)}, \boldsymbol{\lambda}\right)$ *is,*

$$f_{\mathbf{Y}}(\mathbf{y}) = 2\,|\boldsymbol{\Omega}|^{-\frac{1}{2}} \int_{-\infty}^{\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})} h^{(n+1)}\left[u^2 + q(\mathbf{y})\right] du,$$

where $q(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu})^t\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})$.

Proof.

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= 2f_{h^{(n)}}(\mathbf{y})\, F_{h_{q(\mathbf{y})}}\left[\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})\right] \\
&= 2\left|\boldsymbol{\Omega}\right|^{-\frac{1}{2}} h^{(n)}\left[q(\mathbf{y})\right]\int_{-\infty}^{\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})} h_{q(\mathbf{y})}\left[u^2\right] du \\
&= 2\left|\boldsymbol{\Omega}\right|^{-\frac{1}{2}} h^{(n)}\left[q(\mathbf{y})\right]\int_{-\infty}^{\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})} \frac{h^{(n+1)}\left[u^2+q(\mathbf{y})\right]}{h^{(n)}\left[q(\mathbf{y})\right]} du \\
&= 2\left|\boldsymbol{\Omega}\right|^{-\frac{1}{2}}\int_{-\infty}^{\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})} h^{(n+1)}\left[u^2+q(\mathbf{y})\right] du.
\end{aligned}
$$

■

Note that if $h^{(n+1)}(u) = (2\pi)^{-\frac{n+1}{2}}\exp\left\{-\frac{u}{2}\right\}$, the generator function for a $(n+1)$-variate normal distribution, then, from the previous corollary we obtain

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= 2\left|\boldsymbol{\Omega}\right|^{-\frac{1}{2}}\int_{-\infty}^{\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})} (2\pi)^{-\frac{n+1}{2}}\exp\left\{-\frac{u^2+q(\mathbf{y})}{2}\right\} du \\
&= 2\left|\boldsymbol{\Omega}\right|^{-\frac{1}{2}}(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{q(\mathbf{y})}{2}\right\}\int_{-\infty}^{\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})} (2\pi)^{-\frac{1}{2}}\exp\left\{-\frac{u^2}{2}\right\} du \\
&= 2N_n(\mathbf{y}\,|\boldsymbol{\mu},\boldsymbol{\Omega})\,\Phi\left[\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})\right].
\end{aligned}
$$

That is to say, we obtain the multivariate skew-normal distribution defined in Azzalini and Dalla-Valle (1996), which we will denote by $SN_n(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda})$.

5.2 Bayes Factor

In this section we assume that have a data set coming from the following linear regression model,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{5.2.1}$$

where $\boldsymbol{\beta}\in\mathbb{R}^k$, $\boldsymbol{\epsilon}\sim SE_n\left(\mathbf{0},\phi^{-1}\mathbf{I}_n,h^{(n)},\lambda\mathbb{1}_n\right)$, $\phi>0$ and $\lambda\in\mathbb{R}$. Our goal will be to search for evidences in the data that allow us to choose between a symmetrical or asymmetric model for the errors. In other words, evidences will be looked from the data to favor of $\lambda = 0$ or, $\lambda \neq 0$.

Note that if $\lambda = 0$, the data coming from a distribution with p.d.f. given by

$$\phi^{\frac{n}{2}} h^{(n)} \left(\phi \left\|\mathbf{y} - \mathbf{X}\ \boldsymbol{\beta}\right\|^2\right),$$

this case were studied in Chapter 1, and if $\lambda \neq 0$, the data coming from

$$2\phi^{\frac{n}{2}} h^{(n)} \left[q\left(\mathbf{y}\right)\right] \int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} \frac{h^{(n+1)}\left[u^2 + q\left(\mathbf{y}\right)\right]}{h^{(n)}\left[q\left(\mathbf{y}\right)\right]} du,$$

where $q\left(\mathbf{y}\right) = \phi \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right\|^2$. So that, for the data $\mathbf{y} = (y_1, \ldots, y_n)$, the Bayes factor in favor of $\lambda = 0$ is,

$$BF = \frac{\int \phi^{\frac{n}{2}} h^{(n)} \left[q\left(\mathbf{y}\right)\right] \pi\left(\boldsymbol{\beta}, \phi\right) d\boldsymbol{\beta} d\phi}{2\int \left\{\phi^{\frac{n}{2}} h^{(n)} \left[q\left(\mathbf{y}\right)\right] \int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right) du\right\} \pi\left(\boldsymbol{\beta}, \phi, \lambda\right) d\boldsymbol{\beta} d\phi d\lambda}, \quad (5.2.2)$$

where $\pi\left(\cdot\right)$ represents the prior distribution for the respecting parameters. As we can see, by expression (5.2.2), a closed form to Bayes factor is not possible to obtain in a general way. Also, the numerical computation is complex. However, when $\boldsymbol{\beta}$ and ϕ are known, Equation (5.2.2) has a better expression:

$$BF = \frac{1}{2\int \left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right) du\right] \pi\left(\lambda\right) d\lambda}.$$

The next proposition gives a particular case which is an example where the Bayes factor can not discriminate between a symmetric and asymmetric model.

Proposition 5.2.1. *Let* $\mathbf{y} = (y_1, \ldots, y_n)$ *a random sample from the model (5.2.1), where* $\boldsymbol{\epsilon} \sim SN_n\left(\mathbf{0}, \phi^{-1}\mathbf{I}_n, \lambda\ \mathbb{1}_n\right)$. *If* $\lambda \perp\!\!\!\perp (\boldsymbol{\beta}, \phi)$ *and* $\lambda \sim N\left(0, v^2\right)$. *Then the Bayes factor (5.2.2) is equal to 1.*

Proof. Notice that

$$\begin{aligned}\int\left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right)du\right]\pi\left(\lambda\right)d\lambda &= \int\left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du\right]\pi\left(\lambda\right)d\lambda \\ &= \int\mathbb{P}\left[Z\leq\lambda\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\pi\left(\lambda\right)d\lambda,\end{aligned}$$

where $Z\sim N\left(0,1\right)$. But since $Z\perp\!\!\!\perp\lambda$ and

$$Z-\lambda\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\sim N\left(0,1+\phi v^2\left[\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]^2\right),$$

then

$$\begin{aligned}\int\mathbb{P}\left[Z\leq\lambda\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\pi\left(\lambda\right)d\lambda &= \int\mathbb{P}\left[Z\leq\lambda\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\middle|\lambda\right]\pi\left(\lambda\right)d\lambda \\ &= \mathbb{P}\left[Z\leq\lambda\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right] \\ &= \mathbb{P}\left[Z-\lambda\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\leq 0\right]=2.\end{aligned}$$

Therefore, $\int\left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right)du\right]\pi\left(\lambda\right)d\lambda=2$ and the Bayes factor given by (5.2.2) is equal to 1. ■

The previous proposition is telling us to assume a normal distribution with zero mean as prior distribution for λ, is equivalent to suppose a normal linear regression model for the data $\mathbf{y}$. Others particular cases of the Bayes factor to compare a symmetric distribution with an asymmetric one are the next propositions, where it is necessary to know the sign of λ. The knowledge of the skewness parameter sign is feasible in practice. Before presenting those propositions we must recall the following lemma that will be us useful for the proof of the next propositions.

Lemma 5.2.2. *If* $\mathbf{Y}\left|\boldsymbol{\mu},\ \boldsymbol{\Sigma},\tau\right.\sim N_n\left(\ \boldsymbol{\mu},\tau^{-1}\boldsymbol{\Sigma}\right)$ *and* $\tau\left|a,b\right.\sim Ga\left(a,b\right)$*, then* $\mathbf{Y}\left|\boldsymbol{\mu},\boldsymbol{\Sigma},a,b\right.\sim t_n\left(\boldsymbol{\mu},\frac{b}{a}\boldsymbol{\Sigma},2a\right)$.

Proof. If $q(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$, then

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, a, b) &= \int_0^\infty N_n\left(\mathbf{y} \mid \boldsymbol{\mu}, \tau^{-1}\boldsymbol{\Sigma}\right) Ga(\tau \mid a, b)\, d\tau \\ &= \frac{b^a}{\Gamma(a)} (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \int_0^\infty \tau^{\frac{n}{2}+a-1} \exp\left\{-\tau\left[\frac{1}{2}q(\mathbf{y}) + b\right]\right\} d\tau \\ &= \frac{b^a}{\Gamma(a)} (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \Gamma\left(a + \frac{n}{2}\right) \left[\frac{1}{2}q(\mathbf{y}) + b\right]^{-\left(a+\frac{n}{2}\right)} \\ &= \left|\frac{b}{a}\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \frac{(2a)^a}{\Gamma(a)} \pi^{-\frac{n}{2}} \Gamma\left(\frac{n+2a}{2}\right) \left[2a + \frac{a}{b}q(\mathbf{y})\right]^{-\frac{n+2a}{2}}. \end{aligned}$$

■

Proposition 5.2.3. *Let* $\mathbf{y} = (y_1, \ldots, y_n)$ *a random sample from (5.2.1), where* $\boldsymbol{\epsilon} \sim SN_n\left(\mathbf{0}, \phi^{-1}\mathbf{I}_n, \lambda\mathbb{1}_n\right)$ *with* $\lambda \perp\!\!\!\perp (\boldsymbol{\beta}, \phi)$, $\lambda^2 \sim Ga(a, b)$ *and* $\mathbb{P}(\lambda > 0) = 1$. *Then, the Bayes factor (5.2.2) to compare the models specified by the hypotheses* $H_0 : \lambda = 0$ *and* $H_1 : \lambda > 0$ *is given by*

$$BF = \frac{\int \phi^{\frac{n}{2}} h^{(n)}[q(\mathbf{y})]\, \pi(\boldsymbol{\beta}, \phi)\, d\boldsymbol{\beta} d\phi}{2\int \phi^{\frac{n}{2}} h^{(n)}[q(\mathbf{y})]\, F_T\left[\mathbb{1}_n^t(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\sqrt{\frac{a\phi}{b}}\right] \pi(\boldsymbol{\beta}, \phi)\, d\boldsymbol{\beta} d\phi},$$

where F_T *is the c.d.f. of the* $t(0, 1, 2a)$ *and* $q(\mathbf{y}) = \phi\,\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$.

Proof. Notice that if $\lambda^2 \sim Ga(a, b)$, then λ has probability density function $f(\lambda \mid a, b) = \frac{2b^a}{\Gamma(a)} \left(\lambda^2\right)^{a-\frac{1}{2}} \exp\left(-b\lambda^2\right)$. Therefore,

$$\begin{aligned} \int \left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right) du\right] \pi(\lambda)\, d\lambda &= \frac{2b^a}{\sqrt{2\pi}\Gamma(a)} \int_0^\infty \lambda^{2a-1} e^{-b\lambda^2} \\ &\quad \int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} e^{-\frac{u^2}{2}} du d\lambda \\ &= \frac{2b^a}{\sqrt{2\pi}\Gamma(a)} \int_0^\infty \lambda^{2a} e^{-b\lambda^2} \\ &\quad \int_{-\infty}^{\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} e^{-\frac{\lambda^2 r^2}{2}} dr d\lambda. \end{aligned}$$

Then, with the change of variable $l = \lambda^2$, we obtain

$$\int \left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right) du \right] \pi\left(\lambda\right) d\lambda = \frac{b^a}{\sqrt{2\pi}\Gamma\left(a\right)} \int_0^{\infty} l^{a-\frac{1}{2}} e^{-bl} \int_{-\infty}^{\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} e^{-\frac{lr^2}{2}} dr dl,$$

and, from Lemma 5.2.2,

$$\int \left[\int_{-\infty}^{\lambda\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right) du \right] \pi\left(\lambda\right) d\lambda = F_T\left[\sqrt{\phi}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right],$$

where $T \sim t\left(0, \frac{b}{a}, 2a\right)$. Then, the result of the proposition is immediate. ■

If in the previous proposition we assume $\boldsymbol{\beta}$ and ϕ known, then

$$BF = \frac{1}{2F_T\left[\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\sqrt{\frac{a\phi}{b}}\right]}.$$

In a similar way the following result is also obtained.

Proposition 5.2.4. *Let* $\mathbf{y} = (y_1, \ldots, y_n)$ *a random sample from (5.2.1), where* $\boldsymbol{\epsilon} \sim SN_n\left(\mathbf{0}, \phi^{-1}\mathbf{I}_n, \lambda\ \mathbb{1}_n\right)$ *with* $\lambda \perp\!\!\!\perp (\boldsymbol{\beta}, \phi)$*,* $\lambda^2 \sim Ga\left(a, b\right)$ *and* $\mathbb{P}\left(\lambda < 0\right) = 1$*. Then the Bayes factor (5.2.2) to compare the models defined by the hypotheses* $H_0 : \lambda = 0$ *and* $H_1 : \lambda < 0$ *is given by*

$$BF = \frac{\int \phi^{\frac{n}{2}} h^{(n)}\left[q\left(\mathbf{y}\right)\right] \pi\left(\boldsymbol{\beta}, \phi\right) d\boldsymbol{\beta} d\phi}{2\int \phi^{\frac{n}{2}} h^{(n)}\left[q\left(\mathbf{y}\right)\right] F_T\left[\mathbb{1}_n^t\left(\mathbf{X}\boldsymbol{\beta}-\mathbf{y}\right)\sqrt{\frac{a\phi}{b}}\right] \pi\left(\boldsymbol{\beta}, \phi\right) d\boldsymbol{\beta} d\phi},$$

where F_T *is the c.d.f. of the* $t\left(0, 1, 2a\right)$ *and* $q\left(\mathbf{y}\right) = \phi\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2$.

