University of London

Simulation-based methods for time series diagnostics

Maria Teresa Catarino Leitao

London School of Economics and Political Science

1

UMI Number: U615246

Dissertation Publishing

THESES

F

8288

Abstract

We present sampling-based methodologies for the estimation of structural time series in the presence of outliers and structural shifts. We start by considering a simple structural model: a local level model, in the presence of outliers and level shifts. The existence of shocks is accounted for by including a product of intervention variables in the measurement and transition equations. These factors are composed of the product of an indicator variable and a parameter for the magnitude of the intervention variable, defining the size of the shocks. The Gibbs sampler is the Markov chain Monte Carlo method used for estimating the intervention model. Our contribution is in the use of a uniform prior distribution for the size of intervention variables. We show that this choice provides advantages over the usual multinomial and normal prior assumptions. The methodology is extended to a basic structural model. Using this model formulation, we consider 4 types of shocks: outliers, level, slope and seasonal shifts. The use of simulation based methods for this range of different breaks in structural models is not dealt with in the existing literature. By using the Gibbs sampler, we simultaneously estimate all the hyperparameters, detect the position of the shocks and estimate their size. Finally, we consider the local level model in the presence of outliers and level shifts for the case where one of the hyperparameters is equal to zero. In this situation, simulation based methods usually assume a multinomial prior distribution for the size of the intervention variables. We use a uniform prior, and present a two stages sampling scheme. In this two stage process the Gibbs sampler is first run on an auxiliary data set which has the same shocks as the original data set. For all the methods presented, performance is assessed by Monte Carlo studies and empirical applications to real data sets.

Acknowledgments

To my parents. To Anita and Tó.

4

Contents

List of Tables

List of Figures

11

Chapter 1

Introduction

A time series is a sequence of observations recorded over time. In general, the observations are serially correlated. The purpose of modeling time series is to determine the structure that best explains this correlation, and use it to forecast the future behaviour of the series.

For state space models, the observations evolve over time as a linear function of a state variable. The state variable is often latent. State space models are widely used after the Kalman filter was proposed in Kalman (1960). These models present a high degree of flexibility in the type of dynamics they can model. For example, any linear time series process has a state space representation.

Structural models are a class of unobserved components models. They decompose the time series in the sum of several unobserved effects, commonly irregular, trend, seasonal or cyclical effects. Structural models fit naturally into the state space framework. In Harvey (1989), an extensive study of the properties and algorithms for structural time series is presented. The basic structural model is defined as consigning irregular, trend and seasonal components.

Shocks, aberrant observations, extreme values, unexpected observations; these are some of the descriptive terms often associated with outliers. An extensive study of their characterization and implications for the estimation of a statistical model is given in Barnett and Lewis (1984). The implication of shocks in time series data is dependent on the way the shocks are included in the dynamics of the model. An outlier is the result of a shock that affects one single observation. A shock with a

more persistent effect results in a structural shift. One of the first characterizations of shocks in time series models is given in Fox (1972).

Bayesian methods for estimating time series models, are based on the statistical properties of the posterior samples of the parameters of the model. The seminal work of Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) draws attention to the potential statistical applications of these methods and lay the foundations for Markov chain Monte Carlo techniques. As access to computers is widespread and more powerful machines are available, the popularity of Bayesian techniques has increased. Bayesian methodologies in time series are well established and have accumulated a large literature. One of the more frequently used tools is the Metropolis-Hastings algorithm (Hastings, 1970). The Gibbs sampler, Geman and Geman (1984), which can be viewed as a composition of several Metropolis-Hastings steps, is the technique we shall focus our attention on.

The aim of this work is to derive and implement sampling-based methods for the estimation of structural time series, in the presence of outliers and structural shifts. Our contribution has three main components:

- the prior distribution for the size of the shocks variable use of a flat uninformative distribution.

- a Bayesian methodology for the estimation of a basic structural model (BSM), in the presence of outliers, level shifts, slope shifts and seasonal shifts.

- a Bayesian method for the detection of outliers and level shifts, for a local level model, when one of the hyperparameters is equal to zero, assuming a continuous prior distribution for the size of the shocks variables.

McCulloch and Tsay (1994) describe a method for detection of outliers and structural shifts in time series models. Gerlach, Carter, and Kohn (2000) focus on application of the Gibbs sampler to diagnostics on state space models. The prior assumptions commonly used, for the size of the shocks variables are multinomial distributions and Gaussian distributions. The main advantage of using a flat distribution is that it requires less prior knowledge about the characteristics of the shocks

16

that might be present in the data. When assuming a multinomial prior, we have to define *a priori* the set of values from which to sample the size of the shocks. Assuming a normal distribution, the prior mean and variance are set *a priori*. We show that the posterior samples obtained for the size of the shocks variables, present some undesirable sensitivity to the choice of the distribution parameters, namely to the choice of the variance. Assuming a flat prior distribution, the parameters u and v for a $U[u, v]$ have to be set *a priori*. We show, by means of a sensitivity study, that the posterior samples present less sensitivity to the choice of these parameters, than to the choice of the normal distribution parameters. Therefore, by assuming a uniform distribution we obtain an estimation method that is less dependent on prior assumptions than the existent methodologies.

Generalizing the method we propose for detecting outliers and level shifts, we present a Bayesian algorithm for detection of outliers and structural shifts, for the basic structural model. The BSM defines the dynamics of a time series as the sum of several components: an irregular, trend and seasonal component. We consider the case of a trend with a stochastic slope component. Given this decomposition of the time series, two main types of shocks might occur: outliers and structural shifts. The structural shifts can be of three types: level, slope and seasonal shift. We formulate a model allowing for these four type of shocks. Using the Gibbs sampler, we estimate the hyperparameters and simultaneously detect the position and estimate the size of shocks in the data. An additional innovation is that we assume a uninformative uniform prior for the size of all the type of shocks considered. The choice of this prior distribution has the aim of making the process of detection and characterization of the shocks has independent as possible, from prior knowledge of the behaviour of the time series data.

The methods we have so far described, are valid if none of the hyperparameters, for the equations through which shocks affect the data, is equal to zero. Consider the case of the local level model where outliers and level shifts might be present. If either of the two hyperparameters is equal to zero, the methods we have proposed imply sampling from a degenerate normal distribution when sampling from the size of the shocks full conditional distribution. The case of one of the hyperparameters

being equal to zero is often overlooked in the literature. Gerlach, Carter, and Kohn (2000), deal with this case, but they assume a discrete prior distribution for the intervention variables. As we have argued before, this assumption is restrictive and demands a considerable prior knowledge of the data set.

To overcome the problem posed by having one of the hyperparameters equal to zero in the uniform prior case, we propose a two stage sampling scheme. In the first stage we apply the Gibbs sampler to an auxiliary data set. This data set is constructed in such a way that it has the same shocks as the original data set. It is modeled as a local level model, but with both hyperparemeters different from zero. It is used to detect the shocks, that affect the observations through the equation with the null hyperparameter. In the second stage, we run a second Gibbs sampler to detect the other type of shocks and estimate the non null hyperparameter.

The results presented for the empirical applications and Monte Carlo studies are obtained using Ox (Doornik, 1999). The Ox package SsfPack, (Koopman, Shephard, and Doornik, 1999), for estimation of state space models, was also used extensively. The results obtained using maximum likelihood, which are reported for comparison with the results we obtain by using our Bayesian approach, were generated using the structural time series package STAMP (Koopman, Harvey, Doornik, and Shephard, 2000).

We briefly describe the structure that this work follows, by giving a summary of the main topics covered in each chapter. In Chapter 2, we present a review of the results available in the literature concerning the two main aspect of our work: Bayesian methods and state space models. We present some of the main definitions and results needed to apply Markov chain Monte Carlo methods. From this class of methods we will use the Gibbs sampler. We describe this algorithm and conditions that ensure its convergence. Next, we present the formulation of state space models. The Kalman filter and smoother, and a simulation smoother are explained. To illustrate the application of the Gibbs sampler for the estimation of state space models, we present the results obtained by applying the sampler to an artificial data set generated from a local level model. Furthermore, we perform an analysis of sensitivity to assess the effect, of the choice of the parameters of the

prior distributions of the hyperparameters, on the parameters estimated.

In Chapter 3 we study the detection of outliers and level shifts, for the local level model, using the Gibbs sampler. We assume that both hyperparameters are different from zero. The model is formulated for the detection of shocks by including intervention variables for the presence of outliers and level shifts. For each type of shock this amounts to defining two intervention variables: an indicator variable for the presence of a shock, and a size of shock variable. The innovation in the method we propose is that the size of the shock is assumed to have a flat prior distribution. When assuming a flat distribution, the boundaries of a bounded uniform distribution have to be set *a priori*. To compare the performance of the sampler when assuming a flat prior with a normal prior for the size of the shocks variable, we perform an analysis of sensitivity to the choice of the parameters of the normal and uniform distributions. We show that using the flat prior, the estimation results are less sensitive to the choice of the prior distribution parameters. Having established the benefits of the uninformative prior, a sampling algorithm is proposed for estimating a local level model, in the presence of outliers and level shifts. A Monte Carlo study is presented, for assessing the performance of the sampler. The artificial data sets are generated from a local level model, where two outliers and two level shifts are input. To illustrate the application of the method to a real data set, we consider the data composed of the coal consumption in the UK, from the first quarter of 1960 to the fourth quarter of 1986.

In Chapter 4 we consider a more general model formulation; the basic structural model. The time series is decomposed as the sum of an irregular, trend and seasonal components. The trend component has a stochastic slope. The seasonal component is a formulated through dummy variables. A Gibbs sampler is presented. From running the sampler we obtain posterior samples, that allow us to estimate the four hyperparameters, detect the position of the outliers and structural shifts, and estimate their sizes. The uninformative uniform distribution is chosen as prior distribution for the size of the four type of shocks that might be present in the data being modeled: outliers, level shifts, slope shifts and seasonal shifts. We assume that the hyperparameters, for equations where intervention variables are included, are

different from zero. Our contribution, is that we present a method for simultaneously estimate the hyperparameters and detect any type of shock, for a basic structural model. Furthermore, with the uniform prior assumptions the shock detection is done without requiring an extensive prior analysis of the data. The results from a Monte Carlo study are presented. The data are generated from a basic structural model. All the four type of shocks are input to each data set: an outlier, a level shift, a slope shift and a seasonal shift. Our methodology is applied to the real data set of the number of marriages in the UK, from the first quarter of 1958 to the fourth quarter of 1984.

In Chapter 5, we consider the problem of estimating a local level model, when outliers and level shifts are present in the data, and one of the hyperparameters is equal to zero. Our aim is to be able to keep the assumption of a prior uniform for the size of the shocks, given that we have established the advantages of using this distribution in the previous chapters. To do so, we present a sampling scheme made up of two stages. These stages depend on which hyperparameter is equal to zero. When the irregular hyperparameter is equal to zero, we generate an auxiliary data set, which has the same outliers has the original data set, follows a local level model, but has an irregular variance different from zero. The sampling scheme described in Chapter 3, is applied to the auxiliary data set, and we detect the position and estimate the size of the outliers. Inputting this information, in a second stage, the sampling scheme is run, for the original data set, without sampling from the variables related to the detection of the outliers. In this way, we overcome the problem of detecting the outliers when the irregular hyperparameter is equal to zero. From the second stage we obtain posterior samples that will allow us to estimate the level variance, detect the level shifts and estimate their sizes. We propose an analogous method for the case when the level variance is equal to zero. To detect the level shifts, given that the level variance is equal to zero, we generate an artificial data set, that has the same level shifts as the original data, but has a level hyperparameter different from zero. The sampling scheme from Chapter 3 is run for this auxiliary data set, to detect and characterize the level shifts. Having detected this type of shocks, their position and estimated sizes are input, and the sampling scheme is

20

run, for the original data set. It delivers an estimate for the irregular variance, the position of the outliers and estimates of their sizes. The performance of the sampling schemes proposed is analyzed by two Monte Carlo studies, for each of the hyperparameters equal to zero. As an empirical application, when the irregular variance is equal to zero, we model the real data set of monthly quotes of bonds issued by the Greek government, from August 1916 to June 1930. The volume of the Nile data set, from 1871 to 1970, is used as an empirical application, for the case when the level variance is equal to zero.

