A Robust Bayesian Approach to Portfolio Selection

Katia Passarin

Submitted for the degree of Ph.D. in Economics

University of Lugano, Switzerland

Accepted on the recommendation of

Prof. Giovanni Barone-Adesi, University of Lugano

Prof. Antonietta Mira, Insubria University

Prof. Elvezio Ronchetti, advisor, University of Geneva and University of Lugano

Prof. Fabio Trojani, University of Lugano

July 2004

To all my friends who help me

to look at the Truth

Abstract

This thesis aims to study the local robustness properties of Bayesian posterior summaries and to derive a robust procedure to estimate such quantities. Such results are then applied to in the Bayesian Mean-Variance portfolio selection problem. In the first part, we study the local robustness of Bayesian estimators. In particular we build a framework where any Bayesian quantity can be seen as a posterior functional. This point of view allows us to construct different robustness measures. We derive local influence measures for posterior summaries with respect both to prior and sampling distributions and to observations. Afterwards we address the issue of efficient implementation of the derived measures through MCMC algorithms. In the second part, we deal with the problem of robust estimation in a Bayesian context, providing a useful result to generalize univariate robust distributions to the multivariate case. We also propose criteria to assess when a robust model is recommended and how to choose among estimates obtained with different distributions. The third part finally considers the Mean-Variance portfolio selection problem. We give evidence that the Bayesian approach works better than the Certainty Equivalence approach whenever data are normally distributed, although this is no longer true when data contain few outlying observations. Moreover, we compute useful measures of sensitivity of Bayesian weights and we construct and implement a new estimator, which is robust to the presence of 'extreme' observations.

Keywords: Bayesian Mean-Variance approach, Estimation risk, Posterior summaries, Robustness measures, Robust estimation, MCMC methods.

Acknowledgments

I wish to thank all the people who have helped me during my PhD studies. First of all, I wish to thank: my advisor, Professor Elvezio Ronchetti, for his precious time and guidance, Professor Giovanni Barone-Adesi and Professor Fabio Trojani for their kind availability and for the several profitable discussions on many different topics, Professor Antonietta Mira for her encouragement and her careful supervision of my use of MCMC methodologies, Fondazione Daccò and expecially Mrs. Cele Daccò for the financial support over these years.

I also wish to thank many colleagues of mine at the University of Lugano: Francesco Fierli, Patrick Gagliardini, Loriano Mancini, Claudia Ravanelli, Paolo Tenconi for the many enlightening pieces of advice on my thesis and for helping me to reach efficient programming. I also wish to thank Paolo Schmidt for solving many computer problems for me, Gisèle Nodiroli and Stefano Vizzola for their kindness, Rossana Galli, Federica De Rossa and Gianluca Cassese for their valuable companionship.

I wish then to thank my family and Professor Bernardo Colombo, who always believed in me. Thanks to Chiara Piccini, the Toti family, Sabrina Lurati, Claudio Del Ponte, all of whose friendship is so important in my life.

Finally, I thank with gratitude Ettore Codevilla for his endless love and support and for his encouraging presence, without which I would never discover what happiness is.

Contents

List of Figures

List of Tables

Introduction

The Mean-Variance approach to portfolio theory (Markowitz, 1952; Markowitz, 1959) states that a risk-averse investor should choose the amount to invest among a set of assets relying just on the first and second moments of return distribution. Such approach relies on assumptions which may appear too simplistic with respect to the empirical evidence on asset returns distribution (Cont, 2001), e.g. the normality of asset returns distribution and independence. Furthermore, it is not able to account for important characteristics such as the presence of extremely high or low returns. Nevertheless, this approach is important for both practitioners and researchers in finance (Britten-Jones, 1999) and it is central to many asset pricing theories. Its popularity in practical applications is due to the fact that in the classical implementation of this theory (known as Certainty Equivalence or naive approach) unknown parameters are simply replaced by their sample estimates. However practitioners find that the derived optimal portfolio allocation is often unreasonable (Black and Litterman, 1992).

A first reason for this fact is that the naive approach does not consider the *estimation risk,* i.e. the risk due to the fact that the estimated parameters display a sampling error which cannot be ignored. There is evidence in the literature that not taking into account parameters uncertainty leads to suboptimal portfolios (Barry, 1974; Brown, 1979; Jorion, 1986; Cavadini, Sbuelz and Trojani, 2002). A Bayesian approach to Mean-Variance portfolio selection problem is a possible solution (see Bawa, Brown and Klein, 1979; for a more general reference on Bayesian statistics see Berger, 1985; Cifarelli and Muliere, 1989; Bernardo and Smith, 1994). In this approach Bayesian Mean-

Variance portfolio weights turn out to be a function of the moments of the predictive distribution of future returns. The Bayesian point of view not only considers parameter uncertainty but also satisfies the axiomatic paradigm of Neumann-Morgenstern-Savage's expected utility maximization.

A second reason for this inefficiency is that Mean-Variance portfolio weights are extremely sensitive to observations. It is found in practice that time series of asset returns are characterized by the presence of several extremely low or high returns. Such outlying values are due for instance to market crashes that can hardly be included in the data generating process. This fact can induce a bias in the estimates and leads to the so called *model risk* problem. Recent papers deal with model risk in the Certainty Equivalence approach (Victoria-Feser, 2000; Cavadini, Sbuelz and Trojani, 2002; Perret-Gentil and Victoria-Feser, 2003). They show that few outlying returns have a strong influence on the composition of the resulting optimal portfolio. Moreover Cavadini, Sbuelz and Trojani (2002) find that model risk plays a greater role than estimation risk. However no evidence of this fact is given for Bayesian weights.

This thesis aims to study the robustness properties of Bayesian Mean-Variance weights and proposes a new estimator which is not too much affected by the presence of extreme observations in the sample. In order to do this we first review the field of robustness in Bayesian statistics and we propose a simple and unified framework that helps to construct useful measures of sensitivity of Bayesian quantities and to build robust Bayesian estimators.

In recent years the question about Bayesian procedures sensitivity to their inputs has become more and more popular, and all contributions in this sense fall into the category of the so-called Bayesian robust statistics. Any Bayesian quantity depends on two distributional assumptions and on a sample of observed data. Most efforts concentrate on building measures of sensitivity to changes in the distributional assumptions (prior or/and sampling distributions).

One of the first attempts in this direction is due to Box and Tiao (Box and Tiao, 1964; Box and Tiao, 1992), who distinguish between criterion robustness and inference robustness. When we make inference from a sample of data, the criterion to draw inferential conclusions (e.g. a

statistics) depends on the assumption we made about the data generating process. Criterion robustness evaluates how the chosen statistics changes as the distributional assumptions change. If the statistics does not differ substantially under different distributions, it is said to be (criterion) robust. However, if distributional assumptions are known to be different from the ones believed, the fixed criterion would also be different. Inference robustness evaluates how inferential conclusions change as the criterion changes.

More recent literature within the field of criterion robustness studies the sensitivity of Bayesian quantities to questionable distributional assumptions. Usually, this uncertainty is represented by varying the suspected source (either the prior or the sampling model) within a class of distributions. The global approach to robustness considers large classes of different distributions and evaluates the range of variation of the quantity under study. A good review on this topic can be found in Berger (1994). A second direction is the local approach to robustness. It assesses the effects of small perturbations of the assumed distributions represented by neighborhoods of the base models. The sensitivity to small deviations from the base model is evaluated with suitable derivatives (Ruggeri and Wasserman, 1993; Sivaganesan, 1993; Dey *et al.,* 1996; Gustafson *et al.,* 1996; Moreno *et al.,* 1996; Peña and Zamar, 1997; Gustafson, 2000). Little attention has been paid in the Bayesian literature to the sensitivity to observations. However this is a well-known matter in the Theory of Robust Statistics developed in Hampel, Ronchetti, Rousseeuw and Stahel (1986). Here any statistics is seen as a functional and different quantities can be defined in order to assess the influence of a single observation in the sample.

Once the sensitivity of a Bayesian quantity has been checked, the next step is to build robust Bayesian procedures. We find two main directions in the literature. The first direction is developed within the global approach and applies when a large range of variation is obtained for the quantity under study. It aims to narrow the class of prior and/or sampling distributions down to the point where a satisfactory range is reached (see Berger, 1994; Liseo *et al.,* 1996; Moreno *et al.,* 1996). A second direction applies when normality is adopted for the sampling distribution and

this assumption may appear inadequate because of the presence of few atypical observations. In Bayesian analyses the normality assumption is often convenient in order to obtain analytical results for the posterior distribution. However, in this case it is well known that the sensitivity of posterior quantities to observations is more pronounced and that only few atypical values in the sample heavily influence estimates. The reason for this fact has been found by many authors in light tails of the normal model adopted (Box and Tiao, 1992; Dawid, 1973; Zellner, 1976). Robustness with respect to atypical observations is achieved by choosing a so-called *robust model*, i.e. a location-scale family of symmetric unimodal distributions enriched with 'robustness' parameters that control its shape (see Box and Tiao, 1962; Ramsay and Novick, 1980; West, 1984; Albert *et al.,* 1991).

This thesis follows the local approach to Bayesian robustness. Such approach would consider the fact that asset returns display sometimes 'extreme' values which can be hardly reflected in a normal data generating model and we may be interested to capture the structure of the stochastic process that generates the bulk of the data. In the first chapter we build a framework where any Bayesian quantity can be seen as a posterior functional and its sensitivity to all inputs is checked. Moreover, we derive local of influence measures for posterior summaries with respect both to distributional assumptions and to observations and we consider the issue of efficient implementation of the derived measures. In the second chapter we deal with the problem of robust estimation in a Bayesian context, providing a useful result to generalize univariate robust distributions to the multivariate case. We also propose criteria to assess when a robust model is recommended and how to choose among estimates obtained with different distributions. Finally, the third chapter considers the Mean-Variance portfolio selection problem. We give evidence that when data are normally distributed the Bayesian approach works better than the Certainty Equivalence approach, but this is no longer true when data contain few outlying observations . Moreover, we computed useful measures of sensitivity of Bayesian weights and we construct a new estimator which is robust to the presence of 'extreme' observations.

Chapter 1

Local robustness measures for posterior summaries

1.1 Abstract

This paper deals with measures of local robustness for particular Bayesian quantities, i.e. posterior summaries. We build a framework where any Bayesian quantity can be seen as a posterior functional and its sensitivity to all inputs is checked. First, we use the Gateaux derivatives to measure the impact on posterior summaries of perturbations of prior or sampling models, giving some general expressions. Such quantities capture both a 'data effect' and a 'model effect' on the functional. Secondly, we check the sensitivity to one observation in the sample, once a particular combination of prior/sampling models has been chosen. Moreover, we propose a new estimator of the Bayes factor for efficient implementation. Finally, illustrative examples of sensitivity analyses are provided and discussed.

1.2 Introduction

Any Bayesian quantity depends strongly on the modeling assumptions and on the sample of observed data. Bayesian Robust Statistics evaluates the sensitivity of this quantity to their inputs and in recent years there has been a growing literature in the field (D. Rìos Insua and F. Ruggeri, 2000). Most efforts concentrate on global robustness, in particular with respect to prior specification. Such approach consists in calculating the range of the quantity of interest as the model varies within a class of distributions. If this range is small, the quantity is declared to be robust. If not, further analysis is needed. For more details on this issue see Lavine (1991), Berger (1994), Basu (1999), Sivaganesan (1999, 2000), Berger *et al.* (2000), Moreno (2000) and Shyamalkumar (2000).

A second approach - named local - assesses the sensitivity to deviations only in a neighborhood of the reference model. Measures of local robustness are obtained by suitable derivatives of the functional (Ruggeri and Wasserman, 1993; Sivaganesan, 1993; Dey *et al.,* 1996; Gustafson *et al.,* 1996; Moreno *et al.,* 1996; Peña and Zamar, 1997). The functional is said to be robust if the measure is small. Also in this case, most contributions are only concerned with local prior influence (Gustafson, 2000).

In this paper we deal with local robustness. It is interesting to note that the same approach is used in robust statistics as developed in the frequentist framework (Huber, 1981; Hampel *et al.*, 1986). However the robustness perspective slightly differs in a frequentist and in a Bayesian context. We discuss this point in Section 1.3, introducing the concept of functional and looking at any Bayesian quantity as a function of three distinct elements, i.e. the prior, the sampling model and the data. Such point of view constitutes a simple and unified framework for robustness evaluation in Bayesian statistics. In particular we consider the posterior expectation of a generic function $\rho(\theta)$, called posterior summary. The goal of this paper is to check the sensitivity of posterior summaries to a given input, all the rest remaining fixed. Different diagnostic tools for distributional assumptions -called local influence measures- are derived in Section 1.4. Such measures capture the impact on the functional of contaminations of the reference model in different

directions. The sensitivity of a Bayesian functional to observations is addressed in Section 1.5. Section 1.6 deals with the matter of implementation of local influence measures when analytical calculations are not feasible. Starting from the work of Chen and Shao (1997), we propose a new estimator for the Bayes Factor which is more efficient in terms of computational time. Illustrative examples are given in Section 1.7. Finally, Section 1.8 gives a summary of the findings and Section 1.9 suggests possible directions for future research.

1.3 Frequentist and Bayesian robustness

In this section we underline some common and different features of the robustness concept in a Bayesian and in a frequentist framework.

First let us introduce some notation. We will use capital letters for both a probability distribution and its corresponding cumulative distribution function. Moreover, we denote with small letters the corresponding density, when it exists. We consider i.i.d. one-dimensional random variables $X = (X_1, .., X_n)$ generated by a reference distribution F_{θ_0}, which belongs to the set $F^* = \{F_\theta : \theta \in \Theta\}$. Each observation in sample $x = (x_1, .., x_n)$ takes value in a sample space $\Xi \subseteq \mathbb{R}$.

We denote by $F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{x_i}(y)$ the empirical distribution where $\Delta_x(y)$ is the Dirac distribution which puts mass 1 at x. In a Bayesian setting we also define $\Pi(\theta)$ and $P(\theta|x)$ to be an element respectively of the set Π^* of all possible priors and of the set P^* of all possible posteriors on the parameter space Θ.

In frequentist statistics observed data are used to make inference on the true parameter value θ_0, which is assumed to be a fixed constant (Cox and Hinkley, 1974). The approach of classical robust theory based on influence functions (Hampel, 1974; Hampel *et al.*, 1986) deals with estimators that can be (at least asymptotically) expressed as functionals, i.e. $T_n(F_n) = T(F_n)$ for all n and F_n. Such functional $T : F^* \to \mathbb{R}^k$ is such that it converges to the asymptotic value of the estimator $(T(F_n) \underset{n\to\infty}{\longrightarrow} T(F_{\theta_0}))$ and that Fisher consistency holds $(T(F_{\theta_0}) = \theta_0)$.

Measures of robustness to small deviations from the reference model are obtained by computing the influence function (IF), which is the Gateaux derivative of the functional under a locally perturbed distribution in direction of a point mass. Therefore the evaluation of robustness properties of the estimator occurs at an asymptotic level. In the sample one can calculate some empirical version of the IF such as the Empirical Influence Function and the Sensitivity Curve.

In Bayesian statistics the parameter θ is not a fixed quantity, but a random variable, whose entire probability distribution has to be computed (Bernardo and Smith, 1994). Two distributions are matched with the observed data: Π that represents our knowledge *a priori* on θ and F_θ that expresses the parametric model we believe generated observations x. Using the Bayes theorem, the posterior distribution for parameter θ is obtained:

$$\begin{aligned} P(\theta|x) &= \frac{\Pi(\theta) L_F(x|\theta)}{m(x;\Pi,F_\theta)} \\ &= \frac{\widetilde{P}(\theta|x)}{m(x;\Pi,F_\theta)}, \end{aligned} \tag{1.1}$$

where $L_F(x|\theta) = \prod_i f_\theta(x_i)$ is the likelihood and $m(x;\Pi,F_\theta) = \int \widetilde{p}(\theta|x)\, d\theta$ is the marginal likelihood. Inferential conclusions on the value of θ are based on (1.1).

Any Bayesian quantity can be expressed as a functional of type

$$T_B : F_n^* \times \Pi^* \times F^* \to \Upsilon,$$

where $F_n^* = \{$all discrete distributions with probability $p_1, .., p_n$ at the points $x_1, .., x_n$, $p_i \geqslant 0$, $\sum_i p_i = 1\}$ and Υ is a suitable space. For example, one can be interested in the entire posterior distribution ($\Upsilon = P^*$) or in some posterior summaries ($\Upsilon = \mathbb{R}^k, k \geqslant 1$).

When the number of observations increases, the impact of Π on (1.1) disappears because the likelihood dominates the prior distribution and the posterior collapses to a point mass on the true parameter value θ_0. Therefore, Bayesian functionals satisfy $T_B(F_n, \Pi, F_\theta) \underset{n\to\infty}{\longrightarrow} T(F_{\theta_0})$. Asymptotic functionals do not allow to capture the sensitivity of posterior quantities to perturbations in the prior. Hence, we will work with sample-based functionals. In particular we will focus on

robustness evaluation for posterior summaries of type

$$T_B(F_n, \Pi, F_\theta) = \int \rho(\theta)\, p(\theta|x)\, d\theta. \tag{1.2}$$

In the sequel we will in short denote T_B and $m(x)$ respectively the posterior summary and the marginal likelihood under reference models Π and F_θ.

1.4 Sensitivity to distributional assumptions

In this section we deal with the sensitivity of a Bayesian estimator to small departures from the assumed model, either the prior or the sampling distribution. In order to simplify the notation we will denote the posterior functional only as a function of the distribution under study, say a distribution H, keeping the remainder fixed. We represent these deviations through $\varepsilon-$contamination classes of type:

$$I_\varepsilon(H) = \{H_\varepsilon = (1-\varepsilon)\, H + \varepsilon C \mid 0 \leq \varepsilon \leq 1, C \in C^*\}. \tag{1.3}$$

Set (1.3) represents the perturbation of the reference distribution H in the direction of C and ε is the contamination's amount (assumed to be small in local analysis). Clearly, the wider the set of contaminating distribution C^* is, the richer the neighborhood we are considering. As in Sivaganesan (1993) and Peña and Zamar (1997), we measure the impact of such contaminations on functional (1.2) by the Gateaux derivative:

$$\begin{aligned} LI(C; T_B, H) &= \left[\frac{\partial T_B(H_\varepsilon)}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &= \int \rho(\theta) \left[\frac{\partial p_\varepsilon(\theta|x)}{\partial \varepsilon}\right]_{\varepsilon=0} d\theta. \end{aligned} \tag{1.4}$$

We refer to this quantity as *local influence* (LI) of T_B when H is perturbed in the direction of C. Note that measure (1.4) is a sample-based quantity. We will see that it captures both a 'data effect', i.e. the effect on the functional of choosing a contaminating model which is more adequate than the reference one with respect to observed data, and a 'model effect', i.e. the effect on the functional value of perturbing the reference model in some directions. The strong dependence of measure

(1.4) on the sample is the reason why Sivaganesan (1993) looks at it only to compare whether a functional is more sensible to prior or sampling model specifications and does not evaluate its magnitude. For this purpose we define

$$LI^*(C^*; T_B, H) = \sup_{C \in C^*} \left| \frac{LI\,(C; T_B, H)}{T_B\,(H)} \right|, \tag{1.5}$$

which gives the maximum relative effect on the functional as the distribution moves locally around H in different directions. Measure (1.5) evaluates the magnitude of the sensitivity of the functional and can be used to compare robustness properties among different functionals. In the following sections we derive local influence measures for both the prior and the sampling model.

1.4.1 Prior distribution

Many papers in Bayesian robustness are concerned with the assessment of the sensitivity with respect to the prior (Ruggeri and Wasserman, 1993; Gustafson *et al.,* 1996; Moreno *et al.,* 1996; Peña and Zamar, 1997). The main reason for this widespread interest is probably due to the feeling that prior knowledge formalized by the researcher is the most subjective source of the analysis. Much work has been done in the direction of global robustness. A good review on the topic is provided by Berger (1994).

Local robustness assesses effects of small prior perturbations on the functional. We consider a neighborhood of the reference prior Π of type (1.3), with Q the contaminating distribution. The local influence of T_B when Π is perturbed in the direction of Q is given by:

$$\begin{aligned}
LI(Q; T_B, \Pi) &= \left[\frac{\partial T_B(\Pi_\varepsilon)}{\partial \varepsilon} \right]_{\varepsilon=0} \\
&= \left[\int \rho\,(\theta)\, \frac{\partial}{\partial \varepsilon} \left(\frac{L_F\,(x|\theta)\,\pi_\varepsilon(\theta)}{m\,(x\,; \Pi_\varepsilon, F_\theta)} \right) d\theta \right]_{\varepsilon=0} \\
&= \int \rho\,(\theta)\, \frac{L_F\,(x|\theta)\,[q\,(\theta) - \pi(\theta)]}{m\,(x)} d\theta + \\
&\quad + \int \rho\,(\theta)\, \frac{[m\,(x\,; Q, F_\theta) - m\,(x\,)]\, L_F\,(x|\theta)\,\pi(\theta)}{m\,(x\,)} d\theta \\
&= \frac{m(x; Q, F_\theta)}{m(x)}\, [T_B(Q) - T_B]\,,
\end{aligned} \tag{1.6}$$

where $m(x; Q, F_\theta)$ and $T_B(Q)$ are respectively the marginal likelihood and the posterior summary obtained when the prior is Q. Measure (1.6) depends on two factors. The first is the ratio of marginal likelihoods under contaminating and reference distribution respectively (Bayes factor). This can be regarded as a measure of data supporting degree for different contaminating priors that compares the researcher's subjectivity and the objectiveness of the data. If this amount is greater (smaller) than one, data may be said to support more (less) the contaminating prior then the reference one. For this reason the Bayes factor can be said to capture a 'data effect' on the functional. The second factor is the difference between the functional value computed under the contaminating and the reference prior respectively. It captures the effect on the functional when choosing a different model for the prior and we refer to this as 'model effect'. For example, if the value of $T_B(Q)$ is very different from the value of T_B, the model effect turns out to be large. However the total effect on the functional will be large itself only if model Q is not completely discarded by the data, i.e. the Bayes factor does not go to zero.

1.4.2 Sampling distribution

Another source of possible misspecification is the data-generating model. Robustness with respect to sampling model specification is referred in the literature as *model or likelihood robustness.* In most scenarios inference will depend much more heavily on the model than on the prior (see Section 1.3). However, few contributions in assessing likelihood robustness can be found in the literature (see Sivaganesan, 1993; Dey et *al.*, 1996; Gustafson, 1996; Shyamalkumar, 2000).

This fact can be explained by considering the non linearity of the posterior with respect to the sampling distribution. Indeed when regarded as a function of the prior, (1.1) is a ratio of two linear functionals, or briefly is said to be *ratio-linear*. This is not true when considered as a function of the sampling model, as the sampling density enters through the likelihood function. This often leads to intractable global analysis from an analytical point of view. However, in local analysis this problem can be tackled by taking the derivative with respect to the quantity of contamination

ε when ε is small.

Assume we represent uncertainty about the reference sampling model F_θ by (1.3) with G the contaminating distribution. The obtained perturbed likelihood will be differently combined with the prior according to the information G brings on θ.

If G is a distribution still depending on parameter θ, we denote the contaminating distribution by G_θ. For example G_θ can be an unimodal distribution around θ. In this case the local influence of T_B when F_θ is perturbed in the direction of G_θ is given by

$$\begin{aligned} LI(G_\theta; T_B, F_\theta) &= \left[\frac{\partial T_B(F_{\theta,\varepsilon})}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &= \left[\int \rho(\theta) \frac{\partial}{\partial \varepsilon}\left(\frac{L_{F_\varepsilon}(x|\theta)\cdot\pi(\theta)}{m(x\,;\Pi, F_{\theta,\varepsilon})}\right) d\theta\right]_{\varepsilon=0} \\ &= \sum_j \frac{m_j(x;\Pi,F_\theta,G_\theta)}{m(x)}\left[T_{B,j}(F_\theta,G_\theta) - T_B\right], \end{aligned} \tag{1.7}$$

where

$$m_j(x;\Pi,F_\theta,G_\theta) = \int \widetilde{p_j}(\theta|x)\, d\theta$$

and

$$T_{B,j}(F_\theta;G_\theta) = \frac{\int \rho(\theta)\,\widetilde{p_j}(\theta|x)\, d\theta}{m_j(x;\Pi,F_\theta,G_\theta)}$$

are respectively the marginal likelihood and the posterior functional obtained when the sampling distribution is G_θ only for observation x_j and F_θ for the others, the quantity $\widetilde{p_j}$ is defined as

$$\widetilde{p_j}(\theta|x) = g_\theta(x_j)\, L_F\left(x_{(-j)}|\theta\right)\pi(\theta),$$

and $x_{(-j)}$ is the sample x without observation x_j.

If G does not depend on θ we denote the contaminating distribution by V. The local influence of T_B when F_θ is perturbed in the direction of V is then given by:

$$\begin{aligned} LI(V; T_B, F_\theta) &= \left[\frac{\partial T_B(F_{\theta,\varepsilon})}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &= \left[\int \rho(\theta) \frac{\partial}{\partial \varepsilon}\left(\frac{L_{F\varepsilon}(x|\theta)\cdot\pi(\theta)}{m(x\,;\Pi, F_{\theta,\varepsilon})}\right) d\theta\right]_{\varepsilon=0} \\ &= \sum_j \frac{m_j(x;\Pi,F_\theta,V)}{m(x)}\left[T_B^{(-j)} - T_B\right], \end{aligned} \tag{1.8}$$

where $m_j(x;\Pi,F_\theta,V) = v\,(x_j)\cdot m(x_{(-j)};\Pi,F_\theta)$ and $m(x_{(-j)};\Pi,F_\theta)$ and $T_B^{(-j)}$ are respectively the marginal likelihood and the posterior functional under reference models using sample $x_{(-j)}$. For detailed calculations see Appendix A.

For any observation x_j the local influence measure for the sampling distribution is still a function of two factors and it captures both a 'data effect' and a 'model effect'. The Bayes factor plays the important role of increasing (decreasing) the difference when data support (do not support) the contaminating distribution more than the reference distribution for observation j ('data effect'). The second factor is the difference between the value of the functional computed when model G is assumed only for observation x_j and the base functional T_B. Note that observation x_j enters in the calculation of the former value only if G depends on θ. Otherwise, x_j cannot give any information for updating our prior knowledge and the resulting functional has the form of the reference one where one observation has been dropped out. The total effect on the functional of perturbations of the sampling model turns out to be the sum of the effect for each observation.

1.5 Sensitivity to observations

In the previous section we assessed the influence on posterior summaries of a perturbation of the assumed model in some direction. In this section we measure the influence of a given observation in the sample (*outlier robustness*). It is worth stressing the difference between model robustness and outlier robustness. Model robustness evaluates the impact on the functional of a small contamination of the reference sampling model (see section 1.4.2). Outlier robustness evaluates the effect of moving one observation in the sample once prior and sampling distributions are fixed. In this section we still denote the Bayesian functional as a function of the distribution under study, i.e. the empirical distribution.

Little attention has been paid in Bayesian literature to the impact of outliers and mainly focused on the posterior distribution. Ramsay and Novick (1980), for example, propose to look at the rate of change of the sampling model density with respect to an observation value. A similar idea is

used by West (1984) on Bayesian regression. However such approach is hardly applicable because it involves derivatives which are difficult to compute apart from particular family of distributions. The same problem is addressed by Chen and Fournier (1999). Their influence measure summarizes the difference between posterior distributions computed with original data and with an additional observation. Such posterior distributions are obtained through the use of numerical techniques and are therefore always applicable.

In this paper, however, we do not deal with posterior distributions directly, but with posterior summaries. Studying the sensitivity of such a quantity to observations is a well known matter in frequentist robust statistics. It is done by means of the Sensitivity Curve (see Hampel *et al.*, 1986), defined as

$$SC(z) = \frac{[T_B(F_n^z) - T_B(F_{n-1})]}{\frac{1}{n}}, \tag{1.9}$$

where $F_{n-1} = (x_1, .., x_{n-1})$ is the empirical distribution of the sample of $(n-1)$ observations and $F_n^z = (x_1, .., x_{n-1}, z)$ is the sample in which observation z has been added. In a Bayesian context this measure captures the influence of moving just one observation under a certain prior/sampling model combination. If this measure diverges as z becomes larger, the functional is said to be non robust with respect to observations. Typically this curve is useful to identify observations with a large influence, such as outliers and loosely speaking an outlier is defined to be an observation that is unlikely to have been generated by the assumed sampling model. For its simple definition (1.9) can be implemented even when analytical calculations are not feasible by means of numerical algorithms.

In the next section we will discuss the practical implementation of local sensitivity measures derived in the previous sections when analytical results are not available.

1.6 Implementation of local sensitivity measures

Posterior distribution and local influence measures are analytically tractable when conjugate prior and sampling models are assumed. However, often this is not the case and we need to use numerical

procedures to compute them. Typically MCMC algorithms are used to generate a sample from complicated distributions. Local influence measures can be then easily obtained by estimating the Bayes factor and the functionals under reference and contaminating distributions. In this section we concentrate on implementation of (1.7) by means of Metropolis-Hastings algorithm and we propose a way to speed up its computation.

Local influence measures for the sampling distribution involve the computation of Bayes factors and of posterior summaries (see Section 1.4.2). We first consider the estimation of the former quantity (shortly denoted by r_j), which is given by

$$\begin{aligned} r_j &= \frac{m_j(x;\Pi,F_\theta,G)}{m(x)} \\ &= \frac{\int \widetilde{p}_j\left(\theta|x\right)d\theta}{\int \widetilde{p}\left(\theta|x\right)d\theta}. \end{aligned} \tag{1.10}$$

Different bridge estimators (Meng and Wong, 1996; Chib and Jeliazkov, 2001; Mira and Nicholls, 2001) are available. However, to compute such local influence measures we would have to run $n+1$ simulations, where n is the number of observations. Clearly, the estimation procedure will take a long time when n is large.

We need a way to be more efficient in terms of computational time. A good starting point is the two-stage estimator proposed by Chen and Shao (1997). Ratio (1.10) can be written as

$$r_j = \frac{\int \frac{\widetilde{p}_j(\theta|x)}{\xi(\theta)}\xi\left(\theta\right)d\theta}{\int \frac{\widetilde{p}(\theta|x)}{\xi(\theta)}\xi\left(\theta\right)d\theta}, \tag{1.11}$$

where $\xi\left(\theta\right)$ is an arbitrary importance sampling density. When observations are i.i.d. from ξ, the importance density which minimizes the relative mean square error of the estimator is given by

$$\begin{aligned} \xi_j^{opt}\left(\theta\right) &= \frac{\left|p_j\left(\theta|x\right)-p\left(\theta|x\right)\right|}{\int\left|p_j\left(\theta|x\right)-p\left(\theta|x\right)\right|d\theta} \\ &= \frac{\left|\widetilde{p}_j\left(\theta|x\right)-r_j\cdot\widetilde{p}\left(\theta|x\right)\right|}{\int\left|\widetilde{p}_j\left(\theta|x\right)-r_j\cdot\widetilde{p}\left(\theta|x\right)\right|d\theta}, \end{aligned} \tag{1.12}$$

where $p_j=\widetilde{p}_j/m_j$ and $p=\widetilde{p}/m$.