Proof. Notice that if $\lambda^2 \sim Ga\left(a, b\right)$, then $\lambda < 0$ has density function $f\left(\lambda\left|a, b\right.\right) =$

$\frac{2b^a}{\Gamma(a)}\left(\lambda^2\right)^{a-\frac{1}{2}}\exp\left(-b\lambda^2\right)$. Therefore, making the change of variable $l=-\lambda$,

$$\begin{aligned}\int\left[\int_{-\infty}^{\lambda\sqrt{\phi}\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right)du\right]\pi\left(\lambda\right)d\lambda &= -\int_{\infty}^{0}\left[\int_{-\infty}^{-l\sqrt{\phi}\;\mathbb{1}_n^t(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} h_{q(\mathbf{y})}\left(u^2\right)du\right] \\ &\quad f\left(-l\left|a,b\right.\right)dl \\ &= \int_{0}^{\infty}\left[\int_{-\infty}^{l\sqrt{\phi}\mathbb{1}_n^t(\mathbf{X}\boldsymbol{\beta}-\mathbf{y})} h_{q(\mathbf{y})}\left(u^2\right)du\right] \\ &\quad f\left(l\left|a,b\right.\right)dl.\end{aligned}$$

Then, it is enough to follow the proof of the previous proposition. ■

When the parameters $\boldsymbol{\beta}$ and ϕ are not known, the previous propositions are not a solution to the problem described to the beginning of this chapter due to the great analytic and numeric complexity that presents the calculation of the Bayes factors given by these propositions. A more tractable case is presented in the next subsection.

5.2.1 Bayes Factor for Representable Skew Elliptical Linear Model

The propositions of the previous section use skew-normal distributions for the error distribution of the model (5.2.1). In this section we will work with a wider class of skew-elliptical distributions than the skew-normal class, however the analytic exertion is not much more complicated than in the skew-normal case.

Definition 5.2.5. *We will say that $\mathbf{Y}|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda}$ has multivariate skew-elliptical representable distribution under the c.d.f. H if its p.d.f. can be written by*

$$f_{\mathbf{Y}|\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda}}\left(\mathbf{y}\right)=2\int_0^{\infty}N_n\left(\mathbf{y}\left|\boldsymbol{\mu},\omega\boldsymbol{\Omega}\right.\right)\Phi\left[\omega^{-\frac{1}{2}}\boldsymbol{\lambda}^t\boldsymbol{\Omega}^{-\frac{1}{2}}\left(\mathbf{y}-\boldsymbol{\mu}\right)\right]dH\left(\omega\right),$$

where H is the c.d.f. of a random variable ω, which is non-negative and such that $\omega \perp\!\!\!\perp (\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda})$.

An equivalent definition is given by: $\mathbf{Y}|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda}$ is skew representable if and only if there is $\omega \sim H$ and $\omega \perp\!\!\!\perp (\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda})$ such that $\mathbf{Y}|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda},\omega \sim SN_n\left(\boldsymbol{\mu},\omega\boldsymbol{\Omega},\boldsymbol{\lambda}\right)$. $\mathbf{Y}|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda} \sim RSE_n\left(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda}\right)$ will be used to denote that $\mathbf{Y}|\,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda}$ has multivariate skew-elliptical representable distribution. Properties and examples of this class of distributions can be found in Branco and Dey (2001).

Proposition 5.2.6. *Let* $\mathbf{y} = (y_1,\ldots,y_n)$ *a random sample from the model (5.2.1), where* $\boldsymbol{\epsilon} \sim RSE_n\left(\mathbf{0},\phi^{-1}\mathbf{I}_n,\lambda\mathbb{1}_n\right)$, $\lambda \perp\!\!\!\perp (\boldsymbol{\beta},\phi)$, $\boldsymbol{\beta}|\,\phi \sim N_k\left(\mathbf{m},\phi^{-1}\mathbf{B}\right)$ *and* $\phi \sim Ga\left(a,b\right)$. *Then the Bayes factor (5.2.2) is given by*

$$BF\left(\mathbf{y}\right) = \frac{\int \omega^{\frac{k-n}{2}} \left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{-\frac{1}{2}} r^{-n-2a} dH\left(\omega\right)}{2\int \omega^{\frac{k-n}{2}} \left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{-\frac{1}{2}} r^{-n-2a} g\left(\omega\right) dH\left(\omega\right)},$$

where

$$g\left(\omega\right) = \int F_T\left[\sqrt{\frac{\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)}{r^2\omega\left(n+2a\right)^{-1}\lambda^{-2}}}\,\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right]\pi\left(\lambda\right)d\lambda,$$

F_T *is the c.d.f. of the* $t\left(0,1,n+2a\right)$, $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\left(\mathbf{X}^t\mathbf{y}+\omega\mathbf{B}^{-1}\mathbf{m}\right)$, $\mathbf{W} = \mathbf{X}^t\left(\mathbf{I}_n+\lambda^2\mathbb{1}_n\mathbb{1}_n^t\right)\mathbf{X}+\omega\mathbf{B}^{-1}$ *and* $r^2 = \frac{\mathbf{y}^t\mathbf{y}}{\omega}+\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\omega^{-1}\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}+2b$.

Proof. See Appendix E. ■

The previous proposition presents an expression of the Bayes factor to detect evidence of the data with respect to the skewness in a representable skew-elliptical model. In the particular case where H is degenerated in $\omega = 1$, we obtain skew-normal distribution for the errors. In this case the Bayes factor is given by

$$BF\left(\mathbf{y}\right) = \frac{1}{2\int F_T\left[\sqrt{\frac{\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)}{r^2\left(n+2a\right)^{-1}\lambda^{-2}}}\,\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right]\pi\left(\lambda\right)d\lambda} \tag{5.2.3}$$

where $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}\right)^{-1}\left(\mathbf{X}^t\mathbf{y}+\mathbf{B}^{-1}\mathbf{m}\right)$, $\mathbf{W} = \mathbf{X}^t\left(\mathbf{I}_n+\lambda^2\mathbb{1}_n\mathbb{1}_n^t\right)\mathbf{X}+\mathbf{B}^{-1}$ and $r^2 = \mathbf{y}^t\mathbf{y}+\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}+2b$. The previous expression is easier to calculate through numerical methods.

5.3 Simulation Results

In this section we describe the simulation results in order to study the behavior of the Bayes factor given by (5.2.3). We used a integration subroutine of MATLAB package, version 6.0.0.88, based on the recursive adaptive Simpson quadrature method.

In this simulation study we generated data from the linear regression model (5.2.1), where $\boldsymbol{\beta} = (2,1)^t$, $\boldsymbol{\epsilon} \sim SN_n\left(\mathbf{0}, \phi^{-1}\mathbf{I}_n, \lambda \mathbb{1}_n\right)$ with the following values variety: $n = 50$ and 100; $\phi = 0.01$, 1 and 100; and $\lambda = 0, \ldots, 5$. Also, the design matrix is given by

$$\mathbf{X}^t = \mathbb{1}^t_{\frac{n}{10}} \otimes \begin{pmatrix} & & \mathbb{1}^t_{10} & & \\ -4 & -3 & \cdots & 4 & 5 \end{pmatrix}.$$

For each one of the 36 previously described models, we made 100 replicates and, for each one of these, calculated the Bayes factor given by (5.2.3) under the following prior distributions, $\boldsymbol{\beta}|\,\phi \sim N_2\left[(2,1)^t, \phi^{-1}v\mathbf{I}_2\right]$, $\phi \sim Ga\left(a,b\right)$ and $\lambda \sim Ga\left(a_\lambda, b_\lambda\right)$. The Tables 5.1 to 5.4 display the first, second and third quartile of the Bayes factor estimated for the 100 replicates in each one of 36 models. Each table shows the results for different prior conditions where a_λ and b_λ are chosen such that $\mathbb{E}\left(\lambda\right) = \frac{a_\lambda}{b_\lambda}$ and $\mathbb{V}\left(\lambda\right) = \frac{a_\lambda}{b_\lambda^2}$, where $\mathbb{E}\left(\lambda\right)$ and $\mathbb{V}\left(\lambda\right)$ are given initially.

Results for concentrated prior distributions are shown in Table 5.1. In this case, we are assuming $\boldsymbol{\beta} \sim t_2\left((2,1)^t, \mathbf{I}_2, 2\right)$, $\phi \sim Ga(1,100)$ and $\lambda \sim Ga\left(62.5, 0.025\right)$. In general, the results of this table are good, although the values of the Bayes factor are not very different from 1. In this sense, the results for the case $\phi = 100$ are the worst, but this was expected because the prior distribution taken for ϕ has mean equal to 0.01 and variance, 10^{-4}.

Prior conditions of Table 5.2 are similar to the previous table, except that $\lambda \sim Ga\left(0.625, 2.5\right)$. Due to this, nice results are obtained even when $\phi = 100$. We

Table 5.1: Simulation results with $v = 0.01$, $a = 1$, $b = 100$, $\mathbb{E}(\lambda) = 2.5$ and $\mathbb{V}(\lambda) = 0.1$.

		$\phi = 0.01$			$\phi = 1$			$\phi = 100$		
n	λ	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.
50	0	0.6552	0.9870	2.0963	0.8498	1.0306	1.4500	0.9715	1.0034	1.0351
	1	0.5868	0.7353	1.0429	0.6795	0.7850	0.8893	0.9523	0.9680	0.9812
	2	0.5883	0.7492	1.0438	0.6894	0.7806	0.8626	0.9429	0.9682	0.9832
	3	0.5799	0.7445	0.9679	0.6826	0.7830	0.8802	0.9585	0.9733	0.9869
	4	0.5868	0.7381	1.0663	0.7355	0.8237	0.9177	0.9524	0.9694	0.9842
	5	0.5544	0.7225	0.9950	0.6664	0.7786	0.8759	0.9558	0.9677	0.9843
100	0	0.7211	1.0683	1.7372	0.8510	1.0087	1.3086	0.9662	0.9922	1.0253
	1	0.6365	0.7553	0.8552	0.7183	0.8058	0.8997	0.9568	0.9734	0.9906
	2	0.6209	0.7392	0.8509	0.7157	0.8105	0.9029	0.9597	0.9797	0.9907
	3	0.6393	0.7240	0.8196	0.7281	0.8096	0.8928	0.9523	0.9766	0.9856
	4	0.6168	0.7031	0.8688	0.7444	0.8201	0.9035	0.9567	0.9750	0.9874
	5	0.6193	0.7119	0.8347	0.7217	0.8411	0.9125	0.9574	0.9734	0.9875

expected this improvement since this prior distribution for λ is more in agreement with the true values of λ which we used in order to generate the data.

In Table 5.3 we increased the prior variance of λ with the purpose of observing the behavior of Bayes factor with respect to a vague informative prior distribution for λ. As we expected, the Bayes factor does not work when we use a vague informative prior distributions (see, for example, Kass and Raftery (1995)). Also, in order to obtain these calculations, the computational time was great.

Later, we wanted a less informative prior distribution on any parameter that is not λ, then we took $\phi \sim Ga(1, 0.01)$ that has mean equal to 100 and variance, 10^4. Table 5.4 displays the results. Same to the previous case, as expected, the Bayes factor held a wrong behavior. Another interesting aspect of the results of this table is the little variability in the estimates of the Bayes factor, in spite of possessing so different data sets.

Table 5.2: Simulation results with $v = 0.01$, $a = 1$, $b = 100$, $\mathbb{E}(\lambda) = 2.5$ and $\mathbb{V}(\lambda) = 1$.

		$\phi = 0.01$			$\phi = 1$			$\phi = 100$		
n	λ	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.
50	0	0.7112	1.0213	2.8531	0,8502	1,0305	1,4481	0,9715	1,0005	1,0268
	1	0.5492	0.6399	0.7831	0,6801	0,7854	0,8896	0,9528	0,9683	0,9812
	2	0.5502	0.6361	0.7390	0,6899	0,7811	0,8630	0,9430	0,9673	0,9832
	3	0.5551	0.6342	0.7696	0,6832	0,7835	0,8805	0,9586	0,9743	0,9873
	4	0.5893	0.6832	0.8349	0,7361	0,8241	0,9179	0,9532	0,9700	0,9842
	5	0.5387	0.6156	0.7588	0,6670	0,7791	0,8762	0,9556	0,9678	0,9844
100	0	0,7549	1,0765	1,9039	0,7951	1,0213	1,2740	0.9735	1.0041	1.0318
	1	0,6087	0,6974	0,8090	0,7496	0,8332	0,9082	0.9560	0.9763	0.9882
	2	0,6506	0,7441	0,8546	0,7240	0,7896	0,8909	0.9540	0.9728	0.9869
	3	0,6611	0,7486	0,8435	0,7116	0,7989	0,8917	0.9570	0.9718	0.9844
	4	0,6258	0,7617	0,8715	0,7579	0,8314	0,9249	0.9521	0.9696	0.9894
	5	0,6241	0,7215	0,8494	0,7209	0,8204	0,8876	0.9540	0.9711	0.9860

Table 5.3: Simulation results with $v = 0.01$, $a = 1$, $b = 100$, $\mathbb{E}(\lambda) = 2.5$ and $\mathbb{V}(\lambda) = 100$.