In Chapter 6, conclusions are drawn on the methodologies we have proposed and the results obtained. Some considerations are made on possible future lines of research using the approaches proposed.

Chapter 2

Simulation methods for state space models

2.1 Introduction

A broad literature is available on classical methods for estimating parametric time series models, from ARMA models, (Box and Jenkins, 1970) to state space models (SSM), (Harvey, 1989). The aim of this chapter is to review Bayesian methodologies for estimating time series which fit in the space state modeling framework.

The basic difference between Bayesian and classical approaches for parametric inference is that Bayesian methods make no distinction between observations Y and parameters θ, in that they consider all of them random variables. Parametric inference is based on the *posterior distribution* $P(\theta|Y)$. Means, quantiles, confidence intervals or any other statistical properties for θ are obtained from samples of that posterior. In most cases the posterior distribution is not available in a closed form; the methodology used to overcome this problem is what distinguishes the majority of Bayesian methods.

The impact of the parameters on the observable Y is measured by the *likelihood function* $P(Y|\theta)$. Information, if any, on the parameters distribution previous to the observation of Y is summarized in the *prior distribution* $P(\theta)$. Using Bayes' theorem we have the following expression that relates these three distributions:

$$P(\theta|Y) = \frac{P(\theta)P(Y|\theta)}{\int P(\theta)P(Y|\theta)d\theta}. \tag{2.1}$$

As $\int P(\theta)P(Y|\theta)d\theta$ integrates out the dependence on the parameters, expression

(2.1) states that the posterior distribution $P(\theta|Y)$ is proportional to the product of the likelihood $P(Y|\theta)$, and the prior distribution $P(\theta)$:

$$P(\theta|Y) \propto P(\theta)P(Y|\theta).$$

Markov chain Monte Carlo (MCMC) methods rely on constructing a process which is Markovian and has as its limiting distribution the posterior distribution. MCMC methods are distinguished by the way in which these processes are constructed. When applying this methodology to time series, Y will be a set of time series data $\{Y_1, \ldots, Y_T\}$ and θ the set of parameters in the model, together with any missing observations.

In section 2.2 the main Markov chain notations will be presented, together with results that are the basis for MCMC methods. The characterization of these methods will be done with special reference to the Metropolis-Hasting algorithm and the Gibbs sampler. Comments are made on problems arising in the use of the general MCMC framework.

The main objective of the results presented for MCMC methods is their application to SSM. These models are introduced in section 2.3, together with examples, and algorithms.

For simplicity, we consider the estimation of SSM, in the context of the Gibbs sampler, divided in two steps: sampling from the full conditional of the state vector and sampling from the full conditional of the parameters. To sample from the full conditional of the states we use a simulation smoother algorithm, and this method is explained for a general SSM. The methodology for sampling from the full conditional of the parameters is presented for an unobserved components model (UCM).

The estimation of SSM using the Gibbs sampler is exemplified for an artificial data set, generated from a local level (random walk plus noise) model. A Monte Carlo experiment is conducted to study the sensitivity of the posterior samples of the hyperparameters to the prior distributions parameters.

2.2 Markov chain Monte Carlo methods

2.2.1 Definitions and ergodic results

A sequence of random variables $X^{(0)}, X^{(1)}, \ldots, X^{(t)}, \ldots$ is a *Markov chain* $\left(X^{(t)}\right)$ if the distribution of $X^{(t+1)}$ given all the previous states of the chain $X^{(0)}, \ldots, X^{(t)}$ depends only on $X^{(t)}$. For any t,

$$P\left(X^{(t+1)} \in A | X^{(0)}, X^{(1)}, \ldots, X^{(t)}\right) = P\left(X^{(t+1)} \in A | X^{(t)}\right),$$

for any given $A \in B(S)$, the set of all the subsets of S, the state space where the chain is defined. Typically A is a subset of $\Re^k$. The *transition kernel* $P(x, A)$ of a Markov chain is defined as:

$$P(x, A) = P\left\{X^{(t+1)} \in A | X^{(t)} = x\right\}. \tag{2.2}$$

We are interested only in time homogeneous Markov chains and for that reason $P(x, A)$ in (2.2) is independent of t.

The nth-iterate of (2.2) is

$$P^n(x, A) = P\left\{X^{(n)} \in A | X^{(0)} = x\right\},$$

and $P^n(\cdot | x)$ the conditional distribution of $X^{(n)}$ given the initial state of the chain $X^{(0)} = x$.

A distribution π is an *invariant* distribution for a Markov chain if:

$$\pi(A) = \int P(x, A) \pi(x) dx,$$

for all measurable sets A. Under certain general conditions, which are stated later, an invariant distribution π is also an *equilibrium* distribution, that is, for π-almost x[1]:

$$\lim_{n \to \infty} P^n(x, A) \longrightarrow \pi(A),$$

In other words, under some conditions, if we run the chain for long enough, the distribution of the states converges to the invariant distribution independently of the initial state distribution.

---

[1]The set where the condition does not hold has null measure.

Suppose we want to estimate $E_\pi(g)$:

$$E_\pi(g) = \int g(x)\pi(x)dx,\qquad(2.3)$$

where $g(\cdot)$ is a real-valued function. If we construct a Markov chain $(X^{(t)})$ that converges to the target distribution $\pi(\cdot)$, the sample analogue to the expectation in (2.3) is:

$$\hat{E}_\pi(g) = \bar{g}_N = \frac{1}{N}\sum_{i=1}^{N}g(X^{(i)}).\qquad(2.4)$$

In order to be able to use Markov chains so that (2.4) converges to the expected value in (2.3) several questions must be addressed. In particular, we require conditions that insure convergence of the chain to a unique limit distribution, and a method to construct the desired chain. We start by presenting some general definitions and results for Markov chains. Afterwards we explain some of the methods available for constructing those chains, which are a subset of the available MCMC techniques.

Let P be the transition kernel of a Markov chain, defined in a finite σ-algebra $(S, B(S))$. Given a σ-finite measure π, P is π-*irreducible* if, independently of the initial state $X^{(0)}$, the probability of achieving any measurable set $A \in B(S)$, with $\pi(A) > 0$, in a finite number of steps is positive; that is, for each $x \in S$ there exists an n such that $P^n(x, A) > 0$.

An π-irreducible chain is called *recurrent* if it will visit A an infinite number of times: for any measurable set A with $\pi(A) > 0$,

$$P(X^{(n)} \in A \text{ infinitely often}| X^{(0)} = x) > 0 \quad \text{for all } x,$$

$$P(X^{(n)} \in A \text{ infinitely often}| X^{(0)} = x) = 1 \quad \text{for } \pi\text{-almost } x.$$

If $P(X^{(n)} \in A \text{ infinitely often}| X^{(0)} = x) = 1$, for all x, the chain is called *Harris recurrent*.

A recurrent π-irreducible chain is *positive recurrent* if there is a finite invariant measure for P. Otherwise the chain is *null recurrent*.

An π-irreducible chain is said to be *aperiodic* if it can not oscillate between measurable non-empty sets in periodic movements. Otherwise it is called *periodic*

and there exists $n > 2$ and a sequence of non-empty sets $\{A_0, A_1, \ldots, A_{n-1}\}$ in $B(S)$ such that for all $i = 0, \ldots, n-1$ and all $x \in A_i$:

$$P(x, A_j) = 1, \quad \text{for } j = i + 1 \pmod{n}.$$

MCMC methods rely on constructing Markov chains where the invariant distribution coincides with the target distribution π. If the chain is irreducible, positive recurrent and aperiodic the next theorem, in Tierney (1994), implies that π is also the equilibrium distribution:

Theorem 2.2.1 *Suppose P is π-irreducible and π is an invariant distribution for P. Then P is positive recurrent and π is the unique invariant distribution of P. If P is also aperiodic then, for π-almost x,*

$$\|P^n(x, .) - \pi\|_{TV} \rightarrow 0. \tag{2.5}$$

If P is Harris recurrent, then the convergence occurs for all x.

The *Total Variation* distance $\| \cdot \|_{TV}$ is defined, for any bounded measure ϕ on $(S, B(S))$, as

$$\|\phi\|_{TV} = \sup_{A \in B(S)} \phi(A) - \inf_{A \in B(S)} \phi(A)$$

By defining a kernel with invariant distribution π, such that it is irreducible, positive recurrent and aperiodic, after a long enough run of the chain, states from the equilibrium distribution π will be generated, independently from the starting state. The question of how long should the chain be run for is related to the rate of convergence of (2.5) and will be addressed later. In order to present results concerning the asymptotic properties of the estimator in (2.4) we start by presenting some definitions of ergodicity. A Markov chain is called *ergodic* if it is Harris recurrent and aperiodic. A stronger form of ergodicity, is uniform ergodicity. A Markov chain is *uniformly ergodic* if there exists a constant $M > 0$ and $0 < r < 1$ such that:

$$\sup_{x \in X} \|P^n(x, \cdot) - \pi\|_{TV} \leq Mr^n.$$

Uniform ergodicity is a sufficient condition for establishing the central limit result in the next theorem (Tierney 1994):

Theorem 2.2.2 *Suppose $(X^{(t)})$ is uniformly ergodic with equilibrium distribution π and suppose g is a real valued function and $\boldsymbol{E}_\pi(g^2) < \infty$. Then there exists a real number $\sigma(g)$ such that the distribution of*

$$\sqrt{N}(\bar{g}_N - \boldsymbol{E}_\pi(g))$$

converges in distribution to a normal distribution with mean 0 and variance $\sigma(g)^2$ for any initial distribution.

By construction, if $\left(X^{(t)}\right)$ is a random sample obtained from a stationary Markov chain, after convergence is achieved, all the $X^{(t)}$ will have the same distribution, but they are not independent. And although that does not affect the ergodic result in Theorem 2.2.2, it will have implications in obtaining a consistent estimator of the variance $\sigma(g)^2$, the MCMC variance of the posterior sample mean. From the time series literature a method to obtain a consistent estimate of the variance for a correlated sample is by using a smoothed estimate of the periodogram at frequency 0, with an appropriate choice of a lag window, see for example Brockwell and Davis (1991).

Suppose we are interested in estimating $\boldsymbol{E}_\pi (g)$ where $g \in L^2$ [2], using a simulated sample $\{x^{(1)}, \ldots, x^{(N)}\}$ from a Markov chain, which has reached equilibrium, by the sample mean $\bar{g}_N$ using the formula in (2.4). For estimating the variance of $\bar{g}_N$ we use a Parzen window $w(\cdot)$, defined as in Priestley (1981), in the following way:

$$\text{vâr}(\bar{g}_N) = \frac{1}{N}\left[\hat{\gamma}(0) + \frac{2N}{N-1}\sum_{i=1}^{M_N} w\left(\frac{i}{M_N}\right)\hat{\gamma}(i)\right], \tag{2.6}$$

where

$$\hat{\gamma}(i) = \frac{1}{N}\sum_{j=i+1}^{N}\left(g(x^{(j)}) - \bar{g}_N\right)\left(g(x^{(j-i)}) - \bar{g}_N\right).$$

The bandwidth is chosen according to the size of the sample being considered and its correlation structure . The weights of the Parzen window are defined as:

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3, & |x| < \frac{1}{2}, \\ 2(1 - |x|)^3, & \frac{1}{2} \leq |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

---

[2]Let (Ω, F, μ) be a measurable space. $g \in L^p \subseteq F$, $p \in [1, +\infty]$ if $\int_\Omega |g|^p d\mu < \infty$.