The corresponding estimator $\widehat{r}_j^{opt}$ is implemented in two stages. First, a Monte Carlo estimate of (1.11) is computed with a random sample from an arbitrary distribution. Then a random draw

from (1.12) can be obtained by means of a MCMC simulation. One advantage of $\widehat{r}_j^{opt}$ is that its estimate is available with a single random sample from ξ_j^{opt} rather than two samples respectively from p_j and p. However, we are still expected to generate n samples to compute (1.7).

In order to run a single MCMC simulation we propose to use an importance sampling density with a form similar to the optimal one, but which does not depend on j. Such a density is given by

$$\xi^*(\theta) = \frac{|\widetilde{p}^*(\theta|x) - r^* \cdot \widetilde{p}(\theta|x)|}{\int |\widetilde{p}^*(\theta|x) - r^* \cdot \widetilde{p}(\theta|x)|\, d\theta}, \tag{1.13}$$

where $\widetilde{p}^*(\theta|x) = \frac{1}{n}\sum_{j=1}^{n} \widetilde{p}_j(\theta|x)$ and $r^* = \frac{\int \widetilde{p}^*(\theta|x)d\theta}{\int \widetilde{p}(\theta|x)d\theta}$. Figure 1.1 compares density (1.13) with the posterior densities p and $p_j's$.

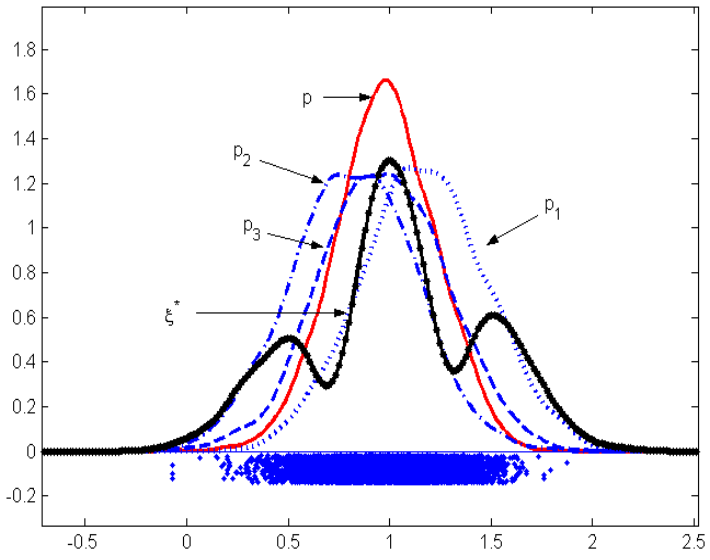

Figure 1.1: Importance sampling densities ξ^* and posterior densities p and p_j's.

The sampling density displays fatter tails which is a crucial characteristic for a good importance sampling. The corresponding modified two-stages estimator is given by

$$\widehat{r}_j^* = \frac{\sum_{i=1}^{n_\xi} \frac{\widetilde{p}_j(\theta_i|x)}{\xi^*(\theta_i|x)}}{\sum_{i=1}^{n_\xi} \frac{\widetilde{p}(\theta_i|x)}{\xi^*(\theta_i|x)}}, \tag{1.14}$$

where $[\theta_i]_{i=1}^{n_\varepsilon}$ is the output of a MCMC simulation for (1.13). We tested the performance of the new estimator by running $K = 30$ independent simulations of length s $(s = 1000, 2000, .., 5000)$ under the normal sampling model. For each chain we estimate (1.14) and we compute its mean value with the corresponding confidence interval. Figure 1.2 shows that estimator (1.14) behaves well with a mean value of $\widehat{r}_j^*$ close to the analytical value and smaller variability with increasing number of simulations.

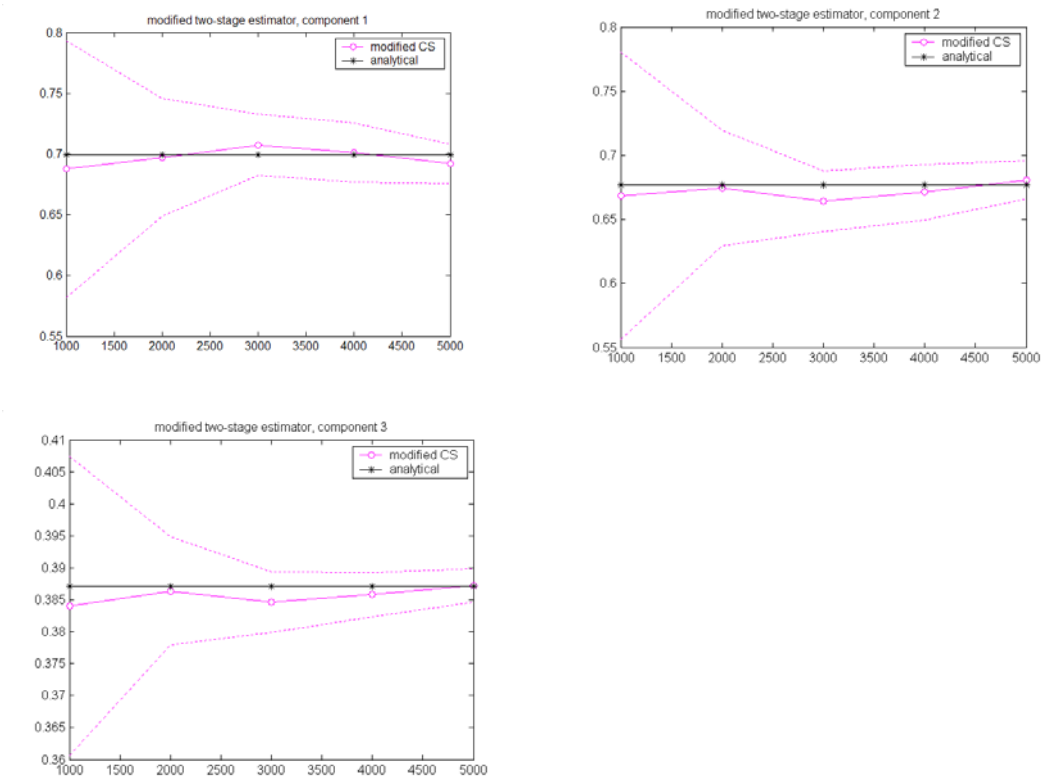

Figure 1.2: Analytical and estimated value of r_j $(j = 1, 2, 3)$ with confidence intervals.

To estimate the local influence measure for the sampling distribution, we still need to compute T_B and $T_{B,j}(F_\theta, G)$. The former quantity can be obtained by running a MCMC simulation for posterior p. The latter can be obtained using importance sampling technique with different sampling

densities. If ξ^* is chosen as importance density, measure (1.7) can be written as

$$\begin{aligned} LI(G;T_B,F_\theta) &= \sum_j \frac{m_j(x;\Pi,F_\theta,G)}{m(x)}\left[T_{B,j}\left(F_\theta,G\right)-T_B\right] \\ &= \sum_j r_j\left[\int \rho\left(\theta\right)p_j\left(\theta|x\right)d\theta-\int \rho\left(\theta\right)p\left(\theta|x\right)d\theta\right] \\ &= \sum_j r_j\cdot\left[\frac{m_\xi}{m_j}\cdot\int \rho\left(\theta\right)\frac{\widetilde{p}_j\left(\theta|x\right)}{\widetilde{\xi}^*\left(\theta|x\right)}\xi^*\left(\theta|x\right)d\theta-\int \rho\left(\theta\right)p\left(\theta|x\right)d\theta\right] \\ = \sum_j\left[r_\xi\cdot\int \rho\left(\theta\right)\frac{\widetilde{p}_j\left(\theta|x\right)}{\widetilde{\xi}^*\left(\theta|x\right)}\xi^*\left(\theta|x\right)d\theta-r_j\cdot\int \rho\left(\theta\right)p\left(\theta|x\right)d\theta\right], & \qquad (1.15) \end{aligned}$$

where $r_\xi=\frac{m_\xi}{m}$. Denoting by $[\theta_s]_{s=1}^{n_p}$ and $[\theta_i]_{i=1}^{n_\xi}$ respectively the samples from $p\left(\theta|x\right)$ and from $\xi^*\left(\theta\right)$, the ratio r_ξ can be estimated using optimal Meng and Wong's bridge estimator given by

$$\widehat{r}_\xi^{t+1}=\frac{\frac{1}{n_p}\sum_{s=1}^{n_p}\frac{\widetilde{\xi}^*(\theta_s)}{n_\xi\cdot\widetilde{\xi}^*(\theta_s)+n_p\cdot\widehat{r}_\xi^t\cdot\widetilde{p}(\theta_s)}}{\frac{1}{n_\xi}\sum_{i=1}^{n_\xi}\frac{\widetilde{p}(\theta_i)}{n_\xi\cdot\widetilde{\xi}^*(\theta_i)+n_p\cdot\widehat{r}_\xi^t\cdot\widetilde{p}(\theta_i)}}.$$

An estimator of (1.15) is then obtained as

$$\widehat{LI}(G;T_B,F_\theta)=\sum_{j=1}^{n}\left[\widehat{r}_\xi\left(\frac{1}{n_\xi}\sum_{i=1}^{n_\xi}\rho\left(\theta_i\right)\frac{\widetilde{p}_j\left(\theta_i|x\right)}{\widetilde{\xi}^*\left(\theta_i|x\right)}\right)-\widehat{r}_j^*\left(\frac{1}{n_p}\sum_{s=1}^{n_p}\rho\left(\theta_s\right)\right)\right]. \qquad (1.16)$$

If p is chosen as importance density, measure (1.7) can be written as

$$\begin{aligned} LI(G;T_B,F_\theta) &= \sum_j \frac{m_j(x;\Pi,F_\theta,G)}{m(x)}\left[T_{B,j}\left(F_\theta,G\right)-T_B\right] \\ &= \sum_j r_j\cdot\left[\frac{1}{r_j}\cdot\int \rho\left(\theta\right)\frac{\widetilde{p}_j\left(\theta|x\right)}{\widetilde{p}\left(\theta|x\right)}p\left(\theta|x\right)d\theta-\int \rho\left(\theta\right)p\left(\theta|x\right)d\theta\right] \end{aligned}$$

and its estimator is given by

$$\widehat{LI}(G;T_B,F_\theta)=\sum_{j=1}^{n}\left[\left(\frac{1}{n_p}\sum_{s=1}^{n_p}\rho\left(\theta_i\right)\frac{\widetilde{p}_j\left(\theta_s|x\right)}{\widetilde{p}\left(\theta_s|x\right)}\right)-\widehat{r}_j^*\left(\frac{1}{n_p}\sum_{s=1}^{n_p}\rho\left(\theta_s\right)\right)\right]. \qquad (1.17)$$

In the next section we will provide some examples on how to perform a Bayesian sensitivity analysis.

1.7 Examples of local sensitivity analyses

In the following simple examples we perform sensitivity analyses of the functional of interest. We keep the same notation as in previous sections. We first consider the Bayes estimator given by the

mean of the posterior distribution. For this example we simulate a sample of $n = 3$ observations from a standard univariate normal given by $(0.5375, 1.4221, 1.0946)$. Then we consider a Bayesian regression model using real data. In both case we perform conjugate analyses in order to obtain analytical results.

1.7.1 Posterior mean

The posterior mean is a frequently used estimator of the parameter of interest. We now illustrate how a sensitivity analysis on this functional can be carried out. We assume that prior Π is $N(\theta_0, \sigma_0^2)$ with $\theta_0 = 0.5$ and $\sigma_0^2 = 1$. Moreover sampling distribution F_θ is $N\left(\theta, \sigma^2\right)$ with $\sigma^2 = 0.2$. The posterior mean and the marginal likelihood can be computed analytically and turn out to be respectively

$$T_B = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\overline{x} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\theta_0$$

and

$$\begin{aligned} m(x) &= (2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{(n-1)}{2}} \left(n\sigma_0^2 + \sigma^2\right)^{-\frac{1}{2}} \cdot \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_i (x_i - \overline{x})^2\right\} \exp\left\{-\frac{n\left(\theta_0 - \overline{x}\right)^2}{2\left(n\sigma_0^2 + \sigma^2\right)}\right\}. \end{aligned}$$

First, we assume we are not very confident about the value of prior mean θ_0. We express our uncertainty through the set of possible contaminating prior distribution $\widetilde{Q} = \left\{N(\lambda, \sigma_0^2) : \lambda \in [-4.5, 5.5]\right\}$. In this case the local influence measure is given by (1.6) with

$$T_B\left(Q\right) = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\overline{x} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\lambda$$

and

$$\begin{aligned} m(x; Q, F_\theta) &= (2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{(n-1)}{2}} \left(n\sigma_0^2 + \sigma^2\right)^{-\frac{1}{2}} \cdot \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_i \left(x_i - \overline{x}\right)^2\right\} \exp\left\{-\frac{n\left(\lambda - \overline{x}\right)^2}{2\left(n\sigma_0^2 + \sigma^2\right)}\right\}. \end{aligned}$$

Table 1.1 and Figure 1.3 show such a measure for different values of σ_0^2. The magnitude of LI decreases with increasing prior variances, meaning that flatter priors are less influenced by perturbations. The two factors of measure (1.6) are displayed in Figure 1.4.

	σ_0^2			
	0.5	**1**	10	100
T_B	0.9571	**0.9857**	1.0146	1.0177
$LI^*\left(\widetilde{Q};T_B,\Pi\right)$	0.1270	**0.0702**	0.0148	0.0029
λ for $LI^*\left(\widetilde{Q};T_B,\Pi\right)$	1.6	**1.8**	3.9	5.5
λ for max $m(x;Q,F_\theta)/m\left(x\right)$	1	**1**	1	1

Table 1.1: Relative local influence measures of the posterior mean with respect to the prior model with different prior precision.

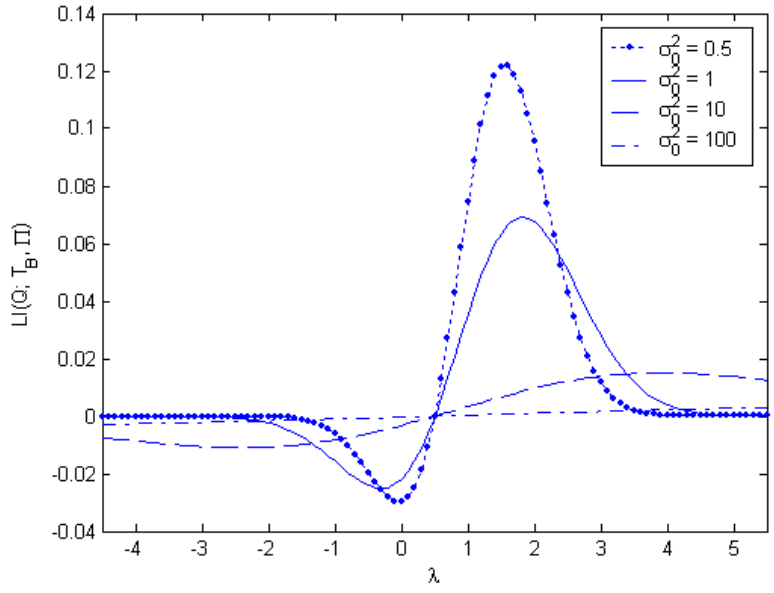

Figure 1.3: $LI(Q;T_B,\Pi)$ measure for the posterior mean with different values of prior variance σ_0^2.

The effect on the functional of choosing prior Q instead of prior Π ('model effect') is linear and smaller with decreasing prior precision. Moreover, priors with λ around the value of the sample mean ($\overline{x} = 1.01$) appear to be more adequate than Π for small value of σ_0^2. As long as the reference prior becomes flatter, the Bayes factor approaches to 1 for all possible contaminating distributions.

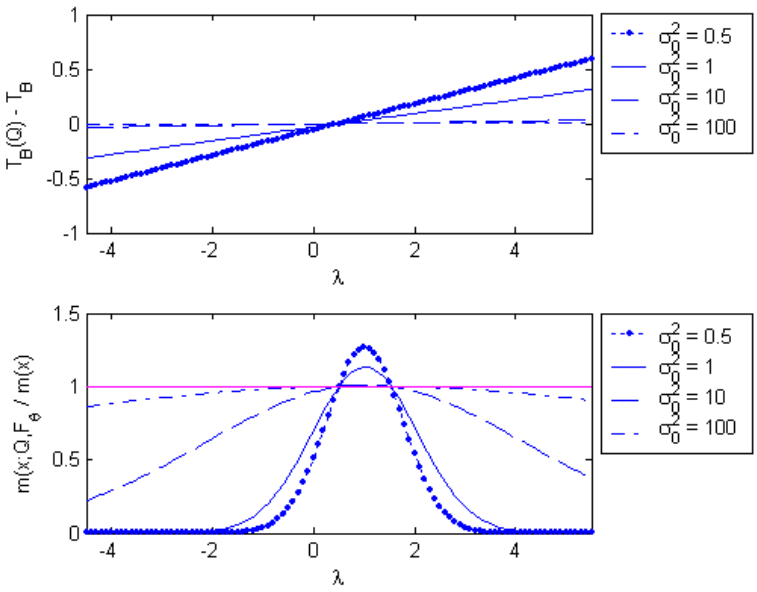

Figure 1.4: Difference $T_B(Q) - T_B$ and Bayes factor for different values of prior variance σ_0^2.

We turn now to the sampling model. We account for perturbations of the reference distribution in the direction of flatter ones. The chosen contaminating set is $\widetilde{G}_\theta = \left\{N(\theta, \eta^2) : \eta^2 \in [0.2, 2]\right\}$ .Clearly this contamination is quite restrictive, but it leads to analytical results. LI measure for the sampling model is given by (1.7) with

$$\begin{aligned} m_j(x; \Pi, F_\theta, G_\theta) &= (2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{(n-2)}{2}} \left(\sigma^2\eta^2 + (n-1)\,\eta^2\sigma_0^2 + \sigma^2\sigma_0^2\right)^{-\frac{1}{2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i\neq j}\left(x_i - \overline{x}_{(-j)}\right)^2 - \frac{\sigma^2\left(x_j - \theta_0\right)^2}{2\left(\sigma^2\eta^2 + (n-1)\,\eta^2\sigma_0^2 + \sigma^2\sigma_0^2\right)}\right\} \\ &\quad \cdot \exp\left\{-\frac{(n-1)\eta^2\left(\overline{x}_{(-j)} - \theta_0\right)^2 + (n-1)\,\sigma_0^2\left(\overline{x} - \overline{x}_{(-j)}\right)^2}{2\left(\sigma^2\eta^2 + (n-1)\,\eta^2\sigma_0^2 + \sigma^2\sigma_0^2\right)}\right\}, \end{aligned}$$

where $\overline{x}_{(-j)}$ is the mean of the sample without observation x_j. Calculations can be found in

Appendix B.

Table 1.2 and Figure 1.5 show measures (1.7) for different values of σ^2. LI measure is very small when $\sigma^2 = 0.2$, which corresponds to the value of the sample variance, and LI^* shows its minimum value which is around 0.009. As long as σ^2 moves away from 0.2, LI^* increases up to around 0.065.

	σ^2			
	0.1	**0.2**	1	4
T_B	1.0014	**0.9857**	0.8885	0.7220
$LI^*\left(\widetilde{G}_\theta; T_B, F_\theta\right)$	0.0650	**0.0096**	0.0544	0.0651
η^2 for $LI^*\left(\widetilde{G}_\theta; T_B, F_\theta\right)$	1.0	**0.6**	4.0	13.6

Table 1.2: Relative local influence measures of the posterior mean with respect to the sampling model with different sampling precision.

To better understand such a result, each row of Figure 1.6 plots the two factors of measure (1.7) for observation j ($j = 1, 2, 3$). The 'model effect' on the functional is increasing with increasing variance of the contaminating model, but it is no longer linear as in the prior case. When $\sigma^2 = 0.1$ or $\sigma^2 = 0.2$, data support at least few contaminating models more than the reference one. This is not true in other cases where the Bayes factor declines rapidly. Therefore the plot of the Bayes factor helps also to check whether the assumed sampling model is reasonable with respect to the data we have in the hand.

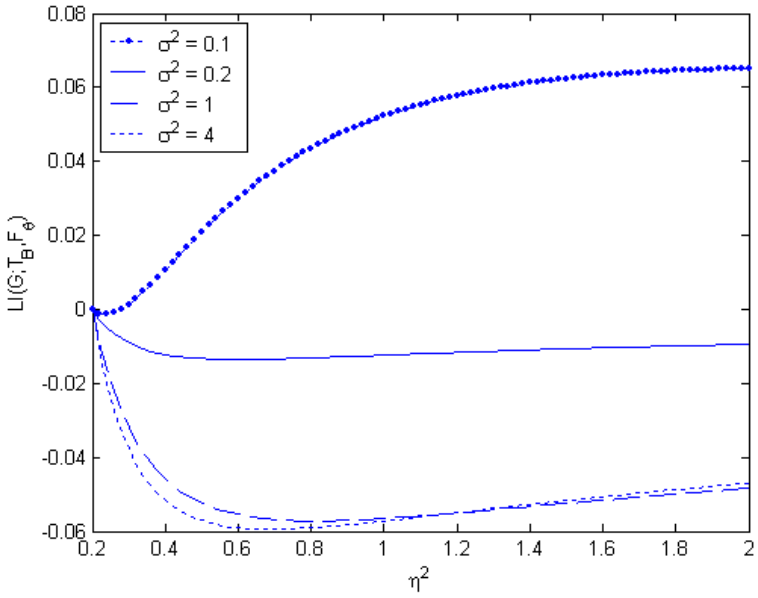

Figure 1.5: $LI\left(G;T_B,F_\theta\right)$ measure for the posterior mean with different values of sampling variance σ^2.

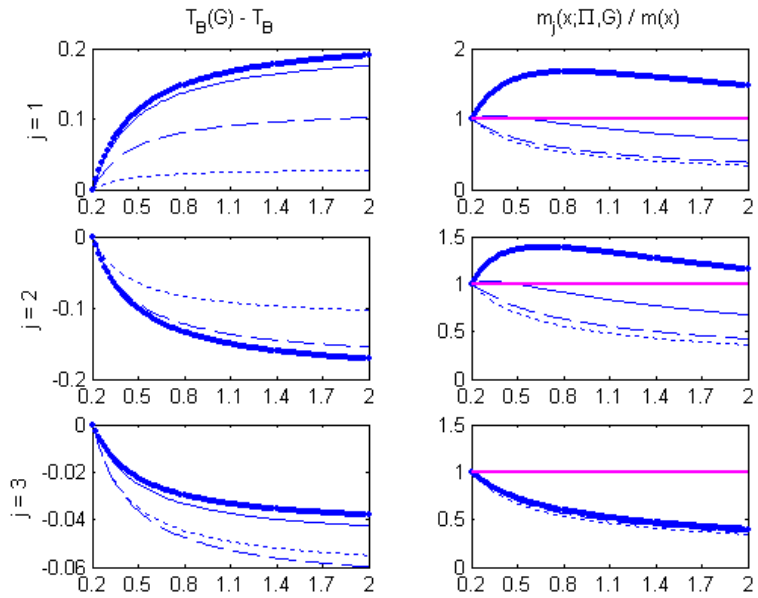

Figure 1.6: Difference $T_{B,j}\left(F_\theta,G\right)-T_B$ and ratio $m_j\left(x;\Pi,G\right)/m\left(x\right)$ for different values of sampling variance σ^2.

Comparing now the two bold columns in Table 1.1 and Table 1.2, we conclude that with these data the posterior mean is more sensible to perturbations in the prior model specification ($LI^*\left(\widetilde{Q};T_B,\Pi\right) = 0.0702 > LI^*\left(\widetilde{G}_\theta;T_B,F_\theta\right) = 0.0096$). However both measures are small and the estimate is evaluated locally robust with respect to our distributional assumptions.

Finally Figure 1.7 plots the $SC\,(z)$. We let observation z move in the range $[-5,5]$. The effect of an extreme observation on the posterior mean with a normal prior/normal sampling model combination is linear and therefore potentially unbounded.

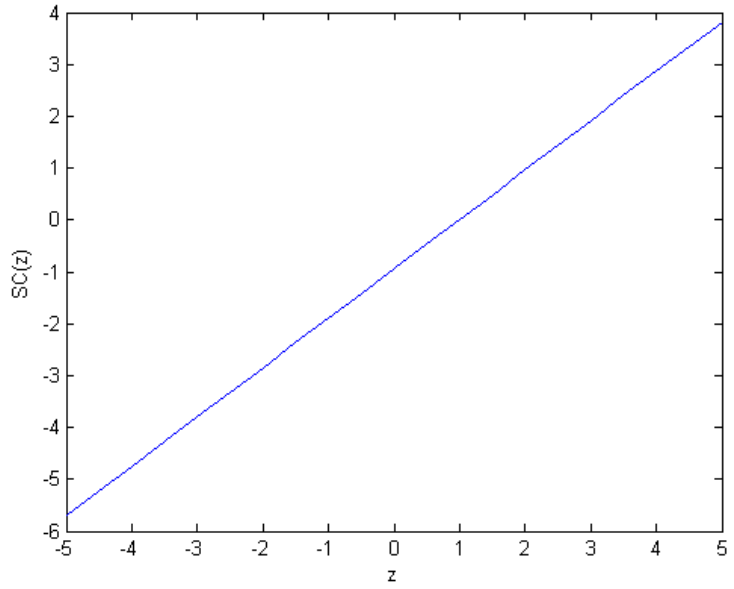

Figure 1.7: SC for the posterior mean under normality of both prior and sampling distributions.

Hence, it is crucial to assess whether some extreme observations are present in the sample. We expect that in such a case measure (1.7) increases since data would support sampling models with higher variance more than the reference one and model effect would also display a greater value. In order to investigate this point we introduce the observation $x_4 = -5$ in the sample and we compute LI measures again. Results given in Table 1.3 support our hypothesis. Therefore in presence of outliers measure (1.7) takes into account the fact that the normal distribution becomes

inadequate.

T_B	-0.4395
$LI^*\left(\widetilde{Q}; T_B, \Pi\right)$	0.2303
$LI^*\left(\widetilde{G}_\theta; T_B, F_\theta\right)$	$7.41 \cdot 10^{25}$

Table 1.3: Relative local influence measures of the posterior mean with respect to the base prior and sampling models. Contaminated sample.

1.7.2 Linear Bayesian Regression

We now consider the Bayesian linear model $y = X\beta + u$. For simplicity, we assume that the error distribution F is a $N\left(0, \sigma^2 I\right)$ with known variance σ^2. We further adopt a normal prior distribution $\Pi\left(\beta\right)$ of type $N\left(\beta_0, \sigma^2\Sigma_0\right)$. Under the assumed models, the Bayes estimator of β is given by

$$\widehat{\beta}_{Bayes} = \left(\Sigma_0^{-1} + X'X\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + X'y\right).$$

If $\widetilde{Q}$ is the family $\left\{N\left(\alpha_0, \sigma^2\Sigma_0\right) : \alpha_0^{\inf} \leq \alpha_0 \leq \alpha_0^{\sup}\right\}$ that accounts for uncertainty in the prior mean, measure (1.6) is given by

$$\begin{aligned} LI\left(Q; T_B, \Pi\right) &= \exp\left\{-\frac{\left(\alpha_0 - \beta_0\right)'\left[\Sigma_0^{-1} - \Sigma_0^{-1}V'\Sigma_0^{-1}\right]\left(\alpha_0 - \beta_0\right)}{2\sigma^2}\right\} \cdot \\ &\quad \cdot \left[\left(X'X + \Sigma_0^{-1}\right)^{-1}\Sigma_0^{-1}\left(\alpha_0 - \beta_0\right)\right]. \end{aligned} \tag{1.18}$$

Furthermore, assuming a contaminating family $\widetilde{G}$ for the sampling distribution of type $\left\{N\left(0, c^2 I\right) : c^{\inf} \leq c^2 \leq c^{\sup}\right\}$, measure (1.7) becomes

$$LI\left(G; T_B, F\right) = \sum_{j=1}^{n}\left(\frac{c^2\left|V\right|}{\sigma^2\left|V_j\right|}\right)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\left(\frac{\sigma^2}{c^2} - 1\right)y_j^2 + \widehat{\beta}'_{Bayes}V^{-1}\widehat{\beta}_{Bayes}}{2\sigma^2}\right\}.$$

$$\cdot \exp\left\{\frac{\widehat{\beta}^{(j)\prime}_{Bayes} V_j^{-1} \widehat{\beta}^{(j)}_{Bayes}}{2\sigma^2}\right\} \cdot \left(\widehat{\beta}^{(j)}_{Bayes} - \widehat{\beta}_{Bayes}\right), \tag{1.19}$$

where c^2 is the variance of the contaminating distribution, $V_j = \left[X'_{(-j)}X_{(-j)} \quad + \frac{\sigma^2}{c^2} x_j x'_j + \Sigma_0^{-1}\right]^{-1}$ and $\widehat{\beta}^{(j)}_{Bayes} = V_j \cdot \left(X'_{(-j)} y_{(-j)} + \frac{\sigma^2}{c^2} x_j y_j + \Sigma_0^{-1}\beta_0\right)$ are respectively the posterior variance and mean when distribution G is assumed only for observation j, x'_j is the row of matrix X corresponding to observation j, $X_{(-j)}$ and $y_{(-j)}$ are respectively matrix X and vector y without observation j. For detailed calculations see Appendix C.

Relative measures of local influence are given respectively by

$$LI^*\left(\widetilde{Q}; T_B; \Pi\right) = \sup_{Q \in \widetilde{Q}} \left| diag^{-1}\left(\widehat{\beta}_{Bayes}\right) \cdot LI\left(Q; T_B, \Pi\right)\right|$$

and

$$LI^*\left(\widetilde{G}; T_B; F\right) = \sup_{G \in \widetilde{G}} \left| diag^{-1}\left(\widehat{\beta}_{Bayes}\right) \cdot LI\left(G; T_B, F\right)\right|,$$

where $diag^{-1}\left(\widehat{\beta}_{Bayes}\right)$ is the inverse of the diagonal matrix with diagonal elements given by $\widehat{\beta}_{Bayes}$.

Bayesian estimation and local influence measures in the normal linear model are now illustrated. We use the same data set employed by Ramsay and Novick (1980). These are observations on 29 children on 3 psychological variables: a test of verbal intelligence (VI), a test of performance intelligence (PI) and $sin^{-1}\left(\sqrt{p_i}\right)$, where p_i is the proportion correct on a dichotic listening task (DL). We regress DL on remaining variables including a constant term. β_1 and β_2 are the coefficient corresponding to VI and PI respectively, whereas β_3 is the intercept. We also adopt the same values for both prior parameters and sampling variance which have been discussed at length by the authors. Analytical Bayes estimate of regression coefficients $\widehat{\beta}_{Bayes}$ equals $(0.7458, -0.0734, 38.3505)'$.

Plots of measure (1.18) and (1.19) are shown in Figure 1.8 and Figure 1.9. Each component of contaminating prior mean α_0 varies in the range $(-2, 2)$ with respect to the corresponding component of β_0. The impact on the Bayes estimate of contaminations in the prior is negligible. However, this is probably more a proof of the disappearing impact of the prior as the number of observations increases than a sign of robustness itself.