		$\phi = 0.01$			$\phi = 1$			$\phi = 100$		
n	λ	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.
50	0	0.9821	1.1157	1.3673	1.0198	1.0986	1.1782	1.0855	1.0967	1.1062
	1	0.8714	0.9344	1.0135	0.9746	1.0112	1.0447	1.0811	1.0894	1.0939
	2	0.8750	0.9366	0.9924	0.9544	1.0095	1.0490	1.0803	1.0855	1.0924
	3	0.8790	0.9343	1.0049	0.9872	1.0256	1.0609	1.0812	1.0886	1.0925
	4	0.9046	0.9631	1.0362	0.9757	1.0144	1.0518	1.0790	1.0859	1.0920
	5	0.8653	0.9245	1.0029	0.9794	1.0097	1.0529	1.0808	1.0864	1.0923
100	0	0.9735	1.0639	1.2652	1.0185	1.0861	1.1996	1.0885	1.0949	1.1050
	1	0.9154	0.9755	1.0450	0.9818	1.0140	1.0586	1.0787	1.0864	1.0909
	2	0.9416	0.9988	1.0414	0.9980	1.0272	1.0616	1.0778	1.0864	1.0912
	3	0.9092	0.9747	1.0281	0.9921	1.0349	1.0648	1.0804	1.0862	1.0893
	4	0.9237	0.9874	1.0326	0.9881	1.0203	1.0567	1.0779	1.0851	1.0895
	5	0.9240	0.9626	1.0349	0.9801	1.0095	1.0630	1.0777	1.0855	1.0890

Table 5.4: Simulation results with $v = 100$, $a = 1$, $b = 0.01$, $\mathbb{E}(\lambda) = 2.5$ and $\mathbb{V}(\lambda) = 1$.

		$\phi = 0.01$			$\phi = 1$			$\phi = 100$		
n	λ	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.	1^{th} c.	Median	3^{th} c.
50	0	0.8014	0.9999	1.0003	0.9999	1.0000	1.0001	0.9999	1.0000	1.0001
	1	0.8014	0.9999	1.0000	0.9998	0.9999	0.9999	0.9998	0.9999	0.9999
	2	0.8014	0.9999	1.0000	0.9998	0.9999	0.9999	0.9998	0.9999	0.9999
	3	0.8014	0.9998	1.0000	0.9998	0.9999	0.9999	0.9998	0.9999	0.9999
	4	0.8014	0.9999	1.0000	0.9998	0.9999	1.0000	0.9999	0.9999	1.0000
	5	0.8014	0.9999	1.0000	0.9998	0.9999	0.9999	0.9998	0.9999	0.9999
100	0	0.9999	1.0000	1.0000	0.9999	1.0000	1.0001	0.9999	1.0000	1.0001
	1	0.9999	0.9999	1.0000	0.9999	1.0000	1.0000	0.9999	1.0000	1.0000
	2	0.9999	0.9999	1.0000	0.9999	0.9999	1.0000	0.9999	0.9999	1.0000
	3	0.9999	0.9999	1.0000	0.9999	0.9999	1.0000	0.9999	1.0000	1.0000
	4	0.9999	0.9999	1.0000	0.9999	0.9999	1.0000	0.9999	1.0000	1.0000
	5	0.9999	0.9999	1.0000	0.9999	1.0000	1.0000	0.9999	0.9999	1.0000

These results corroborate the well-known result about the great influence that the prior distributions exert on the Bayes factor. Indeed, in these results the prior distributions almost determine the good behavior of this Bayes factor. Nice results are only obtained when a priori, a correct idea is had about the true values of the parameters.

5.4 An Application

In this section we illustrate the calculus from equation (5.2.3) using the real data set described in the example of Section 4.4. The data of the explanatory variable was the values of IPSA observed between March, 1990 and April, 1999.

For each company, Table 5.5 displays the Bayes factor given by (5.2.3) under the following prior distributions $\lambda \sim Ga\,(0.4, 0.2)$, $\boldsymbol{\beta}|\,\phi \sim N_2\left[(0,1)^t\,, 10\phi^{-1}\mathbf{I}_2\right]$ and $\phi \sim Ga\,(100, 1)$. Then, the prior mean and variance for λ are 2 and 10 respectively.

Table 5.5: Bayes factors by company in linear regression model.

Company	$\widehat{BF}$
Cementos	0.9998
Cervezas	0.9993
Chilquinta	0.9990
Copec	0.9998
Iansa	1.0004

From Table 5.5, it can be seen that no company present a strong evidence of the Bayes factor and the only data set in favor of the symmetry assumption is the Iansa company, similar to those results in Section 4.4. However, the results of Section 4.4 presented strong evidence. The simulations in the previous section as well as these results with real data show a weak evidence of the Bayes factor, probably because the inclusion of the explanatory variable produces a better explanation of the response variable behavior. Thus, we think that in order to study these real data it is not worthwhile to use an asymmetric model.

Chapter 6

Concluding Remarks

In Chapter 1 we presented a class of prior distributions that make the posterior $\boldsymbol{\beta}$ independent of the choice of $h^{(n)}$. Although it is not established explicitly, the same invariance holds for the predictive distribution of $\mathbf{y}$, as pointed out in Osiewalski and Steel (1993). We also specified a conditional distribution for $\mathbf{y}|\,\boldsymbol{\beta},\phi,h^{(n)}$ and a prior for $(\,\boldsymbol{\beta},\phi)|\,h^{(n)}$, in such a way that $\mathbf{y}\perp\!\!\!\perp h^{(n)}$, considering $h^{(n)}$ as random (i.e., $h^{(n)}$ is marginally ancillary). Thus, any procedure for model comparisons that is based on the predictive distributions would be not useful to discriminate among different density generators. Even if we introduce a prior for $h^{(n)}$, this would not be updated under the hypotheses imposed in Chapter 1. On the other hand, it becomes clear that Bayesian model comparisons should include not only the predictive distribution, but also all the model components. We point out that the comparisons should refer to alternatives for the joint distribution of $(\mathbf{y},\boldsymbol{\theta})$. In particular, it is suggested there to study the ratio

$$BF\left(\boldsymbol{\theta}\right)=\frac{f_1\left(\mathbf{y},\ \boldsymbol{\theta}\right)}{f_2\left(\mathbf{y},\ \boldsymbol{\theta}\right)}$$

as a function of $\boldsymbol{\theta}$, which is called Global Bayes Factor (GBF) and may be rewritten as

$$BF\left(\boldsymbol{\theta}\right)=\frac{\pi_1\left(\,\boldsymbol{\theta}\right|\mathbf{y})}{\pi_2\left(\,\boldsymbol{\theta}\right|\mathbf{y})}BF,$$

where BF is the usual Bayes factor. Thus

$$\mathbb{E}_{\pi_2(\,\cdot|\mathbf{y})}\left[BF\left(\boldsymbol{\theta}\right)\right]=BF.$$

Others alternative procedures for model comparison were discussed in Section 2.6.

Model comparison and selection in MEM models could still be developed in several directions, for example the comparison and selection of models inside the class of skew-elliptical distributions, the selection of variables, development of the topics seen in this thesis extending the range of distributions for the errors, etc.

When obtaining samples from the posterior distribution of a MEM, it is important to dedicate efforts to improve the transition probability functions in the M-H algorithm. In our case turned out to be quite inefficient: around a 10% of the candidates were accepted, although convergence existed. It is thought that this inefficiency was been due, in great measure, to the fact there were 4 unknown parameters. We made simulations assuming two unknown parameters, and the convergence was much quicker.

With the purpose of eliminating the influence of prior distributions, it is important to improve objective Bayesian methods for model comparison in MEM. We mentioned some possible paths to follow in Section 2.6.

In Chapter 3 we present a strategy for examining the sensitivity of posterior distribution with respect to the deletion of sets of observations. We find interesting to study the sensitivity of posterior distribution with respect to departures from normality relative to other multivariate elliptical models and also using noninformative prior distributions.

In Section 4.1, two expressions were presented for the L_1 distance between posterior distributions of μ and σ. Then, to determine the sensitivity of the λ parameter concerning these posterior distributions, the numeric calculation would be most useful. On the other hand, in Section 4.2, different prior conditions were assumed obtaining different Bayes Factor expressions. Similar results could be obtained for other prior conditions, for instance, we could assume that $|\lambda| \sim Ga\,(a, b)$ or $\lambda \sim$ Half-Normal, etc.

Extensions in other directions are to study the usage of default Bayes factors for testing skewness. We could consider noninformative prior distribution $\pi\,(\lambda)$ in (4.2.5). In this case, a first obstacle is, for example, to consider an improper prior distribution $\pi\,(\lambda) \propto 1$, since $\int_{-\infty}^{\infty} \left[\prod_{i=1}^{n} G\,(\lambda y_i)\right]^b d\lambda$ diverge with $0 < b \leq 1$, then the Intrinsic Bayes factor and Fractional Bayes factor could not be calculated. Chapter 5 is a natural generalization of Chapter 1, there we preserve uncorrelation and assume dependence of observations. However, in Chapter 4 we assume independent observations. Thus, a possible alternative of Chapter 5 is considering independent observations. The simulation results of Chapter 5 induces us to believe that the covariates presence in multivariate case could subtract importance to the asymmetry of the errors. This can be noticed when comparing the monthly rentability example results of five Chilean companies as detailed in Sections 4.4 and 5.4.

Appendix A

Bayes Factors in MEM

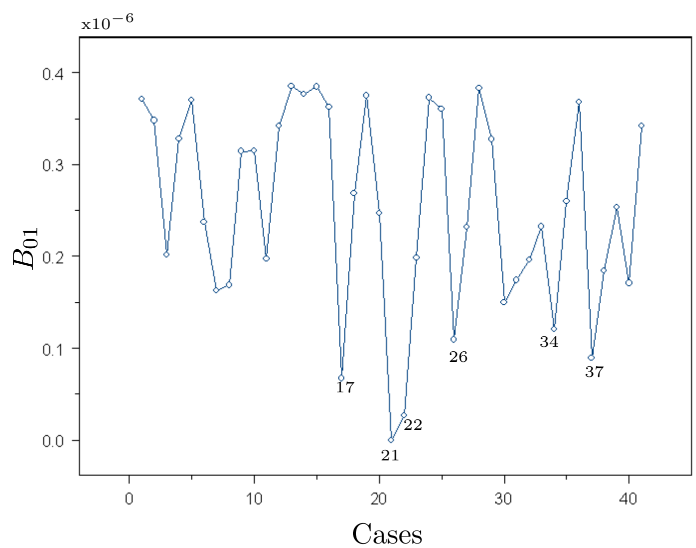

Figure A.1: B_{01} given by (3.1.2) under MEM in deletion case.

Table A.1: Optimal Bayes factor values for simulated data in functional MEM.

σ_u	$\mu = 0$, $\tau^2 = 1$	$\mu = 0$, $\tau^2 = 3$	$\mu = 0$, $\tau^2 = 5$	$\mu = 3$, $\tau^2 = 1$	$\mu = 3$, $\tau^2 = 3$	$\mu = 3$, $\tau^2 = 5$
0	7.815e-039	9.477e+078	7.766e+072	6.284e+051	7.846e+105	1.180e+087
0.2	2.269e-085	4.769e+071	3.131e+064	6.179e+048	1.671e+092	1.876e+080
0.4	2.448e-053	1.007e+063	2.884e+054	1.441e+069	1.476e+073	4.298e+067
0.6	4.323e-096	5.320e+051	9.901e+054	4.835e+067	3.164e+078	5.644e+059
0.8	2.229e-100	6.062e+026	1.331e+053	3.530e+055	9.421e+052	8.939e+063
1.0	7.227e-025	8.155e+031	7.625e+054	2.702e+067	2.304e+048	1.108e+061
1.2	1.896e-113	2.879e-029	7.428e+040	8.060e+059	1.952e+047	6.135e+056
1.4	1.918e-059	3.450e+012	3.397e+015	6.527e+038	1.770e+043	2.151e+057
1.6	1.361e-184	4.547e-100	7.733e-007	2.798e+049	4.009e+032	5.862e+021
1.8	1.344e-204	2.023e-034	1.017e+018	1.351e+059	734.5918	8.839e-015
2.0	4.321e-137	2.108e-112	5.243e-005	4.273e+070	6.444e+035	2.797e-013
2.2	4.113e-206	1.164e-024	7.841e+011	2.853e+050	3.296e+015	3.609e-004
2.4	1.512e-201	3.340e-053	3.233e-133	7.941e+053	9.752e+031	1.150e+009
2.6	0	1.608e-025	0	5.426e+069	8.992e+032	6.585e+043
2.8	8.096e-146	6.008e-154	2.536e-049	1.464e+077	1.095e+014	10.5466
3.0	2.344e-169	1.022e-128	5.010e-132	1.914e+054	4.741e+026	6.582e-055

Table A.2: Optimal Bayes factor values for simulated data in structural MEM.

σ_u	$\delta = 0.1$	$\delta = 1$	$\delta = 5$
0	1.5960e+025	0.0093	0.6591
0.2	1.0768e+009	2.1851e+004	4.1210e-005
0.4	25.1171	0.3261	3.4341e-004
0.6	0.0023	0.0165	0.1086
0.8	0.0060	1.7242e-005	0.1667
1.0	0.0010	0.0063	0.6042
1.2	9.5568e-005	9.8342e-005	0.4622
1.4	3.6166e-006	0.0055	0.1263
1.6	1.5987e-008	0.0334	0.5573
1.8	5.7717e-005	1.1722e-006	0.0368
2.0	4.1454e-018	4.3199e-056	0.0159
2.2	3.2095e-004	6.2342e-093	0.3534
2.4	1.5946e-093	4.5352e-066	0.2880
2.6	4.2793e-004	9.8820e-041	0.1595
2.8	1.6547e-050	0	0.0534
3.0	1.0853e-014	9.0370e-060	0.0145

Table A.3: Corn hectares determined by two methods.