In Geyer (1992) several other methods for consistent estimation of the MCMC variance are discussed.

The *relative numerical efficiency*, Geweke (1989), is defined as the ratio between the lag window estimator of the variance and the estimate of the variance if assuming an independent sample. This quantity is a measure of the speed of the mixing of the chain.

In the next section we see how the previous results can be applied when considering Markov chains constructed using the Metropolis-Hastings algorithm and the Gibbs sampler.

2.2.2 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm was proposed by Hastings (1970). Suppose we are interested in sampling from a target distribution $\pi(\cdot)$, and that is easy to sample from a proposal distribution $q(x|y)$, known up to a normalizing constant. A Markov chain is constructed by using as generator density $q(y|x^{(t)})$, where $x^{(t)}$ is the present state of the chain. Let $Y \sim q(y|x^{(t)})$. The new state of the chain will be given by:

$$X^{(t+1)} = \begin{cases} Y, & \text{with probability } \alpha(x^{(t)}, Y) \\ x^{(t)}, & \text{with probability } 1 - \alpha(x^{(t)}, Y), \end{cases} \qquad (2.7)$$

where

$$\alpha(x, y) = \min\left\{ \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}, 1 \right\}.$$

The transition kernel of the chain constructed in this way is given by:

$$P(y|x) = \alpha(x, y)q(y|x) + (1 - r(x))\, \delta_x(y),$$

with $r(x) = \int \alpha(x, y)q(y|x)dy$.

In order to ensure that the target distribution π is also the invariant distribution of the chain, a minimal condition is imposed:

$$\bigcup_{x \in \text{supp } \pi} \text{supp } q(\cdot|x) \supset \text{supp } \pi = \mathcal{S},$$

as it is stated in the next theorem (Robert and Casella 1999):

Theorem 2.2.3 *For every conditional distribution q, whose support includes S, π is an invariant distribution of the chain $\left(X^{(t)}\right)$ produced by (2.7).*

The following result, in Roberts and Tweedie (1996), gives sufficient conditions for irreducibility and aperiodicity of the Markov chain obtained through the Metropolis-Hastings algorithm (assuming S is connected):

Theorem 2.2.4 *Assume π is bounded and positive on every compact set of its support S. If there exist positive numbers ε and δ such that*

$$q\left(y|x\right) > \varepsilon \ \ if \ |x - y| < \delta,$$

then the Metropolis-Hastings Markov chain $\left(X^{(t)}\right)$ is π-irreducible and aperiodic. Moreover, any nonempty compact set is a small set.

Using Theorem 2.2.1, the two previous theorems provide sufficient conditions for the Metropolis-Hastings chain to converge to a unique limit distribution, that coincides with the target distribution π.

The next theorem (see Robert and Casella, 1999 for proof), states conditions that ensure convergence of the posterior sample mean estimator.

Theorem 2.2.5 *Suppose that for the Metropolis-Hasting Markov chain $\left(X^{(t)}\right)$, we have that $q(y|x) > 0$ for every $(x,y) \in S \times S$ (sufficient condition for being irreducible).*
If $g \in L^1$, then

$$\lim_{N\to\infty} \bar{g}_N = E_\pi\left(g\right)$$

for π-almost everywhere.

These results do not provide any information regarding the rate of convergence of the chain. That is an important aspect as the chain may remain in the same state for a long time. The proportion of time t for which $X^{(t+1)} = X^{(t)}$ is defined as the rejection rate.

Several MCMC algorithms are classified as being of the Metropolis-Hastings type. Examples include the Metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth,

Teller, and Teller, 1953), where $q(X|Y) = q(Y|X)$, and the Gibbs sampler (Geman and Geman, 1984, Gelfand and Smith, 1990) where the proposal distribution coincides with the full conditional distribution.

2.2.3 Gibbs sampler

Suppose that $X \sim \pi(\cdot)$, is composed of several blocks $X = (X_1, \ldots, X_p)$, $p \geq 2$. Additionally suppose that it is possible to sample from all the full conditional distributions $\pi_i(X_i|X_{-i})$, where $X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p)$, for $i = 1, \ldots, p$. Let $X_{-i}^{(t)} = \left(X_1^{(t+1)}, \ldots, X_{i-1}^{(t+1)}, X_{i+1}^{(t)}, \cdots X_p^{(t)}\right)$. The Gibbs sampling constructs a Markov chain via the following algorithm. Given the present state of the chain $X^{(t)}$ it will move, with probability one, to the state $X^{(t+1)}$, sequentially generated by:

$$X_i^{(t+1)} \sim \pi_i\left(X_i|X_{-i}^{(t)}\right), \quad i = 1, \ldots, p.$$

This algorithm generates a Markov chain with transition kernel:

$$P(y|x) = \prod_{i=1}^{p} \pi(y_i|x_j, j > i, y_j, j < i)$$

It is easy to check that the updating of each component can be done by application of the Metropolis-Hastings algorithm. Consider the updating of the component i, $X_i^{(t)}$. Let $\pi_{-i}(\cdot)$ be the distribution for the vector X_{-i}. A candidate Y_i for the updating of $X_i^{(t)}$ will be accepted with probability:

$$
\begin{aligned}
\alpha(X_i^{(t)}, X_{-i}^{(t)}, Y_i) &= \min\left(\frac{\pi(Y_i, X_{-i}^{(t)})\pi_i(X_i^{(t)}|X_{-i}^{(t)})}{\pi(X_i^{(t)}, X_{-i}^{(t)})\pi_i(Y_i|X_{-i}^{(t)})}, 1\right) \\
&= \min\left(\frac{\pi_{-i}(X_{-i}^{(t)})}{\pi_{-i}(X_{-i}^{(t)})}, 1\right) \\
&= \min(1, 1) = 1.
\end{aligned}
$$

In conclusion, the Gibbs sampler is obtained by composing several Metropolis-Hastings algorithms, with acceptance rates uniformly equal to one and such that for each of those algorithms the proposal distributions are the full conditional distributions.

By construction the chain obtained with the Gibbs sampler has π as an invariant distribution (see Chan, 1993). The several sub-chains obtained, by sampling from

the full conditional distributions are not irreducible and so we can not use the result in Theorem 2.2.1. In general to ensure that the chain is π-irreducible and aperiodic is sufficient that:

$$\pi_i(X_i|X_{-i}) > 0, \quad \text{for all } i \text{ and } X, \tag{2.8}$$

as is proved in Geman and Geman (1984) and Chan (1993). Under very mild conditions (Roberts and Smith, 1994), namely the condition in (2.8) the following convergence and ergodic results hold:

Convergence in distribution: As $t \to \infty$

$$\left(X_1^{(t)}, X_2^{(t)} \ldots, X_p^{(t)}\right) \xrightarrow{d} \pi\left(X_1, X_2 \ldots, X_p\right)$$

and hence for each $i = 1, \ldots, p$

$$\left(X_i^{(t)}\right) \xrightarrow{d} \pi_i\left(X_i\right)$$

Geometric rate of convergence:

Using the Total Variation norm $\| \cdot \|_{TV}$, $\left(X_1^{(t)}, X_2^{(t)} \ldots, X_p^{(t)}\right)$ converges to the true distribution in a geometric rate of convergence in t.

Ergodic theorem: For any measurable function T of $(X_1, X_2 \ldots, X_p)$ whose expectation exists,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} T\left(X_1^{(t)}, X_2^{(t)} \ldots, X_p^{(t)}\right) \xrightarrow{a.s.} E_\pi(T).$$

2.2.4 Some practical issues

The Gibbs sampler is an example of an MCMC algorithm in, at least, a bidimensional space. In fact in most of the practical applications of MCMC methods the interest lies in features of multivariate random variables. A sample from the chain obtained after reaching equilibrium, will not be independent and the existence of correlation poses more serious problems when we are interested in samples from different random variables, which might also be correlated among themselves. Several techniques can be used for variance reduction, which reduces the length of the run of the chain, after convergence is achieved, for a desired level of accuracy of the

estimates. For example, the use of antithetic variables, Green and Han (1992), or blocking schemes where the variables are sampled in blocks. In Liu, Wong, and Kong (1994) it is showed that, for the Gibbs sampler, grouping random variables can result in more efficient sampling schemes.

Another question concerning these methods is how to assess the convergence of the chain and for how long should it be run (the determination of the *burn-in period*), before we start storing results. Some theoretic results are available on the rates of convergence and provide lower bounds for the burn in period. In Roberts and Sahu (1997) bounds for the rate of convergence of the Gibbs sampler, when the target distribution is Gaussian, are derived. However these bounds are usually difficult to obtain and depend on the target distribution.

In practice, detection of convergence can be done by analyzing the simulated chain. Techniques include inspection of the path simulated, analysis of the correlation structure, which should present a rapid convergence towards zero. These are not exact methods of assessing convergence, and might be misleading in some cases. However, they present the advantage of being easily implemented. In Brooks and Gelman (1998) a classification of the different methods of assessing convergence is given, together with a review of the different methodologies.

2.3   State space models

Many time series models can be represented in state space form. This formulation is quite unrestricted and it allows for the inclusion of unobservable effects such as level, trend, seasonality or cycles. It also allow us to model latent variables, such as the volatility of a financial asset. The general formulation of SSM considers a N-dimensional time series, $\mathbf{y}_t = (y_t^1, \ldots, y_t^N)'$ which is related to the m-dimensional *state space vector* $\boldsymbol{\alpha}_t = (\alpha_t^1, \ldots, \alpha_t^m)'$, through the *measurement* equation:

$$\mathbf{y}_t = \mathbf{c}_t + \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{G}_t \mathbf{u}_t, \qquad (2.9)$$

and the state space vector evolution is determined by the *transition* equation:

$$\boldsymbol{\alpha}_{t+1} = \mathbf{d}_t + \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{H}_t \mathbf{u}_t, \qquad (2.10)$$

for $t = 1, \ldots, T$.

The innovations process can be assumed to have any distribution. As we are interested in the case of Gaussian SSM, we assume that:

$$\mathbf{u}_t \sim NID\left(0, \sigma^2 \mathbf{I}_r\right).\tag{2.11}$$

The matrices $\mathbf{Z}_t$ $(N \times m), \mathbf{T}_t$ $(m \times m), \mathbf{G}_t$ $(N \times r)$, and $\mathbf{H}_t$ $(m \times r)$ are deterministic, but not necessarily constant over time. The parameters responsible for the stochastic movements of the state variables are called *hyperparameters*. With the formulation above, they correspond to σ^2 and any non zero element in the matrices $\mathbf{G}_t$ and $\mathbf{H}_t$. For the above model to be completely specified we have to impose an initial condition:

$$\boldsymbol{\alpha}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1),\tag{2.12}$$

together with $\boldsymbol{\alpha}_1$ being uncorrelated with the innovation vector $\mathbf{u}_t$, $E(\boldsymbol{\alpha}_1, \mathbf{u}_t) = 0$, for $t = 1, \ldots, T$.