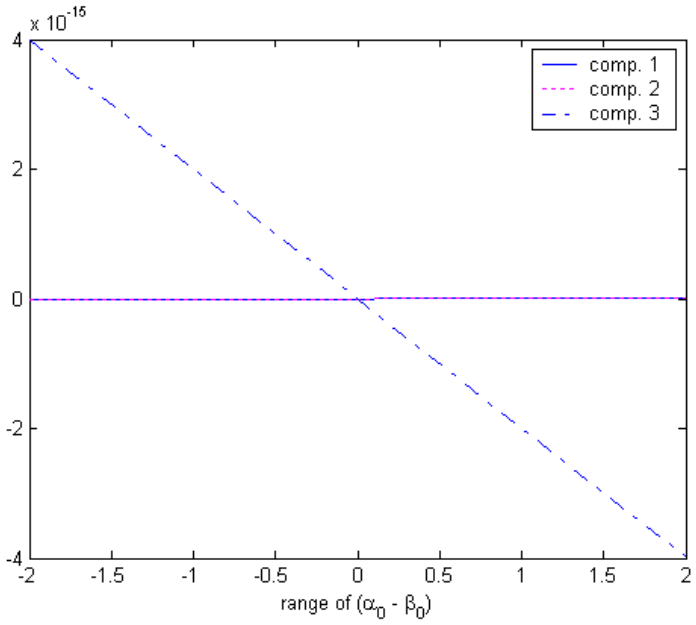

Figure 1.8: $LI\left(Q;T_B,\Pi\right)$ measure for regression coefficients.

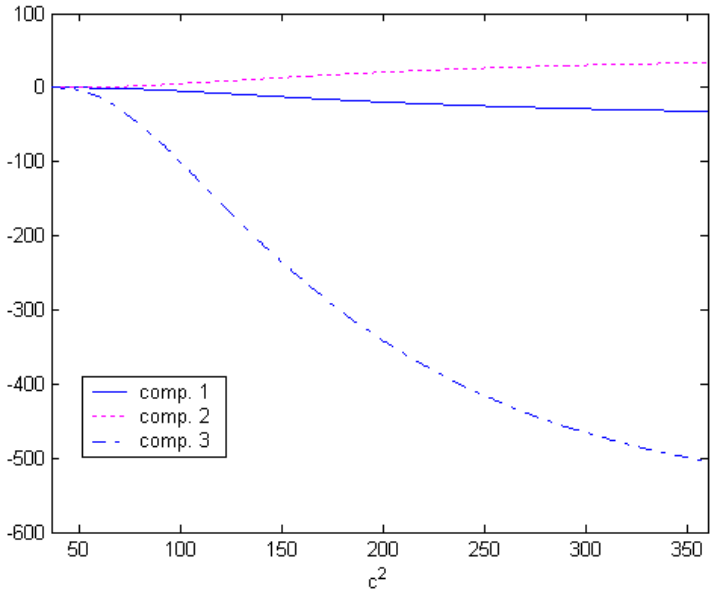

Figure 1.9: $LI\left(G;T_B,F\right)$ measure for regression coefficients.

Contaminating variance c^2 moves in the range $\left(\sigma^2, 10 \cdot \sigma^2\right)$. Perturbations of the sampling distribution play an important role on the estimates. The effect seems more pronounced for intercept β_3, but relative measures of Table 1.4 reveal a stronger impact for β_2. The size of LI^* measure for the sampling model is not negligible at all. Coefficient estimates turn out to be very sensitive to the assumption of a normal model for the data generating process.

component	1	2	3
$LI^*\left(\widetilde{Q}; T_B; \Pi\right)$	$2.2 \cdot 10^{-19}$	$2.9 \cdot 10^{-18}$	$1.0 \cdot 10^{-18}$
α_0 for $LI^*\left(\widetilde{Q}; T_B; \Pi\right)$	-1.69	-1.69	41
$LI^*\left(\widetilde{G}; T_B; F\right)$	42.93	458.18	13.14
c^2 for $LI^*\left(\widetilde{G}; T_B; F\right)$	360	360	360

Table 1.4: Relative local influence measures of regression coefficient estimates with respect to the base prior and sampling models.

We now concentrate on the sensitivity to observations. We move the value of the first two regressors in the range[1] $(65, 135)$ as represented by asterisks in Figure 1.10 and we look at the effect on the estimates. Figure 1.11 measures whether the added observation is an influential point through the Cook's distance. As the value moves away from the mean value of the regressors ($\overline{VI} = 99.75$ and $\overline{PI} = 104.89$), the added point becomes more and more influential. The same pattern is found in Figure 1.12 where the SC of β is displayed. Coefficient estimates are strongly dependent on the value of just one observation. In normal regression, hence, coefficients turn out to be so sensible that we do not necessary have to observe "extreme" value before estimates are influenced.

[1] This interval represents the theoretical values of the regressors.

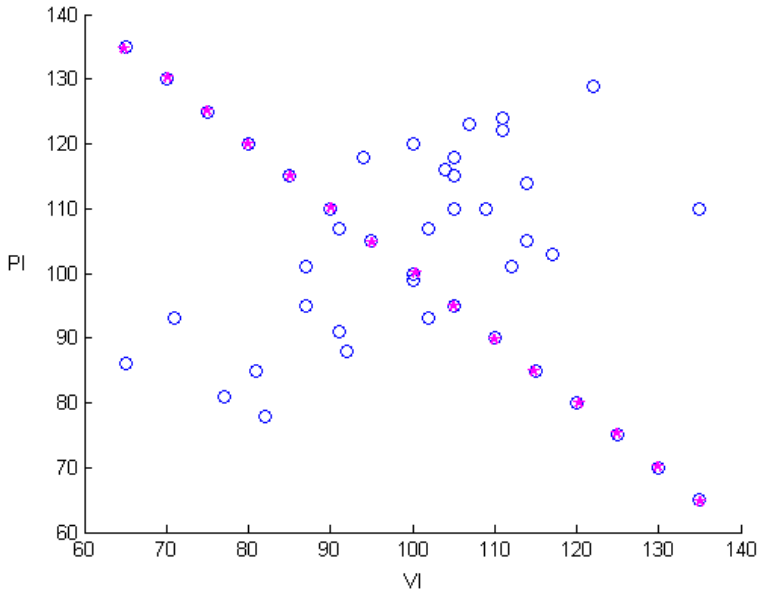

Figure 1.10: Scatterplot of VI towards PI. Asteriscs represent the observations which have been added.

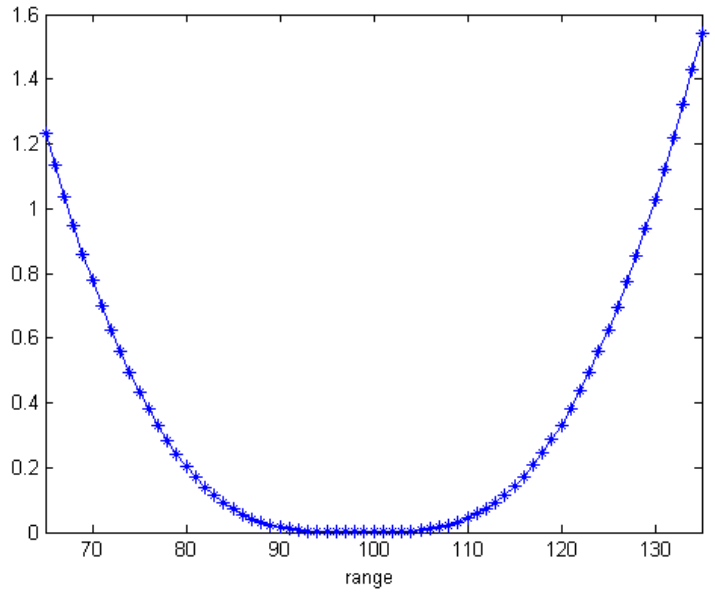

Figure 1.11: Cook's distance for observations which have been added.

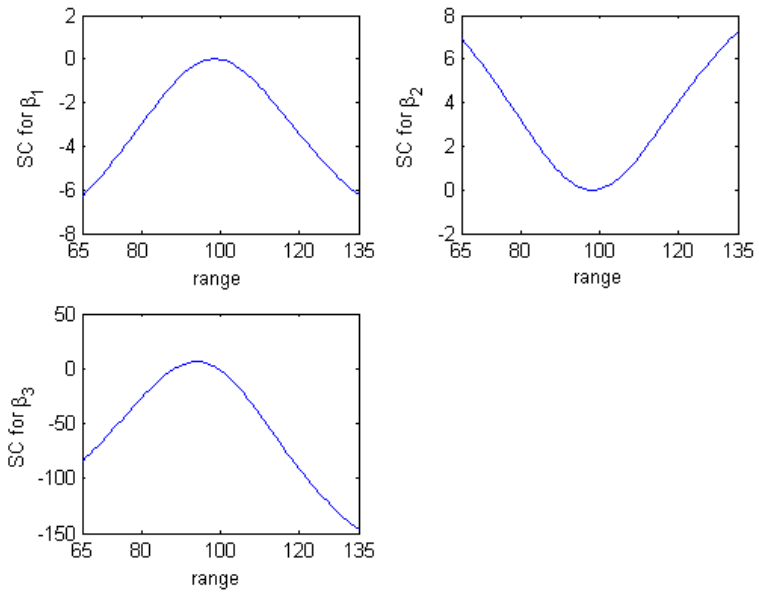

Figure 1.12: *SC* of regression coefficients moving the first two regressors in the range $(65, 135)$.

1.8 Summary

In this paper we construct a framework to perform the sensitivity analysis of any Bayesian quantity to all inputs. Bayesian robustness literature considers the sensitivity mainly to the prior distribution only. In our framework the sensitivity to all inputs is considered, giving a picture of the whole robustness properties of the functional itself. We concentrate on posterior summaries and we measure the impact of perturbations of prior or sampling models in different directions by means of local influence measures. Such impact is the product of two effects: a 'data effect', i.e. the effect on the functional of choosing a contaminating model which is more adequate than the reference one with respect to observed data, and a 'model effect', i.e. the effect on the functional value of perturbing the reference model in some directions. In some special cases we also derive analytical formulations for these quantities. Local influence measure for the prior model decreases with flatter (less informative) prior and with increasing number of observations. However, the latter is probably simply an effect of the disappearing impact of the prior as the number of observations

increases.

Then we check the sensitivity of a Bayesian functional to observations by means of the Sensitivity Curve. Typically this curve is useful to identify observations with a large influence, such as outliers and loosely speaking an outlier is defined to be an observation that is unlikely to have been generated by the assumed sampling model. Therefore when the influence on the functional of a single observation is potentially unbounded, it is crucial to determine whether some outliers are present in the sample. We show that the local influence measure for the sampling model can be used for this purpose. In this case, indeed, it assumes huge values revealing that reference sampling model is very sensitive to perturbations and hence probably inadequate for the presence of some outlying observations.

Finally we deal with the issue of practical implementation. We concentrate on the local influence measure for the sampling model and we propose a new estimator for the Bayes factor which speeds up computations. Such estimator performs well, giving precise estimates with small confidence intervals.

1.9 Outlook on future research

In this final section we suggest some possible directions for future research. First, it would be interesting to extend the local influence measures proposed in this paper to more general measures, e.g. measures that consider the sensitivity to more than one input a time. This would help to assess the combined effect on a Bayesian quantity of perturbing a particular prior/sampling model combination. A second direction would be to consider the intrinsic discrepancy measure between probability distributions (Bernardo and Ruenda, 2002). Such measure has been shown to have many attractive properties (Bernardo and Juárez, 2003) and it may be used to define a new type of sensitivity measures.

Chapter 2

Robust Bayesian estimation

2.1 Abstract

This paper deals with the problem of robust estimation in a Bayesian context. We present an overview on some families of so-called robust distributions and we show that they belong to the family of elliptical distributions. According to this result, extensions to the multivariate case can be easily obtained. Moreover we propose criteria to assess when using a robust model is recommended and how to choose among estimates obtained with different models.

2.2 Introduction

The problem of building robust estimation procedures in a Bayesian context is an intriguing issue. In 1980 Box argues that to build efficient models, model robustification is required, "where by robustification I mean judicious and grudging elaboration of the model to ensure against particular hazard (..). Robustification becomes necessary when it is known that likely, but not easy detectable, model discrepancies can yield badly misleading analyses."

In the Bayesian literature we find two ways to build robust procedures. The first one is used within the global approach and applies when a large range is obtained for the functional. It aims

to narrow the class of prior and/or sampling distributions down to the point where a satisfactory range is reached. We refer for example to Berger (1994), Liseo *et al.* (1996) and Moreno *et al.* (1996) for different ways of reducing the width of a class. A second way applies when normality is adopted for the sampling distribution. In Bayesian analyses this assumption is often convenient in order to obtain analytical results for the posterior. However, in this case it is well known that the sensitivity of posterior quantities to observations is more pronounced and that only few atypical values in the sample heavily influence estimates. The reason for this fact has been found by many authors in light tails of the normal model adopted (Box and Tiao, 1992; Dawid, 1973; Zellner, 1976). Robustness with respect to outliers is achieved by choosing a so-called *robust model*. A robust model consists in a location-scale family of symmetric unimodal distributions enriched with 'robustness' parameters that control its shape. Therefore different univariate unimodal heavy-tailed models have been proposed to replace the normal model (Box and Tiao, 1962; Ramsay and Novick, 1980; West, 1984; Albert *et al.,* 1991) and the resulting posterior distribution becomes analytically intractable. However nowadays this is not a limitation since the availability of faster personal computers allow us to easily obtain estimates by means of Monte Carlo Markov Chain algorithms. Alternatively, normal approximations of the posterior distribution can be used (see for example Box and Tiao, 1992).

The goal of this paper is to propose Bayesian estimates which are robust against outliers, where we define an outlier[1] to be an observation which is unlikely to have been generated by the assumed sampling model. For this purpose we follow the second way and we concentrate on posterior summaries with a normal sampling model assumption. However, many points have to be discussed. First, in many situations the presence of influential observations is not easily detectable (e.g. for the multivariate nature of data) and we may fail to recognize the need of a robust sampling model. Is it possible to define measures that help us in deciding whether a robust model has to be adopted? Secondly, once we judge that a robust distribution is needed, how do we choose between

[1] We use the term *outlying observation* as a synonymous.

different models? This paper discusses such points and it is organized as follows. In Section 2.3 we present an overview of some univariate robust models and we show that they fall into the more general elliptical family. The main contribution of the paper is to provide criteria to assess when adopting a robust model is recommended and how to choose between different distributions. We do this in Section 2.4. Different examples of robust Bayesian estimation are then implemented in Section 2.5. Finally a brief summary of the findings and suggestions for future research are given in Section 2.6.

2.3 Robust models

In this section we present different models which have been proposed in the literature. First, we briefly introduce the class of elliptical distributions. Then we present an overview of some families of robust models. We show that such distributions fall into the class of elliptical distributions. Detailed proofs are given in Appendix D. This result helps to easily generalize univariate distributions to the multivariate case and it is useful in many practical situations. Finally we propose criteria to assess the need of adopting robust models and to choose among them.

2.3.1 Elliptical distributions

The class of Elliptical Distribution (ED) is a family of symmetric distributions which includes among others the normal and the student$-t$. Moreover, it offers a simple way to generalize a univariate distribution to the multivariate case. It was first introduced by Kelker (1970) and then studied by several authors (e.g. Fang and Anderson, 1990 and Gupta and Varga, 1993).

Definition 1 *Let X be a $k \times 1$ dimensional random vector whose distribution is absolutely continuous. Then, $X \sim ED_k(\theta, \Sigma)$ if and only if the p.d.f. of X has the form*

$$f(X) = c \cdot |\Sigma|^{-1/2} g\left(\frac{1}{2}(X-\theta)' \Sigma^{-1} (X-\theta)\right) \tag{2.1}$$

where g is an univariate function called density generator. Moreover the characteristic function of X can be written as

$$\varphi_X(t) = \exp(it'\theta) \cdot \Psi\left(\frac{1}{2}t'\Sigma t\right),$$

where Ψ is an univariate function.

The condition $\int_0^\infty u^{\frac{k}{2}-1} g(u)du < \infty$ guarantees g to be a density generator. Moreover, the normalizing constant can be obtained using the polar coordinates in several dimensions and is given by

$$c = \frac{\Gamma(k/2)}{(2\pi)^{k/2}} \left[\int_0^\infty u^{\frac{k}{2}-1} g(u)du\right]^{-1}.$$

A detailed prove of this result can be found in the paper by Landsman and Valdez (2003).

2.3.2 Main robust distributions

We now present location-scale families of distributions with tails decreasing to zero more slowly than in the normal case. Parameters (μ, σ) represent the mean and the standard deviation of the distribution. We give the form of the density generator when a distribution belongs to the elliptical family. Moreover in Appendix D we show that the condition $\int_0^\infty u^{\frac{k}{2}-1} g(u)du < \infty$ holds for the densities where this result has never been proved.

In 1962 Box and Tiao introduce the family of exponential power-series distributions (EPS). Such a family is given by

$$\left\{ f(x|\mu,\sigma,\delta) = k_\delta \cdot \sigma^{-1} \cdot \exp\left(-c_\delta \cdot \left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\delta+1}}\right), x \in \Re, -1 < \delta \leq 1 \right\}, \tag{2.2}$$

where

$$\begin{aligned} c_\delta &= \left[\frac{\Gamma\left(\frac{3}{2}(\delta+1)\right)}{\Gamma\left(\frac{1}{2}(\delta+1)\right)}\right]^{\frac{1}{\delta+1}} \quad \text{and} \\ k_\delta &= \frac{\left[\Gamma\left(\frac{3}{2}(\delta+1)\right)\right]^{1/2}}{(\delta+1)\left[\Gamma\left(\frac{1}{2}(\delta+1)\right)\right]^{3/2}}. \end{aligned}$$

In EPS family μ is the location parameter, σ the scale parameter and δ can be regarded as a non-Normality parameter. For $\delta > 0$ the distributions have heavier tails, for $\delta < 0$ the distributions

have flatter tails than the normal form. This family is quite large including as special cases the normal ($\delta = 0$), the double exponential ($\delta = 1$) and the rectangular ($\delta \to -1$) distributions. Since we are interested in distributions which are only slightly different from the normal one, we will choose small values for δ. The EPS distribution belongs to the elliptical family with density generator $g(u) = \exp\{-c_\delta \cdot (2u)^{\frac{1}{\delta+1}}\}$.

Ramsay and Novick (1980) measure the influence of a single observation x_j on the posterior distribution by considering the derivative of the logposterior with respect to x_j. The latter turns out to be a function of a particular quantity, named influence function of the likelihood (IF_{lik}). They show that for a certain symmetric family of distributions, which includes the normal, the IF_{lik} is unbounded. Hence, they propose a new family with bounded IF_{lik} given by

$$\left\{ f(x|\mu,\sigma,a,b) = k_{a,b} \cdot \ r(x) \ \cdot s(\mu,\sigma) \cdot \exp\left\{-\eta_{a,b}(d)\right\},\ a>0, b>0 \right\}, \tag{2.3}$$

where $\eta_{a,b}(d) = \frac{1}{ba^{2/b}}\gamma(2/b, a|d|^b)$, d is a measure of the distance of x from the location parameter μ, $\gamma(p,z)$ is the incomplete gamma function and $k_{a,b}$ is the normalizing constant. The normal distribution is obtained for $a \to 0$. Therefore we would consider small values of this parameter. A peculiarity of this distribution is that its tails do not decrease to zero as x tends to ∞. Indeed in this case $\eta_{a,b}(d) \to \frac{1}{ba^{2/b}}\Gamma(2/b)$, which is a fixed quantity. The consequence is that $k_{a,b}$ has to be computed in a region of integration with finite fixed limits. The choice of such limits is not so important as long as they are sufficiently far away from observed data. The RN distribution belongs to the elliptical family with density generator $g(u) = \exp\{-\left(ba^{2/b}\right)^{-1} \cdot \int_0^{a\cdot(2u)^{b/2}} e^{-t}t^{2/b-1}dt\}$.

In 1991 Albert, Delampady and Polasek propose an extension of the EPS distribution, called extended power distribution (EP). This family is given by

$$\left\{ f(x|\mu,\phi,c,\lambda) = k_{c,\lambda} \ \cdot \phi^{1/2} \cdot \exp\left\{-\frac{c}{2}\cdot\rho_\lambda\left(1+\frac{\phi(x-\mu)^2}{c-1}\right)\right\}, c>1, \lambda \geq 0 \right\}, \tag{2.4}$$

where

$$\rho_\lambda(v) = \begin{cases} \frac{v^\lambda - 1}{\lambda} & \text{if } \lambda > 0 \\ \\ \lim_{\lambda \to 0} \frac{v^\lambda - 1}{\lambda} = \log v & \text{if } \lambda = 0. \end{cases},$$

(μ, ϕ) are the location-scale parameters, (c, λ) are the robustness parameters and $k_{c,\lambda}$ is the normalizing constant. The main advantage of (2.4) with respect to (2.2) is that the former is differentiable everywhere. For this density we know that a relation of type $\sigma^2 = \nu(\phi)$ between the variance σ^2 and parameter ϕ holds. Therefore we may alternatively express (2.4) as

$$f(x|\mu, \sigma, c, \lambda) = k_{c,\lambda} \cdot \sigma^{-1} \cdot \left[\nu^{-1}\left(\sigma^2\right) \ \cdot \sigma^2\right]^{\frac{1}{2}} \cdot \exp\left\{-\frac{c}{2} \cdot \rho_\lambda\left(1 + \frac{\nu^{-1}\left(\sigma^2\right) \cdot \ \sigma^2}{c-1}\left(\frac{x-\mu}{\sigma}\right)^2\right)\right\}.$$

If $\lambda = 0$ the relation is given by $\sigma^2 = \frac{(c-1)\sqrt{2}}{(c-3)\phi}$. If $\lambda > 0$ the relation can be found only numerically. Different location-scale densities are included in this family, like the normal and the Student-t. The tails behavior is controlled by the parameter λ. For $0 \leq \lambda < 1$ we get fatter tails, whereas for $\lambda > 1$ we get sharper tails than the normal case. For our purpose, we consider only the case $\lambda = 0$ and we choose the scale parameter ϕ so that the variance σ^2 equals the variance of the other distributions. Also this density belongs to the elliptical family with density generator $g(u) = \left[\nu^{-1}\left(\sigma^2\right) \cdot \ \sigma^2\right]^{\frac{1}{2}} \cdot \exp\left\{-\frac{c}{2} \cdot \rho_\lambda\left(1 + \frac{2\ \nu^{-1}\left(\sigma^2\right) \cdot \sigma^2}{c-1} u\right)\right\}$.

Another well known heavy-tailed distribution is the Student-t. The advantage of considering the previous families rather than the Student-t may be found in the larger choice of the elements in the class. In particular for distributions (2.3) and (2.4) two robustness parameters control the shape of the density function better. Furthermore the fact that for such models the normalizing constant has to be computed numerically does not represent a limitation. Indeed, robust estimation under these distributions is implemented through MCMC algorithms like Metropolis-Hastings (Hastings, 1970). This way turns out to be very convenient because the normalizing constant cancels out in the acceptance probability. Figure 2.1 and 2.2 show plots of the densities we presented in this section for different values of the robust parameters.

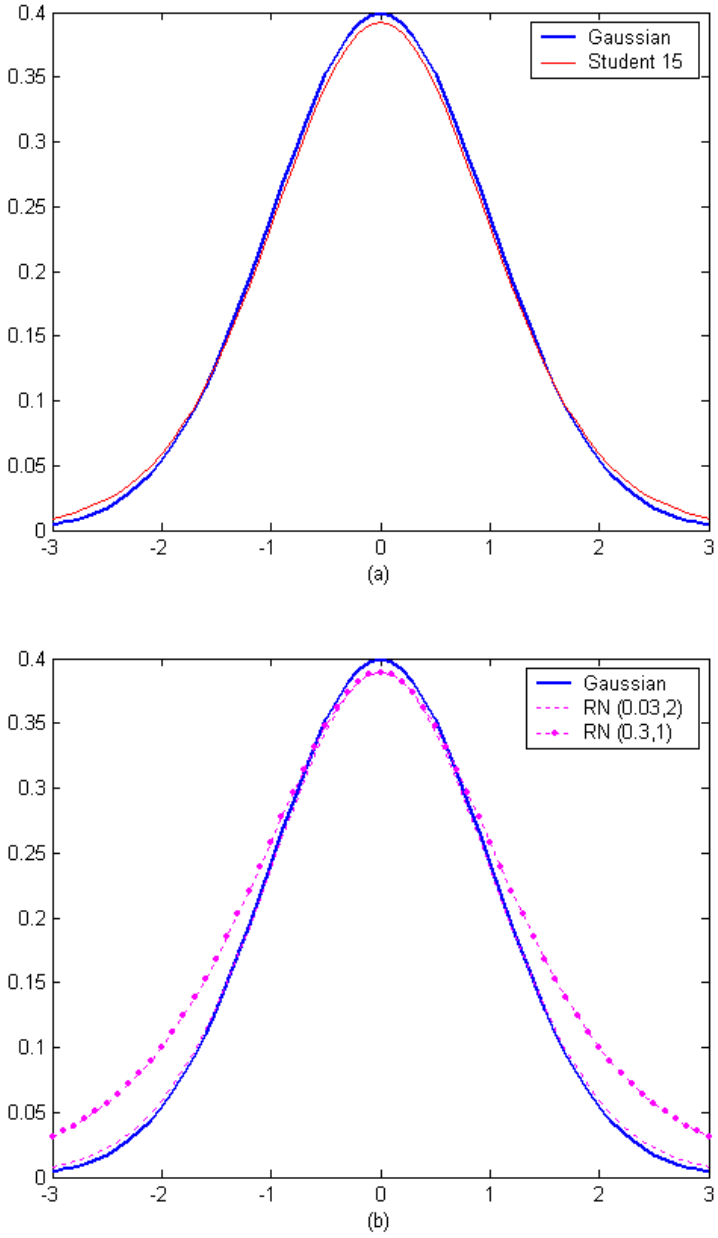

Figure 2.1: Plots of normal and robust densities: (a) Student-t and (b) RN distributions.

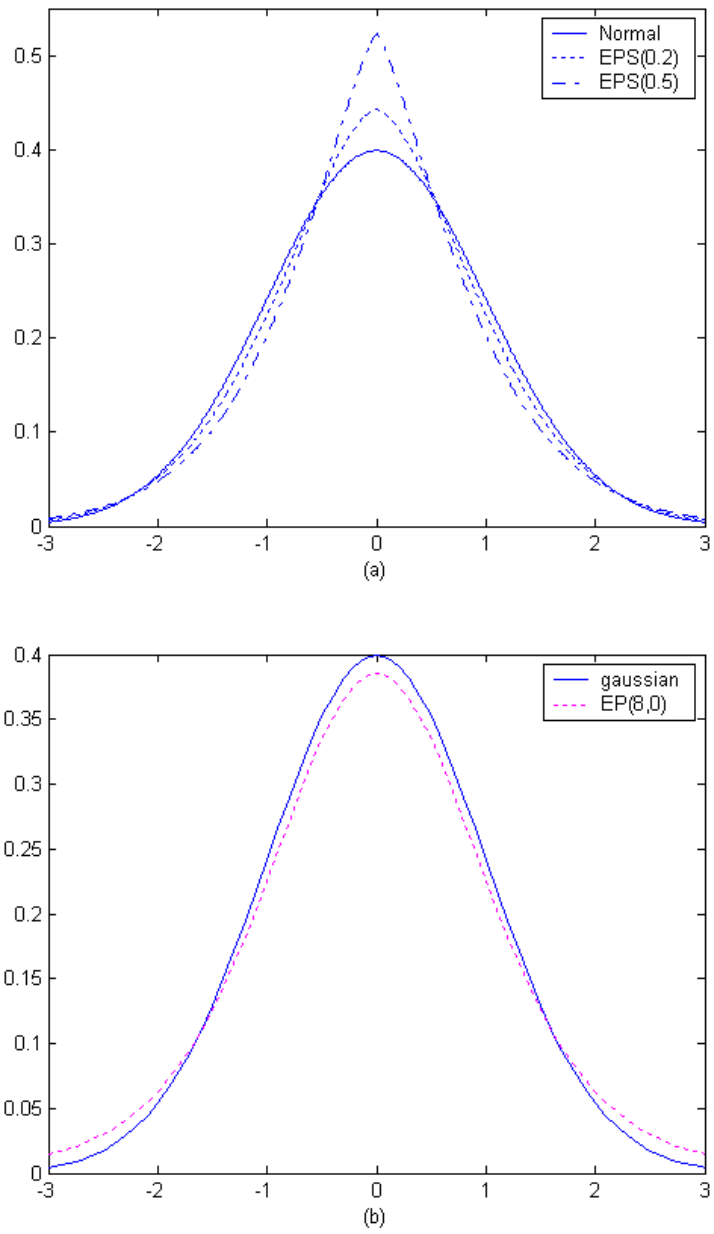

Figure 2.2: Plots of normal and robust densities: (a) EPS and (b) EP distributions.

2.4 Criteria for robust estimation procedures

In this Section we describe how understanding when robust models have to be used and how choosing among different distributions.

In the literature robust models are adopted when the normal assumption appears inadequate. Typically this conclusion is drawn on the basis of a visual inspection of the data, which is straightforward in the univariate or bivariate dimensions. When the dimension increases up to k ($k > 2$) checking up the adequacy of normal model assumption is not so straightforward. For this purpose local robustness measures described in Chapter 1 are useful tools. Indeed we have shown that such measures reveal the presence of observations that have been unlikely generated by the assumed sampling model (outlier).

In this paper we concentrate on robust estimation of posterior summaries of type (1.2). To establish the need of using robust models, the first thing to do is to compute the SC defined in Section 1.5. Such quantity evaluates the effect of moving one observation in the sample once a particular combination of prior/sampling distributions is fixed. If this measure diverges as z becomes bigger, a single observation has a potentially unbounded influence on the functional estimated value. However, this is not a sufficient reason to justify the use of a robust rather than a normal sampling model because influential observations may not be present in the sample. In order to detect outliers we have to compute the local influence measure for the sampling model (see Section 1.4.2). Such measure assesses the so-called *model or likelihood robustness* and evaluates the impact on the functional of a small contamination of the base sampling model. For posterior summaries, it can be written as

$$LI(G; T_B, F_\theta) \; = \sum_j \frac{m_j(x; \Pi, F_\theta, G)}{m(x)} \left[T_{B,j}\left(F_\theta, G\right) - T_B \right], \tag{2.5}$$

where m_j/m is the Bayes factor and $T_{B,j}\left(F_\theta; G\right)$ is the posterior functional obtained when the sampling distribution is G only for observation x_j and F_θ for the others. Alternatively, a relative local influence measure can be defined for the purpose of comparing different functionals (see

Chapter 1) and it will be denoted by LI^*. The Bayes factor in measure (2.5) captures the effect on the functional of choosing a contaminating model for observation j which is more adequate or less adequate than the base one with respect to observed data. If an outlier is present in the sample we expect that this quantity would assume values greater than 1 and the difference $[T_{B,j}(F_\theta, G) - T_B]$ would be not negligible, leading to a substantial value of (2.5). In this case the use of a robust model is recommended for dumping the effect of extreme observations on the estimate.

Finally, measure (2.5) can also help in choosing the most appropriate robust model. If we adopt one of the distributions presented in Section 2.3.2, we guess that the corresponding LI measure for the sampling model displays quite a small value. Therefore a criterion of choice is to adopt the distribution which displays the smallest value for (2.5). Furthermore, if such value is small, we achieve robustness both with respect to outliers and with respect to the sampling model. In the next Section we provide some examples on robust estimation procedures.