Segment	Photograph y_i	Interview x_i	Segment	Photograph y_i	Interview x_i
1	167.14	165.76	20	120.19	121.00
2	159.04	162.08	21	115.74	109.91
3	161.06	152.04	22	125.45	122.66
4	163.49	161.75	23	99.96	104.21
5	97.12	96.32	24	99.55	92.88
6	123.02	114.12	25	163.09	149.94
7	111.29	100.60	26	60.30	64.75
8	132.33	127.88	27	101.98	99.96
9	116.95	116.90	28	138.40	140.43
10	89.84	87.41	29	94.70	98.95
11	84.17	88.59	30	129.50	131.04
12	88.22	88.59	31	132.74	127.07
13	161.87	165.35	32	133.55	133.55
14	106.03	104.00	33	83.37	77.70
15	87.01	88.63	34	78.51	76.08
16	159.85	153.70	35	205.98	206.39
17	209.63	185.35	36	110.07	108.33
18	122.62	116.43	37	134.36	118.17
19	93.08	93.48			

Table A.4: Concrete compressive strength measurements (psi).

Sample	Day 28	Day 2	Sample	Day 28	Day 2	Sample	Day 28	Day 2
1	4470	2830	15	4690	2985	29	4650	3335
2	4740	3295	16	4880	3135	30	4680	3800
3	5115	2710	17	3425	2750	31	5165	2680
4	4880	2855	18	4265	3205	32	5075	3760
5	4445	2980	19	4485	3000	33	4710	3605
6	4080	3065	20	5220	3035	34	4200	2005
7	5390	3765	21	7695	4245	35	4645	2495
8	4045	3265	22	3330	1635	36	4725	3205
9	4370	3170	23	4065	2270	37	4695	2060
10	4955	2895	24	4715	2895	38	5470	3425
11	3835	2630	25	4735	2845	39	4330	3315
12	4290	2830	26	3605	2205	40	4950	3825
13	4600	2935	27	4670	3590	41	4460	3160
14	4605	3115	28	4720	3080			

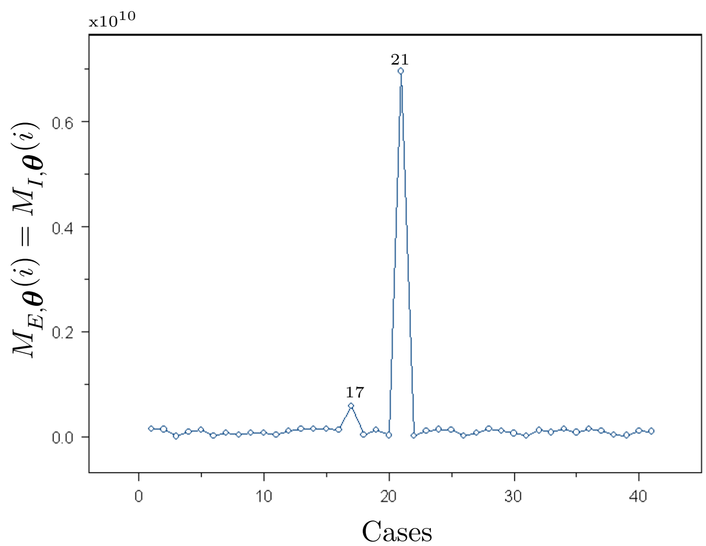

Figure A.2: $M_{E,\boldsymbol{\theta}}(i) = M_{I,\,\boldsymbol{\theta}}(i) = \left\|\mathbb{E}\left[(1-h(\boldsymbol{\theta}))\ \boldsymbol{\theta}\,|\mathbf{y},\mathbf{x}\right]\right\|^2_{\mathbf{I}_4}$, where $\boldsymbol{\theta} = (\alpha,\beta,\sigma^2_\epsilon,\sigma^2_u)$.

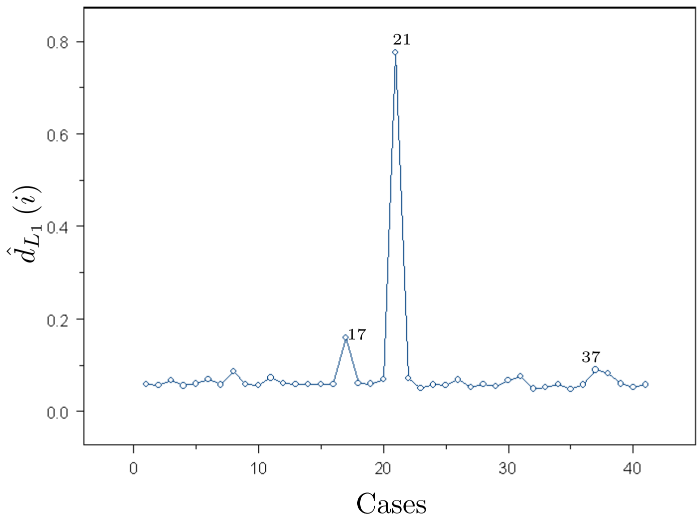

Figure A.3: L_1-influence under MEM in deletion case.

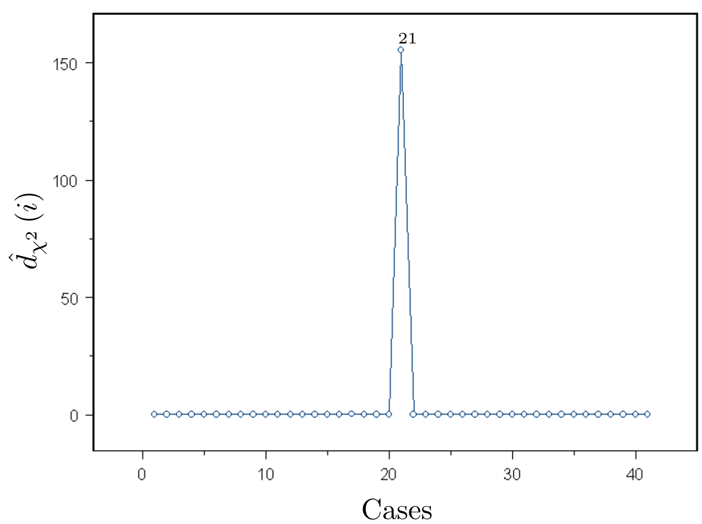

Figure A.4: χ^2-influence under MEM in deletion case.

Appendix B

Descriptive Statistics by Company

Table B.1: Some descriptive statistics for the monthly rentability of five Chilean companies, measured between March, 1990 and April, 1999.

Company	Mean	s.d.	Skewness	Kurtosis
Cementos	0.01347727	0.13283164	1.33268186	3.31334209
Cervezas	0.02022545	0.11423326	0.51903086	0.66034736
Chilquinta	0.02778182	0.13638980	0.68802875	2.21338495
Copec	0.01201909	0.10339547	0.71021655	0.98810570
Iansa	0.0006200	0.1131153	0.3621623	0.2541416

Appendix C

Proof of Proposition 4.1.6

We should calculate

$$\begin{aligned}\pi\left(\mu,\sigma|\,\lambda,\mathbf{x}\right) &= \frac{f\left(\mathbf{x}|\,\lambda,\mu,\sigma\right)\pi\left(\mu,\sigma\right)}{\int f\left(\mathbf{x}|\,\lambda,\mu,\sigma\right)\pi\left(\mu,\sigma\right)d\left(\mu,\sigma\right)} \\ &= \frac{\sigma^{-n}\prod_{i=1}^{n}\phi\left(\frac{x_i-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)\pi\left(\mu,\sigma\right)}{\int\sigma^{-n}\prod_{i=1}^{n}\phi\left(\frac{x_i-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)\pi\left(\mu,\sigma\right)d\left(\mu,\sigma\right)}.\end{aligned}$$

Since

$$\pi\left(\mu,\sigma\right)=\pi\left(\mu|\,\sigma\right)\pi\left(\sigma\right)=\frac{b^a\sqrt{2v}}{\Gamma\left(a\right)\pi^{\frac{1}{2}}}\frac{1}{\left(\sigma^2\right)^{a+1}}\exp\left\{-\frac{1}{2\sigma^2}\left[v\left(\mu-m\right)^2+2b\right]\right\}$$

and

$$\sigma^{-n}\prod_{i=1}^{n}\phi\left(\frac{x_i-\mu}{\sigma}\right)=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\},$$

then

$$\begin{aligned}\pi\left(\mu,\sigma|\,\lambda,\mathbf{x}\right) &\propto \frac{2b^a\sqrt{v}}{\Gamma\left(a\right)\left(2\pi\right)^{\frac{n+1}{2}}}\frac{1}{\left(\sigma^2\right)^{\frac{n}{2}+a+1}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}\left(x_i-\mu\right)^2+v\left(\mu-m\right)^2+2b\right]\right\} \\ &\quad \prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right) \\ &= \frac{2b^a\sqrt{v}}{\Gamma\left(a\right)\left(2\pi\right)^{\frac{n+1}{2}}}\frac{1}{\left(\sigma^2\right)^{\frac{n}{2}+a+1}}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(n+v\right)\left(\mu-\hat{\mu}\right)^2+r^2\right]\right\} \\ &\quad \prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right),\end{aligned}$$

where $\hat{\mu} = \frac{n\bar{x}+mv}{n+v}$, $r^2 = ns^2 + \frac{nv}{n+v}(m-\bar{x})^2 + 2b$ and $s^2 = n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

On the other hand, the normalizing constant is

$$\begin{aligned} m(\mathbf{x}) &= \int \sigma^{-n} \prod_{i=1}^{n} \phi\left(\frac{x_i-\mu}{\sigma}\right) \Phi\left(\lambda \frac{x_i-\mu}{\sigma}\right) \pi(\mu,\sigma)\, d(\mu,\sigma) \\ &= \frac{2b^a\sqrt{v}}{\Gamma(a)(2\pi)^{\frac{n+1}{2}}} \int_0^\infty \frac{1}{(\sigma^2)^{\frac{n}{2}+a+1}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} h(\sigma)\, d\sigma, \end{aligned}$$

where

$$h(\sigma) = \int_{-\infty}^{\infty} \exp\left\{-\frac{n+v}{2\sigma^2}(\mu-\hat{\mu})^2\right\} \prod_{i=1}^{n} \Phi\left(\lambda \frac{x_i-\mu}{\sigma}\right) d\mu.$$

Making the change of variable $y = \sqrt{n+v}\frac{\mu-\hat{\mu}}{\sigma}$, we obtain

$$\begin{aligned} h(\sigma) &= \frac{\sigma}{\sqrt{n+v}} \int_{-\infty}^{\infty} \exp\left\{-\frac{y^2}{2}\right\} \prod_{i=1}^{n} \Phi\left[\lambda\left(\frac{x_i-\hat{\mu}}{\sigma} - \frac{y}{\sqrt{n+v}}\right)\right] dy \\ &= \frac{\sigma(2\pi)^{-\frac{n}{2}}}{\sqrt{n+v}} \int_{-\infty}^{\infty} \exp\left\{-\frac{y^2}{2}\right\} \int \cdots \int_{-\infty}^{\lambda\left(\frac{x_i-\hat{\mu}}{\sigma} - \frac{y}{\sqrt{n+v}}\right)} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} dz_i dy. \end{aligned}$$

Making now the change of variables $u_i = z_i + y\frac{\lambda}{\sqrt{n+v}}$ and exchanging the integration order, we obtain

$$\begin{aligned} h(\sigma) &= \frac{\sigma(2\pi)^{-\frac{n}{2}}}{\sqrt{n+v}} \int \cdots \int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{\sigma}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n} u_i^2 - \frac{\lambda^2}{v+n(1+\lambda^2)}\left(\sum_{i=1}^{n} u_i\right)^2\right]\right\} \\ &\quad \int_{-\infty}^{\infty} \exp\left\{-\frac{v+n(1+\lambda^2)}{2(n+v)}\left[y - \frac{\lambda\sqrt{n+v}}{v+n(1+\lambda^2)}\sum_{i=1}^{n} u_i\right]^2\right\} dy du_i \\ &= \frac{\sigma(2\pi)^{-\frac{n-1}{2}}}{\sqrt{v+n(1+\lambda^2)}} \int \cdots \int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{\sigma}} \exp\left\{-\frac{1}{2}\mathbf{u}^t\left[\mathbf{I}_n - \frac{\lambda^2}{v+n(1+\lambda^2)}\mathbb{1}\mathbb{1}^t\right]\mathbf{u}\right\} du_i \\ &= \frac{\sigma\sqrt{2\pi}\,|\boldsymbol{\Sigma}|^{\frac{1}{2}}}{\sqrt{v+n(1+\lambda^2)}} F_{\mathbf{U}}\left(\lambda\frac{x_1-\hat{\mu}}{\sigma}, \ldots, \lambda\frac{x_n-\hat{\mu}}{\sigma}\right), \end{aligned}$$

where $\mathbf{U} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}^{-1} = \mathbf{I}_n - \frac{\lambda^2}{v+n(1+\lambda^2)}\mathbb{1}\mathbb{1}^t$. Using the well known identities (see Fang and Zhang (1990, Chapter 1))

$$\left(\mathbf{A} - \mathbf{u}\mathbf{v}^t\right)^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^t\mathbf{A}^{-1}}{1 - \mathbf{v}^t\mathbf{A}^{-1}\mathbf{u}}, \tag{C.0.1}$$

for some squared matrix $\mathbf{A}$ and vectors $\mathbf{u}$ and $\mathbf{v}$, and

$$|\mathbf{I}_p + \mathbf{AB}| = |\mathbf{I}_q + \mathbf{BA}|, \tag{C.0.2}$$

where $\mathbf{A}_{p\times q}$ and $\mathbf{B}_{q\times p}$, we obtain $\mathbf{\Sigma} = \mathbf{I}_n + \frac{\lambda^2}{v+n}\mathbb{1}\mathbb{1}^t$ and $|\mathbf{\Sigma}| = 1 + \frac{\lambda^2 n}{v+n}$.