We will focus our attention on the unidimensional time series case ($N = 1$), for which we will start by presenting two examples of models that have the representation defined by (2.9) to (2.12): a stationary AR(p) model and unobserved component model. Suppose we have a stationary AR(p) model:

$$y_t - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p} = \varepsilon_t,$$

where all the solutions of $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p = 0$ are outside the unit circle, and the innovations are independent and identically distributed with $\varepsilon_t \sim N(0, \sigma^2)$, for $t = 1, \ldots, T$. This model is a SSM defined by:

$$
\begin{aligned}
\boldsymbol{\alpha}_t &= \left(y_t, \ldots, y_{t-p+1}\right)', \\
\mathbf{Z}_t &= (1, 0, \cdots, 0), \\
\mathbf{T}_t &= \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ & & \cdots & & \\ & & \cdots & 1 & 0 \end{bmatrix}, \\
\mathbf{G}_t &= 0,
\end{aligned}
$$

$$\mathbf{H}_t = [1 \cdots 0]',$$

$$\mathbf{c}_t = 0,$$

$$\mathbf{d}_t = [0 \ldots 0]',$$

$$u_t \sim N(0, \sigma^2).$$

Another type of models that has an SSM representation are the unobserved component models. For this class of models, y_t is decomposed as the sum of several effects, most commonly the sum of irregular, trend, cyclical and seasonal components. A example of UCM is the basic structural model (BSM) defined as:

$$
\begin{array}{llll}
y_t & = \mu_t + \gamma_t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma_\varepsilon^2), & \\
\mu_{t+1} & = \mu_t + \beta_t + \eta_t, & \eta_t \sim N(0, \sigma_\eta^2), & \\
\beta_{t+1} & = \beta_t + \zeta_t, & \zeta_t \sim N(0, \sigma_\zeta^2), & (2.13) \\
\sum_{i=0}^{s-1} \gamma_{t+1-i} & = \omega_t, & \omega_t \sim N(0, \sigma_\omega^2), &
\end{array}
$$

where we have considered a trend component μ_t, a dummy seasonal component γ_t, with seasonal periodicity of s, and an irregular component ε_t. The trend component has a stochastic slope component β_t. The different disturbances are taken to be mutually uncorrelated, and normally distributed. For simplicity we consider $s = 4$, which corresponds to quarterly data. The vectors for the SSM representation are then:

$$\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \gamma_t, \gamma_{t-1}, \gamma_{t-2})',$$

$$\mathbf{Z}_t = (1\ 0\ 1\ 0\ 0),$$

$$\mathbf{T}_t = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\mathbf{G}_t = [\sigma_\varepsilon\ 0\ 0\ 0],$$

$$\mathbf{H}_t = \begin{bmatrix} 0 & \sigma_\eta & 0 & 0 \\ 0 & 0 & \sigma_\zeta & 0 \\ 0 & 0 & 0 & \sigma_\omega \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{c}_t = 0,$$

$$\mathbf{d}_t = [0 \ldots 0]',$$

$$\mathbf{u}_t \sim NID(0, \mathbf{I}_4).$$

Basic structural models have the property that the components of the state space vector have diffuse initial conditions, $\boldsymbol{\alpha}_1 \sim N(0, \kappa \mathbf{I})$ with $\kappa \to \infty$. In this particular case:

$$\mu_1 \sim N(0, \kappa I), \tag{2.14}$$

$$\beta_1 \sim N(0, \kappa I), \tag{2.15}$$

$$\gamma_i \sim N(0, \kappa I), \ i = -1, 0, 1. \tag{2.16}$$

For models in SSM several algorithms have been developed of which the best known is the Kalman filter. Originally from the engineering literature, (Kalman, 1960) the potential statistical applications were put forward in works such as Jazwinski (1970) and Harvey (1989). The Kalman filter is a recursive process that gives the optimal linear estimate of the state space vector $\boldsymbol{\alpha}_t$ given all the information available at time $t - 1$, which we represent by $\mathbf{a}_t = E(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1})$. Suppose we have $\mathbf{a}_t$ and its variance-covariance matrix $\mathbf{P}_t$. Then, when observation $\mathbf{y}_t$ is available, the one step-ahead prediction of the state vector and its variance are obtained using the equations that define the Kalman filter:

$$\mathbf{v}_t = \mathbf{y}_t - \mathbf{c}_t - \mathbf{Z}_t \mathbf{a}_t, \tag{2.17}$$

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{G}_t \mathbf{G}_t', \tag{2.18}$$

$$\mathbf{K}_t = \left(\mathbf{T}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H}_t \mathbf{G}_t'\right) \mathbf{F}_t^{-1}, \tag{2.19}$$

$$\mathbf{a}_{t+1} = \mathbf{d}_t + \mathbf{T}_t \mathbf{a}_t + \mathbf{K}_t \mathbf{v}_t, \tag{2.20}$$

$$\mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \left(\mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t\right)' + \mathbf{H}_t \left(\mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t\right)'. \tag{2.21}$$

Starting with initial conditions $\mathbf{a}_1$ and $\mathbf{P}_1$, given by (2.12), these equations will output: $\mathbf{a}_t = E(\boldsymbol{\alpha}_t | \mathbf{Y}_{t-1})$ and $\mathbf{P}_t = E\left[(\boldsymbol{\alpha}_t - \mathbf{a}_t)(\boldsymbol{\alpha}_t - \mathbf{a}_t)'\right]$ for $t = 2, \ldots, T$. The Kalman gain $\mathbf{K}_t$ represents the decrease in the error variance from the information contained in $\mathbf{y}_t$. The one-step ahead prediction error is given by $\mathbf{v}_t = \mathbf{y}_t - E(\mathbf{y}_t | \mathbf{Y}_{t-1})$ and var $(\mathbf{v}_t) = \mathbf{F}_t$.

For a Gaussian SSM, $\mathbf{a}_t$ is the minimum mean square error estimator (MMSE) of the state space vector given all the information available until $t - 1$ (assuming

the parameters of the model are known). In the non-Gaussian framework, a_t is the MMSE in the class of linear estimators (MMSLE).

When considering nonstationary time series models, for example the model defined in (2.13), the initial condition in (2.12) is diffuse, that is the initial state has an arbitrarily high variance-covariance matrix. This reflects a non-informative prior knowledge on the state vector. De Jong (1991) and Koopman (1997) provide an analytical treatment of this question. Another way of dealing with the diffuse initial conditions in the state vector is by defining:

$$\mathbf{P}_1 = \mathbf{P}_* + k\mathbf{P}_\infty,$$

where $\mathbf{P}_*$ is a symmetric $m \times m$ matrix, $\mathbf{P}_\infty$ is a diagonal $m \times m$ matrix with ones and zeros on the diagonal and k is taken big enough to reflect the diffuse distribution of some of the components of $\boldsymbol{\alpha}_1$. This is the approach taken in Koopman, Shephard, and Doornik (1999) with $k = 10^7$. As an example of the later approach, for an unobserved components model with trend and stochastic slope we have $\mathbf{P}_* = 0$ and $\mathbf{P}_\infty = \mathbf{I}_2$.

In the framework of estimation of SSM by maximum likelihood, the one-step ahead prediction error and its variance-covariance matrix can be used to obtain recursive expressions of the scores and using an optimization algorithm estimate the parameters of the model. An example of this can be found in Harvey (1989).

Smoothing algorithms have the purpose of prediction at time t given information available after t. Such algorithms have been proposed in Anderson and Moore (1979), Ansley and Kohn (1985), De Jong (1988) and Koopman (1993). They are composed of a set of backward recursions that take as input the output from the Kalman filter defined in (2.17) to (2.21). We present the disturbance smoother in Koopman (1993).

Starting with $\mathbf{r}_T = 0$ and $\mathbf{N}_T = 0$, for $t = T - 1, \ldots, 1$ the backward recursions are given by:

$$\mathbf{e}_t = \mathbf{F}_t^{-1}\mathbf{v}_t - \mathbf{K}_t'\mathbf{r}_t, \qquad (2.22)$$

$$\mathbf{D}_t = \mathbf{F}_t^{-1} + \mathbf{K}_t'\mathbf{N}_t\mathbf{K}_t, \qquad (2.23)$$

$$\mathbf{r}_{t-1} = \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{v}_t + \mathbf{L}'\mathbf{r}_t, \qquad (2.24)$$

$$\mathbf{N}_{t-1} = \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t + \mathbf{L}_t'\mathbf{N}_t\mathbf{L}_t, \tag{2.25}$$

$$\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t\mathbf{Z}_t, \tag{2.26}$$

where $\mathbf{v}_t$, $\mathbf{K}_t$ and $\mathbf{F}_t$ are stored after running the Kalman filter. From the smoothing algorithm we get the smoothed predictions of the innovation process, and the correspondent variance-covariance matrix :

$$E(\mathbf{u}_t|\mathbf{Y}_T) = \mathbf{G}_t'\mathbf{e}_t + \mathbf{H}_t'\mathbf{r}_t,$$

$$\text{var}(\mathbf{u}_t|\mathbf{Y}_T) = \sigma^2\left(\mathbf{I}_r - \mathbf{G}_t'\left(\mathbf{D}_t\mathbf{G}_t - \mathbf{K}_t'\mathbf{N}_t\mathbf{H}_t\right) - \mathbf{H}_t'\left(\mathbf{N}_t\mathbf{H}_t - \mathbf{N}_t\mathbf{K}_t\mathbf{G}_t\right)\right).$$

The disturbance smoother is used in Koopman and Shephard (1992) to obtain the exact scores for SSM, and estimate the hyperparameters by maximum likelihood.

When working with UCM, the disturbance smoother can be used to obtain the auxiliary residuals, Harvey and Koopman (1992). These are smoothed estimates of the components disturbances and are used as diagnostic tools for the detection of shocks not accounted for by the model. In Durbin and Koopman (2001) is discussed the use of auxiliary residuals for diagnostic checking for a general state space model. For the BSM in (2.13) the auxiliary residuals are given by

$$\hat{\varepsilon}_t = E(\varepsilon_t|Y) = \sigma_\varepsilon^2 e_t, \tag{2.27}$$

$$\hat{\eta}_t = E(\eta_t|Y) = \sigma_\eta^2 \mathbf{r}_t^1, \tag{2.28}$$

$$\hat{\zeta}_t = E(\zeta_t|Y) = \sigma_\zeta^2 \mathbf{r}_t^2, \tag{2.29}$$

$$\hat{\omega}_t = E(\omega_t|Y) = \sigma_\omega^2 \mathbf{r}_t^3, \tag{2.30}$$

$$\tag{2.31}$$

with variances

$$\sigma_{\hat{\varepsilon}}^2 = \text{var}(\hat{\varepsilon}_t) = \sigma_\varepsilon^4 D_t, \tag{2.32}$$

$$\sigma_{\hat{\eta}}^2 = \text{var}(\hat{\eta}_t) = \sigma_\eta^4 \mathbf{N}_t^{1,1}, \tag{2.33}$$

$$\sigma_{\hat{\zeta}}^2 = \text{var}\left(\hat{\zeta}_t\right) = \sigma_\zeta^4 \mathbf{N}_t^{2,2}, \tag{2.34}$$

$$\sigma_{\hat{\omega}}^2 = \text{var}(\hat{\omega}_t) = \sigma_\omega^4 \mathbf{N}_t^{3,3}, \tag{2.35}$$

$$\tag{2.36}$$

where $\mathbf{r}_t^i$ is the i-th component of the vector $\mathbf{r}_t$ and $\mathbf{N}_t^{i,i}$ is the i-th diagonal component of the matrix $\mathbf{N}_t$, for $i = 1, 2, 3$.