2.5 Examples of robust estimates

In this section we continue the examples considered in the previous Chapter. We first consider the mean of the posterior distribution in the univariate case with a sample drawn from a Gaussian distribution and we evaluate the effect of assuming a robust model when it is necessary and when it is not. Then we consider a Bayesian regression model using Ramsay and Novick's data and we produce robust estimates of regression coefficients. We use different heavy-tailed models for the sampling distribution to illustrate the robust estimation procedure. When MCMC algorithms are used we check the convergence of the chain and of the averages by means of BOA library in R language.

2.5.1 Posterior mean

In Chapter 1 we found that the posterior mean was not robust with respect to observations. However, the small size of LI measures suggested that atypical observations were not present and

robust estimation was not necessary. What would be the effect on the estimate of assuming a robust model in such a situation? We answer this question considering different sampling models whose densities are shown in Figures 2.1 and 2.2. The choice of robustness parameters has been made so that robust densities show heavier tails than the normal case (Figure 2.3).

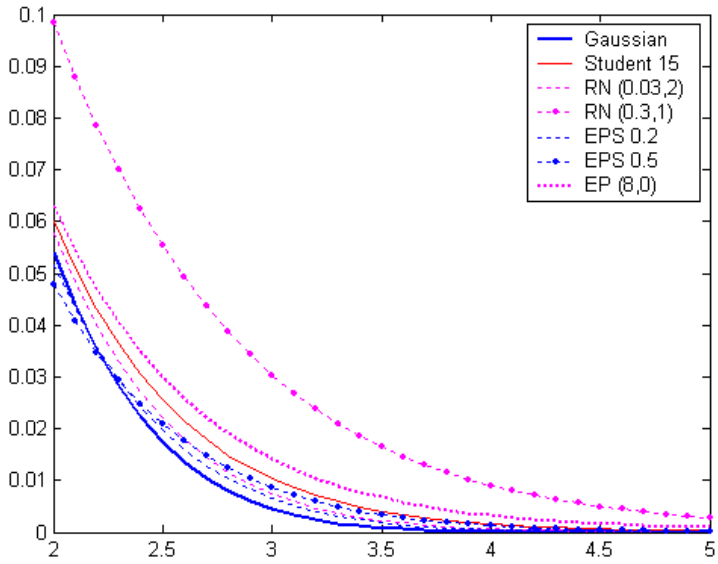

Figure 2.3: Plot of the thickness of tails in normal and robust models.

By means of Random Walk Metropolis-Hastings algorithm the posterior distribution has been computed. For each simulation we run a chain of 100.000 steps. The prior is chosen to be $N(0.5, 1)$ and different sampling models are used.

Estimates of posterior quantities are shown in Table 2.1. Analytical estimates, computed for the normal case, are reported in the bottom line. The concordance between analytical and numerical results supports convergence of our algorithm. Estimates of posterior mean do not differ as much under different sampling models. However, the more we move away from normality, the more we loose in efficiency, since posterior variance increases. This trade-off between efficiency and dumping

effect of outliers is typical of robust estimates. Moreover the concordance of posterior mean and median together with a visual insight show that the posterior distribution is still symmetric and unimodal for all distributions considered.

	$T_B(F_\theta)$		$median$	σ^2_{post}
Normal	0.9862	(0.0021)	0.9858	0.0625
Student-t (15)	0.9875	(0.0022)	0.9872	0.0697
$EPS\,(0.2)$	0.9996	(0.0021)	1.0037	0.0639
$EPS\,(0.5)$	1.0038	(0.0022)	1.0149	0.0669
$EP\,(8;0)$	0.9953	(0.0023)	0.9966	0.0764
$RN\,(0.03;2)$	0.9798	(0.0022)	0.9803	0.0699
$RN\,(0.3;1)$	0.9663	(0.0027)	0.9737	0.1020
Analytical	0.9857			0.0625

Table 2.1: Posterior estimates (standard error) under different sampling models. $MCMC$ simulations with 100.000 runs.

Table 2.2 shows relative local influence measures under different sampling models, computed by perturbing the base sampling distribution in the direction of a $N\left(\theta, 10\cdot\sigma^2\right)$. Derived LI^* measures are small, supporting the fact that all models are approximately adequate to our data. Looking at all these elements together, we conclude that using a robust family of distributions when no extreme observations are present let us still correctly estimate the posterior mean.

In Table 2.3 and 2.4 we reproduce the same analysis introducing the observation $x_4 = -5$ in the sample. Numerically estimated posterior expectations are now very different and change according to the robust model adopted. Tails inflation permits controlling the impact of the outlier on the estimate. Again the efficiency of estimates decreases as we move away from the normal case.

	$LI^*\left(\widetilde{G}_\theta; T_B, F_\theta\right)$	range $T_{B,j}(F_\theta, G_\theta)$	range r_j
Normal	0.0065	$[0.8111; 1.1584]$	$[0.3864; 0.6978]$
Student-t (15)	0.0123	$[0.8151; 1.1606]$	$[0.4049; 0.7203]$
$EPS\,(0.2)$	0.0183	$[0.8083; 1.1654]$	$[0.3869; 0.7553]$
$EPS\,(0.5)$	0.0249	$[0.8119; 1.1785]$	$[0.3872; 0.8356]$
$EP\,(8; 0)$	0.0162	$[0.8129; 1.1571]$	$[0.3434; 0.6910]$
$RN\,(0.03; 2)$	0.0173	$[0.8112; 1.1583]$	$[0.4023; 0.7017]$
$RN\,(0.3; 1)$	0.0553	$[0.8058; 1.1146]$	$[0.4920; 0.7440]$
Analytical	0.0096	$[0.8149; 1.1611]$	$[0.3871; 0.6995]$

Table 2.2: Relative local influence measures of the posterior mean with respect to the sampling distribution under different sampling models. $MCMC$ simulations with 100.000 runs.

	$T_B\,(F_\theta)$		$median$	σ^2_{post}
Normal	-0.4384	(0.0018)	-0.4392	0.0469
Student-t (15)	0.8032	(0.0023)	0.8067	0.0754
$EPS\,(0.2)$	0.0407	(0.0024)	0.0464	0.0825
$EPS\,(0.5)$	0.5207	(0.0024)	0.5295	0.0832
$EP\,(8; 0)$	0.8805	(0.0024)	0.8857	0.0823
$RN\,(0.03; 2)$	0.9776	(0.0022)	0.9761	0.0703
$RN\,(0.3; 1)$	0.9270	(0.0027)	0.9309	0.1028
Analytical	-0.4395			0.0476

Table 2.3: Posterior estimates (standard error) under different sampling models $MCMC$ simulations with 100.000 runs. Contaminated sample.

	$LI^*\left(\widetilde{G}_\theta; T_B, F_\theta\right)$	range $T_{B,j}(F_\theta, G_\theta)$	range r_j
Normal	$7.02 \cdot 10^{25}$	$[-0.943; 0.804]$	$[0; 2.49] \cdot 10^{25}$
Student-t (15)	91.9854	[0.572; 0.921]	$[0; 3.07] \cdot 10^{3}$
$EPS\,(0.2)$	$2.52 \cdot 10^{16}$	$[-0.460; 0.801]$	$[0; 1.43] \cdot 10^{15}$
$EPS\,(0.5)$	$6.50 \cdot 10^{7}$	[0.104; 0.965]	$[0; 1.13] \cdot 10^{8}$
$EP\,(8; 0)$	3.4226	[0.6841; 1.0256]	[0.19; 21.88]
$RN\,(0.03; 2)$	147.8905	$[-0.071; 0.582]$	[0.39; 878.17]
$RN\,(0.3; 1)$	0.3508	[0.6721; 1.0482]	[0.39; 2.07]
Analytical	$7.41 \cdot 10^{25}$	$[-0.947; 0.804]$	$[0; 2.62] \cdot 10^{25}$

Table 2.4: Relative local influence measures of the posterior mean with respect to the sampling distribution under different sampling models. $MCMC$ simulations with 100.000 runs. Contaminated sample.

As expected, the relative measure of local influence for the normal sampling model explodes, revealing inadequacy of the model to the data ($LI^* = 7.02 \cdot 10^{25}$). This explosion is due to the huge value that ratio r_j assumes in correspondence to the outlier ($j = 4$). Marginal likelihood $m_4\,(x; \Pi, F_\theta, G)$ is much bigger than the base marginal $m\,(x)$, which means that data support more distributions with heavy tails for observation x_4. In all robust models considered ranges both for $T_{B,j}$ and for r_j are narrowed and local influence measure is reduced up to 0.35. Therefore to compute robust estimation we would adopt the RN distribution with parameters $(0.3; 1)$. Robust estimate of the posterior mean is given by 0.9270.

In the previous section we say that to achieve robustness with respect to outliers a robust sampling model has to be adopted. Therefore, in such a situation we expect the SC of posterior mean to be bounded for extreme observations. In Figure 2.4 we compute the SC for the selected robust model. The curve shows the expected behavior.

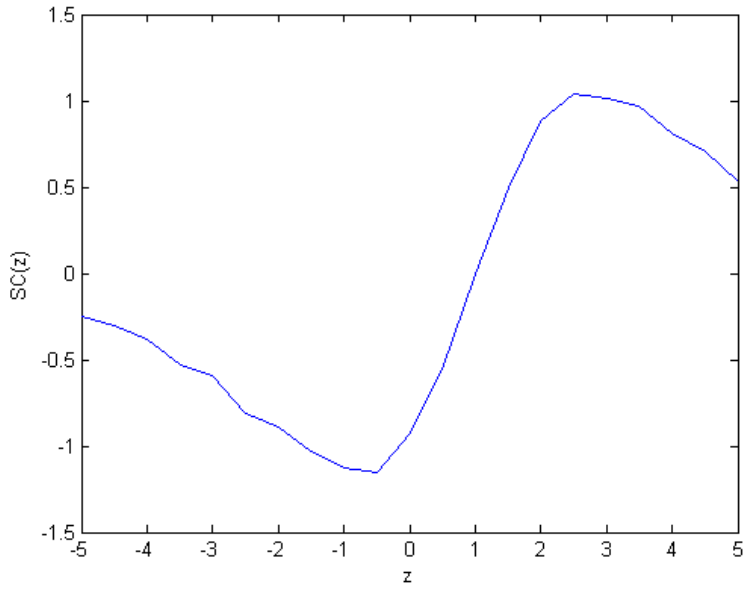

Figure 2.4: Sensitivity curve of the posterior mean under a $RN\,(0.3, 1)$ sampling model. $MCMC$ simulations with 100.000 runs.

2.5.2 Bayesian Linear Regression

We now consider a Bayesian linear regression. We use the same data set employed by Ramsay and Novick (1980) and we study the impact that both a test of verbal intelligence and a test of performance intelligence have on dichotic listening task[2]. Bayes estimate of regression coefficients are found to be extremely sensitive to observations. Moreover the local influence measure with respect to the sampling model reveals that the normal distribution is not so adequate (see Section 1.7.2). In this section we will derive robust estimates of regression coefficients.

We consider different robust sampling models and compute the posterior distribution with 200.000 runs of the Metropolis-Hastings algorithm. Computed Bayes estimates are shown in Table 2.5.

[2]We choose a normal prior distribution with the same parameters used by Rambsay and Novick (1980).

	Regression coefficients		
	β_1	β_2	β_3
normal	0.7455	-0.0716	38.1791
	(0.0006)	(0.0006)	(0.0492)
RN $(0.05, 2)$	0.6637	0.0082	37.6231
	(0.0008)	(0.0009)	(0.0577)
EPS (0.2)	0.7051	-0.0290	37.6690
	(0.0007)	(0.0007)	(0.0515)
student (15)	0.6677	0.0067	37.4549
	(0.0022)	(0.0023)	(0.1449)
EP $(5, 0)$	0.5836	0.0716	38.6109
	(0.0009)	(0.0008)	(0.0691)
Analytical	0.7458	-0.0734	38.3505

Table 2.5: Bayes Estimates (standard error) of regression coefficients under different sampling models. $MCMC$ simulations with 200.000 runs.

The value of coefficients changes substantially according to different models while the standard error increases only a little. The substantial difference with normal estimates is clear for $\widehat{\beta}_2$. In this case the relation between the dichotic listening task and the test of performance intelligence changes from negative to slightly positive.

In order to choose robust estimates we compute local influence measures of regression coefficients for the sampling distribution. Table 2.6 shows the results. All robust models lead an improvement in terms of reducing the value of LI^* measures, in particular the density proposed by Ramsay and Novick. Robust estimates of regression coefficients are therefore given by $\widehat{\beta}^{rob}_{Bayes} = (0.6637, 0.0082, 37.6231)'$. Such values are expected to be robust against influential

points.

	$LI^*\,(G;T_B,F)$ relative to		
	β_1	β_2	β_3
normal	43.14	471.41	5.82
$RN\ (0.05,2)$	$0.14\cdot10^{-5}$	$5.94\cdot10^{-5}$	$0.15\cdot10^{-5}$
$EPS\ (0.2)$	$0.11\cdot10^{-2}$	$0.79\cdot10^{-2}$	$0.12\cdot10^{-2}$
student (15)	7.32	173.14	0.30
$EP\ (5,0)$	0.91	2.34	0.14

Table 2.6: Relative local influence measures of regression coefficients with respect to the sampling distribution under different models. $MCMC$ simulations with 100.000 runs.

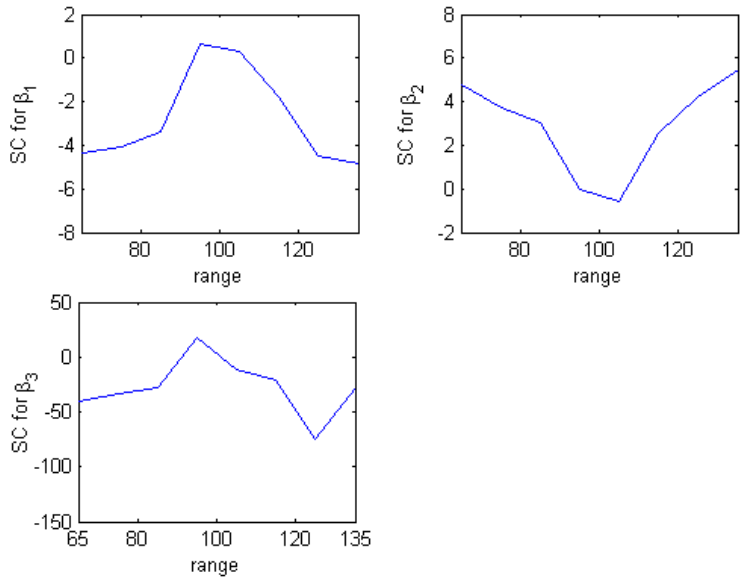

Figure 2.5: Sensitivity Curve of regression coefficients under a $RN\,(0.05,2)$ sampling model. $MCMC$ simulations with 100.000 runs.

Figure 2.5 shows the SC of $\widehat{\beta}^{rob}_{Bayes}$. The improvement for regression coefficients estimates is clear since all the curves become bounded.

2.6 Summary and outlook on future research

In this paper we review some families of so-called robust distributions and we show that they belong to the more general elliptical family. According to this result, multivariate robust distributions can be easily obtained. Moreover we propose criteria to assess when the use of a robust model is recommended and how to choose between different distributions. First, the SC has to be computed for the estimator of interest. If a single observation plays a potentially unbounded influence, it is crucial to determine whether influential observations are present in the sample. For this purpose we use the local influence measure for the sampling model proposed in the previous Chapter. The examples show both that the size of this measure becomes substantial when outliers are present and that adopting a robust model leads to estimates on which the effect of outlying observations is dumped. Moreover, the use of a robust family of distributions when no extreme observations are present let us still obtain correct estimates. Obtained robust estimates behave well since the corresponding SC is bounded for extreme observations. Finally, the local influence measure for the sampling model provides also a criterion for choosing among different robust estimates. An interesting matter for future research on the field would be to include prior distributions also for robustness parameters of robust models.

Chapter 3

Robust Bayesian mean-variance portfolio selection

3.1 Abstract

It is well known that the Bayesian approach to mean-variance portfolio selection problem accounts for estimation risk. However, no results are present on the effects of model risk in this case. This paper aims to study the robustness properties of the Bayesian mean-variance weights. We first perform a simulation study to explore the effect of model risk on Bayesian weights. Then we compute their measures of sensitivity both to distributional assumptions and to observations. Moreover, we propose a robust estimation procedure which dampens the effect of 'extreme' observations. We study the performance of computed measures through a simulation study and we obtain robust Bayesian mean-variance weights using real market data.

3.2 Introduction

According to Markowitz's modern portfolio theory (Markowitz, 1952; 1959), a risk-averse investor should choose the amount to invest among a set of assets relying just on the first and second moments of return distribution[1]. However practitioners found that the derived optimal portfolio allocation are often unreasonable (Black and Litterman, 1992). Two main reasons can be given for this fact.

The first one is that means and variances are generally unknown parameters that have to be estimated. Not accounting for this fact can induce a bias in the estimated weights and lead to the so called *estimation risk* problem. In mean-variance portfolio selection problem observed returns are assumed to be i.i.d. random drawn from a multivariate normal distribution[2]. There are essentially two different approaches for the implementation of portfolio theory: the *Certainty Equivalence or naive approach*, where parameters uncertainty is ignored and unknown parameters are simply replaced by their sample estimates, and the *Bayesian approach*, where parameters are treated as random variables and unknown parameters are estimated by summary measures of the predictive distribution. There is evidence in the literature that not taking into account parameters uncertainty leads to suboptimal portfolios (Barry, 1974; Brown, 1979; Jorion, 1986; Cavadini, Sbuelz and Trojani, 2002). The Bayesian point of view not only considers parameter uncertainty but also satisfies the axiomatic paradigm of Neumann-Morgenstern-Savage's expected utility maximization (see Bawa, Brown and Klein, 1979).

The second reason is that mean-variance portfolio weights are extremely sensitive to observations. It is found in practice that securities display sometimes extremely low or high returns. We

[1] It may be argued that such moments can be not sufficient in describing portfolio returns characteristics and that including higher moments can be more efficient. However we choose to deal with the mean-variance framework because of its wide use in practical applications.

[2] We known that the i.i.d. assumption can be unrealistic since observed series of returns may exihibit autocorrelation. However in practice the i.i.d. multivariate normal model is widely used and we do not explore this matter here.

refer to them as outlying observations. This fact can induce a bias in the estimates and leads to the so called *model risk* problem. Indeed mean-variance weights are derived under the assumption of a normal data generating process. Nothing is assured when the assumed mechanism is only approximately true. Recent papers deal with model robustness in the CE approach (Victoria-Feser, 2000; Cavadini, Sbuelz and Trojani, 2002; Perret-Gentil and Victoria-Feser, 2003). They show that few outlying returns have a strong influence on the composition of the resulting optimal portfolio. Moreover Cavadini, Sbuelz and Trojani (2002) find that model risk plays a greater role than estimation risk. However no evidence of this fact is given for Bayesian weights.

This paper focuses on Bayesian mean-variance portfolio selection and aims to assess its robustness properties. It is organized as follow. In Section 3.3 we briefly describe two different approaches to portfolio theory implementation and we compare them. Measures of sensitivity of Bayesian weights both to distributional assumptions and to observations are computed in Section 3.4. The behavior of such measures is then explored by means of a simulation study. In Section 3.5 we propose a Bayesian estimation procedure for investment decisions which dampens the effect of 'extreme' observations. An example on estimates of robust Bayesian weights with real market data is given in Section 3.6. Finally Section 3.7 concludes.

3.3 Certainty Equivalence and Bayesian portfolio selection

In mean-variance portfolio theory the optimal portfolio is the one that minimizes portfolio risk for a given level of portfolio expected return or, viceversa, maximizes portfolio expected return for a given level of portfolio risk. Assuming both a multivariate normal distribution for future returns r and a negative exponential utility function[3], this approach can be set in an expected utility maximization paradigm. We would consider a standard one-period model in which investors use portfolio to transfer wealth from one period to the next.

Maximizing the expected utility of the end-of-period wealth is equivalent to maximizing the

[3] An alternative assumption would be to assume a quadratic utility function for investors.

expected utility of portfolio return. Under some constraints[4], investors would choose the weights q such that

$$\begin{aligned} \max_q E\left[U(r_{_P})\right] &= \max_q \int U(r_{_P}) p(r_{_P}|\vartheta) dz \\ &= \max_q \left\{ E\left(r_{_P}\right) - \frac{b}{2} Var\left(r_{_P}\right) \right\} \\ &= \max_q \left\{ q'\theta - \frac{b}{2} q'\Sigma q \right\} \end{aligned} \tag{3.1}$$

with $r_{_P} = q'r$ the portfolio return, b the risk aversion coefficient and $\vartheta = (\theta, \Sigma)$ the first and second moments of the assumed future returns distribution. The optimal weights are given by:

$$q = \frac{1}{b}\Sigma^{-1}(\theta - \lambda \cdot \mathbf{1}), \tag{3.2}$$

where q is the vector of the proportion invested on the risky assets, $\mathbf{1}$ is the vector of ones and $\lambda = \frac{\mathbf{1}'\Sigma^{-1}\theta - b}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$.

In the classical application of Markowitz's theory, known as *Certainty Equivalence* (CE) or *naive approach*, the unknown parameters are simply replaced by their sample estimates and portfolio selection problem is solved by finding weights q such that

$$\begin{aligned} \max_q E_{r_{_P}|\vartheta}\left[U(r_{_P})|\hat{\vartheta}\right] &= \max_q \int U(r_{_P}) p(r_{_P}|\hat{\vartheta}) dz \\ &= \max_q \left\{ q'\widehat{\theta} - \frac{b}{2} q'\widehat{\Sigma} q \right\}. \end{aligned} \tag{3.3}$$

This way completely ignores the *estimation risk*, that is the risk linked with the variability of parameter estimates.

Some works in this direction (Best and Grauer, 1991; Victoria-Feser, 2000) show that CE portfolio weights and moments are very sensitive to changes in parameters value, especially when non-negative weights constraint is absent. An evaluation of the relative impact of errors in parameter estimates is also provided by Chopra and Ziemba (1993), who find that for risk-tolerant

[4] We consider the contraint $1'q = 1$.

investors the impact of errors in mean is much greater than the impact of errors in variance and covariance parameters.

Alternatively weights q can be estimated using a *Bayesian approach*, in which prior informations about the unknown parameters are matched with observed asset returns through the posterior distribution. Bayesian portfolio selection problem would choose weights q such that:

$$\begin{aligned} \max_q E_{\vartheta|x}\left[E_{r_P|\vartheta}\left[U(r_P)|\vartheta\right]\right] &= \max_q E_{r_P|x}[U(r_P)] \\ &= \max_q \int U(r_P)p(r_P|x)dr_P \\ &= \max_q \left\{q'\theta_{r|x} - \frac{b}{2}q'\Sigma_{r|x}q\right\}, \end{aligned} \tag{3.4}$$

where $p(r_P|x) = \int p(r_P|\vartheta)p(\vartheta|x)d\vartheta$ is the predictive density of portfolio returns and $\left(\theta_{r|x}, \Sigma_{r|x}\right)$ are the moments of the predictive density of future returns $r|x$.

Bawa, Brown and Klein (1979) are the first that explore in deep the Bayesian approach and set it in a Neumann-Morgenstern-Savage paradigm. Brown (1979) provides numerical evidence of the impact of estimation risk. By means of a measures of utility loss due to the estimation process he shows that CE rule is dominated by Bayes rule. Similar results are found by Jorion (1986), who compares the risk linked with different estimators of asset expected return and gives a Bayesian interpretation to the proposed Stein estimator.

These results point out that in the presence of normally distributed data the Bayesian approach outperforms the CE one. But what happens if we move away from the normality assumption of model distribution? Will the portfolio weights be still near the optimal solution (3.2)? This matter, known in the literature as model risk, has recently been addressed for the CE estimator of portfolio weights. Victoria-Feser (2000) shows that the efficient frontier and portfolio composition can be seriously biased when data contain 'extreme' observations. In a later work Perret-Gentil and Victoria-Feser (2003) study the (asymptotic) stability properties in an neighborhood of the model, proving that bias of portfolio composition only depends on bias of estimated moments.

Moreover Cavadini, Sbuelz and Trojani (2002) prove that in classical estimated portfolios model risk generates greater bias than estimation risk. All these papers clearly show that in such a situation robust estimation procedures should be used. However no results are present for the Bayesian case. In the next section we perform a simulation study to compare CE and Bayesian estimators and we explore the effect of model risk on Bayesian weights.

3.3.1 Comparison of the two approaches through a simulation study

In this section we compare CE and Bayesian approach when the normality assumption of data is both satisfied and not satisfied. We generate T sets ($T = 1000$) of n observations ($n = 100$) from a multivariate normal with parameters given in Table 3.1

Sample estimates of monthly returns (January 1977-December 1981)

	Mean	Covariance Matrix						
Canada	1.287	42.18						
France	1.096	20.18	70.89					
Germany	0.501	10.88	21.58	25.51				
Japan	1.524	5.30	15.41	9.60	22.33			
Switz.	0.763	12.32	23.24	22.63	10.32	30.01		
U.K.	1.854	23.84	23.80	13.22	10.46	16.36	42.23	
U.S.	0.620	17.41	12.62	4.70	1.00	7.20	9.90	16.42

Table 3.1: Parameters of the sampling model in the first simulation study.

Dollar returns in percent per month.

The parameters[5] used for the exercise consist in sample estimates from monthly stock market returns for seven major countries calculated over a 60-month period (January 1977-December

[5] The parameters are the same as in the study of Jorion (1986). For calculations in the Bayesian case we choose a normal prior with mean $0.005 \cdot \mathbf{1}$ and covariance matrix $0.0025 \cdot I$.

1981). Contaminated samples are obtained by substituting randomly the 5% of the observations as specified in the sequel.

We are first interested in detecting the effect of the presence of 'extreme' returns on portfolio composition. For this purpose we estimated K optimal weights using non contaminated and contaminated samples. The latter are generated by substituting randomly the 5% of the observations with random drawn from a Dirach distribution $\Delta_{\theta^*}(y)$ which puts mass 1 at $\theta^* = \theta + \Sigma^{\frac{1}{2}} \cdot 3 \cdot \mathbf{1}$. Vector θ^* has components that are 15 to 40 time larger than the ones of θ. We obtained the boxplot for each of its component (Figure 3.1 and 3.2) and computed summary measures of the accuracy of each estimate (Table 3.2). The circle represents the true weight components. The risk aversion coefficient is set equal to 2 but similar results are achieved for different values of this parameter[6].

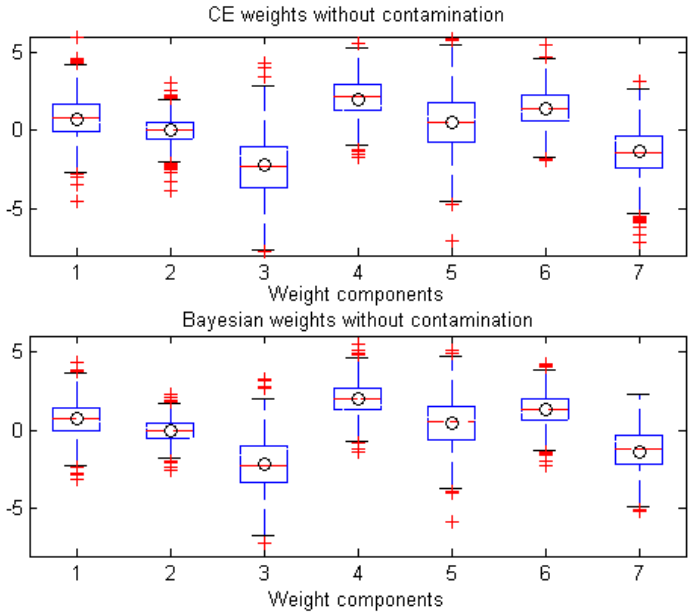

Figure 3.1: Boxplot of CE and Bayesian weights with non contaminated data.

Figure 3.1 displays the boxplot of CE and Bayesian weights when the data are not contaminated.

[6]The risk aversion coefficient has been set equal to 0.1, 2, 10 and 23 respectively.

From a visual insight the precision of estimates looks like the same, but the size of the box appears smaller under the Bayesian approach. This is confirmed by numerical measures of the difference between estimated and true components in Table 3.2. While the biases are similar, variability measures support the fact that in the presence of normally distributed data the Bayesian method works better.

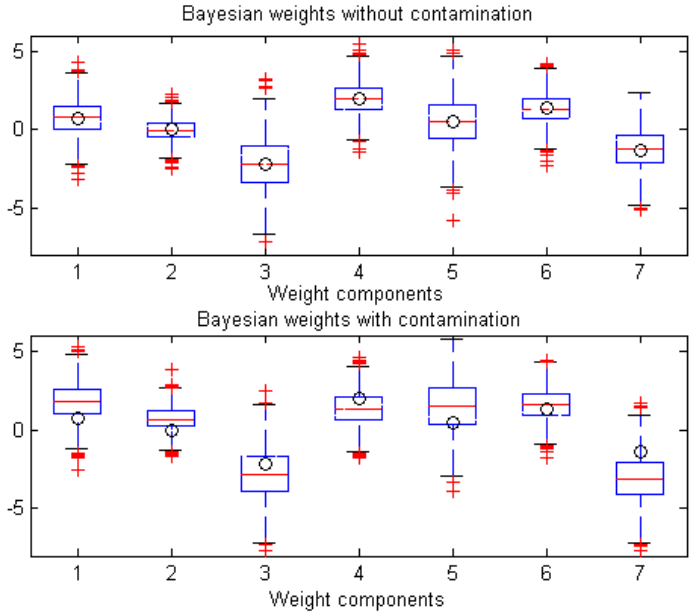

Figure 3.2: Boxplot of Bayesian weights with non contaminated and contaminated data.

A similar analysis for Bayesian weights with non contaminated and contaminated data leads to Figure 3.2. The more evident effect of the presence of outliers is that estimated components are far away from the true ones. As a consequence, summary measures of Table 3.2 increase. Component 1, 5 and 7 show the greatest bias and variability measures rise up to three times the one in the non contaminated case. Curiously, these components do not correspond to the greatest components of outlier θ^*.

Measure	q_1	q_2	q_3	q_4	q_5	q_6	q_7
$med\,(q_{B,j} - q_{true,j})$	0.01	-0.02	+0.04	0.02	0.04	-0.05	0.13
$med\left(q_{B,j}^{(\varepsilon)} - q_{true,j}\right)$	1.04	0.73	-0.67	-0.62	1.04	0.24	-1.73
$med\,(q_{CE,j} - q_{true,j})$	0.05	0.01	-0.13	0.14	0.06	0.03	-0.05
$med\left((q_{B,j} - q_{true,j})^2\right)$	0.55	0.20	1.34	0.51	1.16	0.41	0.84
$med\left((q_{B,j}^{(\varepsilon)} - q_{true,j})^2\right)$	1.39	0.59	1.43	0.83	1.68	0.50	3.09
$med\left((q_{CE,j} - q_{true,j})^2\right)$	0.75	0.25	1.71	0.70	1.68	0.61	1.06

Table 3.2: Summary measures of estimated weights with CE and Bayesian approach with and without contamination.

We want to investigate also the effect of the presence of 'extreme' observations on the maximum expected utility of investors. We use the loss measure proposed by Jorion[7] (1986), given by:

$$L(q, \widehat{q}) = \frac{EU_{MAX} - EU(\widehat{q})}{|EU_{MAX}|}, \tag{3.5}$$

where EU_{MAX} is the maximum expected utility when everything is known, and $EU(\widehat{q})$ is the maximum expected utility when the weights are estimated using either the CE or the Bayesian approach. Figure 3.3 shows such measure for different length of the sample computed both with non contaminated and contaminated data. For each possible sample size, measure (3.5) is calculated 1000 times and then averaged. The contaminated sample is obtained by replacing the 5% of the observations with random drawn from a multivariate normal with the same mean but the variance 100 time larger than parameters of Table 3.1.