Therefore,

$$m(\mathbf{x}) = \frac{2b^a}{\Gamma(a)}\left(\frac{v}{(v+n)(2\pi)^n}\right)^{\frac{1}{2}} \int_0^\infty \frac{1}{\sigma^{n+2a+1}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} F_{\mathbf{U}}\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{\sigma}\right) d\sigma.$$

Making the change of variable $\sigma^2 = \frac{r^2}{s}$, we obtain

$$\begin{aligned} m(\mathbf{x}) &= \frac{b^a}{\Gamma(a)\, r^{n+2a}}\left(\frac{v}{(v+n)(2\pi)^n}\right)^{\frac{1}{2}} \\ &\quad \int_0^\infty s^{\frac{n}{2}+a-1}\exp\left\{-\frac{s}{2}\right\} F_{\mathbf{U}}\left(\lambda\sqrt{s}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right) ds \\ &= \frac{b^a(2\pi)^{-n}}{\Gamma(a)\, r^{n+2a}}\left(\frac{v}{v+n(1+\lambda^2)}\right)^{\frac{1}{2}} \\ &\quad \int_0^\infty s^{\frac{n}{2}+a-1}\exp\left\{-\frac{s}{2}\right\}\int\cdots\int_{-\infty}^{\lambda\sqrt{s}\frac{x_i-\hat{\mu}}{r}} \exp\left\{-\frac{1}{2}\mathbf{u}^t\mathbf{\Sigma}^{-1}\mathbf{u}\right\} du_i ds. \end{aligned}$$

Making the change of variables $t_i\sqrt{s} = u_i$ and exchanging the integration order, we obtain

$$\begin{aligned} m(\mathbf{x}) &= \frac{b^a(2\pi)^{-n}}{\Gamma(a)\, r^{n+2a}}\left(\frac{v}{v+n(1+\lambda^2)}\right)^{\frac{1}{2}} \\ &\quad \int\cdots\int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{r}}\int_0^\infty s^{n+a-1}\exp\left\{-\frac{s}{2}\left[1+\mathbf{t}^t\mathbf{\Sigma}^{-1}\mathbf{t}\right]\right\} ds dt_i \\ &= \frac{(2b)^a\,\Gamma(n+a)}{\Gamma(a)\, r^{n+2a}\pi^n}\left(\frac{v\,|\mathbf{\Sigma}|}{v+n(1+\lambda^2)}\right)^{\frac{1}{2}}\int\cdots\int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{r}} |\mathbf{\Sigma}|^{-\frac{1}{2}}\left[1+\mathbf{t}^t\mathbf{\Sigma}^{-1}\mathbf{t}\right]^{-(n+a)} dt_i \\ &= \frac{(2b)^a\,\Gamma\left(\frac{n}{2}+a\right)}{\Gamma(a)\, r^{n+2a}\pi^{\frac{n}{2}}}\left(\frac{v}{v+n}\right)^{\frac{1}{2}} F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right), \end{aligned}$$

where $\mathbf{T} \sim t_n(\mathbf{0}, \mathbf{\Sigma}, n+2a)$.

Therefore,

$$\begin{aligned}\pi\left(\mu,\sigma\right|\lambda,\mathbf{x}) &= \frac{r^{n+2a}\sqrt{v+n}}{2^{\frac{n-1}{2}+a}\Gamma\left(\frac{n}{2}+a\right)\sqrt{\pi}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)} \\ &\quad \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}+a+1}}\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[\left(n+v\right)\left(\mu-\hat{\mu}\right)^2+r^2\right]\right\}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\pi\left(\mu\right|\lambda,\mathbf{x}) &= \int_0^\infty \pi\left(\mu,\sigma\right|\lambda,\mathbf{x})\,d\sigma \\ &= \frac{r^{n+2a}\sqrt{v+n}}{2^{\frac{n-1}{2}+a}\Gamma\left(\frac{n}{2}+a\right)\sqrt{\pi}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)} \\ &\quad \int_0^\infty \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}+a+1}}\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma}\right)\exp\left\{-\frac{1}{2\sigma^2}\left[\left(n+v\right)\left(\mu-\hat{\mu}\right)^2+r^2\right]\right\}d\sigma \\ &= \frac{r^{n+2a}\sqrt{v+n}}{2^{n+a-\frac{1}{2}}\Gamma\left(\frac{n}{2}+a\right)\pi^{\frac{n+1}{2}}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)} \\ &\quad \int_0^\infty \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}+a+1}}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(n+v\right)\left(\mu-\hat{\mu}\right)^2+r^2\right]\right\} \\ &\quad \int\cdots\int_{-\infty}^{\lambda\frac{x_i-\mu}{\sigma}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}z_i^2\right\}dz_i d\sigma.\end{aligned}$$

Making the change of variables $z_i = \frac{y_i}{\sigma}$ and exchanging the integration order, we obtain

$$\begin{aligned}\pi\left(\mu\right|\lambda,\mathbf{x}) &= \frac{r^{n+2a}\sqrt{v+n}}{2^{n+a-\frac{1}{2}}\Gamma\left(\frac{n}{2}+a\right)\pi^{\frac{n+1}{2}}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)} \\ &\quad \int\cdots\int_{-\infty}^{\lambda(x_i-\mu)}\int_0^\infty \frac{1}{\left(\sigma^2\right)^{n+a+1}}\exp\left\{-\frac{\left(n+v\right)\left(\mu-\hat{\mu}\right)^2+r^2+\mathbf{y}^t\mathbf{y}}{2\sigma^2}\right\}d\sigma dy_i,\end{aligned}$$

but if we make the change of variable $\sigma^{-2} = s$, then

$$\begin{aligned}
\pi\left(\mu|\,\lambda, \mathbf{x}\right) &= \frac{r^{n+2a}\sqrt{v+n}}{2^{n+a+\frac{1}{2}}\Gamma\left(\frac{n}{2}+a\right)\pi^{\frac{n+1}{2}}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)} \\
&\quad \int\cdots\int_{-\infty}^{\lambda(x_i-\mu)}\int_0^{\infty} s^{n+a-\frac{1}{2}}\exp\left\{-\frac{s}{2}\left[(n+v)\left(\mu-\hat{\mu}\right)^2+r^2+\mathbf{y}^t\mathbf{y}\right]\right\}ds dy_i \\
&= \frac{r^{n+2a}\sqrt{v+n}\Gamma\left(n+a+\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}+a\right)\pi^{\frac{n+1}{2}}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)} \\
&\quad \int\cdots\int_{-\infty}^{\lambda(x_i-\mu)}\left[r^2+(n+v)\left(\mu-\hat{\mu}\right)^2+\mathbf{y}^t\mathbf{y}\right]^{-\left(n+a+\frac{1}{2}\right)}dy_i \\
&= \frac{r^{n+2a}\sqrt{v+n}\Gamma\left(\frac{n+1}{2}+a\right)}{\Gamma\left(\frac{n}{2}+a\right)\sqrt{\pi}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)}\frac{F_{\mathbf{Y}}\left(\lambda\left(x_1-\mu\right),\ldots,\lambda\left(x_n-\mu\right)\right)}{\left[r^2+(n+v)\left(\mu-\hat{\mu}\right)^2\right]^{\left(\frac{n+1}{2}+a\right)}},
\end{aligned}$$

where $\mathbf{Y}\sim t_n\left(\mathbf{0},\frac{r^2+(n+v)(\mu-\hat{\mu})^2}{n+2a+1}\mathbf{I}_n, n+2a+1\right)$.

Moreover, if we consider the function $h\left(\sigma\right)$, we get

$$\begin{aligned}
\pi\left(\sigma|\,\lambda, \mathbf{x}\right) &= \int_{-\infty}^{\infty}\pi\left(\mu,\sigma|\,\lambda,\mathbf{x}\right)d\mu \\
&= \frac{r^{n+2a}\sqrt{v+n}}{2^{\frac{n-1}{2}+a}\Gamma\left(\frac{n}{2}+a\right)\sqrt{\pi}F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)}\frac{h\left(\sigma\right)}{\left(\sigma^2\right)^{\frac{n}{2}+a+1}}\exp\left\{-\frac{r^2}{2\sigma^2}\right\} \\
&= \frac{r^{n+2a}\sigma^{-n-2a-1}}{2^{\frac{n}{2}+a-1}\Gamma\left(\frac{n}{2}+a\right)F_{\mathbf{T}}\left(\sqrt{n+2a}\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{r}\right)}F_{\mathbf{U}}\left(\lambda\frac{\mathbf{x}-\hat{\mu}\mathbb{1}}{\sigma}\right)\exp\left\{-\frac{r^2}{2\sigma^2}\right\}.
\end{aligned}$$

Appendix D

Proof of Proposition 4.2.6

Notice that

$$\begin{aligned}
f\left(\mathbf{x}\middle|\,\lambda,\mu,\sigma,\omega_1,\ldots,\omega_n\right) &= \prod_{i=1}^{n}\frac{2}{\sigma\sqrt{\omega_i}}\phi\left(\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right)\Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right) \\
&= \left(\frac{2}{\pi\sigma^2}\right)^{\frac{n}{2}}\left(\prod_{i=1}^{n}\omega_i\right)^{-\frac{1}{2}} \\
&\quad \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sqrt{\omega_i}}\right)^2\right]\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right) \\
&= \left(\frac{2}{\pi\sigma^2}\right)^{\frac{n}{2}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}}\exp\left(-\frac{\eta}{2\sigma^2}S_\omega^2\right) \\
&\quad \exp\left[-\frac{\eta}{2\sigma^2}\left(\mu-\sum_{i=1}^{n}\nu_i x_i\right)^2\right]\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right),
\end{aligned}$$

where $\mathbf{D}\left(\boldsymbol{\omega}\right) = \operatorname{diag}\left(\omega_1,\ldots,\omega_n\right)$, $\eta = \sum_{i=1}^{n}\omega_i^{-1}$, $S_\omega^2 = \sum_{i=1}^{n}\nu_i x_i^2 - \left(\sum_{i=1}^{n}\nu_i x_i\right)^2 = \sum_{i=1}^{n}\nu_i\left(x_i-\sum_{j=1}^{n}\nu_j x_j\right)^2$ and $\nu_i = \frac{\omega_i}{\eta}$ for each $i = 1,\ldots,n$.

Now, considering $\pi\left(\mu|\,\sigma\right) = N\left(\mu|\,m, \frac{\sigma^2}{v}\right)$, we obtain

$$\begin{aligned}
f\left(\mathbf{x}|\,\lambda,\sigma,\omega_1,\ldots,\omega_n\right) &= \sqrt{\frac{2^{n-1}v}{\left(\pi\sigma^2\right)^{n+1}}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}}\exp\left(-\frac{\eta}{2\sigma^2}S_\omega^2\right) \\
&\quad \int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2\sigma^2}\left[\eta\left(\mu-\sum_{i=1}^{n}\nu_i x_i\right)^2+v\left(\mu-m\right)^2\right]\right\} \\
&\quad \prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right)d\mu \\
&= \sqrt{\frac{2^{n-1}v}{\left(\pi\sigma^2\right)^{n+1}}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \\
&\quad \exp\left\{-\frac{\eta}{2\sigma^2}\left[S_\omega^2+\frac{v}{\eta+v}\left(m-\sum_{i=1}^{n}\nu_i x_i\right)^2\right]\right\} \\
&\quad \int_{-\infty}^{\infty}\exp\left[-\frac{\eta+v}{2\sigma^2}\left(\mu-\hat{\mu}\right)^2\right]\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right)d\mu, \quad \text{(D.0.1)}
\end{aligned}$$

where

$$\hat{\mu} = \frac{\eta\sum_{i=1}^{n}\nu_i x_i + vm}{\eta+v}.$$

Making the change of variable $y = \sqrt{\eta+v}\frac{(\mu-\hat{\mu})}{\sigma}$, we obtain

$$\begin{aligned}
h\left(\lambda,\sigma,\omega_1,\ldots,\omega_n\right) \quad &:= \int_{-\infty}^{\infty}\exp\left[-\frac{\eta+v}{2\sigma^2}\left(\mu-\hat{\mu}\right)^2\right]\prod_{i=1}^{n}\Phi\left(\lambda\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right)d\mu \\
&= \frac{\sigma}{\sqrt{\eta+v}}\int_{-\infty}^{\infty}\exp\left(-\frac{y^2}{2}\right)\prod_{i=1}^{n}\Phi\left[\lambda\left(\frac{x_i-\hat{\mu}}{\sigma\sqrt{\omega_i}}-\frac{y}{\sqrt{\omega_i}\sqrt{\eta+v}}\right)\right]dy \\
&= \frac{\sigma}{\sqrt{\eta+v}}\int_{-\infty}^{\infty}\exp\left(-\frac{y^2}{2}\right) \\
&\quad \int\cdots\int_{-\infty}^{\lambda\left(\frac{x_i-\hat{\mu}}{\sigma\sqrt{\omega_i}}-\frac{y}{\sqrt{\omega_i}\sqrt{\eta+v}}\right)}\left(2\pi\right)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}z_i^2\right)dz_i dy.
\end{aligned}$$