The auxiliary residuals are standardized before use for diagnostics purpose, as the estimated variances at the beginning and end of the sample are different from the variances at the middle of the sample. If the model is well specified they should be normally distributed, although they are serially correlated (Harvey and Koopman, 1992). The normality of these processes is the basis for the diagnostic tools proposed in Harvey and Koopman (1992), by plotting them to detect outstanding values, and using normality tests, corrected for the existence of correlation. In Harvey and Koopman (1992), when analyzing the plots of the standardized auxiliary residuals, values in absolute value greater than 2 are indication of outliers or structural shifts. This is approximately the critical value for a two side individual test, for a size of 0.05. The detection of outliers or structural shifts by inspection of outlying values for the standardized auxiliary residuals is a simultaneous test problem. In Penzer (2001) critical values are derived for simultaneous testing, when the statistics are independent and identically distributed. For example, if the individual statistic is distributed as a standard normal, for simultaneous test the significance of 100 statistics, the critical value is 3.28, for a one side test of size 0.05. The critical value for performing a one side test for a single statistic is 1.64, for a test of size 0.05. In Penzer (2001) several simulation studies are conducted to evaluate the impact in the position of the critical values of the existence of correlation in the statistics. The conclusion is that, even in the presence of correlation, in many cases the statistics can be treated as independent. In these cases the critical values obtained assuming independence, provide an accurate estimate of the critical values obtained by simulation of the correlated statistics. From the tables in Penzer (2001), using the critical values, obtained under independence assumption, the critical value for a two side test of size 0.05 for the significance of the standardized auxiliary residuals, for a sample of size $T = 100$, is approximately 3.5.

2.4 MCMC methods for state space models

Suppose we have a SSM like the one defined by equations (2.9) to (2.12), and let Ψ be the vector of all unknown parameters in the model, and $\alpha = (\alpha_1, \ldots, \alpha_T)$. We are interested in sampling from $f(\alpha, \Psi | Y)$, the joint posterior distribution of the state vector and parameters. Sampling from that posterior will enable us to estimate Ψ using a similar expression to (2.4), and to make inference about the statistical properties of the state space vector, which we recall is very often a latent variable. Using the Gibbs sampler, and defining $X = (\alpha, \Psi)$, we construct a chain that converges to the target joint posterior by iteratively sampling:

$$\alpha \sim f(\alpha | \Psi, Y),$$
$$\Psi \sim f(\Psi | \alpha, Y).$$

We begin by presenting a method for sampling from the conditional distributions $f(\alpha | \Psi, Y)$. As the composition of Ψ depends essentially on the model in consideration, the method for sampling from its full conditional will be illustrated for the basic structural model.

2.4.1 Simulating from $f(\alpha | \Psi, Y)$

When using Gibbs sampler for sampling from the states full conditional distribution two sampling strategies can be used: single-state and multi-state Gibbs sampler. A single-state sampling scheme was proposed in Carlin, Polson, and Stoffer (1992). For each draw of $f(\alpha | \Psi, Y)$ we sequentially sample from $f(\alpha_t | \alpha_{\setminus t}, \Psi, Y)$, where $\alpha_{\setminus t} = (\alpha_1, \ldots, \alpha_{t-1}, \alpha_{t+1}, \ldots, \alpha_T,)$ for $t = 1, \ldots, T$. This scheme has the drawback that the samples for the states will be highly correlated, which implies a slower convergence of the chain to its equilibrium distribution. Alternatively a multi-state approach can be used. An example of this technique is given in Carter and Kohn (1994) and Fruhwirth-Schnatter (1994). Given that

$$f(\alpha | \Psi, Y) = f(\alpha_T | \Psi, Y) f(\alpha_{T-1} | \alpha_T, \Psi, Y) \ldots f(\alpha_0 | \alpha_1, \Psi, Y),$$

and that each conditional distribution on the right hand side is Gaussian, the draws from $f(\alpha | \Psi, Y)$, are obtained by running all the subdraws.

The simulation smooother, was proposed in De Jong and Shephard (1995), and it is the method we will use for sampling from the states' full conditional distribution. Instead of simulating directly from the full conditional of the states, it draws from the full conditional of the innovations, and from there, using the fact that the states are a linear combination of the innovations, obtains the desired samples from the states' full conditional distribution. The advantage of this method is that we sample from a multivariate distribution of uncorrelated variables, the innovation processes, and therefore increase the speed of convergence of the chain to its equilibrium distribution.

Let $\mathbf{S}_t$ be a selection matrix , which defines the subset of the vector of innovations we wish to sample from,

$$\eta_t = \mathbf{S}_t \mathbf{u}_t. \tag{2.37}$$

For example, if $\mathbf{S}_t = \mathbf{H}_t$ we sample from the joint full conditional of the transition equation innovations, which allow us to get draws from $f(\alpha|\Psi, Y)$. We assume, for simplicity of exposition, that $\mathbf{c}_t = \mathbf{d}_t = 0$, for all t.

The simulation smoother starts by running the Kalman filter once and storing the quantities $\mathbf{v}_t, \mathbf{F}_t, \mathbf{K}_t$ for $t = 1, 2, \ldots, T$, present on equations (2.17) to (2.19) . Then, setting $\mathbf{r}_T = 0$ $(N \times 1)$ and $\mathbf{U}_T = 0$ $(N \times N)$, and defining $\mathbf{L}_t = \mathbf{T}_t - \mathbf{K}_t \mathbf{Z}_t$, $\mathbf{J}_t = \mathbf{H}_t - \mathbf{K}_t \mathbf{G}_t$, for all $t = 1, \ldots, T$, the following recursions are run for $t = T-1, \ldots, 1$:

$$\mathbf{C}_t = \mathbf{S}_t \left(\mathbf{I} - \mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{G}_t - \mathbf{J}_t' \mathbf{U}_t \mathbf{J}_t \right) \mathbf{S}_t', \tag{2.38}$$

$$\varepsilon_t \sim N\left(0, \sigma^2 \mathbf{C}_t\right), \tag{2.39}$$

$$\mathbf{V}_t = \mathbf{S}_t \left(\mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{J}_t' \mathbf{U}_t \mathbf{L}_t \right), \tag{2.40}$$

$$\mathbf{r}_{t-1} = \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}_t' \mathbf{r}_t - \mathbf{V}_t' \mathbf{C}_t^{-1} \varepsilon_t, \tag{2.41}$$

$$\mathbf{U}_{t-1} = \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t + \mathbf{L}_t' \mathbf{U}_t \mathbf{L}_t + \mathbf{V}_t' \mathbf{C}_t^{-1} \mathbf{V}_t. \tag{2.42}$$

From the output of this recursive process we are interested in storing:

$$\eta_t = \mathbf{S}_t \left(\mathbf{G}_t' \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{J}_t' \mathbf{r}_t \right) + \varepsilon_t,$$

for $t = 0, 1, \ldots, T$ (for which $\mathbf{G}_0 = 0$ is set), as they are a sample from $f(\eta|\Psi, \mathbf{y})$ The proof of this statement can be found in De Jong and Shephard (1995).

The above expressions are simpler when we consider the case of $\mathbf{S}_t = \mathbf{H}_t$. Suppose for that case that $\mathbf{H}_t\mathbf{G}'_t = 0$ (uncorrelated measurement and transition innovations), and define $\boldsymbol{\Omega}_t$, apart from the factor σ^2, the variance-covariance matrix for the transition innovations, $\boldsymbol{\Omega}_t = \mathbf{H}'_t\mathbf{H}_t$. In this case equations (2.38) and (2.40) are replaced by:

$$\mathbf{C}_t = \boldsymbol{\Omega}_t - \boldsymbol{\Omega}_t\mathbf{U}_t\boldsymbol{\Omega}_t, \tag{2.43}$$

$$\mathbf{V}_t = \boldsymbol{\Omega}_t\mathbf{U}_t\mathbf{L}_t. \tag{2.44}$$

In order to obtain $\boldsymbol{\alpha} \sim f(\boldsymbol{\alpha}|\boldsymbol{\Psi}, \mathbf{y})$, we run the recursion:

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t\boldsymbol{\alpha}_t + \boldsymbol{\eta}_t,$$

for $t = 1, \ldots, T$, using the initial condition $\boldsymbol{\alpha}_1 = \mathbf{a}_1$.

2.4.2 Simulating from $f(\boldsymbol{\Psi}|Y, \boldsymbol{\alpha})$ for the BSM

Consider the basic structural model, defined by equations in (2.13), for a general seasonal period s. As we have seen before the state space vector is composed by $\boldsymbol{\alpha}_t = (\mu_t, \beta_t, \gamma_t, \ldots, \gamma_{t-(s-1)})'$, with initial conditions as in equations (2.14), (2.15) and $\gamma_i \sim N(0, \kappa I)$, $i = -s + 3, \ldots, 0, 1$. The set of unknown parameters of the model is given by the vector of hyperparameters $\boldsymbol{\Psi} = (\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2)'$. Let $\mathbf{x} = (x_1, \ldots, x_T)$, for $x = \mu, \beta, \gamma$.

Sampling from $f(\boldsymbol{\Psi}|Y, \boldsymbol{\alpha})$ is done by sampling from the full conditional for each of the hyperparameters. Given the assumption of uncorrelation between the different components this is achieved by sampling:

$$\sigma_\varepsilon^2 \sim f(\sigma_\varepsilon^2|Y, \boldsymbol{\mu}, \boldsymbol{\gamma}), \tag{2.45}$$

$$\sigma_\eta^2 \sim f(\sigma_\eta^2|\boldsymbol{\mu}, \boldsymbol{\beta}), \tag{2.46}$$

$$\sigma_\zeta^2 \sim f(\sigma_\zeta^2|\boldsymbol{\beta}), \tag{2.47}$$

$$\sigma_\omega^2 \sim f(\sigma_\omega^2|\boldsymbol{\gamma}). \tag{2.48}$$

$$\tag{2.49}$$

For each of the hyperparameters we assume an inverse gamma prior:

$$\sigma_\varepsilon^2 \sim IG\left(\frac{c_1}{2}, \frac{s_1}{2}\right), \tag{2.50}$$

41

$$\sigma_\eta^2 \sim IG\left(\frac{c_2}{2}, \frac{s_2}{2}\right),$$ $$\tag{2.51}$$

$$\sigma_\zeta^2 \sim IG\left(\frac{c_3}{2}, \frac{s_3}{2}\right),$$

$$\sigma_\omega^2 \sim IG\left(\frac{c_4}{2}, \frac{s_4}{2}\right),$$

$$\tag{2.52}$$

with $c_i, s_i > 0$ for $i = 1, \ldots, 4$.

The reason for this choice is the conjugacy property for this distribution, as it is stated in next lemma.