[7] The analysis has been performed setting the risk aversion coefficient equal to 23, as in Jorion (1986).

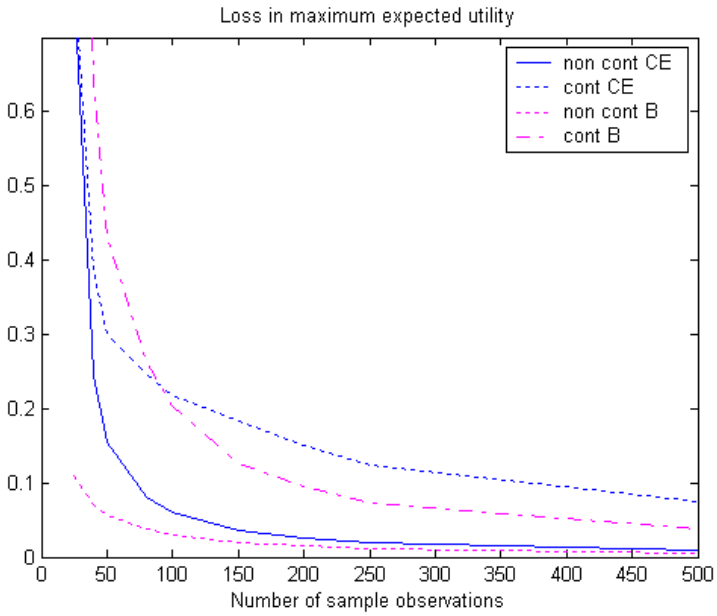

Figure 3.3: Loss function for the CE and B approach with and without contamination.

For both methods the loss function increases under contaminated data. However, for small sample sizes the gain of Bayesian over CE approach is no longer evident. We repeated the same analysis with different values of the risk aversion coefficient and we found that for more risk tolerant investors the loss in expected utility is even worst for the Bayesian case under contaminated data. Therefore if the data do not satisfies the normality assumption it is not guaranteed that the Bayesian approach gives an improvement in terms of loss in expected utility.

To explore the effect of different investor attitudes to the risk on the maximum expected utility we compared the mean loss over 1000 simulated samples of fixed size for different methods and values of b (Table 3.3). For each level of contamination Bayesian loss results very stable over different values of risk aversion coefficient. Incorporation of estimation risk in the Bayesian approach seems to preserve from the increasing loss that the CE approach displays as b increases. Therefore we could say for example that with a sample of 100 observations the Bayesian approach leads a loss in expected utility around the 3% or the 20% depending on how well the data satisfies

the normality assumption. If the data does not contain outliers the Bayesian approach always outperforms the CE one, but if it does we find that for small sample size or more risk tolerant investors the loss is even greater.

Number of observations	Approach and size of contamination		value of b 0.1	2	10	23
	CE	0%	7.55	7.92	8.62	11.86
50	B	0%	5.67	5.72	5.64	5.85
	CE	5%	6.11	6.34	9.56	31.03
	B	5%	42.32	43.37	44.98	45.98
	CE	0%	3.61	3.49	3.95	4.90
100	B	0%	2.93	2.94	2.87	3.03
	CE	5%	3.32	3.61	6.59	21.72
	B	5%	19.68	20.41	19.63	19.44
	CE	0%	1.39	1.42	1.52	2.01
250	B	0%	1.19	1.18	1.17	1.20
	CE	5%	2.70	2.76	4.41	12.41
	B	5%	7.52	7.60	7.31	7.20
	CE	0%	0.32	0.33	0.36	0.46
1000	B	0%	0.30	0.29	0.30	0.29
	CE	5%	2.83	2.82	3.22	4.92
	B	5%	1.80	1.80	1.74	1.81

Table 3.3: Percent loss (%) in expected utility using CE and Bayesian approach with and without contamination.

These results highlight that the normality assumption is necessary for the Bayesian method to work well. Otherwise a procedure is needed that still guarantees its property of optimal tool. In the next sections we formalize the problem and derive some measures of sensitivity which are helpful to check the robustness properties of Bayesian weights.

3.4 Robustness of Bayesian mean-variance portfolio selection

3.4.1 Defining the problem

In portfolio selection problem the multivariate normal distribution is assumed to be the returns generating process. However it is well known that observed returns are often not normal and securities have sometimes unexpected high or low values. If this is the case we may be interested that such outlying observations do not play a strong influence on portfolio composition, that is we would like a robust estimator of portfolio weights. More generally with the word *robust* we define the insensitivity of a statistical procedure to deviation from the assumptions.

In this paper we focus on Bayesian mean-variance portfolio selection. In Chapter 1 we showed that robustness evaluation in a Bayesian setting involves the prior, the sampling distribution and the data. As first we are interested in developing measures of sensitivity of portfolio composition to distributional assumptions. Such quantities reveal the effect on Bayesian weights of perturbing the base model in different directions. For the sampling model it also turns out to be a useful tool for detecting the presence of outlying observations. Secondly we assess the influence that a single observation plays on the portfolio composition derived under a specific choice of prior/sampling models. Then we would like to use the Bayesian estimation procedure presented in Chapter 2 to obtain Bayesian weights that work well even if some outlying observations are present.

The first step is to see the Bayesian weights as a function of the three distinct elements we previously mentioned. We will do it in the next section.

3.4.2 Bayesian weights as functional

In this section we write the Bayesian weights in a functional form. We will denote a probability distribution and its corresponding cumulative distribution function by capital letters. If it exists, we use small letters for the density function.

In Section 3.3 we assume the normality of future returns r. In order to allow analytical calculations, we suppose covariance matrix Σ to be known[8]. Therefore we denote by F_θ the multivariate normal model $N(\theta, \Sigma)$. We assume that the same distribution generated also past observations (sampling model). The prior Π for parameter θ is chosen to be $N(\theta_0, \Sigma_0)$ and the empirical distribution is denoted by $F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \Delta_{x_i}(y)$ where $\Delta_x(y)$ is the Dirach distribution which puts mass 1 at x.

Under these assumptions, the predictive distribution of future returns turns out to be $N\left(\theta_{r|x}, \Sigma_{r|x}\right)$ with

$$\begin{aligned} \theta_{r|x} &= \theta_{\theta|x} \\ &= \Sigma_{\theta|x}\left[n\Sigma^{-1}\overline{x} + \Sigma_0^{-1}\theta_0\right] \end{aligned}$$

and

$$\begin{aligned} \Sigma_{r|x} &= \Sigma + \Sigma_{\theta|x} \\ &= \Sigma + \left[n\Sigma^{-1} + \Sigma_0^{-1}\right]^{-1}. \end{aligned}$$

Such moments depend only on the moments of posterior distribution $P_{\theta|x}$ denoted by $\theta_{\theta|x}$ and $\Sigma_{\theta|x}$. This result holds for any posterior distribution as long as we adopt such normal distribution[9] for future returns. A detailed proof of this result is given in Appendix E. It is worth to notice

[8] Note that this assumption seems quite realistic, as its estimate is more stable over time (Merton, 1980) and plays a minor role with respect to the mean estimate. In pratice, Σ is simply replaced by its sample estimate S.

[9] This hypothesis can also be relaxed and we can compute the moments of the predictive distribution directly.

the structure of predictive moments. First of all, the predictive mean coincides with the posterior mean. This means that the best prediction for expected future returns given past observations is reasonably the expected value of the random variable θ according to the distribution we obtain after updating our prior information with the observed returns x. However, the key point of this approach can be recognized by looking at $\Sigma_{r|x}$. The predictive variance accounts both for the variability of the data generating process Σ (assumed known) and for the variability of expected returns θ measured by its posterior variance $\Sigma_{\theta|x}$. Therefore the investor takes into account the risk of the estimation process into the variance structure.

The solution of the maximization problem (3.4) gives the Bayesian weights

$$\begin{aligned} q_{_B} &= \frac{1}{b}\Sigma_{r|x}^{-1}(\theta_{r|x} - \lambda_{r|x}\cdot\mathbf{1}) \\ &= \frac{1}{b}\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}(\theta_{\theta|x} - \lambda_{\theta|x}\cdot\mathbf{1}), \end{aligned}$$

where $\lambda_{r|x} = \frac{\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{r|x}-b}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}} = \frac{\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\theta_{r|x}-b}{\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\mathbf{1}} = \lambda_{\theta|x}$.

Such quantity can be seen as a function of the posterior distribution and hence, as we show in Section 1.3, as a function of the data, the prior and the sampling model. We have:

$$\begin{aligned} q_{_B} &= q_{_B}(P_{\theta|x}) \\ &= q_{_B}(F_n, \Pi, F_\theta) \\ &= \frac{1}{b}\left[\Sigma + \Sigma_{\theta|x}(F_n, \Pi, F_\theta)\right]^{-1}(\theta_{\theta|x}(F_n, \Pi, F_\theta) - \lambda_{\theta|x}(F_n, \Pi, F_\theta)\cdot\mathbf{1}). \end{aligned}$$

For a shorter notation we will denote only by $q_{_B}$, $\Sigma_{\theta|x}$, $\theta_{\theta|x}$ and $\lambda_{\theta|x}$ the functionals under the base distributions. We can now assess the sensitivity of Bayesian weights both to distributional assumptions and to observations. Since in next sections such sensitivity is considered for one model a time, we will denote the Bayesian functional as a function of the only distribution under study.

3.4.3 Sensitivity measures for Bayesian weights

In this section we derive measures of sensitivity of portfolio composition in a neighborhood of the assumed models. We represent such neighborhood by usual ε-contamination classes and we compute the measures presented in Section 1.4.

We first consider prior distribution Π. The local influence measure of Bayesian weights when the prior is perturbed in the direction of a generic contaminating distribution $Q \in \widetilde{Q}$ is given by

$$\begin{aligned} LI\left(Q; q_{_B}, \Pi\right) &= \left[\frac{\partial q_{_B}(\Pi_\varepsilon)}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &= \frac{1}{b}\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\left[LI\left(Q; \theta_{\theta|x}, \Pi\right) - LI\left(Q; \lambda_{\theta|x}, \Pi\right)\cdot \mathbf{1}\right. \\ &\quad \left. -LI\left(Q; \Sigma_{\theta|x}, \Pi\right)\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\left(\theta_{\theta|x} - \lambda_{\theta|x}\cdot \mathbf{1}\right)\right] \\ &= \frac{1}{b}\Sigma_{r|x}^{-1}\cdot\left(LI\left(Q; \theta_{\theta|x}, \Pi\right) - LI\left(Q; \lambda_{\theta|x}, \Pi\right)\cdot \mathbf{1}\right) \\ &\quad -\Sigma_{r|x}^{-1}\cdot LI\left(Q; \Sigma_{\theta|x}, \Pi\right)\cdot q_{_B}. \end{aligned} \tag{3.6}$$

For detailed calculations and definitions of quantities involved see Appendix F.

A similar structure is found for the local influence measure of Bayesian weights when sampling model F_θ is perturbed in the direction of a generic contaminating distribution $G \in \widetilde{G}$. Such measure is given by:

$$\begin{aligned} LI\left(G; q_{_B}, F_\theta\right) &= \left[\frac{\partial q_{_B}(F_{\theta,\varepsilon})}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &= \frac{1}{b}\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\left[LI\left(G; \theta_{\theta|x}, F_\theta\right) - LI\left(G; \lambda_{\theta|x}, F_\theta\right)\cdot \mathbf{1}\right. \\ &\quad \left. -LI\left(G; \Sigma_{\theta|x}, F_\theta\right)\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\left(\theta_{\theta|x} - \lambda_{\theta|x}\cdot \mathbf{1}\right)\right] \\ &= \frac{1}{b}\Sigma_{r|x}^{-1}\cdot\left(LI\left(G; \theta_{\theta|x}, F_\theta\right) - LI\left(G; \lambda_{\theta|x}, F_\theta\right)\cdot \mathbf{1}\right) \\ &\quad -\Sigma_{r|x}^{-1}\cdot LI\left(G; \Sigma_{\theta|x}, F_\theta\right)\cdot q_{_B}. \end{aligned} \tag{3.7}$$

For detailed calculations and definition of quantities involved see Appendix G.

The stability of Bayesian weights in a neighborhood of the assumed model depends on the

stability of the first two moments of the posterior distribution, which are posterior summaries. We have shown in Chapter 1 that their local influence measures capture both a 'data effect' and a 'model effect' on the functional. For this reason, measures (3.6) and (3.7) are expected to reveal the total effect on the Bayesian weights of perturbations of the prior and sampling model respectively.

Moreover, it may be useful to put in relation the obtained sensitivity measure with the corresponding component of Bayesian weights. We therefore define relative measures of local influence for the prior and the sampling model to be respectively

$$LI^{*}\left(\widetilde{Q};q_{_B};\Pi\right)=\sup_{Q\in\widetilde{Q}}\left|diag^{-1}\left(q_{_B}\right)\cdot LI\left(Q;q_{_B},\Pi\right)\right|$$

and

$$LI^{*}\left(\widetilde{G};q_{_B};F_{\theta}\right)=\sup_{G\in\widetilde{G}}\left|diag^{-1}\left(q_{_B}\right)\cdot LI\left(G;q_{_B},F_{\theta}\right)\right|, \tag{3.8}$$

which give the absolute component by component maximum relative effect as the distribution moves locally around the base model in different directions.

Once a certain combination of prior/sampling model is chosen, portfolio composition becomes a function only of observations. To see the component by component change of $q_{_B}$ as a single observation in the sample is moved, we use the Sensitivity Curve (see Chapter 1). For the Bayesian weights it is defined as

$$\begin{aligned}SC(z) &= \frac{T}{b}\Big\{\left[\Sigma+\Sigma_{\theta|x}\left(F_n^z\right)\right]^{-1}\left(\theta_{\theta|x}\left(F_n^z\right)-\lambda_{\theta|x}\left(F_n^z\right)\cdot\mathbf{1}\right)+\\ &\quad-\left[\Sigma+\Sigma_{\theta|x}\left(F_{n-1}\right)\right]^{-1}\left(\theta_{\theta|x}\left(F_{n-1}\right)-\lambda_{\theta|x}\left(F_{n-1}\right)\cdot\mathbf{1}\right)\Big\},\end{aligned}$$

where $F_{n-1}=(x_1,..,x_{_{T-1}})$ is the empirical distribution of the sample of $(n-1)$ observations and $F_n^z=(x_1,..,x_{_{T-1}},z)$ is the sample in which observation z has been added. If this measure diverges as z becomes larger, the functional is said to be non robust with respect to observations.

In the next section we explore the behavior of such local sensitivity measures.

3.4.4 Performance of sensitivity measures

In the previous sections we derived different measures of sensitivity of the portfolio composition. In this section we assess their performance by means of a simulation study. In order to obtain analytical formulations we consider normal contaminating distributions for both the prior and the sampling model.

We let contaminating prior Q to vary within the set $\{N(\theta_0, \Psi_0) : \Psi_0 = c \cdot \Sigma_0, \quad c \in [0.2, 3]\}$, which allow smaller and greater prior precision around the same prior mean. The measure of local influence of Bayesian weights to perturbations of the prior in the direction of Q is then given by (3.6) with

$$\begin{aligned} \theta_{\theta|x}(Q) &= \Sigma_{\theta|x}(Q)\left[n\Sigma^{-1}\overline{x} + \Psi_0^{-1}\theta_0\right], \\ \Sigma_{\theta|x}(Q) &= \left[n\Sigma^{-1} + \Psi_0^{-1}\right]^{-1} \end{aligned}$$

and

$$\begin{aligned} \frac{m(x; Q, F_\theta)}{m(x)} &= \frac{|\Psi_0|^{-1/2}\left|\Sigma_{\theta|x}(Q)\right|^{1/2}}{|\Sigma_0|^{-1/2}\left|\Sigma_{\theta|x}\right|^{1/2}} \exp\left\{-\frac{1}{2}\theta_0'\left(\Psi_0^{-1} - \Sigma_0^{-1}\right)\theta_0 + \right. \\ &\left. + \frac{1}{2}\theta_{\theta|x}(Q)'\,\Sigma_{\theta|x}^{-1}(Q)\,\theta_{\theta|x}(Q) - \frac{1}{2}\theta_{\theta|x}'\Sigma_{\theta|x}^{-1}\theta_{\theta|x}\right\}. \end{aligned}$$

For the sampling model we consider contaminating distribution G that moves within the set $\{N(\theta, \Omega) \quad : \Omega = d \cdot \Sigma, d \in [1, 3]\}$. Such family allows to increase the volatility of portfolio asset returns without changing the correlation structure between assets. The measure of local influence of Bayesian weights is then given by (3.7) with

$$\begin{aligned} \theta_{\theta|x}^{(j)}(F_\theta, G) &= \Sigma_{\theta|x}^{(j)}(F_\theta, G)\left[(n-1)\Sigma^{-1}\overline{x}_{(-j)} + \Omega^{-1}x_j + \Sigma_0^{-1}\theta_0\right], \\ \Sigma_{\theta|x}^{(j)}(F_\theta, G) &= \left[(n-1)\Sigma^{-1} + \Omega^{-1} + \Sigma_0^{-1}\right]^{-1} \end{aligned}$$

and

$$\begin{aligned}\frac{m_j\left(x\,;\Pi,F_\theta,G\right)}{m\left(x\,;\Pi,F_\theta\right)} &= \frac{|\Sigma|^{\frac{1}{2}}|W|^{\frac{1}{2}}}{|\Omega|^{\frac{1}{2}}|V|^{\frac{1}{2}}}\exp\left\{-\frac{(n-1)}{2}tr(\Sigma^{-1}S_{(-j)})+\right.\\ &\frac{n}{2}tr(\Sigma^{-1}S)-\frac{1}{2}x_j'\Omega^{-1}x_j+\frac{n}{2}\overline{x}'\Sigma^{-1}\overline{x}+\\ &-\frac{(n-1)}{2}\overline{x}_{(-j)}'\Sigma^{-1}\overline{x}_{(-j)}-\frac{1}{2}\theta_{\theta|x}'\Sigma_{\theta|x}^{-1}\theta_{\theta|x}+\\ &\left.+\frac{1}{2}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)'\left[\Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right)\right]^{-1}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\right\},\end{aligned}$$

where $\overline{x}_{(-j)}$ and $S_{(-j)}$ are respectively the sample mean and covariance matrix computed dropping observation j from the sample. Analytical calculations of marginal likelihoods can be found in Appendix H.

We simulate T sets ($T=30$) of n returns ($n=260$) -corresponding to one year of observations- from a k-variate normal ($k=6$) with parameters given in Table 3.4. Parameters for the simulation study are estimated from real market data. We consider daily returns of stock indexes of the six major European countries in the period January 1995-December 2003. We use sample estimates in the period January 1995-December 1997 as parameters for prior distribution Π and sample estimates in the next period (January 1998-December 2003) as parameters for sampling model F_θ. Such choice of prior parameters reflect a positive view with high expected returns and small volatilities and correlations. Contaminated samples are generated from the model $(1-\varepsilon)F_\theta+\varepsilon G$ where $\varepsilon=0.05$ and the contaminating distribution G is a $N(\theta,3\cdot\Sigma)$.

Sample estimates of daily return (January 1995-December 1997)

	Mean	Covariance Matrix					
UK	0.0889	0.0082					
FR	0.0596	0.0062	0.0121				
CH	0.1127	0.0045	0.0056	0.0087			
GE	0.0914	0.0053	0.0068	0.0060	0.0117		
IT	0.0680	0.0054	0.0078	0.0045	0.0058	0.0166	
SP	0.1091	0.0052	0.0072	0.0049	0.0062	0.0068	0.0105

Sample estimates of daily return (January 1998-December 2003)

	Mean	Covariance Matrix					
UK	-0.0127	0.0209					
FR	0.0109	0.0190	0.0269				
CH	-0.0058	0.0159	0.0188	0.0238			
GE	-0.0044	0.0193	0.0251	0.0192	0.0344		
IT	0.0044	0.0166	0.0219	0.0170	0.0225	0.0258	
SP	0.0041	0.0163	0.0219	0.0110	0.0220	0.0208	0.0260

Table 3.4: Parameters of the prior and sampling models respectively in the second simulation study. Euro returns in percent per day.

Figure 3.4 and Figure 3.5 show the average sensitivity measures of Bayesian weights to perturbation in the prior over the T non contaminated and contaminates sets respectively. Perturbations in the direction of more precise prior play a greater influence. However, the LI measure is small over all the set of contaminating models and Bayesian portfolio components are stable to perturbations of the prior[10]. This is confirmed by relative measures in Table 3.5.

[10] We notice that the size of local influence measures decreases with increasing n, i.e. prior information plays a

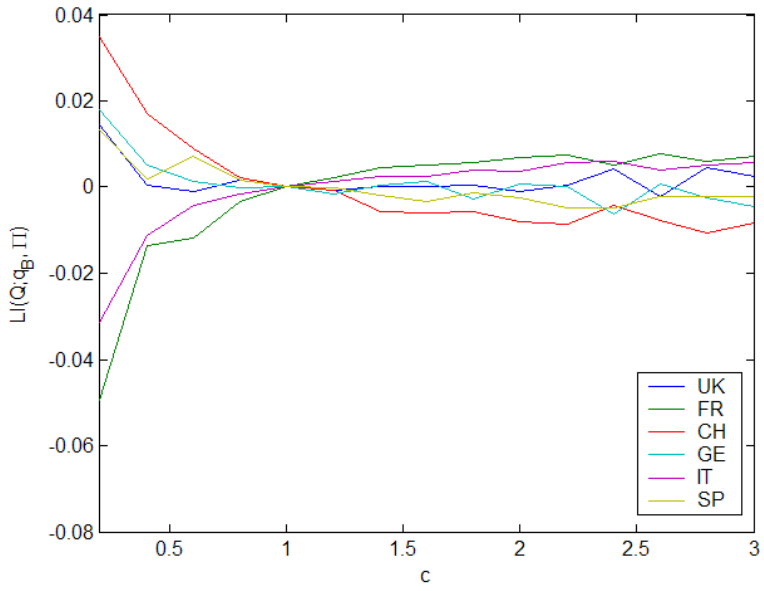

Figure 3.4: $LI\left(Q;q_{_B},\Pi\right)$ measure for Bayesian weights. Non contaminated data.

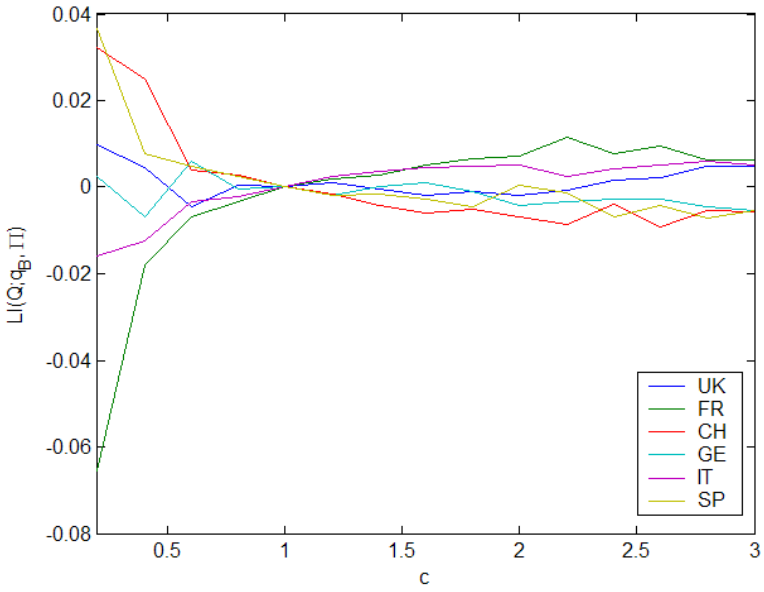

Figure 3.5: $LI\left(Q;q_{_B},\Pi\right)$ measure for Bayesian weights. Contaminated data.

Moreover the plot of the LI measure does not change in the contaminated case. These two

minor role when the number of observations increases.

facts support a more general intuition. When the number of observation is large the value of the Bayesian estimate does not depend heavily on the prior and hence sensitivity to our prior choice is negligible. Therefore with a reasonable amount of past returns such estimates incorporate estimation risk without depending too much on the choice is negligible. Therefore with a reasonable amount of past returns such estimates incorporate estimation risk without depending too much on the choice of the prior.

$\varepsilon\%$			UK	FR	CH	GE	IT	SP
0%	$LI^*\left(\widetilde{Q};T_B;\Pi\right)$		0.0551	0.0705	0.0533	0.0277	0.7984	0.0349
0%	$LI^*\left(\widetilde{G};T_B;F\right)$		9.3968	7.1536	30.2748	13.4007	15.5017	5.1747
5%	$LI^*\left(\widetilde{Q};T_B;\Pi\right)$		0.0592	0.0635	0.6446	0.0732	0.0280	0.0417
5%	$LI^*\left(\widetilde{G};T_B;F\right)$	$10^5\cdot$	[2.3936	0.2572	0.4357	0.3525	0.6610	0.5550]

Table 3.5: Average relative local sensitivity measures of estimated Bayesian weights with respect to the base prior and sampling models under non contaminated and contaminated samples.

This is not the case when considering local influence measures for the sampling model (Figures 3.6-3.7). Portfolio components are now more sensitive to perturbations. Under non contaminated data the local influence measure remains small up to d around 2.4 and increases after that point, leading to quite large relative measures in Table 3.5. We check both the size of the "data effect" and of the "model effect" for each posterior summaries in measure (3.7). We find that this increasing pattern is due to the model effect, since Bayes factors remain reasonably small over all the set of contaminating distributions. Figure 3.7 shows the effect of introducing few outlying observations. The sensitivity measure for each component increases up to order 10^5. Hence the LI measure for

the sampling model is found to be a useful tool for detecting the presence of 'extreme' observations where no visual representation would be possible because of the multivariate nature of the data.

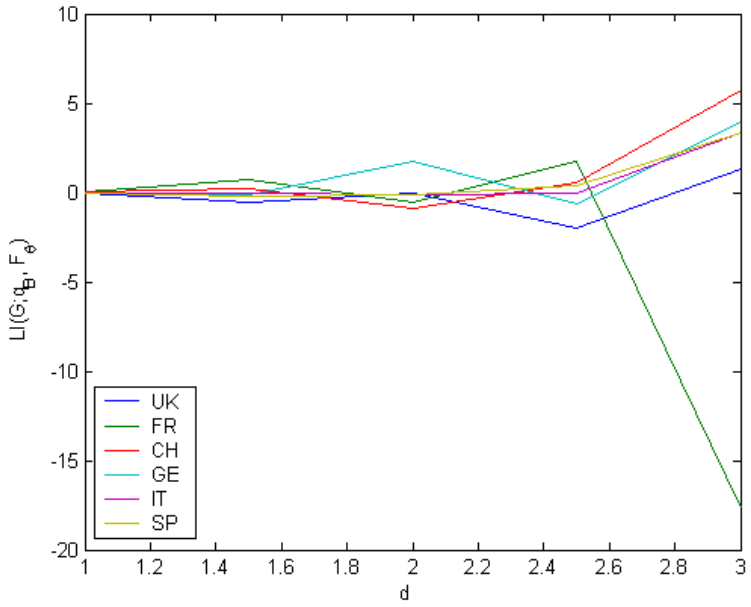

Figure 3.6: $LI\,(G; q_{_B}, F_\theta)$ measure for Bayesian weights. Non contaminated data.

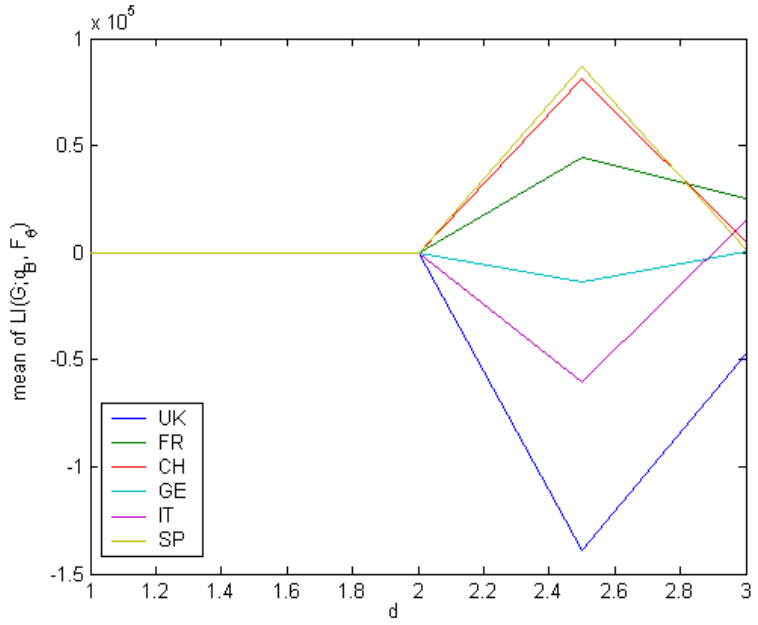

Figure 3.7: $LI\,(G; q_{_B}, F_\theta)$ measure for Bayesian weights. Contaminated data.

Finally we want to measure the influence of a given observation on Bayesian weights. We first consider the case where outlying returns are added only for one security at a time (Figure 3.8). Such observations are chosen to vary between the minimum and the maximum observed return for each security in the period January 1995-December 2001. Of course, the weight which is most influenced corresponds to the one of the perturbed security. However all the other components are changed in the opposite direction because portfolio weights have to sum to one. It is also interesting to note the linear relationship between each weight component and the 'extreme' observation.

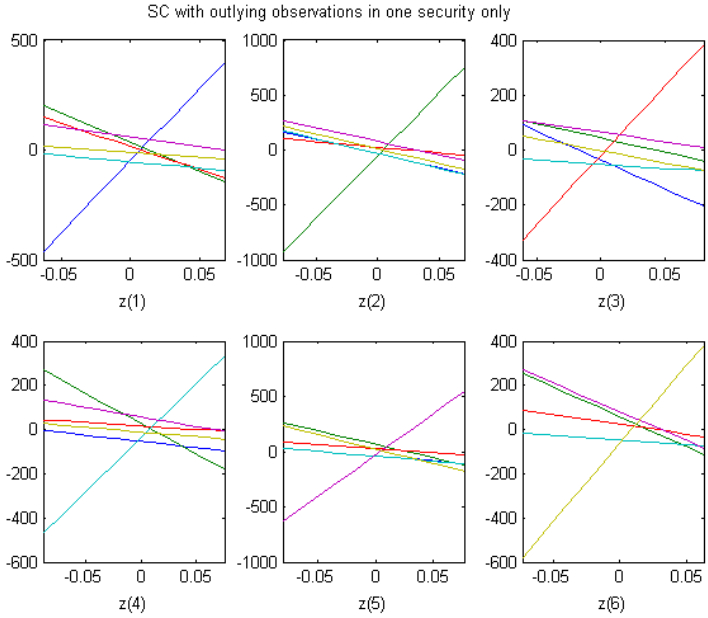

Figure 3.8: *SC* for Bayesian weights under normality of both prior and sampling distributions adding 'extreme' returns only in one security a time.