Then, making the change of variables $u_i = \frac{z_i}{\sqrt{\omega_i}} + y\frac{\lambda}{\omega_i\sqrt{\eta+v}}$ and exchanging the integration order, we obtain

$$
\begin{aligned}
h(\lambda,\sigma,\omega_1,\ldots,\omega_n) &= |\mathbf{D}(\boldsymbol{\omega})|^{\frac{1}{2}} \frac{\sigma(2\pi)^{-\frac{n}{2}}}{\sqrt{\eta+v}} \\
&\quad \int \cdots \int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{\sigma\omega_i}} \exp\left\{-\frac{\left[\sum_{i=1}^n \omega_i u_i^2 - \frac{\lambda^2}{v+\eta(1+\lambda^2)}\left(\sum_{i=1}^n u_i\right)^2\right]}{2}\right\} \\
&\quad \int_{-\infty}^{\infty} \exp\left\{-\frac{v+\eta(1+\lambda^2)}{2(\eta+v)}\left[y-\frac{\lambda\sqrt{\eta+v}}{v+\eta(1+\lambda^2)}\sum_{i=1}^n u_i\right]^2\right\} dy du_i \\
&= |\mathbf{D}(\boldsymbol{\omega})|^{\frac{1}{2}} \frac{\sigma(2\pi)^{-\frac{n-1}{2}}}{\sqrt{v+\eta(1+\lambda^2)}} \\
&\quad \int \cdots \int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{\sigma\omega_i}} \exp\left\{-\frac{1}{2}\mathbf{u}^t\left[\mathbf{D}(\boldsymbol{\omega})-\frac{\lambda^2}{v+\eta(1+\lambda^2)}\mathbb{1}\mathbb{1}^t\right]\mathbf{u}\right\} du_i \\
&= |\mathbf{D}(\boldsymbol{\omega})|^{\frac{1}{2}} \frac{\sigma\sqrt{2\pi}\,|\boldsymbol{\Sigma}|^{\frac{1}{2}}}{\sqrt{v+\eta(1+\lambda^2)}} F_{\mathbf{U}}\left(\lambda\frac{x_1-\hat{\mu}}{\sigma\omega_1},\ldots,\lambda\frac{x_n-\hat{\mu}}{\sigma\omega_n}\right),
\end{aligned}
$$

where $\mathbf{U}$ distributes $N_n(\mathbf{0},\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}^{-1}=\mathbf{D}(\boldsymbol{\omega})-\frac{\lambda^2}{v+\eta(1+\lambda^2)}\mathbb{1}\mathbb{1}^t$. Now, using (C.0.1) we get $\boldsymbol{\Sigma}=[\mathbf{D}(\boldsymbol{\omega})]^{-1}+\frac{\lambda^2}{v+\eta}[\mathbf{D}(\boldsymbol{\omega})]^{-1}\mathbb{1}\mathbb{1}^t[\mathbf{D}(\boldsymbol{\omega})]^{-1}$.

Replacing $h(\lambda,\sigma,\omega_1,\ldots,\omega_n)$ in (D.0.1) we obtain

$$
\begin{aligned}
f(\mathbf{x}|\,\lambda,\sigma,\omega_1,\ldots,\omega_n) &= \sqrt{\frac{2^n v\,|\boldsymbol{\Sigma}|}{(\pi\sigma^2)^n\left[v+\eta(1+\lambda^2)\right]}} F_{\mathbf{U}}\left(\lambda\frac{x_1-\hat{\mu}}{\sigma\omega_1},\ldots,\lambda\frac{x_n-\hat{\mu}}{\sigma\omega_n}\right) \\
&\quad \exp\left\{-\frac{\eta}{2\sigma^2}\left[S_\omega^2+\frac{v}{\eta+v}\left(m-\sum_{i=1}^n \nu_i x_i\right)^2\right]\right\}.
\end{aligned}
$$

Then, since $\sigma^{-2}\sim Ga(a,b)$, one obtains

$$
\begin{aligned}
f(\mathbf{x}|\,\lambda,\omega_1,\ldots,\omega_n) &= \frac{b^a}{\Gamma(a)}\sqrt{\frac{2^{n+2}v\,|\boldsymbol{\Sigma}|}{\pi^n\left[v+\eta(1+\lambda^2)\right]}} \\
&\quad \int_0^{\infty} \sigma^{-n-1-2a}\exp\left(-\frac{r^2}{2\sigma^2}\right) F_{\mathbf{U}}\left(\lambda\frac{x_1-\hat{\mu}}{\sigma\omega_1},\ldots,\lambda\frac{x_n-\hat{\mu}}{\sigma\omega_n}\right) d\sigma,
\end{aligned}
$$

where $r^2=\eta S_\omega^2+\frac{\eta v}{\eta+v}\left(m-\sum_{i=1}^n \nu_i x_i\right)^2+2b$.

A change of variable $s = r^2/\sigma^2$, allows us to write

$$\begin{aligned} f\left(\mathbf{x}\middle|\, \lambda, \omega_1, \ldots, \omega_n\right) &= \frac{b^a}{\Gamma(a)\, r^{n+2a}} \sqrt{\frac{2^n v\, |\boldsymbol{\Sigma}|}{\pi^n \left[v + \eta\left(1+\lambda^2\right)\right]}} \\ &\quad \int_0^\infty s^{\frac{n}{2}+a-1} \exp\left(-\frac{s}{2}\right) F_{\mathbf{U}}\left(\lambda\sqrt{s}\frac{x_1-\hat{\mu}}{r\omega_1}, \ldots, \lambda\sqrt{s}\frac{x_n-\hat{\mu}}{r\omega_n}\right) ds \\ &= \frac{b^a \pi^{-n}}{\Gamma(a)\, r^{n+2a}} \sqrt{\frac{v}{v+\eta\left(1+\lambda^2\right)}} \\ &\quad \int_0^\infty s^{\frac{n}{2}+a-1} \exp\left(-\frac{s}{2}\right) \int \cdots \int_{-\infty}^{\lambda\sqrt{s}\frac{x_i-\hat{\mu}}{r\omega_i}} \exp\left\{-\frac{1}{2}\mathbf{u}^t \boldsymbol{\Sigma}^{-1}\mathbf{u}\right\} du_i ds. \end{aligned}$$

Making the change of variables $t_i = \frac{u_i}{\sqrt{s}}$, we get

$$\begin{aligned} f\left(\mathbf{x}\middle|\, \lambda, \omega_1, \ldots, \omega_n\right) &= \frac{b^a \pi^{-n}}{\Gamma(a)\, r^{n+2a}} \sqrt{\frac{v}{v+\eta\left(1+\lambda^2\right)}} \\ &\quad \int \cdots \int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{r\omega_i}} \int_0^\infty s^{n+a-1} \exp\left\{-\frac{s}{2}\left(1+\mathbf{t}^t\boldsymbol{\Sigma}^{-1}\mathbf{t}\right)\right\} ds dt_i \\ &= \frac{b^a \Gamma(n+a)\, 2^{n+a}}{\Gamma(a)\, r^{n+2a} \pi^n} \sqrt{\frac{v\, |\boldsymbol{\Sigma}|}{v+\eta\left(1+\lambda^2\right)}} \\ &\quad \int \cdots \int_{-\infty}^{\lambda\frac{x_i-\hat{\mu}}{r\omega_i}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left[1+\mathbf{t}^t\boldsymbol{\Sigma}^{-1}\mathbf{t}\right]^{-(n+a)} dt_i \\ &= \frac{b^a \Gamma(n/2+a)\, 2^{n+a}}{\Gamma(a)\, r^{n+2a} \pi^{n/2}} \sqrt{\frac{v\, |\boldsymbol{\Sigma}|}{v+\eta\left(1+\lambda^2\right)}} \\ &\quad F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{x_1-\hat{\mu}}{r\omega_1}, \ldots, \lambda\sqrt{n+2a}\frac{x_n-\hat{\mu}}{r\omega_n}\right), \end{aligned}$$

where $\mathbf{T}$ distributes $t_n\left(\mathbf{0}, \boldsymbol{\Sigma}, n+2a\right)$.

On the other hand,

$$\begin{aligned} f\left(\mathbf{x}\middle|\, \mu, \sigma, \omega_1, \ldots, \omega_n\right) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{\omega_i}} \phi\left(\frac{x_i-\mu}{\sigma\sqrt{\omega_i}}\right) \\ &= \left(\pi\sigma^2\right)^{-\frac{n}{2}} \left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \exp\left(-\frac{\eta}{2\sigma^2} S_\omega^2\right) \exp\left[-\frac{\eta\left(\mu - \sum_{i=1}^n \nu_i x_i\right)^2}{2\sigma^2}\right]. \end{aligned}$$

Considering that $\pi\left(\mu|\,\sigma\right) = N\left(\mu|\,m, \frac{\sigma^2}{v}\right)$, one obtains

$$\begin{aligned} f\left(\mathbf{x}|\,\sigma, \omega_1, \ldots, \omega_n\right) &= \sqrt{\frac{v}{2\left(\pi\sigma^2\right)^{n+1}}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \exp\left(-\frac{\eta}{2\sigma^2} S_\omega^2\right) \\ &\quad \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left[\eta\left(\mu - \sum_{i=1}^n \nu_i x_i\right)^2 + v\left(\mu - m\right)^2\right]\right\} d\mu \\ &= \sqrt{\frac{v}{2\left(\pi\sigma^2\right)^{n+1}}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \exp\left\{-\frac{\eta\left[S_\omega^2 + \frac{v}{\eta+v}\left(m - \sum_{i=1}^n \nu_i x_i\right)^2\right]}{2\sigma^2}\right\} \\ &\quad \int_{-\infty}^{\infty} \exp\left[-\frac{\eta+v}{2\sigma^2}\left(\mu - \hat{\mu}\right)^2\right] d\mu \\ &= \sqrt{\frac{v}{\left(\eta+v\right)\left(\pi\sigma^2\right)^n}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{\eta}{2\sigma^2}\left[S_\omega^2 + \frac{v}{\eta+v}\left(m - \sum_{i=1}^n \nu_i x_i\right)^2\right]\right\}. \end{aligned}$$

Thus, since $\sigma^{-2} \sim Ga\left(a, b\right)$, one gets

$$f\left(\mathbf{x}|\,\omega_1, \ldots, \omega_n\right) = \frac{2b^a}{\Gamma\left(a\right)}\sqrt{\frac{v}{\left(\eta+v\right)\pi^n}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \int_0^{\infty} \sigma^{-n-1-2a} \exp\left(-\frac{r^2}{2\sigma^2}\right) d\sigma.$$

With a new change of variable, $s = r^2/\sigma^2$, one obtains

$$\begin{aligned} f\left(\mathbf{x}|\,\omega_1, \ldots, \omega_n\right) &= \frac{b^a}{r^{n+2a}\Gamma\left(a\right)}\sqrt{\frac{v}{\left(\eta+v\right)\pi^n}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}} \int_0^{\infty} s^{\frac{n}{2}+a-1} \exp\left(-\frac{s}{2}\right) ds \\ &= \frac{2^{\frac{n}{2}+a}\Gamma\left(\frac{n}{2}+a\right) b^a}{\Gamma\left(a\right) r^{n+2a}\pi^{\frac{n}{2}}}\sqrt{\frac{v}{\left(\eta+v\right)}}\left|\mathbf{D}\left(\boldsymbol{\omega}\right)\right|^{-\frac{1}{2}}. \end{aligned}$$

Therefore, the Bayes factor is

$$\begin{aligned} BF\left(\mathbf{x}\right) &= \frac{\int \cdots \int f\left(\mathbf{x}|\,\omega_1, \ldots, \omega_n\right) dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)}{\int\left[\int \cdots \int f\left(\mathbf{x}|\,\lambda, \omega_1, \ldots, \omega_n\right) dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)\right] \pi\left(\lambda\right) d\lambda} \\ &= \frac{\int \cdots \int r^{-n-2a}\left[\left(\eta+v\right)\prod_{i=1}^n \omega_i\right]^{-\frac{1}{2}} dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)}{2^{\frac{n}{2}} \int \cdots \int r^{-n-2a} g\left(\boldsymbol{\omega}\right) dH\left(\omega_1\right) \cdots dH\left(\omega_n\right)}, \end{aligned}$$

where

$$g\left(\boldsymbol{\omega}\right) = \int \sqrt{\frac{\left|\boldsymbol{\Sigma}\right|}{v + \eta\left(1+\lambda^2\right)}} F_{\mathbf{T}}\left(\lambda\sqrt{n+2a}\frac{\mathbf{x} - \hat{\mu}\mathbb{1}}{r}\left[\mathbf{D}\left(\boldsymbol{\omega}\right)\right]^{-1}\right) \pi\left(\lambda\right) d\lambda.$$

But since

$$\begin{aligned} |\boldsymbol{\Sigma}| &= \frac{|\boldsymbol{\Sigma}\mathbf{D}(\boldsymbol{\omega})|}{|\mathbf{D}(\boldsymbol{\omega})|} = \frac{\left|\mathbf{I}_n + \frac{\lambda^2}{v+\eta}\left[\mathbf{D}(\boldsymbol{\omega})\right]^{-1}\mathbb{1}\mathbb{1}^t\right|}{\prod_{i=1}^n \omega_i} \\ &= \frac{\lambda^{2n}\left|\frac{v+\eta}{\lambda^2}\mathbf{I}_n + \left[\mathbf{D}(\boldsymbol{\omega})\right]^{-1}\mathbb{1}\mathbb{1}^t\right|}{(v+\eta)^n \prod_{i=1}^n \omega_i} = \frac{v+\eta\left(1+\lambda^2\right)}{(v+\eta)\prod_{i=1}^n \omega_i}, \end{aligned}$$

the proof is complete.