Lemma 2.4.1 *Suppose that $\sigma^2 \sim IG\left(\frac{c}{2}, \frac{s}{2}\right)$, where c and s are known. Additionally, assume that $u_1, \ldots, u_n$ are independent and identically distributed as a $N(0, \sigma^2)$. Then:*

a)

$$\sigma^2 | u_1, \ldots, u_n \sim IG\left(\frac{n+c}{2}, \frac{\sum\limits_{i=1}^{n} u_i^2 + s}{2}\right).$$

b) If the process $\alpha_1, \ldots, \alpha_n$ is defined recursively by $\alpha_{t+1} = \alpha_t + u_t$, with diffuse initial conditions $\alpha_1 \sim N(0, \kappa I)$, $\kappa \to \infty$, we have that:

$$\sigma^2 | \alpha_1, \ldots, \alpha_n \sim IG\left(\frac{n-1+c}{2}, \frac{\sum\limits_{i=1}^{n-1} u_i^2 + s}{2}\right).$$

Proof : We shall proof only the second part, as the proof of the first one is similar. We have that:

$$f\left(\sigma^2 | \alpha_1, \ldots, \alpha_n\right) \propto f\left(\alpha_1, \ldots, \alpha_n | \sigma^2\right) f\left(\sigma^2\right)$$

$$\propto f\left(\alpha_2, \ldots, \alpha_n | \alpha_1, \sigma^2\right) f\left(\alpha_1 | \sigma^2\right) f\left(\sigma^2\right)$$

$$\propto f\left(\alpha_2, \ldots, \alpha_n | \alpha_1, \sigma^2\right) f\left(\sigma^2\right),$$

where we have used that the distribution of α_1 does not depend on σ^2. On the other hand,

$$f\left(\alpha_2, \ldots, \alpha_n | \alpha_1, \sigma^2\right) = \prod_{i=2}^{n} f\left(\alpha_i | \alpha_{i-1}, \sigma^2\right)$$

42

$$= \prod_{i=2}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\alpha_i - \alpha_{i-1})^2}{2\sigma^2} \right) \right]$$

$$= \prod_{i=1}^{n-1} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u_i^2}{2\sigma^2} \right) \right]$$

Using the prior assumption of an inverse gamma for σ^2 we have that

$$\log\left(f(\sigma^2)\right) \propto -\log\Gamma\left(\frac{c}{2}\right) - \frac{c}{2}\log\frac{s}{2} - \frac{c+2}{2}\log\sigma^2 - \frac{s}{2\sigma^2}, \quad \sigma^2 > 0.$$

Therefore we get:

$$\log\left(f\left(\sigma^2|\alpha_1,\ldots,\alpha_n\right)\right) = const - \frac{n-1}{2}\log\sigma^2 - \frac{\sum_{i=1}^{n-1} u_i^2}{2\sigma^2} - \frac{c+2}{2}\log\sigma^2 - \frac{s}{2\sigma^2}$$

$$= const - \frac{n-1+c+2}{2}\log\sigma^2 - \frac{\sum_{i=1}^{n-1} u_i^2 + s}{2\sigma^2}.$$

Consequently:

$$\sigma^2|\alpha_1\ldots,\alpha_n \sim IG\left(\frac{c+n-1}{2}, \frac{\sum_{i=1}^{n-1} u_i^2 + s}{2} \right).$$

Using Lemma 2.4.1 (a) it is straightforward to sample from the full conditional distribution in (2.45). We have that $y_t - \mu_t - \gamma_t = \varepsilon_t$ are independent and identically distributed with distribution $N\left(0, \sigma_\varepsilon^2\right)$, for $t = 1,\ldots,T$. Hence we have that:

$$\sigma_\varepsilon^2|Y,\mu,\gamma \sim IG\left(\frac{c_1 + T}{2}, \frac{\sum_{t=1}^{T}(y_t - \mu_t - \gamma_t)^2 + s_1}{2} \right). \tag{2.53}$$

By Lemma 2.4.1 (b) we obtain the distributions in expressions from (2.46) and (2.47), respectively:

$$\sigma_\eta^2|\mu,\beta \sim IG\left(\frac{c_2 + T - 1}{2}, \frac{\sum_{t=1}^{T-1}(\mu_{t+1} - \mu_t - \beta_t)^2 + s_2}{2} \right),$$

$$\sigma_\zeta^2|\beta \sim IG\left(\frac{c_3 + T - 1}{2}, \frac{\sum_{t=1}^{T-1}(\beta_{t+1} - \beta_t)^2 + s_3}{2} \right).$$

By a similar derivation used to obtain the result 2.4.1 (b), the distribution in (2.48) is given by

$$\sigma_\omega^2 | \gamma \sim IG \left(\frac{c_4 + T - (s-1)}{2}, \frac{\sum\limits_{t=s-1}^{T-1} \left(\gamma_{t+1} + \gamma_t + \ldots + \gamma_{t-(s-2)} \right)^2 + s_4}{2} \right).$$

After convergence is achieved, if we continue the Gibbs sampler scheme, we obtain N samples from the joint posterior $\left(\mu, \beta, \gamma, \sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2 | Y \right)$, which are used to estimate the different parameters of the model, and obtain their statistical properties.

2.4.3 Numerical examples

To illustrate the application of the Gibbs sampler for estimating SSM, we consider an artificial data set, of size $T = 100$, generated from a local level model,

$$\begin{aligned}
y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N\left(0, \sigma_\varepsilon^2\right), \\
\alpha_{t+1} &= \alpha_t + \eta_t, & \eta_t &\sim N\left(0, \sigma_\eta^2\right),
\end{aligned} \tag{2.54}$$

with $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$.

The Gibbs sampler is run for 2000 iterations, with the first 1000 discarded. The prior assumptions for the hyperparameters are as in expressions (2.50) and (2.51), with all the prior parameters set equal to 5. The state space variable is assumed to have a diffuse prior $\alpha_1 \sim N(0, k)$ with $k \to \infty$. In Table 2.1, we present the results obtained.

Table 2.1: Summary of output from posterior samples for hyperparameters, for a local level model.

	Prior		Posterior sample			
	Mean	SD	Mean	SD	$Q_{0.025}$	$Q_{0.975}$
σ_ε^2	1.67	2.36	0.946	0.0290	0.544	1.506
σ_η^2	1.67	2.36	1.297	0.0303	0.764	1.976

The values reported are the mean, 2.5th and 97.5th percentile of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated with a bandwidth of 100 for the Parzen window). Based on 2,000 Gibbs draws, discarding the first 1,000.

44

The estimated value for σ_ε^2 is closer to the true value of the parameter, than σ_η^2. If estimating this model by maximum likelihood, the hyperparameters are estimated with values $\hat{\sigma}_\varepsilon^2 = 0.914$ and $\hat{\sigma}_\eta^2 = 1.302$. We note that the hyperparameter σ_η^2 is also overestimated, and that the results obtained by the Gibbs sampler algorithm are closer to the true ones.

The convergence of the chain to the target distribution is fast, as it can be concluded from the plots of the path simulated, in Figure 2.1 (a) and (b), where the samples of the burn in period are included. In Figure 2.1 (c) and (d) we plot

Figure 2.1: Plot of simulated paths (a)-(b), and autocorrelation functions for posterior samples (c)-(d), for σ_ε^2 and σ_η^2, for a local level model.
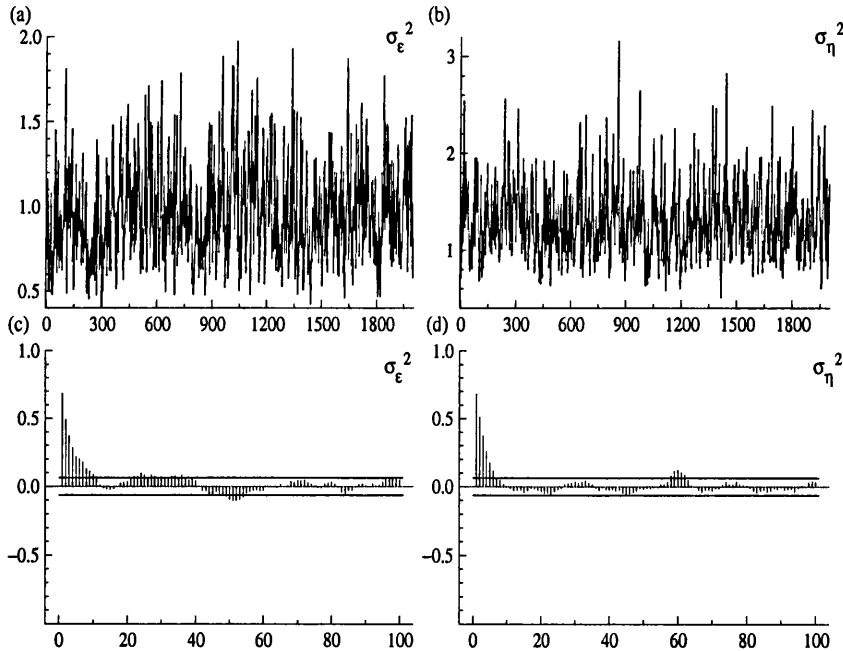
the autocorrelation functions for the 1,000 iterations recorded. The autocorrelation functions have a very rapid convergence towards zero which corroborates that after the burn in period the chain has converged to the target distribution.

To assess the sensitivity of the estimates obtained to the parameters set *a priori* for the inverse gamma distributions, we conduct a Monte Carlo experiment. Using the model defined by equations in (2.54), with $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$, 1,000 artificial data sets, of size $T = 100$, are generated. For each data set, the Gibbs sampler is run

with $s_1 = 1, 3, 5$, for $c_1 = c_2 = s_2 = 5$, and $s_2 = 1, 3, 5$, for $c_1 = s_1 = c_2 = 5$. The results obtained are summarized in Table 2.2 and 2.3

Figure 2.2: Box-plots of posterior sample means of hyperparameters, across 1,000 simulated replications, for prior parameters $\{s_1 = 1, 3, 5, c_1 = c_2 = s_2 = 5\}$ and $\{s_2 = 1, 3, 5, c_1 = c_2 = s_1 = 5\}$, for a local level model.
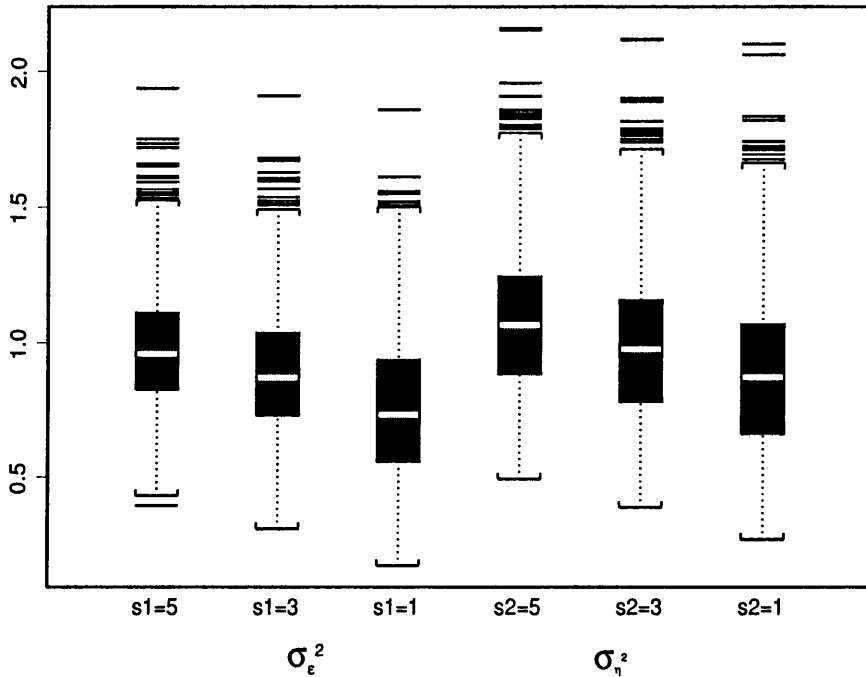
In average, across the simulated replications, the more accurate estimates for the hyperparameters are obtained by taking $s_1 = s_2 = 5$, presenting also the smaller standard deviations across the replications. Comparing the sensitivity of the posterior samples to the values set for the prior distributions, we can see that the hyperparameters present a similar sensitivity to the changes in the correspondent prior parameter. This can be seen also in Figure 2.2.