Figure 3.9 shows the *SC* of q_B when vector z is added. Such observation is chosen so that all its components vary between the minimum and the maximum observed return for each security in the portfolio. The sensitivity of the different components to observations are very different. While the weights of Italy and Germany increase with increasing outlying observations, the opposite

behavior is displayed for UK and Spain. The components of France and Switzerland appear almost insensitive. However, we can conclude that the effect of a single observation on Bayesian portfolio composition is potentially unbounded. In the next Section we propose a procedure that does not suffer from this problem.

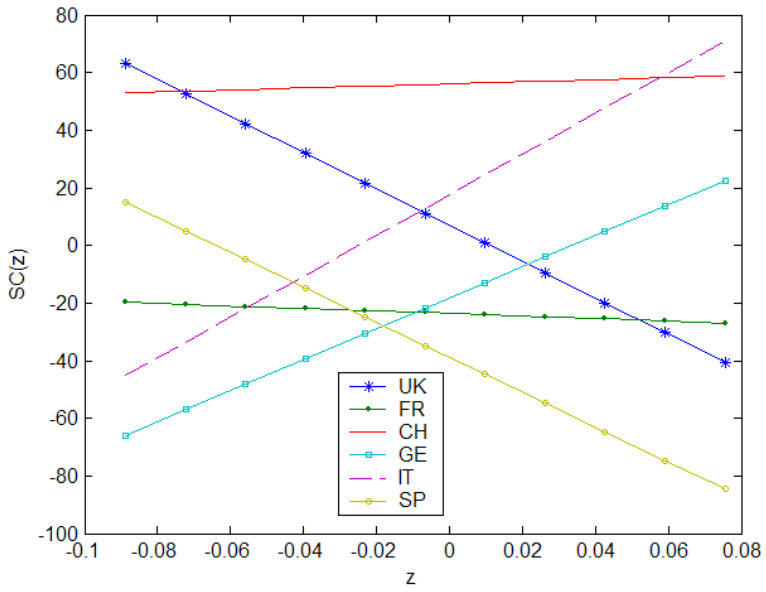

Figure 3.9: *SC* for Bayesian weights under normality of both prior and sampling distributions. Simulated daily returns.

3.5 Robust Bayesian weights

In the previous section we derived measures of sensitivity of Bayesian weights to different inputs. We found that portfolio components are extremely sensitive to observations because of the unboundedness of the SC. Moreover the local influence measure of the sampling model turns out to be a useful tool for detecting the presence of outlying observations. In this section we propose a robust estimation procedure for Bayesian weights, that is a procedure that works well even if only

the majority of the data fits the normal model.

When the presence of outliers is detected robust Bayesian procedure can be built by adopting a robust model for the sampling distribution. Such model consists in a location-scale family of symmetric unimodal distributions enriched with further parameters that control the shape, especially in the tails. Different robust models have been presented in Chapter 2 and generalized to the multivariate case.

Robust Bayesian mean-variance weights $q_B^{(R)}$ are therefore the solution of (3.4) where the robust predictive density of future returns is given by $p^{(R)}(r_P|x) = \int p(r_P|\theta)p^{(R)}(\theta|x)d\theta$ with $p^{(R)}(\theta|x)$ the robust posterior density obtained with a robust sampling model. However the resulting predictive distribution is no longer normal[11]. As we said in Section 3.4.2, robust Bayesian weights basically depend on the first two moments of posterior distribution since the normality assumption of future returns r holds, i.e.

$$q_B^{(R)} = \frac{1}{b}\left[\Sigma + \Sigma_{\theta|x}^{(R)}\right]^{-1}(\theta_{\theta|x}^{(R)} - \lambda_{\theta|x}^{(R)} \cdot \mathbf{1}). \tag{3.9}$$

Therefore $\theta_{\theta|x}^{(R)}$ and $\Sigma_{\theta|x}^{(R)}$ can be computed by means of $MCMC$ algorithms that generate a random drawn from $p^{(R)}(\theta|x)$.

When a robust model is assumed the effect of outlying observations is dampened and measure (3.7) is reduced. This latter fact can be used as a selection criterion among different robust estimates. We choose the robust weights (3.9) that display the smallest value of the LI measure for the sampling distribution. In the next section we will compute robust Bayesian estimates using real market data.

[11] Under the assumption of a quadratic utility function for investors, the Robust Bayesian mean-variance weights still satisfy the expected utility maximization paradigm.

3.6 Application to real data

In this section we implement the robust estimator of Bayesian weights using daily returns of stock indexes of the six major European countries from 1st January 2001 to 31th December 2001.

In order to establish the need for a robust estimation procedure we compute both the SC and measure (3.7) of Bayesian weights under a normal prior/normal sampling model assumptions[12].

Figure 3.10 shows that each component of the Bayesian weights is a linear function of a single observation.

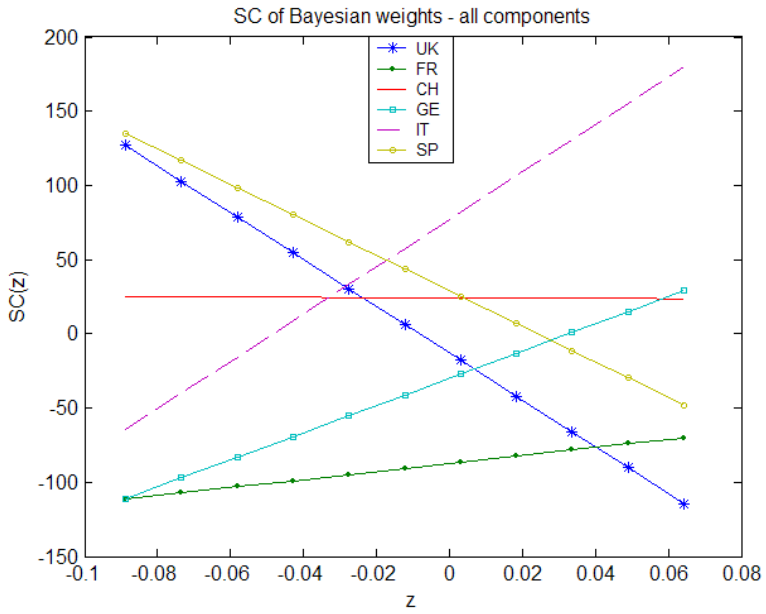

Figure 3.10: SC for Bayesian weights assuming a normal distribution for the sampling model. Daily returns over the period 1.1.2001-31.12.2001.

As expected, q_B are not robust to the presence of outliers. Moreover the computed relative LI

[12]The parameters for the prior distribution are the sample estimates of daily returns from the 1st January 1995 to 31th December 1997 in Table 3.4.

measure for the sampling distribution[13] shown in Table 3.6 is very large compared to the case in which data are generated from a normal distribution (see Table 3.5). Therefore robust estimation is clearly needed.

		UK	FR	CH	GE	IT	SP	Sum
q_B		4.0068	-4.5500	-0.2860	-0.1124	-3.5419	5.4835	1
$LI^*\left(G; q_B, F_\theta\right)$	$10^4 \cdot$	[0.0372	0.3347	0.0079	68.2439	0.2970	0.2280]	691489.61

Table 3.6: Analytical estimates of the Bayesian weights and their relative local influence measures for the sampling distribution.

We use a random walk Metropolis-Hasting algorithm (Hastings, 1970) to compute $\theta^{(R)}_{\theta|x}$ and $\Sigma^{(R)}_{\theta|x}$ under different robust sampling models. We made all the computations in Matlab and we checked the convergence of the chain and of the averages by means of BOA library in R. Estimates of Bayesian weights (3.9) and their relative local influence measures are shown in Table 3.7.

Under the normal model the components of the Bayesian weights are well estimated. LI^* measures are computed using estimator (1.17) (see 1.5) but estimated values are different from analytical results. This is probably due to the fact that (1.17) relies on the importance sampling technique and the choice of the posterior density as sampling density is not adequate. We tried to estimate LI^* using estimator (1.16) that uses density (1.13), but the results were even worse.

[13] Measure (3.8) is calculated assuming a normal contaminating model G with mean θ and variance $\Omega = 3 \cdot \Sigma$.

	UK	FR	CH	GE	IT	SP	
				Normal			Sum
q_B	4.0047	-4.5491	-0.2889	-0.1031	-3.5229	5.4593	1
LI^*	2455.28	2781.77	27299.47	159375.39	3629.39	1257.99	196799.30
				Student-t (15)			
$q_B^{(S)}$	4.0298	-4.8719	-0.7565	-0.3808	-2.0476	5.0271	1
LI^*	4.37	0.06	14.49	30.51	5.84	0.62	55.88
				EPS (0.3)			
$q_B^{(EPS)}$	3.9479	-4.3803	-0.6596	-0.4331	-2.6897	5.2148	1
LI^*	8.67	3.75	33.33	81.74	12.49	0.99	140.97
				EP (8;0)			
$q_B^{(EP)}$	3.9799	-5.2537	-1.1010	-0.4641	-1.6224	5.4614	1
LI^*	1.29	0.32	3.17	4.21	0.63	0.51	10.14
				RN (0.05;2)			
$q_B^{(RN)}$	3.8091	-4.3510	-0.8678	-0.4971	-1.8808	4.7877	1
LI^*	1.31	0.50	3.69	3.08	0.47	0.61	9.67

Table 3.7: Numerical estimates of the Bayesian weights and their relative local influence measures for the sampling distribution under different sampling models. MCMC simulations with 500.000 runs.

However, what it is more important is to look at the effect of assuming a robust model for the sampling distribution on LI^*. For all robust models considered such measure is drastically reduced and the sum of all its components declines up to 9.67. Therefore robust Bayesian weights are obtained under the RN distribution with parameters $(0.05; 2)$. Note that the position held in the robust portfolio is the same (i.e. long in the UK and Spain market indexes and short in the others). Only the proportion invested in each security changes. Moreover as expected the SC of $q_B^{(RN)}$ in Figure 3.11 shows a bounded behavior.

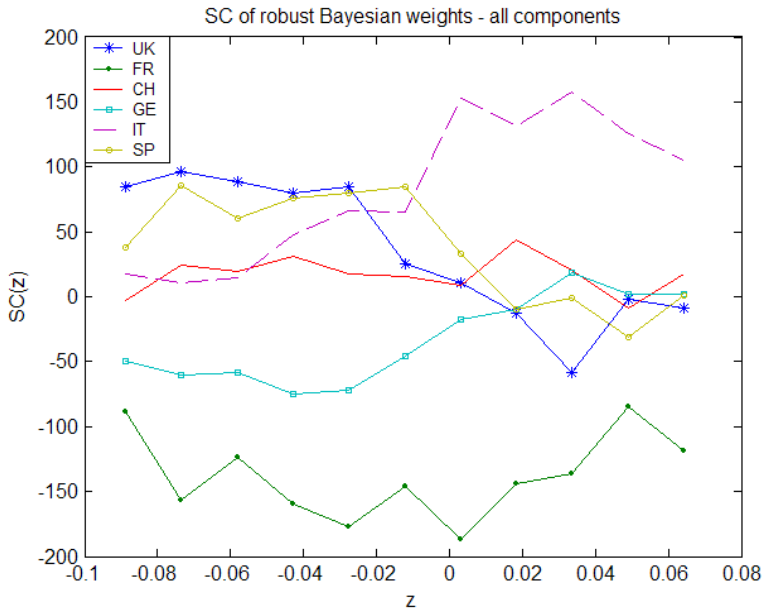

Figure 3.11: SC for Bayesian weights assuming a RN distribution for the sampling model. MCMC simulations with 250.000 runs. Daily returns over the period 1.1.2001-31.12.2001.

We now focus on the estimates of the posterior moments under the normal and the RN sampling distributions (Table 3.8 and Table 3.9). Estimated values in the normal case are very close to the values computed analytically. Under the robust distribution the estimated mean of expected returns increases for Italy and decreases for Switzerland and Spain, whereas the covariance matrix

of the expected return becomes larger. This is a crucial point: using robust estimation procedures for the Bayesian weights does not underestimate the risk of the investment which is given by the predictive variance $\Sigma_{r|x} = \Sigma + \Sigma_{\theta|x}$. Matrix $\Sigma_{\theta|x}$ takes into account the fact that the robust mean of expected returns is obtained dampening the effect of some observations and therefore there is a loss in efficiency which is typical of robust estimation procedures.

	Mean expected returns $\theta_{\theta\|x}$		
	Analytical estimates	Normal MCMC estimates	RN MCMC estimates
UK	-0.000555	-0.000576 $\left(1.19 \cdot 10^{-5}\right)$	-0.000597 $\left(1.43 \cdot 10^{-5}\right)$
FR	-0.000934	-0.000967 $\left(1.28 \cdot 10^{-5}\right)$	-0.000929 $\left(1.39 \cdot 10^{-5}\right)$
CH	-0.000799	-0.000828 $\left(1.18 \cdot 10^{-5}\right)$	-0.000913 $\left(1.44 \cdot 10^{-5}\right)$
GE	-0.000826	-0.000869 $\left(1.45 \cdot 10^{-5}\right)$	-0.000867 $\left(1.77 \cdot 10^{-5}\right)$
IT	-0.001148	-0.001181 $\left(1.31 \cdot 10^{-5}\right)$	-0.000996 $\left(1.64 \cdot 10^{-5}\right)$
SP	-0.000294	-0.000333 $\left(1.32 \cdot 10^{-5}\right)$	-0.000356 $\left(1.60 \cdot 10^{-5}\right)$

Table 3.8: Comparison of analytical and numerical estimates of the posterior mean under the normal and the selected robust sampling models. For the MCMC estimates the error is given in parenthesis. Euro returns in percent per day.

Covariance Matrix of expected returns $\Sigma_{\theta|x}$

	Analytical estimates					
UK	$8.29 \cdot 10^{-7}$					
FR	$7.61 \cdot 10^{-7}$	$9.95 \cdot 10^{-7}$				
CH	$6.35 \cdot 10^{-7}$	$6.93 \cdot 10^{-7}$	$8.62 \cdot 10^{-7}$			
GE	$8.09 \cdot 10^{-7}$	$9.63 \cdot 10^{-7}$	$7.38 \cdot 10^{-7}$	$1.25 \cdot 10^{-6}$		
IT	$7.15 \cdot 10^{-7}$	$8.99 \cdot 10^{-7}$	$6.84 \cdot 10^{-7}$	$9.61 \cdot 10^{-7}$	$1.06 \cdot 10^{-6}$	
SP	$6.52 \cdot 10^{-7}$	$8.84 \cdot 10^{-7}$	$6.40 \cdot 10^{-7}$	$9.04 \cdot 10^{-7}$	$8.72 \cdot 10^{-7}$	$1.07 \cdot 10^{-6}$
	Normal MCMC estimates					
UK	$8.39 \cdot 10^{-7}$					
FR	$7.75 \cdot 10^{-7}$	$1.01 \cdot 10^{-6}$				
CH	$6.36 \cdot 10^{-7}$	$6.96 \cdot 10^{-7}$	$8.62 \cdot 10^{-7}$			
GE	$8.18 \cdot 10^{-7}$	$9.71 \cdot 10^{-7}$	$7.39 \cdot 10^{-7}$	$1.25 \cdot 10^{-6}$		
IT	$7.21 \cdot 10^{-7}$	$9.09 \cdot 10^{-7}$	$6.83 \cdot 10^{-7}$	$9.63 \cdot 10^{-7}$	$1.06 \cdot 10^{-6}$	
SP	$6.66 \cdot 10^{-7}$	$8.97 \cdot 10^{-7}$	$6.43 \cdot 10^{-7}$	$9.09 \cdot 10^{-7}$	$8.74 \cdot 10^{-7}$	$1.08 \cdot 10^{-6}$
	RN MCMC estimates					
UK	$1.17 \cdot 10^{-6}$					
FR	$1.08 \cdot 10^{-6}$	$1.41 \cdot 10^{-6}$				
CH	$8.99 \cdot 10^{-7}$	$9.81 \cdot 10^{-7}$	$1.22 \cdot 10^{-6}$			
GE	$1.14 \cdot 10^{-6}$	$1.36 \cdot 10^{-6}$	$1.05 \cdot 10^{-6}$	$1.78 \cdot 10^{-6}$		
IT	$1.01 \cdot 10^{-6}$	$1.28 \cdot 10^{-6}$	$9.74 \cdot 10^{-7}$	$1.37 \cdot 10^{-6}$	$1.52 \cdot 10^{-6}$	
SP	$9.31 \cdot 10^{-7}$	$1.25 \cdot 10^{-6}$	$9.14 \cdot 10^{-7}$	$1.29 \cdot 10^{-6}$	$1.24 \cdot 10^{-6}$	$1.51 \cdot 10^{-6}$

Table 3.9: Comparison of analytical and numerical estimates of the posterior covariance matrix under the normal and the selected robust sampling models.

3.7 Conclusions and future developments

This paper discusses the robustness properties of the Bayesian mean-variance portfolio weights. We first initially compare the CE and the Bayesian approaches to portfolio selection when data come from a normal distribution or contain some outlying observations. We find that the presence of 'extreme' returns makes the Bayesian method lose its property of outperforming the CE method for small sample sizes or more risk tolerant investors.

Then we compute measures of local sensitivity both to distributional assumptions and to observations. Their behavior is explored by means of a simulation study. We find that LI measure for the prior is very small, which means that results do not depend heavily on prior assumptions with a sample greater than $n = 260$ observations. Moreover the LI measure for the sampling model turns out to be a useful tool for revealing the presence of outlying observations. Our result is useful because the multivariate nature of the data makes this task difficult to achieve otherwise. We also find that Bayesian weights are extremely sensitive to observations when a normal prior/normal sampling models combination is assumed.

Finally we propose a Bayesian estimation procedure for portfolio weights which dampenes the effect of 'extreme' observations. We consider an application to real market data. The unboundedness of the SC and the large size of the LI measure of Bayesian weights reveal that robust estimation is needed. We then compute the robust Bayesian weights and their LI measures turn out to be much smaller than in the normal case. A final important remark: when a robust model is adopted the effect of outlying observations is dampened on the estimated mean of expected returns, but its covariance matrix becomes larger. This means that using robust estimation procedures for the Bayesian weights does not underestimate the risk of the investment.

A possible direction for future research would be to build the robust estimator for Bayesian weights when the hypothesis of known covariance matrix is relaxed and also a prior on this parameter is assumed as in Frost and Savarino (1986).

Appendix A

Consider a linear perturbation of the sampling distribution of type (1.3) with G the contaminating distribution. The perturbed posterior density is given by

$$p_{\varepsilon}\left(\theta|x\right)=\frac{\pi\left(\theta\right)\cdot L_{F_{\varepsilon}}\left(x|\theta\right)}{m\left(x\,;\Pi,F_{\theta,\varepsilon}\right)},$$

and its derivative

$$\begin{aligned}
\left[\frac{\partial p_{\varepsilon}\left(\theta|x\right)}{\partial\varepsilon}\right]_{\varepsilon=0} &= \left[\frac{\left(\pi\left(\theta\right)\frac{\partial L_{F_{\varepsilon}}\left(x|\theta\right)}{\partial\varepsilon}\right)m\left(x\,;\Pi,F_{\theta,\varepsilon}\right)}{m\left(x\,;\Pi,F_{\theta,\varepsilon}\right)^{2}}+\right.\\
&\quad\left.-\frac{\left(\pi\left(\theta\right)\cdot L_{F_{\varepsilon}}\left(x|\theta\right)\right)\left(\frac{\partial m(x\,;\Pi,F_{\theta,\varepsilon})}{\partial\varepsilon}\right)}{m\left(x\,;\Pi,F_{\theta,\varepsilon}\right)^{2}}\right]_{\varepsilon=0}\\
&= \frac{\pi\left(\theta\right)\cdot\sum_{j=1}^{n}\left[\left(g\left(x_{j}\right)-f_{\theta}\left(x_{j}\right)\right)\prod_{i\neq j}f_{\theta}\left(x_{i}\right)\right]}{m\left(x\,;\Pi,F_{\theta}\right)}-\\
&\quad\frac{\pi\left(\theta\right)\;L_{F}\left(x|\theta\right)\sum_{j=1}^{n}\left[m_{j}\left(x\,;\Pi,F_{\theta},G\right)-m\left(x\,;\Pi,F_{\theta}\right)\right]}{m\left(x\,;\Pi,F_{\theta}\right)^{2}}\\
&= \sum_{j=1}^{n}p_{j}\left(\theta|x\right)\frac{m_{j}\left(x\,;\Pi,F_{\theta},G\right)}{m\left(x\,;\Pi,F_{\theta}\right)}-n\cdot p\left(\theta|x\right)\\
&\quad-\pi\left(\theta|x\right)\sum_{j=1}^{n}\frac{m_{j}\left(x\,;\Pi,F_{\theta},G\right)}{m\left(x\,;\Pi,F_{\theta}\right)}+n\cdot p\left(\theta|x\right)\\
&= \sum_{j=1}^{n}\frac{m_{j}\left(x\,;\Pi,F_{\theta},G\right)}{m\left(x\,;\Pi,F_{\theta}\right)}\left[p_{j}\left(\theta|x\right)-p\left(\theta|x\right)\right],
\end{aligned}$$

where

$$p_{j}\left(\theta|x\right)=\frac{\pi\left(\theta\right)\cdot g\left(x_{j}\right)\cdot\prod_{i\neq j}f_{\theta}\left(x_{i}\right)}{m_{j}\left(x\,;\Pi,F_{\theta},G\right)}$$

is the posterior obtained when a sampling distribution G is adopted only for observation j and

$$m_j\left(x\,;\Pi,F_\theta,G\right)=\int g\left(x_j\right)\cdot\prod_{i\neq j}f_\theta\left(x_i\right)\pi\left(\theta\right)d\theta$$

is the corresponding marginal likelihood.

The measure of local influence of the functional to the sampling model is therefore given by

$$\begin{aligned}
LI(G;T_B,F_\theta) &= \int\rho\left(\theta\right)\left[\frac{\partial p_\varepsilon\left(\theta|x\right)}{\partial\varepsilon}\right]_{\varepsilon=0}d\theta \\
&= \int\rho\left(\theta\right)\cdot\sum_{j=1}^{n}\left[\frac{m_j\left(x\,;\Pi,F_\theta,G\right)}{m\left(x\,;\Pi,F_\theta\right)}\cdot\left(p_j\left(\theta|x\right)-p\left(\theta|x\right)\right)\right]d\theta \\
&= \sum_{j=1}^{n}\frac{m_j\left(x\,;\Pi,F_\theta,G\right)}{m\left(x\,;\Pi,F_\theta\right)}\int\rho\left(\theta\right)\cdot\left(p_j\left(\theta|x\right)-p\left(\theta|x\right)\right)d\theta. \qquad \text{(A.1)}
\end{aligned}$$

Expression (A.1) takes different forms depending on the contaminating distribution G. If G is a distribution with parameter θ, we denote it by G_θ. Local influence measure of T_B is then given by:

$$LI(G_\theta;T_B,F_\theta)=\sum_{j=1}^{n}\frac{m_j\left(x\,;\Pi,F_\theta,G_\theta\right)}{m\left(x\,;\Pi,F_\theta\right)}\cdot\left(T_B^{(j)}\left(F_\theta,G_\theta\right)-T_B\left(F_\theta\right)\right)$$

where $m_j(x;\Pi,F_\theta,G_\eta)=\int g_\theta\left(x_j\right)\prod_{i\neq j}f_\theta\left(x_i\right)\,\pi\left(\theta\right)d\theta$.

If G depends on a different known parameter η $(\eta\neq\theta)$, the contaminating distribution is denoted by G_η and (A.1) turns out to be

$$LI(G_\eta;T_B,F_\theta)=\sum_{j}\frac{m_j(x;\Pi,F_\theta,G_\eta)}{m(x)}\left(T_B^{(-j)}-T_B\right)$$

where $m_j(x;\Pi,F_\theta,G_\eta)=g_\eta\left(x_j\right)\cdot\int\prod_{i\neq j}f_\theta\left(x_i\right)\pi\left(\theta\right)d\theta$ and $T_B^{(-j)}$ is the posterior functional under base models using sample x without observation x_j.

Appendix B

Assume prior Π and sampling model F_θ to be respectively $N(\theta_0, \sigma_0^2)$ and $N(\theta, \sigma^2)$. We need to compute the marginal likelihood $m(x) = \int L_F(x|\theta)\,\pi(\theta)\,d\theta$ where $L_F(x|\theta)$ is the likelihood under the reference sampling model. It is well known that in this case the posterior is $N\left(\theta_{\theta|x}; \sigma^2_{\theta|x}\right)$ with $\theta_{\theta|x} = \frac{n\sigma_0^2\overline{x}+\sigma^2\theta_0}{n\sigma_0^2+\sigma^2}$ and $\sigma^2_{\theta|x} = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}$. Our quantity of interest turns out to be:

$$\begin{aligned} m(x) &= \int \pi(\theta)\cdot L_F(x|\theta)\,d\theta \\ &= \int \left(2\pi\sigma_0^2\right)^{-\frac{1}{2}}\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_0^2}(\theta-\theta_0)^2 - \frac{1}{2\sigma^2}\sum_i (x_i-\theta)^2\right\}d\theta \\ &= (2\pi)^{-\frac{(n+1)}{2}}\left(\sigma^2\right)^{-\frac{n}{2}}\left(\sigma_0^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_i (x_i-\overline{x})^2\right\}\cdot\int\exp\{A(\theta)\}\,d\theta, \end{aligned}$$

where $A(\theta) = -\frac{1}{2\sigma_0^2}(\theta-\theta_0)^2 - \frac{n}{2\sigma^2}(\overline{x}-\theta)^2$. Let's work with the exponent $A(\theta)$. We have

$$\begin{aligned} A(\theta) &= -\frac{1}{2}\left[\frac{\left(\theta^2+\theta_0^2-2\theta\theta_0\right)}{\sigma_0^2} + \frac{n\left(\overline{x}^2+\theta^2-2\theta\overline{x}\right)}{\sigma^2}\right] \\ &= -\frac{1}{2\sigma_0^2\sigma^2}\left[\sigma^2\theta^2+\sigma^2\theta_0^2-2\sigma^2\theta\theta_0+n\sigma_0^2\overline{x}^2+n\sigma_0^2\theta^2-2n\sigma_0^2\theta\overline{x}\right] \\ &= -\frac{1}{2}\underbrace{\frac{\sigma^2+n\sigma_0^2}{\sigma_0^2\sigma^2}}_{=\sigma^{-2}_{\theta|x}}\left[\theta^2 - 2\underbrace{\left(\frac{\sigma^2\theta_0+n\sigma_0^2\overline{x}}{\sigma^2+n\sigma_0^2}\right)}_{=\theta_{\theta|x}}\theta + \frac{\sigma^2\theta_0^2+n\sigma_0^2\overline{x}^2}{\sigma^2+n\sigma_0^2}\right]. \end{aligned}$$

Adding and subtracting $\theta^2_{\theta|x}$ we get

$$\begin{aligned} A(\theta) &= -\frac{\sigma^{-2}_{\theta|x}}{2}\left[\left(\theta-\theta_{\theta|x}\right)^2 + \frac{\sigma^2\theta_0^2+n\sigma_0^2\overline{x}^2}{\sigma^2+n\sigma_0^2} - \theta^2_{\theta|x}\right] \\ &= -\frac{1}{2}\left(\frac{\theta-\theta_{\theta|x}}{\sigma_{\theta|x}}\right)^2 - \frac{1}{2}\frac{n}{\sigma^2+n\sigma_0^2}(\theta_0-\overline{x})^2. \end{aligned}$$

Therefore substituting into $m(x)$ we have

$$\begin{aligned} m(x) &= (2\pi)^{-\frac{(n+1)}{2}} \left(\sigma^2\right)^{-\frac{n}{2}} \left(\sigma_0^2\right)^{-\frac{1}{2}} \left(2\pi\sigma_{\theta|x}^2\right)^{\frac{1}{2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_i (x_i-\overline{x})^2 - \frac{1}{2}\frac{n}{\sigma^2+n\sigma_0^2}(\theta_0-\overline{x})^2\right\} \\ &\quad \underbrace{\cdot \int \left(2\pi\sigma_{\theta|x}^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\left(\frac{\theta-\theta_{\theta|x,j}}{\sigma_{\theta|x,j}}\right)^2\right]\right\} d\theta}_{=1} \\ &= (2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{(n-1)}{2}} \left(\sigma^2+n\sigma_0^2\right)^{-\frac{1}{2}} \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_i (x_i-\overline{x})^2 - \frac{1}{2}\frac{n}{\sigma^2+n\sigma_0^2}(\theta_0-\overline{x})^2\right\}. \end{aligned}$$

Consider now the class of contaminating distribution $\widetilde{G}_\theta = \left\{N(\theta,\eta^2) : \eta^2 \in \left[\sigma^2, 10\cdot\sigma^2\right]\right\}$. We need to compute the marginal likelihood in the case where contaminating model G is assumed only for observation j. We denote with $L_{F,G}^{(j)}(\theta|x)$ the likelihood function in this case. The marginal likelihood is now given by

$$\begin{aligned} m_j(x;\Pi,F_\theta,G) &= \int L_{F,G}^{(j)}(\theta|x)\,\pi(\theta)\,d\theta \\ &= \left(2\pi\sigma_0^2\right)^{-\frac{1}{2}} \left(2\pi\sigma^2\right)^{-\frac{(n-1)}{2}} \left(2\pi\eta^2\right)^{-\frac{1}{2}} \cdot \\ &\quad \cdot \int \exp\left\{-\frac{1}{2\sigma_0^2}(\theta-\theta_0)^2 - \frac{1}{2\sigma^2}\sum_i (x_i-\theta)^2 - \frac{1}{2\eta^2}(x_j-\theta)^2\right\} d\theta \\ &= (2\pi)^{-\frac{(n+1)}{2}} \left(\sigma^2\right)^{-\frac{(n-1)}{2}} \left(\sigma_0^2\eta^2\right)^{-\frac{1}{2}} \cdot \\ &\quad \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i\neq j}\left(x_i-\overline{x}_{(j)}\right)^2\right\} \cdot \int \exp\left\{B_j(\theta)\right\} d\theta, \end{aligned}$$

where $B_j(\theta) = -\frac{1}{2\sigma_0^2}(\theta-\theta_0)^2 - \frac{(n-1)}{2\sigma^2}\left(\overline{x}_{(j)}-\theta\right)^2 - \frac{1}{2\eta^2}(x_j-\theta)^2$.