Appendix E

Proof of Proposition 5.2.6

Notice that

$$\begin{aligned}
f\left(\mathbf{y}\right|\phi,\lambda,\omega) &= \int_{\mathbb{R}^k} f\left(\mathbf{y}\right|\boldsymbol{\beta},\phi,\lambda,\omega)\,\pi\left(\boldsymbol{\beta}\right|\phi)\,d\boldsymbol{\beta} \\
&= \int_{\mathbb{R}^k} 2N_n\left(\mathbf{y}\left|\mathbf{X}\boldsymbol{\beta},\frac{\omega}{\phi}\mathbf{I}_n\right.\right)\Phi\left[\left(\frac{\phi}{\omega}\right)^{\frac{1}{2}}\lambda\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]N_k\left(\boldsymbol{\beta}\left|\mathbf{m},\phi^{-1}\mathbf{B}\right.\right)d\boldsymbol{\beta} \\
&= \frac{2\phi^{\frac{n+k}{2}}\left|\mathbf{B}\right|^{-\frac{1}{2}}}{\left(2\pi\right)^{\frac{n+k+1}{2}}\omega^{\frac{n}{2}}}\int_{\mathbb{R}^k}\exp\left\{-\frac{\phi}{2\omega}\left[\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2+\omega\left(\boldsymbol{\beta}-\mathbf{m}\right)^t\mathbf{B}^{-1}\left(\boldsymbol{\beta}-\mathbf{m}\right)\right]\right\} \\
&\quad \int_{-\infty}^{\left(\frac{\phi}{\omega}\right)^{\frac{1}{2}}\lambda\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}\exp\left\{-\frac{z^2}{2}\right\}dz\,d\boldsymbol{\beta},
\end{aligned}$$

but since

$$\begin{aligned}
\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2+\omega\left(\boldsymbol{\beta}-\mathbf{m}\right)^t\mathbf{B}^{-1}\left(\boldsymbol{\beta}-\mathbf{m}\right) &= \left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right) \\
&\quad -\hat{\boldsymbol{\beta}}^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}+\mathbf{y}^t\mathbf{y}+\omega\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m},
\end{aligned}$$

where

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\left(\mathbf{X}^t\mathbf{y}+\omega\mathbf{B}^{-1}\mathbf{m}\right),$$

then

$$f\left(\mathbf{y}\right|\phi,\lambda,\omega) = \frac{2\phi^{\frac{n+k}{2}}\left|\mathbf{B}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{n+k+1}{2}}\omega^{\frac{n}{2}}}h\left(\phi,\lambda,\omega\right) \exp\left\{-\frac{\phi}{2\omega}\left[\mathbf{y}^t\mathbf{y}+\omega\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}\right]\right\}, \tag{E.0.1}$$

where

$$h\left(\phi,\lambda,\omega\right) = \int_{\mathbb{R}^k}\exp\left\{-\frac{\phi}{2\omega}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\} \int_{-\infty}^{\left(\frac{\phi}{\omega}\right)^{\frac{1}{2}}\lambda\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}\exp\left\{-\frac{z^2}{2}\right\}dzd\boldsymbol{\beta}.$$

Now, making the change of variable $u=\frac{z}{\sqrt{\omega}}+\frac{\lambda\sqrt{\phi}}{\omega}\mathbb{1}_n^t\mathbf{X}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)$ and exchanging the integration order, we obtain

$$\begin{aligned} h\left(\phi,\lambda,\omega\right) &= \sqrt{\omega}\int_{-\infty}^{\frac{\lambda\sqrt{\phi}}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)}\int_{\mathbb{R}^k}\exp\left\{-\frac{\omega}{2}\left[u-\frac{\lambda\sqrt{\phi}}{\omega}\mathbb{1}_n^t\mathbf{X}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right]^2\right\} \\ &\quad \exp\left\{-\frac{\phi}{2\omega}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}d\boldsymbol{\beta}du \\ &= \sqrt{\omega}\int_{-\infty}^{\frac{\lambda\sqrt{\phi}}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)}\exp\left\{-\frac{\omega u^2}{2}\right\} \\ &\quad \int_{\mathbb{R}^k}\exp\left\{-\frac{\phi}{2\omega}\left[\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\mathbf{W}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)-2\frac{\omega\lambda u}{\sqrt{\phi}}\mathbb{1}_n^t\mathbf{X}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right]\right\}d\boldsymbol{\beta}du, \end{aligned}$$

where $\mathbf{W}=\mathbf{X}^t\left(\mathbf{I}_n+\lambda^2\mathbb{1}_n\mathbb{1}_n^t\right)\mathbf{X}+\omega\mathbf{B}^{-1}$. But since

$$\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)^t\mathbf{W}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)=\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\mathbf{W}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)-2\frac{\omega\lambda u}{\sqrt{\phi}}\mathbb{1}_n^t\mathbf{X}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)+\tilde{\boldsymbol{\beta}}^t\mathbf{W}\tilde{\boldsymbol{\beta}},$$

where $\tilde{\boldsymbol{\beta}} = \frac{\omega\lambda u}{\sqrt{\phi}}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n$, then

$$\begin{aligned} h(\phi,\lambda,\omega) &= \sqrt{\omega}\int_{-\infty}^{\frac{\lambda\sqrt{\phi}}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)} \exp\left\{\frac{\phi}{2\omega}\tilde{\boldsymbol{\beta}}^t\mathbf{W}\tilde{\boldsymbol{\beta}} - \frac{\omega u^2}{2}\right\} \\ &\quad \int_{\mathbb{R}^k}\exp\left\{-\frac{\phi}{2\omega}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)^t\mathbf{W}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)\right\}d\boldsymbol{\beta}du \\ &= \frac{\omega^{\frac{k+1}{2}}(2\pi)^{\frac{k}{2}}}{\phi^{\frac{k}{2}}|\mathbf{W}|^{\frac{1}{2}}}\int_{-\infty}^{\frac{\lambda\sqrt{\phi}}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)}\exp\left\{-\frac{\omega}{2}\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)u^2\right\}du \\ &= \frac{\omega^{\frac{k}{2}}(2\pi)^{\frac{k+1}{2}}}{\phi^{\frac{k}{2}}|\mathbf{W}|^{\frac{1}{2}}\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)^{\frac{1}{2}}}F_U\left[\frac{\lambda}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right], \end{aligned}$$

where $U \sim N\left(0, \frac{\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)^{-1}}{\phi\omega}\right)$.

Now, if we substitute $h(\phi,\lambda,\omega)$ in (E.0.1), we get

$$\begin{aligned} f(\mathbf{y}|\,\phi,\lambda,\omega) &= \frac{2\phi^{\frac{n}{2}}|\mathbf{B}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\omega^{\frac{n-k}{2}}|\mathbf{W}|^{\frac{1}{2}}\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)^{\frac{1}{2}}}F_U\left[\frac{\lambda}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right] \\ &\quad \exp\left\{-\frac{\phi}{2\omega}\left[\mathbf{y}^t\mathbf{y}+\omega\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}\right]\right\}. \end{aligned}$$

But since $\phi \sim Ga(a,b)$, then

$$\begin{aligned} f(\mathbf{y}|\,\lambda,\omega) &= \int_0^\infty f(\mathbf{y}|\,\phi,\lambda,\omega)\,Ga(\phi\,|a,b)\,d\phi \\ &= \frac{2b^a|\mathbf{B}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\omega^{\frac{n-k}{2}}|\mathbf{W}|^{\frac{1}{2}}\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)^{\frac{1}{2}}\Gamma(a)} \\ &\quad \int_0^\infty F_U\left[\frac{\lambda}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right]\phi^{\frac{n}{2}+a-1}\exp\left\{-\frac{\phi r^2}{2}\right\}d\phi, \end{aligned}$$

where $r^2 = \frac{\mathbf{y}^t\mathbf{y}}{\omega}+\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\omega^{-1}\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}+2b$. Now, from Lemma 5.2.2, we obtain

$$f(\mathbf{y}|\,\lambda,\omega) = \frac{2^{\frac{n}{2}+a+1}b^a\Gamma\left(\frac{n}{2}+a\right)|\mathbf{B}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\omega^{\frac{n-k}{2}}|\mathbf{W}|^{\frac{1}{2}}\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)^{\frac{1}{2}}\Gamma(a)\,r^{n+2a}}F_T\left[\frac{\lambda}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right],$$

where $T \sim t\left(0, \frac{r^2}{\omega\left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\mathbb{1}_n\right)(n+2a)}, n+2a\right)$.

Using the well known identity (see Fang and Zhang (1990, Chapter 1))

$$(\mathbf{A}+\mathbf{CED})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{CE}\left(\mathbf{E}+\mathbf{EDA}^{-1}\mathbf{CE}\right)^{-1}\mathbf{EDA}^{-1}$$

for some matrix $\mathbf{A}_{p\times p}$, $\mathbf{E}_{q\times q}$ $\mathbf{C}_{p\times q}$ and $\mathbf{D}_{q\times p}$, we have

$$\mathbf{W}^{-1} = \left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1} - \frac{\lambda^2\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\mathbf{X}^t\ \mathbb{1}_n\mathbb{1}_n^t\mathbf{X}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}}{1+\lambda^2\mathbb{1}_n^t\mathbf{X}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\mathbf{X}^t\ \mathbb{1}_n}.$$

So that,

$$1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\ \mathbb{1}_n = \left(1+\lambda^2\mathbb{1}_n^t\mathbf{X}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\mathbf{X}^t\ \mathbb{1}_n\right)^{-1}.$$

But, from C.0.2, we have

$$\begin{aligned}
\left|\mathbf{W}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\right| &= \left|\mathbf{I}_k+\lambda^2\mathbf{X}^t\ \mathbb{1}_n\mathbb{1}_n^t\mathbf{X}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\right| \\
&= \left|1+\lambda^2\mathbb{1}_n^t\mathbf{X}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\mathbf{X}^t\ \mathbb{1}_n\right| \\
&= \left(1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\ \mathbb{1}_n\right)^{-1},
\end{aligned}$$

thus

$$|\mathbf{W}| = \frac{\left|\mathbf{W}\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\right|}{\left|\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)^{-1}\right|} = \frac{\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|}{1-\lambda^2\mathbb{1}_n^t\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t\ \mathbb{1}_n}.$$

Therefore,

$$f\left(\mathbf{y}|\,\lambda,\omega\right) = \frac{2^{\frac{n}{2}+a+1}b^a\Gamma\left(\frac{n}{2}+a\right)|\mathbf{B}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\,\omega^{\frac{n-k}{2}}\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{\frac{1}{2}}\Gamma(a)\,r^{n+2a}}F_T\left[\frac{\lambda}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right].$$

On the other hand, when $\lambda = 0$:

$$
\begin{aligned}
f(\mathbf{y}|\,\phi,\omega) &= \int_{\mathbb{R}^k} f(\mathbf{y}|\,\boldsymbol{\beta},\phi,\omega)\,\pi(\boldsymbol{\beta}|\,\phi)\,d\boldsymbol{\beta} \\
&= \int_{\mathbb{R}^k} N_n\left(\mathbf{y}\left|\mathbf{X}\boldsymbol{\beta},\frac{\omega}{\phi}\mathbf{I}_n\right.\right) N_k\left(\boldsymbol{\beta}\left|\mathbf{m},\phi^{-1}\mathbf{B}\right.\right) d\boldsymbol{\beta} \\
&= \frac{\phi^{\frac{n+k}{2}}\left|\mathbf{B}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{n+k}{2}}\,\omega^{\frac{n}{2}}} \int_{\mathbb{R}^k} \exp\left\{-\frac{\phi}{2\omega}\left[\left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|^2+\omega\left(\boldsymbol{\beta}-\mathbf{m}\right)^t\mathbf{B}^{-1}\left(\boldsymbol{\beta}-\mathbf{m}\right)\right]\right\} d\boldsymbol{\beta} \\
&= \frac{\phi^{\frac{n+k}{2}}\left|\mathbf{B}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{n+k}{2}}\,\omega^{\frac{n}{2}}} \exp\left\{-\frac{\phi}{2}\left[\frac{\mathbf{y}^t\mathbf{y}}{\omega}+\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\omega^{-1}\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}\right]\right\} \\
&\quad \int_{\mathbb{R}^k} \exp\left\{-\frac{\phi}{2\omega}\left[\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)^t\left(\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right]\right\} d\boldsymbol{\beta} \\
&= \frac{\phi^{\frac{n}{2}}\left|\mathbf{B}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\,\omega^{\frac{n-k}{2}}\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{\frac{1}{2}}} \\
&\quad \exp\left\{-\frac{\phi}{2}\left[\frac{\mathbf{y}^t\mathbf{y}}{\omega}+\mathbf{m}^t\mathbf{B}^{-1}\mathbf{m}-\hat{\boldsymbol{\beta}}^t\left(\omega^{-1}\mathbf{X}^t\mathbf{X}+\mathbf{B}^{-1}\right)\hat{\boldsymbol{\beta}}\right]\right\}.
\end{aligned}
$$

Considering that $\phi \sim Ga(a,b)$, we obtain

$$
\begin{aligned}
f(\mathbf{y}|\,\omega) &= \frac{b^a\left|\mathbf{B}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\,\omega^{\frac{n-k}{2}}\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{\frac{1}{2}}\Gamma(a)} \int_0^\infty \phi^{\frac{n}{2}+a-1}\exp\left\{-\frac{\phi r^2}{2}\right\} d\phi \\
&= \frac{2^{\frac{n}{2}+a}b^a\Gamma\left(\frac{n}{2}+a\right)\left|\mathbf{B}\right|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}\,\omega^{\frac{n-k}{2}}\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{\frac{1}{2}}\Gamma(a)\,r^{n+2a}}.
\end{aligned}
$$

Therefore, the Bayes factor is

$$
\begin{aligned}
BF(\mathbf{y}) &= \frac{\int f(\mathbf{y}|\,\omega)\,dH(\omega)}{\int\int f(\mathbf{y}|\,\lambda,\omega)\,\pi(\lambda)\,d\lambda dH(\omega)} \\
&= \frac{\int \omega^{\frac{k-n}{2}}\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{-\frac{1}{2}} r^{-n-2a} dH(\omega)}{2\int\int \omega^{\frac{k-n}{2}}\left|\mathbf{X}^t\mathbf{X}+\omega\mathbf{B}^{-1}\right|^{-\frac{1}{2}} r^{-n-2a} F_T\left[\frac{\lambda}{\omega}\mathbb{1}_n^t\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right]\pi(\lambda)\,d\lambda dH(\omega)}.
\end{aligned}
$$

We standardize the random variable T and obtain the result.