The value of, for example, the prior parameter s_1 will also affect the estimate obtained for σ_η^2, as if the variability of the measurement equation is not appropriately taken in account by the estimate of σ_ε^2, it will be partially included in the estimate obtained for σ_η^2. This effect is related to the existence of correlation between the posterior samples of the hyperparameters. The use of less appropriate

Table 2.2: Summary of output from posterior sample means for hyperparameters, across 1,000 simulated replications, for prior parameter $s_1 = 1, 3, 5$ and $c_1 = c_2 = s_2 = 5$, for a local level model.

	s_1	Prior Mean	SD	Posterior sample mean Mean	SD	$Q_{0.025}$	$Q_{0.975}$
σ_ε^2	5	1.67	2.36	0.982	0.218	0.617	1.474
	3	1	1.41	0.898	0.233	0.521	1.420
	1	0.33	0.47	0.758	0.277	0.295	1.363
σ_η^2	5	1.67	2.36	1.080	0.268	0.634	1.665
	3	1.67	2.36	1.150	0.300	0.649	1.805
	1	1.67	2.36	1.319	0.398	0.679	2.142

The values reported are the mean, standard deviation and 2.5th and 97.5th percentile for the samples of posterior sample means obtained across 1,000 simulated replication. Based on 10,000 Gibbs draws, discarding the first 5,000.

Table 2.3: Summary of output from posterior sample means for hyperparameters, across 1,000 simulated replications, for prior parameter $s_2 = 1, 3, 5$ and $c_1 = c_2 = s_1 = 5$, for a local level model.

	s_2	Prior Mean	SD	Posterior sample mean Mean	SD	$Q_{0.025}$	$Q_{0.975}$
σ_ε^2	5	1.67	2.36	0.982	0.218	0.617	1.474
	3	1.67	2.36	1.023	0.232	0.639	1.546
	1	1.67	2.36	1.078	0.254	0.667	1.650
σ_η^2	5	1.67	2.36	1.080	0.268	0.634	1.665
	3	1	1.41	0.988	0.280	0.525	1.603
	1	0.33	0.47	0.882	0.299	0.374	1.531

The values reported are the mean, standard deviation and 2.5th and 97.5th percentile for the samples of posterior sample means obtained across 1,000 simulated replication. Based on 10,000 Gibbs draws, discarding the first 5,000.

prior parameters can imply that the chain will take longer to converge to the target distribution, which is reflected on higher values for the autocorrelation functions of posterior samples, or empirical posterior distribution presenting a high degree of asymmetry. Hence, the analysis of the distribution of the samples obtained after the burn in period is a useful tool for setting the values of the prior distributions parameters efficiently.

Several other numerical examples are presented. The purpose is to assess the impact of the choice of variables as the number of iterations of the sampler, or the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$, on the properties of the posterior samples. In the examples presented the true value of σ_ε^2 is fixed to one. The true value of σ_η^2 is taken equal to 0.1 and 0.01. The number of iterations considered are 2,000, 200,000 and 2,000,000. Half of the iterations are discarded as being burn-in-period. In Table 2.4 we present a summary of the statistical properties of the posterior samples obtained.

Table 2.4: Summary of output from posterior sample of hyperparameters, for 1,000, 100,000 and 1,000,000 runs of the Gibbs sampler , for true parameters $\sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = 0.1, 0.01$.

		Iterations	Mean	SD	$Q_{0.025}$	$Q_{0.975}$
$\sigma_\eta^2 = 0.01$	σ_ε^2	1,000	0.966	0.00475	0.723	1.280
		100,000	0.960	0.000590	0.723	1.270
		1,000,000	0.961	0.000188	0.723	1.272
	σ_η^2	1,000	0.00961	0.000861	0.00338	0.0217
		100,000	0.0108	0.000167	0.00353	0.0295
		1,000,000	0.0108	5.574E-05	0.00351	0.0295
$\sigma_\eta^2 = 0.1$	σ_ε^2	1,000	0.976	0.00998	0.678	1.337
		100,000	0.968	0.00134	0.682	1.337
		1,000,000	0.968	0.000433	0.683	1.335
	σ_η^2	1,000	0.118	0.00802	0.0486	0.241
		100,000	0.124	0.00106	0.0489	0.269
		1,000,000	0.124	0.000339	0.0489	0.272

The values reported are the mean, 2.5th and 97.5th percentile of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated using a Parzen window). Based on n Gibbs draws, discarding the first $\frac{n}{2}$, for $n = 2,000, 200,000$ and $2,000,000$.

The posterior mean of the hyperparameters exhibits a low sensitivity to the number of iterations used. This lack of sensitivity is also present in the values of the

percentiles.

To analyze what the effect of the number of iterations on the posterior samples correlation structure, in Figures 2.3 and 2.4, we plot the autocorrelation functions until lag 100, for $\sigma_\eta^2 = 0.01, 0.1$, and for the different number of iterations used.

Figure 2.3: Autocorrelation functions for posterior samples of the hyperparameters for a local level model, with true parameters $\sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = 0.01$, for 1,000 (a)-(b), 1000,000 (c)-(d) and 1,000,000 (e)-(f) runs of the Gibbs sampler.
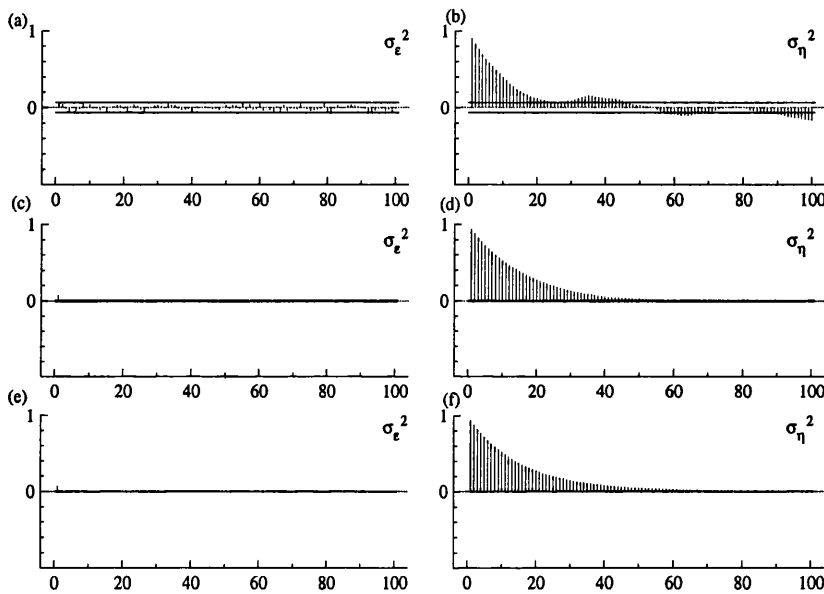
Comparing the autocorrelations structure across number of iterations used, we can see that there is no discernible difference between using 100,000 and 1,000,000. The posterior samples obtained when using 1,000 iterations present a slower convergence to zero. However, even for this number of iterations the rate of convergence of the autocorrelation functions toward zero suggests that this number of iterations is sufficient.

The behaviour of the correlation structure across true value for σ_η^2 is dependent on the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$. The lower that ratio, the slower is the convergence of the auto-correlations functions for σ_η^2, when compared with the behaviour of these functions for the posterior samples of σ_ε^2. From Figure 2.1 (c) and (d), when the ratio is equal

Figure 2.4: Autocorrelation functions for posterior samples of the hyperparameters for a local level model, with true parameters $\sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = 0.1$, for 1,000 (a)-(b), 1000,000 (c)-(d) and 1,000,000 (e)-(f) runs of the Gibbs sampler.
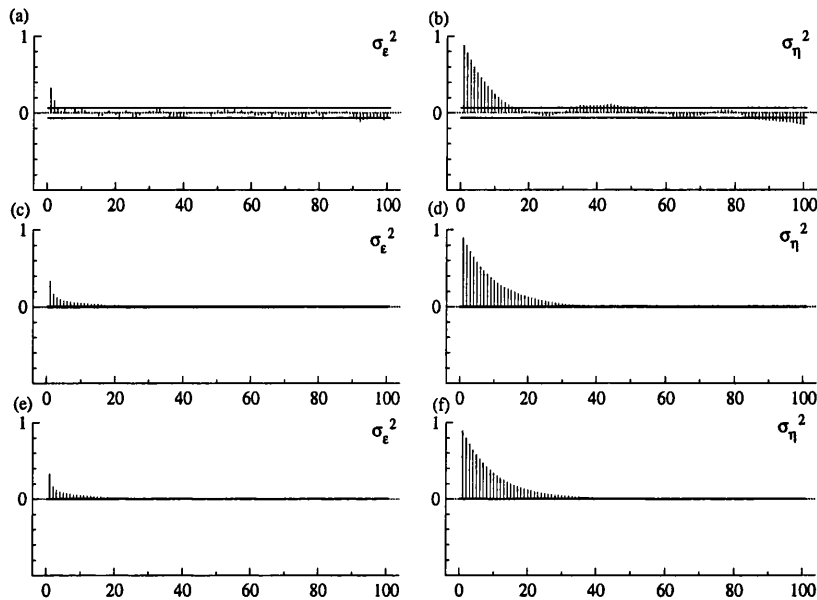
to one, we observe that the autocorrelation functions present a similar converge toward zero, for the posterior samples of the two hyperparameters. In contrast, when the ratio is 0.1 or 0.01, the autocorrelations for the posterior samples of σ_η^2 present a much slower converge to zero, than the posterior samples for σ_ε^2. This difference on the rate of convergence of the autocorrelations increases with the decrease of the ratio of the true values of the hyperparameters.

2.5   Conclusions

We have presented a brief review of the theory and results for Markov chain Monte Carlo methods. The basic notations were introduced with the purpose of presenting results that justify the use of this methodology for parametric estimation. We restricted our attention to the Metropolis-Hastings and Gibbs sampler, as in what follows we will be using the later.

The type of models we will be considering are SSM. They were defined, together

with some examples of times series models that can be written using this formulation. The Kalman filter and smoother, and a simulation smoother were described as they will play a important role in the methodologies used in the forthcoming chapters.

The application of the Gibbs sampler to SSM models was exemplified, by estimating a local level model. Some aspects of the estimation process were considered, namely empirical ways of assessing that the chain has converged, the effect of the choice of the prior distribution parameters and of the number of runs of the Gibbs sampler.

Chapter 3

MCMC methods for shocks detection

3.1 Introduction

Shocks which have not been accounted for in a model can result in misspecification and bias in the parameter estimates. In Fox (1972) a detailed study is conducted on the effect of two types of outliers on the parameters estimates. This is one of the first references to methods of dealing with outliers in the context of time series observations. The two types of outliers considered are: *additive* outlier (AO), when the shock occurring at time t only affects the observation in that period of time; *innovative* outlier (IO), when a shock at time t will affect not only the t-th observation but also subsequent ones. Detection of these effects is done by likelihood ratio test. Tsay (1988) extends this framework considering four type of interventions for univariate time series. In addition to AO and IO, level changes and temporary level changes are considered, with a non-Bayesian approach to detection. One of the first Bayesian methods for detecting AO and IO, in autoregressive models is given by Abraham and Box (1979). The main difference between the Bayesian and non-Bayesian modeling approach is that the former considers the outliers as part of the model formulation whereas the later considers a null model without intervention on which the outliers detection methods are based. The purpose of this chapter is to present a Bayesian method for shocks detection in the context of state space models. Our approach uses more general prior assumptions than those used in

existing methods. We start by presenting the formulation of state space models in the presence of interventions and a brief description of non-Bayesian methodologies available. A review of some Bayesian techniques for this type of problem is also given. The method we have developed is presented, and its correctness established. Finally, some empirical applications are given, for simulated and real data, and conclusions are draw.