Working again with the exponent $B_j(\theta)$ we have:

$$\begin{aligned} B_j(\theta) &= -\frac{1}{2}\left[\frac{\left(\theta^2+\theta_0^2-2\theta\theta_0\right)}{\sigma_0^2} + \frac{(n-1)\left(\overline{x}_{(j)}^2+\theta^2-2\theta\overline{x}_{(j)}\right)}{\sigma^2} + \frac{\left(x_j^2+\theta^2-2\theta x_j\right)}{\eta^2}\right] \\ &= -\frac{1}{2\sigma_0^2\sigma^2\eta^2}\left[\sigma^2\eta^2\theta^2 + \sigma^2\eta^2\theta_0^2 - 2\sigma^2\eta^2\theta\theta_0 + (n-1)\eta^2\sigma_0^2\overline{x}_{(j)}^2\right. \\ &\quad \left. + (n-1)\eta^2\sigma_0^2\theta^2 - 2(n-1)\eta^2\sigma_0^2\theta\overline{x}_{(j)} + \sigma^2\sigma_0^2x_j^2 + \sigma^2\sigma_0^2\theta^2 - 2\sigma^2\sigma_0^2\theta x_j\right] \end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2\sigma_0^2\sigma^2\eta^2}\left[\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)\cdot\theta^2+\sigma^2\sigma_0^2x_j^2+\sigma^2\eta^2\theta_0^2\right.\\
&\left.+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}^2-2\left(\sigma^2\eta^2\theta_0+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}+\sigma^2\sigma_0^2x_j\right)\cdot\theta\right]\\
= \quad & -\frac{1}{2}\underbrace{\frac{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}{\sigma_0^2\sigma^2\eta^2}}_{=\sigma_{\theta|x,j}^{-2}}\left[\theta^2-2\underbrace{\left(\frac{\sigma^2\eta^2\theta_0+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}+\sigma^2\sigma_0^2x_j}{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}\right)}_{=\theta_{\theta|x,j}}\cdot\theta\right.\\
&\left.+\frac{\sigma^2\sigma_0^2x_j^2+\sigma^2\eta^2\theta_0^2+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}^2}{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}\right].
\end{aligned}
$$

Adding and subtracting $\theta_{\theta|x,j}^2$ we get

$$
\begin{aligned}
B_j(\theta) &= -\frac{\sigma_{\theta|x,j}^{-2}}{2}\left[\left(\theta-\theta_{\theta|x,j}\right)^2+\frac{\sigma^2\sigma_0^2x_j^2+\sigma^2\eta^2\theta_0^2+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}^2}{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}-\theta_{\theta|x,j}^2\right]\\
&= -\frac{1}{2}\left(\frac{\theta-\theta_{\theta|x,j}}{\sigma_{\theta|x,j}}\right)^2-\frac{\sigma_{\theta|x,j}^{-2}}{2}\cdot\left(\frac{\sigma^2\sigma_0^2x_j^2+\sigma^2\eta^2\theta_0^2+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}^2}{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}-\theta_{\theta|x,j}^2\right)\\
&= -\frac{1}{2}\left(\frac{\theta-\theta_{\theta|x,j}}{\sigma_{\theta|x,j}}\right)^2-\frac{\sigma_{\theta|x,j}^{-2}}{2}\cdot\left(\frac{\left(\sigma^2\sigma_0^2x_j^2+\sigma^2\eta^2\theta_0^2+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}^2\right)}{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)^2}\cdot\right.\\
&\left.\cdot\frac{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)}{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)^2}-\frac{\left(\sigma^2\eta^2\theta_0+(n-1)\,\eta^2\sigma_0^2\overline{x}_{(j)}+\sigma^2\sigma_0^2x_j\right)^2}{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)^2}\right)\\
&= -\frac{1}{2}\left(\frac{\theta-\theta_{\theta|x,j}}{\sigma_{\theta|x,j}}\right)^2-\frac{\sigma_{\theta|x,j}^{-2}}{2}\cdot\frac{\sigma_0^2\sigma^2\eta^2}{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)^2}\\
&\quad\cdot\left(\sigma^2\left(x_j-\theta_0\right)^2+(n-1)\eta^2\left(\overline{x}_{(j)}-\theta_0\right)^2+(n-1)\,\sigma_0^2\left(\overline{x}_{(j)}-x_j\right)^2\right)\\
&= -\frac{1}{2}\left(\frac{\theta-\theta_{\theta|x,j}}{\sigma_{\theta|x,j}}\right)^2-\frac{1}{2}\cdot\frac{\left(\sigma^2\left(x_j-\theta_0\right)^2+(n-1)\eta^2\left(\overline{x}_{(j)}-\theta_0\right)^2\right)}{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}\\
&\quad-\frac{1}{2}\cdot\frac{(n-1)\,\sigma_0^2\left(\overline{x}_{(j)}-x_j\right)^2}{\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2}.
\end{aligned}
$$

Therefore substituting in $m_j(x;\Pi,F_\theta,G)$ we get

$$
\begin{aligned}
m_j(x;\Pi,F_\theta,G) &= (2\pi)^{-\frac{n}{2}}\left(\sigma^2\right)^{-\frac{(n-1)}{2}}\left(\sigma_0^2\eta^2\right)^{-\frac{1}{2}}\left(2\pi\sigma_{\theta|x,j}^2\right)^{\frac{1}{2}}\\
&\quad\cdot\exp\left\{-\frac{1}{2\sigma^2}\sum_{i\neq j}\left(x_i-\overline{x}_{(j)}\right)^2-\frac{1}{2}\cdot\frac{\sigma^2\left(x_j-\theta_0\right)^2}{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)}\right\}\\
&\quad\cdot\exp\left\{-\frac{1}{2}\cdot\frac{(n-1)\eta^2\left(\overline{x}_{(j)}-\theta_0\right)^2+(n-1)\,\sigma_0^2\left(\overline{x}_{(j)}-x_j\right)^2}{\left(\sigma^2\eta^2+(n-1)\,\eta^2\sigma_0^2+\sigma^2\sigma_0^2\right)}\right\}\\
&\quad\cdot\underbrace{\int\left(2\pi\sigma_{\theta|x,j}^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\left(\frac{\theta-\theta_{\theta|x,j}}{\sigma_{\theta|x,j}}\right)^2\right]\right\}d\theta}_{=1}
\end{aligned}
$$

$$
\begin{aligned}
= \quad & (2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{(n-2)}{2}} \left(\sigma^2\eta^2 + (n-1)\,\eta^2\sigma_0^2 + \sigma^2\sigma_0^2\right)^{-\frac{1}{2}} \\
& \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i \neq j} \left(x_i - \overline{x}_{(j)}\right)^2 - \frac{1}{2} \cdot \frac{\sigma^2 \left(x_j - \theta_0\right)^2}{\left(\sigma^2\eta^2 + (n-1)\,\eta^2\sigma_0^2 + \sigma^2\sigma_0^2\right)} \right\} \\
& \cdot \exp\left\{ -\frac{1}{2} \cdot \frac{(n-1)\eta^2 \left(\overline{x}_{(j)} - \theta_0\right)^2 + (n-1)\,\sigma_0^2 \left(\overline{x}_{(j)} - x_j\right)^2}{\left(\sigma^2\eta^2 + (n-1)\,\eta^2\sigma_0^2 + \sigma^2\sigma_0^2\right)} \right\}.
\end{aligned}
$$

Appendix C

Consider the Bayesian linear regression model where a normal distribution is assumed both for the error model F and for the prior Π. The posterior distribution of regression coefficients turns out to be normal with mean

$$E\left(\beta|y,X\right)=V\left(\Sigma_0^{-1}\beta_0+X'y\right)$$

and variance

$$Var\left(\beta|y,X\right)=\sigma^2V,$$

where $V=\left(\Sigma_0^{-1}+X'X\right)^{-1}$. The corresponding marginal likelihood is given by

$$m\left(y,X\right)=\left(2\pi\sigma^2\right)^{-\frac{k+n}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\exp\left\{-\frac{A}{2}\right\}\left(2\pi\sigma^2\right)^{\frac{k}{2}}\left|V\right|^{\frac{1}{2}},$$

where $A=\sigma^{-2}\left(y'y+\alpha_0'\Sigma_0^{-1}\alpha_0-\beta'_{Bayes}V^{-1}\beta_{Bayes}\right)$. The Bayes estimator β_{Bayes} for regression coefficients is given by the posterior mean $E\left(\beta|y,X\right)$, which is a posterior summary of type (1.2). Therefore measures of local influence of the functional to prior and sampling model perturbations are respectively given by

$$\begin{aligned}LI\left(Q;T_B,\Pi\right) &= \left[\frac{\partial T_B\left(\Pi_\varepsilon\right)}{\partial\varepsilon}\right]_{\varepsilon=0}\\ &= \int\beta\cdot\left[\frac{\partial}{\partial\varepsilon}p\left(\beta|y,X,\Pi_\varepsilon,F\right)\right]_{\varepsilon=0}d\beta\\ &= \frac{m\left(y,X;Q,F\right)}{m\left(y,X\right)}\left[T_B\left(Q\right)-T_B\right],\end{aligned}$$

and

$$\begin{aligned} LI\left(G;T_B,F\right) &= \left[\frac{\partial T_B\left(F_\varepsilon\right)}{\partial\varepsilon}\right]_{\varepsilon=0} \\ &= \int\beta\cdot\left[\frac{\partial}{\partial\varepsilon}p\left(\beta|y,X,\Pi,F_\varepsilon\right)\right]_{\varepsilon=0}\beta \\ &= \sum_{j=1}^{n}\frac{m_j\left(y,X;\Pi,F,G\right)}{m\left(y,X\right)}\left[T_B^{(j)}\left(F,G\right)-T_B\right], \end{aligned}$$

where $m_j\left(y,X;\Pi,F,G\right)=\int L_{F,G}^{(j)}\left(y|X,\beta\right)\pi\left(\beta\right)d\beta$ and $p_j\left(\beta|y,X\right)=\frac{\pi(\beta)\cdot L_{F,G}^{(j)}(y|X,\beta)}{m_j(y,X;\Pi,F,G)}$.

Both measures can be solved analytically only performing a conjugate analysis. Suppose that the uncertainty about the prior distribution on β is represented by the family $\widetilde{Q}=\left\{N\left(\alpha_0,\sigma^2\Sigma_0\right):\right.$ $\left.\alpha_0^{\inf}\leq\alpha_0\leq\alpha_0^{\sup}\right\}$. The posterior derived with such a prior is still normal with mean $\beta_{Bayes}^*=\left(X'X+\Sigma_0^{-1}\right)^{-1}\left(X'y+\Sigma_0^{-1}\alpha_0\right)$ and covariance matrix $\sigma^2V^*=\sigma^2\left(X'X+\Sigma_0^{-1}\right)^{-1}=\sigma^2V$. The corresponding marginal likelihood is given by

$$m\left(y,X;Q,F\right)=\left(2\pi\sigma^2\right)^{-\frac{k+n}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\exp\left\{-\frac{A^*}{2}\right\}\left(2\pi\sigma^2\right)^{\frac{k}{2}}\left|V^*\right|^{\frac{1}{2}},$$

where $A^*=\sigma^{-2}\left(y'y+\alpha_0'\Sigma_0^{-1}\alpha_0-\beta_{Bayes}^{*\prime}V^{-1}\beta_{Bayes}^*\right)$.

Under this assumption the local influence for the prior becomes

$$\begin{aligned} LI\left(Q;T_B,\Pi\right) &= \frac{\exp\left\{-\frac{A^*}{2}\right\}}{\exp\left\{-\frac{A}{2}\right\}}\left(\beta_{Bayes}^*-\beta_{Bayes}\right) \\ &= \exp\left\{-\frac{\left(\alpha_0-\beta_0\right)'\Sigma_0^{-1}\left(\alpha_0-\beta_0\right)}{2\sigma^2}\right\} \\ &= \exp\left\{+\frac{\left(\beta_{Bayes}^*-\beta_{Bayes}\right)'V^{-1}\left(\beta_{Bayes}^*-\beta_{Bayes}\right)}{2\sigma^2}\right\}\left(\beta_{Bayes}^*-\beta_{Bayes}\right) \\ &= \exp\left\{-\frac{\left(\alpha_0-\beta_0\right)'\left[\Sigma_0^{-1}-\Sigma_0^{-1}V'\Sigma_0^{-1}\right]\left(\alpha_0-\beta_0\right)}{2\sigma^2}\right\}\left[V\Sigma_0^{-1}\left(\alpha_0-\beta_0\right)\right]. \end{aligned}$$

Let's now consider the perturbation of the sampling distribution. We will denote by x_j' $(1\times k)$ the row j of matrix X corresponding to observation j and with $X_{(-j)}$ $(n-1\times k)$ and $y_{(-j)}$ respectively the matrix X and the vector y where the observation j has been dropped out. Assuming a contaminating family of type $\widetilde{G}=\left\{N\left(0,c^2\right):c^{\inf}\leq c^2\leq c^{\sup}\right\}$ the marginal likelihood $m_j\left(y,X;\Pi,F,G\right)$

is given by

$$\begin{aligned} m_j\left(y, X; \Pi, F, G\right) &= \int L_{F,G}^{(j)}\left(y|X, \beta\right) \pi\left(\beta\right) d\beta \\ &= \left(2\pi\sigma^2\right)^{-\frac{(k+n-1)}{2}} \left(2\pi c^2\right)^{-\frac{1}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\widetilde{B}\right\}. \end{aligned}$$

The terms $\widetilde{B}$ is given by

$$\begin{aligned} \widetilde{B} &= \sigma^{-2}\left(\beta-\beta_0\right)'\Sigma_0^{-1}\left(\beta-\beta_0\right) + c^{-2}\left(y_j - x_j'\beta\right)'\left(y_j - x_j'\beta\right) \\ &\quad +\sigma^{-2}\left(y_{(-j)} - X_{(-j)}\beta\right)'\left(y_{(-j)} - X_{(-j)}\beta\right) \\ &= \sigma^{-2}\beta'\Sigma_0^{-1}\beta - 2\sigma^{-2}\beta'\Sigma_0^{-1}\beta_0 + \sigma^{-2}\beta_0'\Sigma_0^{-1}\beta_0 + c^{-2}y_j^2 - 2c^{-2}\beta' x_j y_j \\ &\quad +c^{-2}\beta' x_j x_j'\beta + \sigma^{-2}y_{(-j)}'y_{(-j)} - 2\sigma^{-2}\beta' X_{(-j)}'y_{(-j)} + \sigma^{-2}\beta' X_{(-j)}'X_{(-j)}\beta \\ &= \beta'\left[\sigma^{-2}\left(X_{(-j)}'X_{(-j)}\right) + c^{-2}\left(x_j x_j'\right) + \sigma^{-2}\Sigma_0^{-1}\right]\beta \\ &\quad -2\beta'\left(\sigma^{-2}X_{(-j)}'y_{(-j)} + c^{-2}x_j y_j + \sigma^{-2}\Sigma_0^{-1}\beta_0\right) \\ &\quad +\sigma^{-2}\beta_0'\Sigma_0^{-1}\beta_0 + c^{-2}y_j^2 + \sigma^{-2}y_{(-j)}'y_{(-j)} \\ &= \underbrace{\sigma^{-2}\beta_0'\Sigma_0^{-1}\beta_0 + c^{-2}y_j^2 + \sigma^{-2}y_{(-j)}'y_{(-j)} - \sigma^{-2}\mathbf{m}_j V_j^{-1}\mathbf{m}_j}_{=\widetilde{B}_j} \\ &\quad +\sigma^{-2}\left(\beta-\beta_{Bayes}^{(j)}\right)'V_j^{-1}\left(\beta-\beta_{Bayes}^{(j)}\right) \\ &= \widetilde{B}_j + \sigma^{-2}\left(\beta-\beta_{Bayes}^{(j)}\right)'V_j^{-1}\left(\beta-\beta_{Bayes}^{(j)}\right), \end{aligned}$$

where

$$\beta_{Bayes}^{(j)} = \left[X_{(-j)}'X_{(-j)} + \frac{\sigma^2}{c^2}x_j x_j' + \Sigma_0^{-1}\right]^{-1}\left(X_{(-j)}'y_{(-j)} + \frac{\sigma^2}{c^2}x_j y_j + \Sigma_0^{-1}\beta_0\right),$$

and

$$V_j = \left[X_{(-j)}'X_{(-j)} + \frac{\sigma^2}{c^2}x_j x_j' + \Sigma_0^{-1}\right]^{-1}.$$

Marginal $m_j\left(y, X; \Pi, F, G\right)$ becomes

$$m_j\left(y, X; \Pi, F, G\right) = \left(2\pi\sigma^2\right)^{-\frac{(n-1)}{2}}\left(2\pi c^2\right)^{-\frac{1}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\left|V_j\right|^{\frac{1}{2}}\exp\left\{-\frac{\widetilde{B}_j}{2}\right\},$$

and the corresponding posterior distribution turns out to be a $N\left(\beta_{Bayes}^{(j)}, \sigma^2 V_j\right)$. Therefore $T_B^{(j)}\left(F, G_\beta\right) = \beta_{Bayes}^{(j)}$.

Under this assumption the local influence for the sampling model is given by

$$
\begin{aligned}
LI\left(G;T_B,F\right) &= \sum_{j=1}^{n}\left[\frac{\left(2\pi\sigma^2\right)^{-\frac{(k+n-1)}{2}}\left(2\pi c^2\right)^{-\frac{1}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\exp\left\{-\frac{\widetilde{B}_j}{2}\right\}\left(2\pi\sigma^2\right)^{\frac{k}{2}}\left|V_j\right|^{\frac{1}{2}}}{\left(2\pi\sigma^2\right)^{-\frac{k+n}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\exp\left\{-\frac{A^*}{2}\right\}\left(2\pi\sigma^2\right)^{\frac{k}{2}}\left|V\right|^{\frac{1}{2}}}\right. \\
&\quad \left.\cdot\left(\beta^{(j)}_{Bayes}-\beta_{Bayes}\right)\right] \\
&= \sum_{j=1}^{n}\left[\left(\frac{c^2\left|V\right|}{\sigma^2\left|V_j\right|}\right)^{-\frac{1}{2}}\exp\left\{-\frac{\left(\widetilde{B}_j-A\right)}{2}\right\}\right]\left(\beta^{(j)}_{Bayes}-\beta_{Bayes}\right) \\
&= \sum_{j=1}^{n}\left[\left(\frac{c^2\left|V\right|}{\sigma^2\left|V_j\right|}\right)^{-\frac{1}{2}}\exp\left\{-\frac{\left(\frac{\sigma^2}{c^2}-1\right)y_j^2}{2\sigma^2}\right\}\right. \\
&\quad \left.\cdot\exp\left\{-\frac{\beta'_{Bayes}V^{-1}\beta_{Bayes}-\beta^{(j)\prime}_{Bayes}V_j^{-1}\beta^{(j)}_{Bayes}}{2\sigma^2}\right\}\left(\beta^{(j)}_{Bayes}-\beta_{Bayes}\right)\right].
\end{aligned}
$$

Appendix D

In this appendix we give the form of the density generator g for some univariate distributions. Results are summerized in the following table:

Model		$g(u)$
Normal	$N\left(\theta,\sigma^2\right)$	$\exp\left(-u\right)$
Student-t	$t\left(\theta,\sigma^2,p\right)$	$\left(1+\frac{2\cdot u}{p}\right)^{-\frac{(1+p)}{2}}$
EPS	$EPS\left(\theta,\sigma,\delta\right)$	$\exp\left(-c_\delta\cdot\left(2u\right)^{(\delta+1)^{-1}}\right)$
RN	$RN\left(\theta,\sigma,a,b\right)$	$\exp\left(-\left(b\cdot a^{2/b}\right)^{-1}\cdot\int_0^{a\cdot(2u)^{b/2}}e^{-t}t^{\frac{2}{b}-1}dt\right)$
EP	$EP\left(\theta,\sigma,c,\lambda\right)$	$\left[\nu^{-1}\left(\sigma^2\right)\ \cdot\sigma^2\right]^{\frac{1}{2}}\exp\left\{-\frac{c}{2}\cdot\rho_\lambda\left(1+\frac{2\ \nu^{-1}\left(\sigma^2\right)\cdot\ \sigma^2}{c-1}u\right)\right\}$

The paper by Landsman and Valdez (2003) proves the condition

$$\int_0^\infty u^{\frac{k}{2}-1}g(u)du<\infty \tag{D.1}$$

that guarantees g to be a density generator for the normal, student-t ans EPS distributions. In this appendix we prove such condition to be satisfied for the RN and EP distributions. We denote by u the quantity $u=\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2$.

Ramsay and Novick's distribution

If $X \sim RN(\mu, \sigma, a, b)$, then it belongs to elliptical family with

$$g(u) = \exp\left(-\left(b \cdot a^{2/b}\right)^{-1} \cdot \int_0^{a\cdot(2u)^{b/2}} e^{-t} t^{\frac{2}{b}-1} dt\right), a > 0, b > 0.$$

Integral (D.1) is given in this case by

$$I = \int_0^{\infty} u^{-\frac{1}{2}} e^{-\left(b\cdot a^{2/b}\right)^{-1}\cdot\int_0^{a\cdot(2u)^{b/2}} e^{-t} t^{\frac{2}{b}-1} dt} du,$$

and it cannot be computed analitically.

However, considering the integral in the exponent, the following relation holds

$$\int_0^{a\cdot(2u)^{b/2}} e^{-t} t^{\frac{2}{b}-1} dt \leq a \cdot (2u)^{b/2} e^{-t_*} t_*^{\frac{2}{b}-1},$$

where $t_* = \max_t \left(e^{-t} t^{\frac{2}{b}-1}\right) > 0$ because the function is define on $\mathbb{R}^+$.

Therefore (D.1) can be written as

$$\begin{aligned} I &\leq \int_0^{\infty} u^{-\frac{1}{2}} e^{-\left(b\cdot a^{2/b}\right)^{-1}\cdot a\cdot(2u)^{b/2} e^{-t_*} t_*^{\frac{2}{b}-1}} du \\ &= 2^{\frac{1}{2}} \kappa^{-\frac{1}{b}} \Gamma\left(1 + \frac{1}{b}\right) < \infty, \end{aligned}$$

where $\kappa = b^{-1} a^{2/b-1} e^{-t_*} t_*^{\frac{2}{b}-1}$.

It follows that g is a density generator and the RN distribution belongs to the elliptical family.

Extended Power distribution

If $X \sim EP(\mu, \sigma, a, b)$, then it belongs to elliptical family with

$$g(u) = \left[\nu^{-1}\left(\sigma^2\right) \cdot \sigma^2\right]^{\frac{1}{2}} \exp\left\{-\frac{c}{2} \cdot \rho_\lambda\left(1 + \frac{2\,\nu^{-1}\left(\sigma^2\right)\cdot\,\sigma^2}{c-1} u\right)\right\}.$$

Integral (D.1) is given by

$$I = \left[\nu^{-1}\left(\sigma^2\right) \cdot \sigma^2\right]^{\frac{1}{2}} \cdot \left[\int_0^{\infty} u^{-\frac{1}{2}} \exp\left\{-\frac{c}{2} \cdot \rho_\lambda\left(1 + \frac{2\,\nu^{-1}\left(\sigma^2\right)\cdot\,\sigma^2}{c-1} u\right)\right\} du\right],$$

and it can be computed analitically only for $\lambda = 0$.

In this case we have

$$\sigma^2 = \frac{(c-1)\sqrt{2}}{(c-3)\phi} \Longrightarrow \phi = \frac{(c-1)\sqrt{2}}{(c-3)\sigma^2},$$

and

$$\begin{aligned}
\int_0^\infty u^{\frac{k}{2}-1} g(u) du &= \left[\frac{(c-1)\sqrt{2}}{(c-3)\sigma^2}\ \sigma^2\right]^{\frac{1}{2}} \int_0^\infty u^{-\frac{1}{2}} \exp\left\{-\frac{c}{2}\log\left(1 + \frac{2\ \sigma^2}{(c-1)}\frac{(c-1)\sqrt{2}}{(c-3)\sigma^2}\ u\right)\right\} du \\
&= \left[\frac{(c-1)\sqrt{2}}{(c-3)}\right]^{\frac{1}{2}} \int_0^\infty u^{-\frac{1}{2}} \left(1 + \frac{2\sqrt{2}}{(c-3)}\ u\right)^{-\frac{c}{2}} du \\
&= \left[\frac{(c-1)\pi}{4\sqrt{2}}\right]^{\frac{1}{2}} \frac{\Gamma\left(\frac{c}{2}-1\right)}{\Gamma\left(\frac{c}{2}\right)} < \infty.
\end{aligned}$$

Appendix E

In this appendix we show that, as long as future returns r are normally distributed, the moments of the predictive distribution $P_{r|x}$ depend uniquely on the moments of posterior distribution $P_{\theta|x}$, for any posterior distribution considered. Let's assume $r \sim N(\theta, \Sigma)$ with Σ known.

Predictive moments are given by

$$\begin{aligned}
\theta_{r|x} &= \int rp(r|x)\,dr \\
&= \int r\left(\int p(r|\theta)\,p(\theta|x)\,d\theta\right)dr \\
&= \int\left(\int rp(r|\theta)\,dr\right)p(\theta|x)\,d\theta \\
&= \int \theta p(\theta|x)\,d\theta = \theta_{\theta|x}
\end{aligned}$$

and

$$\begin{aligned}
\Sigma_{r|x} &= \int rr'p(r|x)\,dr - \left(\int rp(r|x)\,dr\right)\left(\int rp(r|x)\,dr\right)' \\
&= \int\left(\int rr'p(r|\theta)\,dr\right)p(\theta|x)\,d\theta - \theta_{\theta|x}\theta'_{\theta|x} \\
&= \int\left(\Sigma + \theta\theta'\right)p(\theta|x)\,d\theta - \theta_{\theta|x}\theta'_{\theta|x} \\
&= \Sigma + \Sigma_{\theta|x} + \theta_{\theta|x}\theta'_{\theta|x} - \theta_{\theta|x}\theta'_{\theta|x} \\
&= \Sigma + \Sigma_{\theta|x}.
\end{aligned}$$

which are a function of posterior moments $\left(\theta_{\theta|x}, \Sigma_{\theta|x}\right)$.

Appendix F

The local sensitivity measure of Bayesian weights $q_{_B}$ to perturbation of prior Π in the direction of contaminating distribution Q is given by:

$$\begin{aligned} LI\left(Q;q_{_B},\Pi\right) &= \left[\frac{\partial q_{_B}(\Pi_\varepsilon)}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &= \left[\frac{\partial q_{_B}(\Pi_\varepsilon)}{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right)\right]^{-1})'}\cdot\frac{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right)\right]^{-1})}{\partial vec(\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right))'}\right. \\ &\quad \left.\cdot\frac{\partial vec(\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right))}{\partial \varepsilon}\right]_{\varepsilon=0} + \left[\frac{\partial q_{_B}(\Pi_\varepsilon)}{\partial \theta_{\theta|x}\left(\Pi_\varepsilon\right)'}\cdot\frac{\partial \theta_{\theta|x}\left(\Pi_\varepsilon\right)}{\partial \varepsilon}\right]_{\varepsilon=0} \\ &\quad + \left[\frac{\partial q_{_B}(\Pi_\varepsilon)}{\partial \lambda_{\theta|x}\left(\Pi_\varepsilon\right)}\cdot\frac{\partial \lambda_{\theta|x}\left(\Pi_\varepsilon\right)}{\partial \varepsilon}\right]_{\varepsilon=0}. \end{aligned}$$

Using the properties of operator vec and of Kronecker product $\otimes$, such derivative can be easily computed. The vector of the Bayesian weights $q(\Pi_\varepsilon)$ can be written as

$$\begin{aligned} q_{_B}(\Pi_\varepsilon) &= vec(q_{_B}(\Pi_\varepsilon)) \\ &= \frac{1}{b}\left[\left(\theta_{\theta|x}\left(\Pi_\varepsilon\right)-\lambda_{\theta|x}\left(\Pi_\varepsilon\right)\cdot\mathbf{1}\right)'\otimes I_{_N}\right]\cdot vec(\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right)\right]^{-1}). \end{aligned}$$

Therefore we have:

$$\left[\frac{\partial q(\Pi_\varepsilon)}{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right)\right]^{-1})'}\right]_{\varepsilon=0} = \frac{1}{b}\left[\left(\theta_{\theta|x}-\lambda_{\theta|x}\cdot\mathbf{1}\right)'\otimes I_{_N}\right],$$

$$\left[\frac{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right)\right]^{-1})}{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_\varepsilon\right)\right]^{-1})'}\right]_{\varepsilon=0} = -\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\otimes\left[\Sigma+\Sigma_{\theta|x}\right]^{-1},$$

$$\left[\frac{\partial q_{_B}(\Pi_\varepsilon)}{\partial \theta_{\theta|x}\left(\Pi_\varepsilon\right)'}\right]_{\varepsilon=0} = \frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1},$$

and

$$\left[\frac{\partial q_B(\Pi_\varepsilon)}{\partial \lambda_{\theta|x}(\Pi_\varepsilon)}\right]_{\varepsilon=0} = -\frac{1}{b}\left[\Sigma + \Sigma_{\theta|x}\right]^{-1} \cdot \mathbf{1}.$$

For all remaining quantities it is crucial to compute the derivative of the posterior distribution. If we assume a contaminated prior density of type $\pi_\varepsilon(\theta) = (1-\varepsilon)\pi(\theta) + \varepsilon q(\theta)$, we obtain

$$\begin{aligned}
\left[\frac{\partial}{\partial\varepsilon}\left(\frac{\pi_\varepsilon(\theta) L_F(\theta|x)}{m(x;\Pi_\varepsilon,F_\theta)}\right)\right]_{\varepsilon=0} &= \left[\frac{\frac{\partial\pi_\varepsilon(\theta)}{\partial\varepsilon} L_F(\theta|x)\, m(x;\Pi_\varepsilon,F_\theta)}{m(x;\Pi_\varepsilon,F_\theta)^2} + \right. \\
&\quad \left. \frac{-\pi_\varepsilon(\theta) L_F(\theta|x) \frac{\partial m(x;\Pi_\varepsilon,F_\theta)}{\partial\varepsilon}}{m(x;\Pi_\varepsilon,F_\theta)^2}\right]_{\varepsilon=0} \\
&= \frac{(q(\theta)-\pi(\theta)) L_F(\theta|x)}{m(x)} + \\
&\quad -\frac{\pi(\theta) L_F(\theta|x)(m(x;Q,F_\theta)-m(x))}{m(x)} \\
&= \frac{m(x;Q,F)}{m(x)}\left(\frac{q(\theta) L_F(\theta|x)}{m(x;Q,F)} - \frac{\pi(\theta) L_F(\theta|x)}{m(x)}\right).
\end{aligned}$$

Indeed, we have

$$\begin{aligned}
\left[\frac{\partial\theta_{\theta|x}(\Pi_\varepsilon)}{\partial\varepsilon}\right]_{\varepsilon=0} &= \int\theta\left[\frac{\partial}{\partial\varepsilon}\left(\frac{\pi_\varepsilon(\theta) L_F(\theta|x)}{m(x;\Pi_\varepsilon,F_\theta)}\right)\right]_{\varepsilon=0} d\theta \\
&= \frac{m(x;Q,F)}{m(x)}\left(\theta_{\theta|x}(Q) - \theta_{\theta|x}\right) \\
&= LI\left(Q;\theta_{\theta|x},\Pi\right).
\end{aligned}$$