Bibliography

Abdullah, M. B. (1995). Detection of influential observations in functional errors-in-variables model. *Communications in Statistics: Theory and Methods*, **24**(6), 1585–1595.

Adcock, R. J. (1877). Note on the method of least squares. *Analyst*, **4**, 183–184.

Adcock, R. J. (1878). A problem in least squares. *Analyst*, **5**, 53–54.

Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society B*, **53**, 111–142. with discussion.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Pages 267–281 of: 2nd. Int. Symp. Information Theory*. Budapest: Akademia Kaido.

Aoki, R., Achcar, J. A., Bolfarine, H., and Singer, J. M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. *Journal of Applied Statistics*, **30**, 3–12.

Arellano-Valle, R. B. 1994. *Distribuicões Elípticas: Propiedades e Aplicacões a Modelos de Regressão*. Ph.D. thesis, IME-USP, São Paulo, Brasil. in Portugues.

Arellano-Valle, R. B., Bolfarine, H., and Iglesias, P. (1994). A predictivistic interpretation to multivariate t-distribution. *Test*, **3**, 221–236.

Arellano-Valle, R. B., Galea-Rojas, M., and Iglesias, P. (2000). Bayesian sensitivity analysis in elliptical linear regression models. *Journal of Statistical Planning and Inference*, **86**, 175–199.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. 2001. A general class of univariate asymmetric distributions. Tech. rept. Pontificia Universidad Católica de Chile.

Arellano-Valle, R. B., del Pino, G., and Iglesias, P. 2002a. Bayesian analysis for the spherical linear model. Tech. rept. http://www.mat.puc.cl/ pliz.

Arellano-Valle, R. B., del Pino, G., and Martín, E. San. (2002b). Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters*, **58**, 111–121.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171–178.

Azzalini, A., and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society B*, **61**, 579–602.

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

Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data*. 3rd edn. New York: John Wiley.

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika*, **44**, 533–534.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

Berger, J. O. 1998. Reference on Bayesian hypothesis testing and model selection. *In: VI Valencia International Meeting on Bayesian Statistics*.

Berger, J. O., and Pericchi, L. R. 1993. The Intrinsic Bayes factor for model selection and prediction. Tech. rept. 94-10C. Purdue University.

Berger, J. O., and Pericchi, L. R. (1996b). The Intrinsic Bayes factor for linear models. *Pages 25–44 of:* Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds), *Bayesian Statistics 5*. Oxford University Press. with discussion.

Berger, J. O., and Pericchi, L. R. (1996a). The Intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91**, 109–122.

Berger, J. O., and Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. *In:* Lahiri, P. (ed), *Model Selection*, vol. 38. Institute of Mathematical Statistics, Lecture Notes Monograph Series. with discussion.

Berger, J. O., Pericchi, L. R., and Varshavsky, J. (1998). Bayes factor and marginal distributions in invariant situations. *Sankhyã A*, **60**, 307–321.

Bernardo, J. M. (1999). Nested hypothesis testing: the Bayesian reference criterion. *Pages 101–130 of:* Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds), *Bayesian Statistics 6*. Oxford University Press. with discussion.

Bernardo, J. M., and Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Bolfarine, H., and Cordani, L. K. (1993). Estimation of a structural regression model with know reliability ratio. *Annals of the Institute of Statistical Mathematics*, **45**, 531–540.

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

Branco, M. D., Bolfarine, H., Iglesias, P., and Arellano-Valle, R. B. (2001). Bayesian analysis of the calibration problem under elliptical distributions. *Journal of Statistical Planning and Inference*, **90**, 69–95.

Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptical contoured distributions. *Journal of Multivariate Analysis*, **11**, 368–385.

Carlin, B. P., and Polson, N. G. (1991). An expected utility approach to influence diagnostics. *Journal of the American Statistical Association*, **86**, 1013–1021.

Carrol, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman & Hall.

Chen, M. H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer Series in Statistics. New York: Springer-Verlag.

Cheng, C-L., and Ness, J. W. Van. (1999). *Statistical Regression with Measurement Error*. Kendall's Library of Statistics 6. London: Arnold.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association*, **90**, 1313–1321.

Chib, S., and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, **96**, 270–281.

Chib, S., Tiwari, R. C., and Jammalamadaka, S. R. (1988). Bayes prediction in regressions with elliptical errors. *Journal of Econometrics*, **38**, 349–360.

Chu, K. (1973). Estimation and decision for linear systems with elliptical random processes. *IEEE Trans. Automat. Control*, **18**, 499–505.

Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. London: Chapman & Hall.

Dellaportas, P., and Stephens, D. A. (1995). Bayesian analysis of errors in variable regression models. *Biometrics*, **51**, 1085–1095.

Dongarra, J. J., and Grosse, E. (1987). Distribution of mathematical software via electronic mail. *Communications of the ACM*, **30**, 403–407.

Fang, K. T., and Zhang, Y. T. (1990). *Generalized Multivariate Analysis*. New York/Berlin: Springer-Verlag.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. London/New York: Chapman & Hall.

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

Florens, J. P., Mouchart, M., and Richard, J. F. (1974). Bayesian inference in error-in-variables models. *Journal of Multivariate Analysis*, **4**, 419–452.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley & Sons.

Geisser, S. (1987). Influential observations, diagnostics and predictive discordancy test. *Journal of Applied Statistics*, **14**(2), 133–142.

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

Geman, S., and Geman, D. (1984). Stochastic relaxation, gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–740.

Genton, M. G., and Loperfido, N. M. R. 2001. Generalized skew-elliptical distributions and their quadratic forms. Tech. rept. North Carolina State University.

Geweke, L. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, **57**, 1317–1340.

Girón, F. J., Martínez, M., and Morcillo, C. (1992). A Bayesian justification for the analysis of residual and influence measures. *Pages 651–660 of:* Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds), *Bayesian Statistics 4*. Oxford: University Press.

Guttman, I., and Peña, D. (1988). Outliers and influence: Evaluation by posteriors of parameters in the linear model. *Pages 631–640 of:* Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (eds), *Bayesian Statistics 3*. Oxford: University Press.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57**, 97–109.

Jammalamadaka, S. R., Tiwari, R. C., and Chib, S. (1987). Bayes prediction in the linear model with spherically symmetric errors. *Economics Letters*, **24**, 39–44.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, **31**, 203–222.

Jeffreys, H. (1961). *Theory of Probabilitiy*. 3rd edn. Oxford, U.K.: Oxford University Press.

Johnson, W., and Geisser, S. 1982. Assessing the predictive influence of observations. *Pages 343–348 of:* Kallianpur, G., Krishnaiah, P. R., and Ghosh, J. K. (eds), *Statistics and Probability*. Essay in Honor of C. R. Rao.

Johnson, W., and Geisser, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. *Journal of the American Statistical Association*, **78**, 137–144.

Johnson, W., and Geisser, S. (1985). Estimative influence measures for the multivariate general linear model. *Journal of Statistical Planning and Inference*, **11**, 33–56.

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

Kass, R. E., Tierney, L., and Kadane, J. B. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika*, **76**, 663–674.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyã A*, **32**, 419–430.

Kempthorne, P. J. (1986). Decision-theoretic measures of influence in regression. *Journal of the Royal Statistical Society B*, **48**, 370–378.

Key, J. T., Pericchi, L. R., and Smith, A. F. M. (1999). Bayesian model choice: what and why? *Pages 343–370 of:* Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds), *Bayesian Statistics 6*. Oxford University Press. with discussion.

Kim, M. G. (2000). Outliers and influential observations in the structural errors-in-variables model. *Journal of Applied Statistics*, **27**(4), 451–460.

Lavine, M., and Schervish, M. (1999). Bayes factors: what they are and what they are not. *The American Statistician*, **53**, 119–122.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, **44**, 187–192.

Lindley, D. V., and El-Sayyad, G. M. (1968). The Bayesian estimation of a linear functional relationship. *Journal of Royal Statistical Society B*, **30**, 190–202.

Liseo, B., and Loperfido, N. 2002. Default Bayesian analysis of the skew-normal distribution. Tech. rept. Università di Roma "La Sapienza".

Loschi, R., Iglesias, P., and Arellano-Valle, R. B. 2002. Predictivistic characterization of t distribution with centered spherical symmetry. Tech. rept. http://www.mat.puc.cl/ pliz.

McCulloch, R. E., and Rossi, P. E. (1991). A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics*, **49**, 141–168.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation on state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.

Mudholkar, G. S., and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference*, **83**, 291–309.

Muller, P., and Roeder, K. (1997). A Bayesian semiparametric model for case-control studies with errors in variables. *Biometrika*, **84**, 523–537.

Naylor, J. C., and Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Appl. Statist.*, **31**, 214–225.

Naylor, J. C., and Smith, A. F. M. (1988). Economic illustrations of novel numerical integration methodology for Bayesian inference. *Journal of Econometrics*, **38**, 103–125.

O'Hagan, A. (1994). *Bayesian Inference*. Kendall's Advanced Theory of Statistics, vol. 2B. London: Edward Arnold.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society B*, **57**, 99–138.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test*, **6**, 101–118.

Osiewalski, J., and Steel, M. F. J. (1993). Robust Bayesian inference in elliptical regression models. *Journal of Econometrics*, **57**, 345–363.

Osiewalski, J., and Steel, M. F. J. (1996). Posterior moments of scale parameters in elliptical regression models. *Pages 323–335 of:* D. Berry, et al. (ed), *Bayesian Analysis in Statistics and Econometrics*. New York: Wiley. Essays in Honor of Arnold Zellner.

Peng, F., and Dey, D. K. (1995). Bayesian analysis of outlier problems using divergences measures. *The Canadian Journal of Statistics*, **23**, 199–213.

Pereira, C., and Stern, J. M. (1999). Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy*, **1**, 104–115.

Pettit, L. I. (1992). Bayes factors for outlier models using the device of imaginary observations. *Journal of the American Statistical Association*, **87**, 541–545.

Pettit, L. I., and Smith, A. F. M. (1983). Outliers and influential observations in linear models. *Pages 473–494 of:* Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (eds), *Bayesian Statistics 2*. Elsevier Science Publishers.

Piessens, R., deDoncker Kapenga, E., Uberhuber, C., and Kahaner, D. (1983). *QUADPACK: a Subroutine Package for Automatic Integration*. Berlin: Springer.

Reilly, P., and Patino-Leal, H. (1981). A Bayesian study of the error in variable models. *Technometrics*, **23**(3), 221–231.

Richardson, S. (1996). Measurement error. *Pages 401–414 of:* Gilks, W. R., Richardson, S., and Spiegelhalter, D.I. (eds), *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Richardson, S., and Gilks, W. R. (1993). A Bayesian approach to measurement error problems in epidemiological studies with covariate measurement error. *Statist. Med.*, **12**, 1703–1722.

Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distribution. *Pages 395–402 of:* Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (eds), *Bayesian Statistics 3*. Oxford: University Press. with discussion.

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

Stephens, D. A., and Dellaportas, P. (1992). Bayesian analysis of generalized linear measurement error models. *Pages 813–820 of:* Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds), *Bayesian Statistics 4*. Oxford: University Press.

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

Tierney, L., Kass, R. E., and Kadane, J. B. (1989). Fully exponential laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, **84**, 710–716.

Villegas, C. (1972). Bayesian inference in linear relations. *Ann. Math. Statis.*, **43**, 1767–1791.

Weiss, R. E. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society B*, **58**, 739–750.

Weiss, R. E., and Cook, R. D. (1992). A grafical case statistic for assessing posterior influence. *Biometrika*, **79**, 51–55.

Wellman, M. J., and Gunst, R. F. (1991). Influence diagnostics for linear measurement error models. *Biometrika*, **78**(2), 373–380.

Zellner, A. (1971). *An Introducition to Bayesian Inference in Econometrics*. New York: Wiley.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, **71**, 400–405.

Index