Moreover

$$\begin{aligned}
\left[\frac{\partial vec(\Sigma+\Sigma_{\theta|x}(\Pi_\varepsilon))}{\partial\varepsilon}\right]_{\varepsilon=0} &= \left[\frac{\partial vec(\Sigma_{\theta|x}(\Pi_\varepsilon))}{\partial\varepsilon}\right]_{\varepsilon=0} \\
&= vec\left[\frac{\partial\Sigma_{\theta|x}(\Pi_\varepsilon)}{\partial\varepsilon}\right]_{\varepsilon=0},
\end{aligned}$$

where

$$\begin{aligned}
\left[\frac{\partial \Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)}{\partial \varepsilon}\right]_{\varepsilon=0} &= \left[\frac{\partial}{\partial \varepsilon}\left(\int \theta\theta' \left(\frac{\pi_{\varepsilon}\left(\theta\right) L_F\left(\theta|x\right)}{m\left(x;\Pi_{\varepsilon},F_{\theta}\right)}\right) d\theta - \theta_{\theta|x}\left(\Pi_{\varepsilon}\right)\theta_{\theta|x}\left(\Pi_{\varepsilon}\right)'\right)\right]_{\varepsilon=0} \\
&= \left[\int \theta\theta' \frac{\partial}{\partial \varepsilon}\left(\frac{\pi_{\varepsilon}\left(\theta\right) L_F\left(\theta|x\right)}{m\left(x;\Pi_{\varepsilon},F_{\theta}\right)}\right) d\theta + \right. \\
&\quad \left. -\frac{\partial \theta_{\theta|x}\left(\Pi_{\varepsilon}\right)}{\partial \varepsilon}\theta_{\theta|x}\left(\Pi_{\varepsilon}\right)' - \theta_{\theta|x}\left(\Pi_{\varepsilon}\right)\frac{\partial \theta_{\theta|x}\left(\Pi_{\varepsilon}\right)'}{\partial \varepsilon}\right]_{\varepsilon=0} \\
&= \frac{m\left(x\,;Q,F\right)}{m\left(x\right)}\left[\Sigma_{\theta|x}\left(Q\right) - \Sigma_{\theta|x} + \theta_{\theta|x}\left(Q\right)\theta_{\theta|x}\left(Q\right)' + \right. \\
&\quad \left. -\theta_{\theta|x}\left(Q\right)\theta'_{\theta|x} - \theta'_{\theta|x}\theta_{\theta|x}\left(Q\right)' + \theta'_{\theta|x}\theta'_{\theta|x}\right] \\
&= LI\left(Q;\Sigma_{\theta|x},\Pi\right).
\end{aligned}$$

The last term depends on the previous quantities, and it is given by:

$$\begin{aligned}
\left[\frac{\partial \lambda_{\theta|x}\left(\Pi_{\varepsilon}\right)}{\partial \varepsilon}\right]_{\varepsilon=0} &= \left[\frac{\partial}{\partial \varepsilon}\left(\frac{\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)\right]^{-1}\theta_{\theta|x}\left(\Pi_{\varepsilon}\right) - b}{\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)\right]^{-1}\mathbf{1}}\right)\right]_{\varepsilon=0} \\
&= \frac{\left(\mathbf{1}'\left[\frac{\partial\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)\right]^{-1}}{\partial \varepsilon}\right]_{\varepsilon=0}\theta_{\theta|x}\right)\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)}{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)^2} + \\
&\quad + \frac{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\left[\frac{\partial \theta_{\theta|x}\left(\Pi_{\varepsilon}\right)}{\partial \varepsilon}\right]_{\varepsilon=0}\right)\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)}{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)^2} + \\
&\quad - \frac{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\theta_{\theta|x} - b\right)\left(\mathbf{1}'\left[\frac{\partial\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)\right]^{-1}}{\partial \varepsilon}\right]_{\varepsilon=0}\mathbf{1}\right)}{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)^2} \\
&= \frac{-\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1} LI\left(Q;\Sigma_{\theta|x},\Pi\right)\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\theta_{\theta|x}\right)}{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)} + \\
&\quad + \frac{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1} LI\left(Q;\theta_{\theta|x},\Pi\right)\right)}{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)} + \frac{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\theta_{\theta|x} - b\right)}{\left(\mathbf{1}'\left[\Sigma + \Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)^2} \cdot \\
&\quad \cdot \left(\mathbf{1}' \cdot \left[\Sigma + \Sigma_{\theta|x}\right]^{-1} LI\left(Q;\Sigma_{\theta|x},\Pi\right)\left[\Sigma + \Sigma_{\theta|x}\right]^{-1} \cdot \mathbf{1}\right)
\end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{m\left(x\,;Q,F\right)}{m\left(x\right)}\left[\frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}\left(Q\right)-\theta_{\theta|x}\right)}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}}+\right.\\
& -\frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\Sigma_{\theta|x}\left(Q\right)-\Sigma_{\theta|x}\right)\Sigma_{r|x}^{-1}\theta_{\theta|x}}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}}+\\
& -\frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}\left(Q\right)\theta_{\theta|x}\left(Q\right)'-\theta'_{\theta|x}\theta'_{\theta|x}\right)\Sigma_{r|x}^{-1}\theta_{\theta|x}}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}}+\\
& +\frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}\left(Q\right)\theta'_{\theta|x}+\theta'_{\theta|x}\theta_{\theta|x}\left(Q\right)'\right)\Sigma_{r|x}^{-1}\theta_{\theta|x}}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}}+\\
& +\frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\left(\Sigma_{\theta|x}\left(Q\right)-\Sigma_{\theta|x}\right)\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^{2}}+\\
& +\frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}\left(Q\right)\theta_{\theta|x}\left(Q\right)'-\theta_{\theta|x}\left(Q\right)\theta'_{\theta|x}\right)\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^{2}}+\\
& \left.+\frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\left(\theta'_{\theta|x}\theta'_{\theta|x}-\theta'_{\theta|x}\theta_{\theta|x}\left(Q\right)'\right)\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^{2}}\right]\\
= \quad & LI\left(Q;\lambda_{\theta|x},\Pi\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\left[\frac{\partial\left[\Sigma+\Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)\right]^{-1}}{\partial\varepsilon}\right]_{\varepsilon=0} &= -\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\left[\frac{\partial\Sigma_{\theta|x}\left(\Pi_{\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\\
&= -\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}LI\left(Q;\Sigma_{\theta|x},\Pi\right)\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}.
\end{aligned}
$$

Finally, $LI\left(Q;q_{_B},\Pi\right)$ can be written as:

$$
\begin{aligned}
LI\left(Q;q_{_B},\Pi\right) &= -\frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}LI\left(Q;\Sigma_{\theta|x},\Pi\right)\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\left(\theta_{\theta|x}-\lambda_{\theta|x}\cdot\mathbf{1}\right)+\\
&\quad +\frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}LI\left(Q;\theta_{\theta|x},\Pi\right)-\frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}LI\left(Q;\lambda_{\theta|x},\Pi\right)\cdot\mathbf{1}\\
&= \frac{1}{b}\Sigma_{r|x}^{-1}\cdot\left(LI\left(Q;\theta_{\theta|x},\Pi\right)-LI\left(Q;\lambda_{\theta|x},\Pi\right)\cdot\mathbf{1}\right)+\\
&\quad +\Sigma_{r|x}^{-1}\cdot LI\left(Q;\Sigma_{\theta|x},\Pi\right)\cdot q_{_B}.
\end{aligned}
$$

Appendix G

The local sensitivity measure of Bayesian weights $q_{_B}$ to perturbation of sampling F_θ in the direction of contaminating distribution G is given by:

$$\begin{aligned} LI\left(G;q_{_B},F_\theta\right) &= \left[\frac{\partial q_{_B}(F_{\theta,\varepsilon})}{\partial\varepsilon}\right]_{\varepsilon=0} \\ &= \left[\frac{\partial q_{_B}(F_{\theta,\varepsilon})}{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right)\right]^{-1})'}\cdot\frac{\partial vec(\left[\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right)\right]^{-1})}{\partial vec(\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right))'}\cdot\right. \\ &\quad \left.\cdot\frac{\partial vec(\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right))}{\partial\varepsilon}\right]_{\varepsilon=0}+\left[\frac{\partial q_{_B}(F_{\theta,\varepsilon})}{\partial\theta_{\theta|x}\left(F_{\theta,\varepsilon}\right)'}\cdot\frac{\partial\theta_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0}+ \\ &\quad +\left[\frac{\partial q_{_B}(F_{\theta,\varepsilon})}{\partial\lambda_{\theta|x}\left(F_{\theta,\varepsilon}\right)}\cdot\frac{\partial\lambda_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0}. \end{aligned}$$

As in the previous Appendix, such derivative is computed using the properties of operator *vec* and of Kronecker product $\otimes$. Most of the terms involved have already been obtained. For all remainig quantities it is crucial to compute the derivative of the posterior distribution. If we assume a contaminated sampling density of type $f_\varepsilon(\theta)=(1-\varepsilon)f(\theta)+\varepsilon g(\theta)$, we derive in Appendix A the following result:

$$\left[\frac{\partial}{\partial\varepsilon}\left(\frac{\pi(\theta)\cdot L_{F_\varepsilon}(x|\theta)}{m(x;\Pi,F_{\theta,\varepsilon})}\right)\right]_{\varepsilon=0}=\sum_{j=1}^{n}\frac{m_j(x;\Pi,F_\theta,G)}{m(x;\Pi,F_\theta)}\left(p_j(\theta|x)-p(\theta|x)\right),$$

where $p_j(\theta|x)=\left[\pi(\theta)\cdot g(x_j)\cdot\prod_{i\neq j}f_\theta(x_i)\right]/m_j(x;\Pi,F_\theta,G)$ is the posterior obtained when a sampling distribution G is adopted only for observation j and $m_j(x;\Pi,F_\theta,G)=\int g(x_j)\cdot\prod_{i\neq j}f_\theta(x_i)\,\pi(\theta)\,d\theta$ is the corresponding marginal likelihood.

Indeed, we have

$$\begin{aligned}\left[\frac{\partial\theta_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0} &= \int\theta\left[\frac{\partial}{\partial\varepsilon}\left(\frac{\pi\left(\theta\right)\cdot L_{F_{\varepsilon}}\left(x|\theta\right)}{m\left(x\,;\Pi,F_{\theta,\varepsilon}\right)}\right)\right]_{\varepsilon=0}d\theta \\ &= \sum_{j=1}^{n}\frac{m_{j}\left(x\,;\Pi,F_{\theta},G\right)}{m\left(x\,;\Pi,F_{\theta}\right)}\left(\theta_{\theta|x}^{(j)}\left(F_{\theta},G\right)-\theta_{\theta|x}\right) \\ &= LI\left(G;\theta_{\theta|x},F_{\theta}\right).\end{aligned}$$

Moreover

$$\begin{aligned}\left[\frac{\partial vec(\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right))}{\partial\varepsilon}\right]_{\varepsilon=0} &= \left[\frac{\partial vec(\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right))}{\partial\varepsilon}\right]_{\varepsilon=0} \\ &= vec\left[\frac{\partial\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0},\end{aligned}$$

where:

$$\begin{aligned}\left[\frac{\partial\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0} &= \left[\int\theta\theta'\frac{\partial}{\partial\varepsilon}\left(\frac{\pi\left(\theta\right)\cdot L_{F_{\varepsilon}}\left(x|\theta\right)}{m\left(x\,;\Pi,F_{\theta,\varepsilon}\right)}\right)d\theta\right]_{\varepsilon=0} \\ &\quad -\left[\frac{\partial\theta_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0}\theta_{\theta|x}-\theta_{\theta|x}\left[\frac{\partial\theta_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon}\right]_{\varepsilon=0} \\ &= \sum_{j=1}^{n}\frac{m_{j}\left(x\,;\Pi,F_{\theta},G\right)}{m\left(x\,;\Pi,F_{\theta}\right)}\left[\Sigma_{\theta|x}^{(j)}\left(F_{\theta},G\right)-\Sigma_{\theta|x}+\right. \\ &\quad +\theta_{\theta|x}^{(j)}\left(F_{\theta},G\right)\theta_{\theta|x}^{(j)}\left(F_{\theta},G\right)'+\theta_{\theta|x}'\theta_{\theta|x}'+ \\ &\quad \left.-\theta_{\theta|x}^{(j)}\left(F_{\theta},G\right)\theta_{\theta|x}'-\theta_{\theta|x}'\theta_{\theta|x}^{(j)}\left(F_{\theta},G\right)'\right] \\ &= LI\left(G;\Sigma_{\theta|x},F_{\theta}\right).\end{aligned}$$

The last term to be computed is:

$$\begin{aligned}\frac{\partial\lambda_{\theta|x}\left(F_{\theta,\varepsilon}\right)}{\partial\varepsilon} &= \frac{\partial}{\partial\varepsilon}\left(\frac{\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right)\right]^{-1}\theta_{\theta|x}\left(F_{\theta,\varepsilon}\right)-b}{\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\left(F_{\theta,\varepsilon}\right)\right]^{-1}\mathbf{1}}\right)_{\varepsilon=0} \\ &= \frac{-\left(\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}LI\left(G;\Sigma_{\theta|x},F_{\theta}\right)\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\theta_{\theta|x}\right)}{\left(\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)}+ \\ &\quad +\frac{\left(\mathbf{1}'\Sigma_{r|y}^{-1}LI\left(G;\theta_{\theta|x},F_{\theta}\right)\right)}{\left(\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)}+\frac{\left(\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\theta_{\theta|x}-b\right)}{\left(\mathbf{1}'\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\mathbf{1}\right)^{2}}\cdot \\ &\quad \cdot\left(\mathbf{1}'\cdot\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}LI\left(G;\Sigma_{\theta|x},F_{\theta}\right)\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\cdot\mathbf{1}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} \frac{m_j\left(x\,;\Pi,F_\theta,G\right)}{m\left(x\,;\Pi,F_\theta\right)} \left[\frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)-\theta_{\theta|x}\right)}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}} + \right. \\
&- \frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right)-\Sigma_{\theta|x}\right)\Sigma_{r|x}^{-1}\theta_{\theta|x}}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}} + \\
&- \frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)'+\theta'_{\theta|x}\theta'_{\theta|x}\right)\Sigma_{r|x}^{-1}\theta_{\theta|x}}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}} + \\
&+ \left. \frac{\mathbf{1}'\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\theta'_{\theta|x}+\theta'_{\theta|x}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)'\right)\Sigma_{r|x}^{-1}\theta_{\theta|x}}{\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}} + \right. \\
&+ \frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\left(\Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right)-\Sigma_{\theta|x}\right)\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^2} + \\
&+ \frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\left(\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)'\right)\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^2} + \\
&+ \frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\theta'_{\theta|x}\theta'_{\theta|x}\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^2} + \\
&+ \frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\theta'_{\theta|x}\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^2} + \\
&+ \left. \frac{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\theta_{\theta|x}-b\right)\left(\mathbf{1}'\cdot\Sigma_{r|x}^{-1}\theta'_{\theta|x}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\Sigma_{r|x}^{-1}\cdot\mathbf{1}\right)}{\left(\mathbf{1}'\Sigma_{r|x}^{-1}\mathbf{1}\right)^2} \right] \\
&= LI\left(G;\lambda_{\theta|x},F_\theta\right)
\end{aligned}
$$

Finally, $LI\left(G;q_{_B},F_\theta\right)$ can be written as:

$$
\begin{aligned}
LI\left(G;q_{_B},F_\theta\right) &= -\frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1} LI\left(G;\Sigma_{\theta|x},F_\theta\right)\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\left(\theta_{\theta|x}-\lambda_{\theta|x}\cdot\mathbf{1}\right)+ \\
&\quad +\frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1}\cdot LI\left(G;\theta_{\theta|x},F_\theta\right)-\frac{1}{b}\left[\Sigma+\Sigma_{\theta|x}\right]^{-1} LI\left(G;\lambda_{\theta|x},F_\theta\right)\cdot\mathbf{1} \\
&= \frac{1}{b}\Sigma_{r|x}^{-1}\cdot\left(LI\left(G;\theta_{\theta|x},F_\theta\right)-LI\left(G;\lambda_{\theta|x},F_\theta\right)\cdot\mathbf{1}\right)+ \\
&\quad +\Sigma_{r|x}^{-1}\cdot LI\left(G;\Sigma_{\theta|x},F_\theta\right)\cdot q_{_B}.
\end{aligned}
$$

Appendix H

In this appendix we compute marginal likelihood $m\left(x\,;\Pi,F_\theta\right)$ when the prior and the sampling distribution are respectively $N\left(\theta_0,\Sigma_0\right)$ and $N\left(\theta,\Sigma\right)$. Under the same assumption for Π and F_θ, we then derive marginal likelihood $m_j\left(x\,;\Pi,F_\theta,G\right)$ when contaminating distribution G is $N\left(\theta,\Omega\right)$.

The marginal likelihood $m\left(x\,;\Pi,F_\theta\right)$ is given by

$$\begin{aligned}
m\left(x\,;\Pi,F_\theta\right) &= \int (2\pi)^{-\frac{kn}{2}}(2\pi)^{-\frac{k}{2}}\left|\Sigma\right|^{-\frac{n}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)'\Sigma^{-1}(x_i-\theta)\right\}\cdot \\
&\quad \cdot\exp\left\{-\frac{1}{2}(\theta_0-\theta)'\Sigma_0^{-1}(\theta_0-\theta)\right\}d\theta \\
&= (2\pi)^{-\frac{kn}{2}}(2\pi)^{-\frac{k}{2}}\left|\Sigma\right|^{-\frac{n}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\cdot \\
&\quad \cdot\int\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\theta)'\Sigma^{-1}(x_i-\overline{x}+\overline{x}-\theta)\right\}\cdot \\
&\quad \cdot\exp\left\{-\frac{1}{2}(\theta_0-\theta)'\Sigma_0^{-1}(\theta_0-\theta)\right\}d\theta \\
&= (2\pi)^{-\frac{kn}{2}}(2\pi)^{-\frac{k}{2}}\left|\Sigma\right|^{-\frac{n}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\cdot\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})'\Sigma^{-1}(x_i-\overline{x})\right\}\cdot \\
&\quad \cdot\int\exp\left\{-\frac{n}{2}(\theta-\overline{x})'\Sigma^{-1}(\theta-\overline{x})-\frac{1}{2}(\theta-\theta_0)'\Sigma_0^{-1}(\theta-\theta_0)\right\}d\theta \\
&= (2\pi)^{-\frac{kn}{2}}\left|\Sigma\right|^{-\frac{k}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\left|\Sigma_{\theta|x}\right|^{\frac{1}{2}}\cdot\exp\left\{n\cdot tr(\Sigma^{-1}S)\right\}\cdot \\
&\quad \cdot\exp\left\{-\frac{n}{2}\overline{x}'\Sigma^{-1}\overline{x}-\frac{1}{2}\theta_0'\Sigma_0^{-1}\theta_0+\frac{1}{2}\theta_{\theta|x}'\Sigma_{\theta|x}^{-1}\theta_{\theta|x}\right\}\cdot \\
&\quad \underbrace{\cdot\int(2\pi)^{-\frac{k}{2}}\left|\Sigma_{\theta|x}\right|^{-\frac{1}{2}}\exp\{-\frac{1}{2}((\theta-\theta_{\theta|x})'\Sigma_{\theta|x}^{-1}(\theta-\theta_{\theta|x})\}d\theta}_{=1} \\
&= (2\pi)^{-\frac{kn}{2}}\left|\Sigma\right|^{-\frac{k}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\left|\Sigma_{\theta|x}\right|^{\frac{1}{2}}\cdot\exp\left\{n\cdot tr(\Sigma^{-1}S)\right\}\cdot \\
&\quad \cdot\exp\left\{-\frac{n}{2}\overline{x}'\Sigma^{-1}\overline{x}-\frac{1}{2}\theta_0'\Sigma_0^{-1}\theta_0+\frac{1}{2}\theta_{\theta|x}'\Sigma_{\theta|x}^{-1}\theta_{\theta|x}\right\},
\end{aligned}$$

where

$$\begin{aligned}\theta_{\theta|x} &= \Sigma_{\theta|x}\left[\Sigma_0^{-1}\theta_0 + n\Sigma^{-1}\overline{x}\right], \\ \Sigma_{\theta|x} &= \left[\Sigma_0^{-1} + n\Sigma^{-1}\right]^{-1},\end{aligned}$$

ans S is the sample covariance matrix.

With the same procedure we derive the marginal likelihood $m_j\left(x\,;\Pi,F_\theta,G\right)$ which is given by

$$\begin{aligned}m_j\left(x\,;\Pi,F_\theta,G\right)m_j\left(x\,;\Pi,F_\theta,G\right) &= \int (2\pi)^{-\frac{kn}{2}}(2\pi)^{-\frac{k}{2}}\left|\Sigma\right|^{-\frac{(n-1)}{2}}\left|\Omega\right|^{-\frac{1}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\cdot \\ &\quad \cdot\exp\left\{-\frac{1}{2}\sum_{i\neq j}(x_i-\theta)'\Sigma^{-1}(x_i-\theta)\right\}\cdot \\ &\quad \cdot\exp\left\{-\frac{1}{2}(x_j-\theta)'\Omega^{-1}(x_j-\theta)-\frac{1}{2}(\theta_0-\theta)'\Sigma_0^{-1}(\theta_0-\theta)\right\}d\theta \\ &= (2\pi)^{-\frac{kn}{2}}\left|\Sigma\right|^{-\frac{(n-1)}{2}}\left|\Omega\right|^{-\frac{1}{2}}\left|\Sigma_0\right|^{-\frac{1}{2}}\left|\Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right)\right|^{\frac{1}{2}}\cdot \\ &\quad \cdot\exp\left\{-\frac{1}{2}(n-1)\cdot tr(\Sigma^{-1}S_{(-j)})-\frac{1}{2}x_j'\Omega^{-1}x_j-\frac{1}{2}\theta_0'\Sigma_0^{-1}\theta_0\right\}\cdot \\ &\quad \cdot\exp\left\{-\frac{1}{2}(n-1)\overline{x}_{(-j)}'\Sigma^{-1}\overline{x}_{(-j)}\right\}\cdot \\ &\quad \cdot\exp\left\{\frac{1}{2}\theta_{\theta|x}^{(j)\prime}\left(F_\theta,G\right)\left[\Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right)\right]^{-1}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right)\right\},\end{aligned}$$

where

$$\begin{aligned}\theta_{\theta|x}^{(j)}\left(F_\theta,G\right) &= \Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right)\cdot\left[\Sigma_0^{-1}\theta_0+(n-1)\,\Sigma^{-1}\overline{x}_{(-j)}+\Omega^{-1}x_j\right], \\ \Sigma_{\theta|x}^{(j)}\left(F_\theta,G\right) &= \left[\Sigma_0^{-1}+(n-1)\,\Sigma^{-1}+\Omega^{-1}\right]^{-1}, \\ \overline{x}_{(-j)} &= \frac{1}{n-1}\sum_{i\neq j}x_i,\end{aligned}$$

and $S_{(-j)}$ is the sample covariance matrix computed dropping observation j from the sample.

Bibliography

[1] Albert, J., Delampady, M., Polasek, W. (1991), A class of distributions for robustness studies, *Journal of Statistical and Planning Inference*, **28**, 291-304.

[2] Barry, C. B. (1974), Portfolio analysis under uncertain means, variances and covariances, *Journal of Finance*, **29**, 515-522.

[3] Basu, S. (1999), Posterior sensitivity to the sampling distribution and the prior: more than one observation, *Ann. Inst. Statist. Math.*, **51**, 3, 499-513.

[4] Bawa, V. S., Brown, S. J., Klein, R. W. (1979), *Estimation risk and optimal portfolio choice*, North-Holland, Amsterdam.

[5] Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis*, Springer-Verlag, New York.

[6] Berger, J. O. (1994), An overview of robust Bayesian analysis (with discussion), *Test*, **3**, 1, 5-124.

[7] Berger, J. O., Rìos Insua, J., Ruggeri, F. (2000), Bayesian robustness, in *Robust Bayesian analysis* (D. Rìos Insua and F. Ruggeri, eds), Springer-Verlag, New York.

[8] Bernardo, J. M., Smith, A. F. M. (1994), *Bayesian theory*, Wiley, Chichester.

[9] Bernardo, J. M., Ruenda, R. (2002), Bayesian hypothesis testing: a reference approach, *Internat. Stat. Rev.*, **70**, 351-372.

[10] Bernardo, J. M., Juárez, M. A. (2003), Intrinsic estimation, in *Bayesian statistics 7* (J. M. Bernardo *et al.*, eds), Oxford University Press, Oxford.

[11] Best, M. J., Grauer, R. R. (1991), Sensitivity analysis for mean-variance efficient portfolios to changes in asset means: some analytical and computational results, *The Review of Financial Studies,* **4**, 315-342.

[12] Black, F., Litterman, R. (1992), Global portfolio optimization, *Financial Analysts Journal*, **48**, 28-43.

[13] Box, G. E. P. (1980), Sampling and Bayes'inference in scientific modelling and robustness, *Journal of Royal Statistical Society*, **143**, 4, 383-430.

[14] Box, G. E. P., Tiao, G. (1962), A further look at robustness via Bayes's theorem*, Biometrika,* **49**, 419-432.

[15] Box, G. E. P., Tiao, G. (1964), A note on criterion robustness and inference robustness*, Biometrica,* **51**, 169-173.

[16] Box, G. E. P., Tiao, G. (1992), *Bayesian inference in statistical analysis,* Wiley, New York.

[17] Britten-Jones, M. (1999) The sampling error in estimates of mean-variance efficient portfolio weights, *Journal of Finance,* **54**, 655-671.

[18] Brown, S. J. (1979), The effect of estimation risk on capital market equilibrium*, Journal of Financial and Quantitative Analysis*, **14**, 215-220.

[19] Cavadini, F., Sbuelz, A., Trojani, F. (2001), A simplified way of incorporating model risk, estimation risk and robustness in mean variance portfolio management*, working paper,* University of Lugano, Switzerland.

[20] Chen, M. H., Shao, Q. M. (1997), On Monte Carlo methods for estimating ratios of normalizing constants, *The Annals of Statistics*, **25**, 1563-1594.

[21] Chen, Y., Fournier, D. (1999), Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries, *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 1525-1533.

[22] Chib, S., Jeliazkov, I. (2001), Marginal likelihood from the Metropolis-Hastings output, *Journal of the American Statistical Association*, **96**, 270-281.

[23] Chopra, V., K., Ziemba, W. T. (1993), The effect of errors in means, variances and covariances on optimal portfolio choice, *Journal of Portfolio Management,* **19**, 6-11

[24] Cifarelli, D.M., Muliere, P. (1989), *Statistica Bayesiana. Appunti ad uso degli studenti*, G. Iaculano, Pavia.

[25] Cont, R. (2001), Empirical properties of asset returns: stylized facts and statistical properties, *Quantitative Finance,* **1**, 223-236.

[26] Cox, D. R., Hinkley, D. V. (1974), *Theoretical statistics,* Chapman & Hall, London.

[27] Dawid, A. P. (1973), Posterior expectations for large observations*, Biometrika,* **60**, 664-666.

[28] Dey, D. K., Ghosh, S. K., Lou, K. (1996), On local sensitivity measures in Bayesian analysis (with discussion), in *Bayesian Robustness* (J. O. Berger *et al.*, eds)*,* IMS Lecture Notes - Monograph Series, **29**, Hayward: IMS.

[29] Fang, K. T., Anderson, T. W. (1990), *Statistical inference in elliptically contoured and related distributions*, Allenton Press Inc., New York.

[30] Gupta, A.K., Varga, T. (1993), *Elliptically contoured models in statistics*, Netherlands, Kluwer Academic Publishers.

[31] Frost, P. A., Savarino, J. E. (1986), An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis*, **21** 293-305.

[32] Gustafson, P. (1996), Model influence functions based on mixtures, *The Canadian Journal of Statistics*, **24**, 4, 535-548.

[33] Gustafson, P., Srinivasan, C., Wasserman, L. (1996), Local sensitivity analysis, in *Bayesian Statistics* 5 (J. O. Berger *et al.*, eds), Oxford University Press, Oxford.

[34] Gustafson, P. (2000), Local robustness in Bayesian analysis, in *Robust Bayesian Analysis* (D. Rìos Insua and F. Ruggeri, eds), Springer-Verlag, New York.

[35] Hampel, F. R. (1974), The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, **69**, 383-393.

[36] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986), *Robust statistics: the approach based on influence functions,* Wiley, New York.

[37] Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**, 97-109.

[38] Huber, P. J. (1981), *Robust statistics,* Wiley, New York.

[39] Jorion, P. (1986), Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis*, **21**, 279-292.

[40] Kelker, D. (1970), Distribution theory of spherical distributions and location-scale parameter generalization, *Sankhya*, **32**, 419-430.

[41] Landsman, Z., Valdez, E. (2003), Tail conditional expectation for elliptical distributions, *North American Actuarial Journal*, **7**, 4, 55-71.

[42] Lavine, M. (1991), Sensitivity in Bayesian statistics: the prior and the likelihood, *Journal of the American Statistical Association*, **86**, 396-399.

[43] Liseo, B., Petrella, L., Salinetti, G. (1996), Bayesian robustness: an interactive approach, *Bayesian Statistics* 5, (J. O. Berger *et al.*, eds), Oxford University Press, Oxford.

[44] Markowitz, H. (1952), Portfolio selection, *Journal of Finance*, **7**, 77-91.

[45] Markowitz, H. (1959), *Portfolio selection: efficient diversification of investments,* Yale University Press, New Haven.

[46] Meng, X., Wong, W. (1996), Simulating ratios of normalizing constants via simple identity: a theoretical exploration, *Statistica Sinica*, **6**, 831-860.

[47] Merton, R. C. (1980), On estimating the expected return on the market, *Journal of Financial Economics*, **8**, 323-361.

[48] Mira, A., Nicholls, G. (2001), Bridge estimation of the probability density at a point, *Technical report,* n. 2001.07, University of Insubria, Varese, Italy.

[49] Moreno, E., Martìnez, C., Cano, J. A. (1996), Local robustness and influence of contamination classes of prior distributions (with discussion), in *Bayesian Robustness* (J. O. Berger *et al.*, eds), IMS Lecture Notes - Monograph Series, **29**, 137-154. Hayward: IMS.

[50] Peña, D., Zamar, R. (1997), A simple diagnostic tool for local prior sensitivity, *Statistics and Probability Letters*, **36**, 205-212.

[51] Perret-Gentil, C., Victoria-Feser, M. P. (2003), Robust mean-variance portfolio selection, *Cahiers de recherche du Département d'Econométrie*, n. 2003.02, University of Geneva, Switzerland.

[52] Ramsay, J. O., Novick, M. R. (1980), PLU robust Bayesian decision theory: point estimation, *Journal of the American Statistical Association*, **75**, 901-907.

[53] Rìos Insua, D., Ruggeri, F. (2000), *Robust Bayesian analysis,* Springer-Verlag, New York.

[54] Ruggeri, F., Wasserman, L. (1993), Infinitesimal sensitivity of posterior distributions, *The Canadian Journal of Statistics*, **21**, 2, 195-203.

[55] Shyamalkumar, N., D. (1996), Bayesian robustness with asymptotically nice classes of priors, *Technical Report*, **96-22**, Dept. Stat., Purdue University.

[56] Shyamalkumar, N., D. (2000), Likelihood robustness, in *Robust Bayesian Analysis* (D. Rìos Insua and F. Ruggeri, eds), Springer-Verlag, New York.

[57] Sivaganesan, S. (1993), Robust Bayesian diagnostics, *Journal of Statistical and Planning Inference*, **35**, 2, 171-188.

[58] Sivaganesan, S. (1999), A likelihood based robust Bayesian summary, *Statistics and Probability Letters*, **43**, 5-12.

[59] Sivaganesan, S. (2000), Global and local robustness approaches: uses and limitation, in *Robust Bayesian Analysis* (D. Rìos Insua and F. Ruggeri, eds), Springer-Verlag, New York.

[60] Victoria-Feser, M. P (2000), Robust portfolio selection, *Cahiers de recherche HEC*, n. 2000.14, University of Geneva, Switzerland.

[61] West, M. (1984), Outlier models and prior distributions in Bayesian linear regression, *Journal of Royal Statistical Society (Series B)*, **46**, 3, 431-439.

[62] Zellner, A. (1976), Bayesian and non-Bayesian analysis on the regression model with multivariate Student-t error terms, *Journal of the American Statistical Association*, **71**, 400-405.