3.2   Interventions in state space models

Consider the usual state space models formulation for a univariate time series y_t for $t = 1, \ldots, T$, given by the measurement and transition equations:

$$y_t = c_t + Z_t \alpha_t + G_t u_t, \tag{3.1}$$

$$\alpha_{t+1} = d_t + T_t \alpha_t + H_t u_t, \tag{3.2}$$

where the state vector α_t is of dimension m, $u_t \sim NID(0, \sigma^2 I_r)$, $HG' = 0$ and with initial conditions $\alpha_1 \sim N(a_1, P_1)$. We consider these type of models in the presence of different types of interventions. Following the notation in De Jong and Penzer (1998) the intervention model is formulated as:

$$y_t = c_t + X_t \delta + Z_t \alpha_t + G_t u_t, \tag{3.3}$$

$$\alpha_{t+1} = d_t + W_t \delta + T_t \alpha_t + H_t u_t, \tag{3.4}$$

The *shock design* matrices X_t (1×1) and W_t ($m \times 1$) determine the nature and location of the intervention, whereas δ (1×1) determines its magnitude. Let δD be the effect the intervention has in the observations. $(D_1, \ldots, D_T)$ is the *intervention signature*.

In De Jong and Penzer (1998), a *simple intervention* is defined as an intervention that is generated by taking the shock design matrices equal to zero except at the instant where the intervention occurs. We have a simple additive outlier at instant i, with magnitude δ when:

$$X_t = \begin{cases} 1, & t = i \\ 0, & t \neq i \end{cases},$$
$$W_t = 0, \text{ for all } t.$$

The shock occurs at one single point in time $t = i$ and affects the observations only at its origin i, as it is added directly to the measurement equation. Its signature is represented in 3.1 (a).

An innovative outlier, with origin i, occurs when a shock is added to the transition equation. Its impact on the states vector, and consequently on the observations, is not instantaneous. For simplicity of exposition we will consider $m = 1$, in the following examples of innovative interventions.

Consider a data set $\{y_t\}_{t=1}^T$ generated by a stationary autoregressive model ($|\phi| < 1$) of order one, to which a simple shock of magnitude δ with origin i is applied. The formulation for the intervention model is the following:

$$
\begin{aligned}
y_t &= \alpha_t \\
\alpha_{t+1} &= W_t \delta + \phi \alpha_t + H_t u_t,
\end{aligned}
$$

with

$$
W_t = \begin{cases} 1, & t = i, \\ 0, & t \neq i. \end{cases} \tag{3.5}
$$

Adding the shock to the transition equation has an exponential decaying additive effect $\phi^{t-i-1} \delta$, $t = i+1, \ldots$, as consequence of $| \phi | < 1$ (see Figure 3.1 (b)).

If we consider structural models, instant shocks to one of the component equations will result in a permanent shift. Consider as an example the local level model, where a shock of magnitude δ is added to the equation defining the dynamics of the level component. The resulting level shift of magnitude δ, is a consequence of the non-stationarity of the transition equation (see Figure 3.1 (c)).For this model, a shock to the level component is modeled as:

$$
\begin{aligned}
y_t &= \mu_t + \mathbf{G}_t \mathbf{u}_t, \\
\mu_{t+1} &= W_t \delta + \mu_t + \mathbf{H}_t \mathbf{u}_t,
\end{aligned}
$$

with the intervention defined as in (3.5).

As a last example consider a structural model composed only of a quarterly dummy seasonal component, and with a simple shock to the seasonal component,

Figure 3.1: Shocks in state space models: simple additive outlier (a); innovative outlier in stationary AR(1) (b); level shift (c); seasonal shift (d).
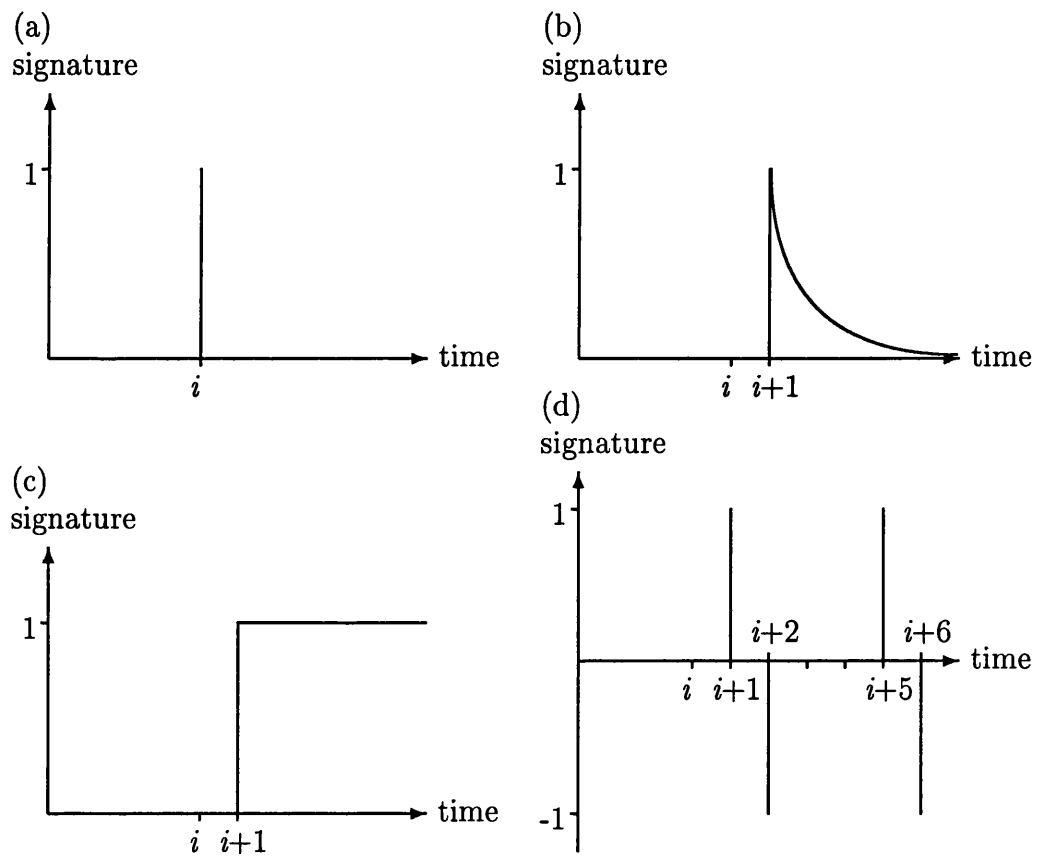
with origin i, and magnitude δ. The intervention model is

$$y_t = \gamma_t,$$
$$\sum_{i=0}^{s-1} \gamma_{t+1-i} = W_t \delta_t + H_t u_t,$$

with W_t defined as in (3.5). The seasonal element, corresponding to $t = i + 1$, will have a permanent shift of δ. The following seasonal elements adjust to this shift. Hence, the result of the shock is a permanent shift in the seasonal pattern (see Figure 3.1 (d)).

3.3 Non-Bayesian shock diagnosis

Classical methods for estimating models with interventions rely on estimating the null model, of no intervention and then analyzing the intervention model:

$$Intervention\ model = Null\ model + Interventions.$$

The main criticism made to these types of approaches is that an initial estimation process will produce biased results as, when shocks are present, this methodology is based on a first stage with a misspecified model.

Several non-Bayesian methods have been proposed for detection and evaluation of the effect of interventions in time series. The benchmark contribution on non-Bayesian detection of interventions is the work in Fox (1972), with the study of additive and innovative outliers, drops the assumptions that the observations were independent and identical distributed, opting instead for an autoregressive structure for the data. In Fox's paper, detection of interventions is done by using likelihood ratios, comparing the null model and the model with interventions. In the context of a more general model, ARMA(p, q), Chen and Liu (1993) propose a method for simultaneous estimation of the model parameters and possible interventions, by an iterative process. Iteratively, the estimation of a model, with a set of known interventions (initialized as empty) is done, followed by the updated detection of the previous set of interventions. This process is repeated until no new interventions are found. The drawback of this method is that the step of interventions detection

is based on the residuals of the previously fitted model, which might not account for all the interventions and therefore induces bias in the parameters estimated.

Atkinson, Koopman, and Shephard (1997) and De Jong and Penzer (1998) present methods which use state space models. Atkinson, Koopman, and Shephard (1997) focus on structural models, and the detection of outliers and structural changes. The methodology presented is based on the change of the scores when considering the null model and the model with interventions. In De Jong and Penzer (1998) a method for detecting shocks for a general SSM is proposed. Through a recursive algorithm, which uses the output from one run of the Kalman filter and smoother, the *smoothations* vector and its covariance matrices are obtained. These quantities are the base of diagnostics tools for detecting the different type of shocks

3.4 Bayesian shock diagnosis

Bayesian methods for detection of shocks have been well reported in the literature. Earlier work focused on linear regression models, for example Box and Tiao (1968). Time series Bayesian methods were initially applied for autoregressive models, in Abraham and Box (1979). Bayesian methods for diagnosing shocks are characterized by considering shocks as a feature of the model, through the inclusion of intervention variables. The objective is to sample from the joint posterior distribution of the parameters and interventions variables.

We initially consider a very simple SSM, namely a local level model, to illustrate a Bayesian approach for estimating intervention time series models. We assume the existence of only one type of intervention, an additive outlier, with origin at $t = i$ and magnitude δ, directly affecting the observations $\{y_t\}_{t=1}^{T}$.

The presence of an outlier can be included in the model in an additive way, through a location parameter; alternatively the presence of the shock can be accounted for by defining the distribution of the error in the measurement equation as a mixture of normals, which is achieved by the inclusion of a scale parameter.

We start by considering the use of a jump in the measurement equation, which

is done by introducing the variable $\mathbf{K} = (K_1, \dots, K_T)$, defined as:

$$K_t = \begin{cases} 0, & \text{for } t \neq i \\ \delta, & \text{for } t = i, \end{cases} \tag{3.6}$$

which is directly included in the measurement equation:

$$y_t = \alpha_t + K_t + \varepsilon_t, \tag{3.7}$$

$$\alpha_{t+1} = \alpha_t + \eta_t, \tag{3.8}$$

where:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{3.9}$$

$$\eta_t \sim N(0, \sigma_\eta^2), \tag{3.10}$$

for $t = 1, \dots, T$ and

$$\alpha_1 \sim N(0, \kappa), \quad \kappa \to \infty. \tag{3.11}$$

A Bernoulli prior is assumed for K_t. Its prior is defined, for $t = 1, \dots, T$ as:

$$P(K_t = j) = \begin{cases} p, & j = \delta \\ 1 - p, & j = 0. \end{cases}$$

If the variable K_t is different from zero, its value δ defines the size of the shock at that instant in time.

An alternative way of accounting for the existence of an outlier is a contamination model formulation. The distribution assumed for the measurement innovation process is:

$$\varepsilon_t \sim N\left(0, K_t^2 \sigma_\varepsilon^2\right), \tag{3.12}$$

where:

$$K_t = \begin{cases} 1, & \text{for } t \neq i \\ \tilde{K}, & \text{for } t = i, \end{cases} \tag{3.13}$$

with $\tilde{K} > 1$ and a value such that the change in the variance accounts for the shock size. If for an instant t $K_t = \tilde{K}$, the measurement innovation is distributed as a

normal with an higher variance than σ_ε^2.

The intervention model using a mixture of distributions is formulated as:

$$y_t = \alpha_t + K_t \varepsilon_t, \tag{3.14}$$

$$\alpha_{t+1} = \alpha_t + \eta_t, \tag{3.15}$$

and the conditions given by expressions (3.9) to (3.11).

The prior distribution for the scale parameter is a Bernoulli distribution:

$$P\left(K_t = j\right) = \left\{ \begin{array}{ll} p, & j = \tilde{K} \\ 1 - p, & j = 1. \end{array} \right.$$

For simplicity we are assuming p fixed *a priori*, an assumption we will drop later. The value of δ, or $\tilde{K}$, is assumed to be prior knowledge. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$.

Using a Bayesian approach we sample from the posterior $f\left(\boldsymbol{\alpha}, \sigma_\varepsilon^2, \sigma_\eta^2, \mathbf{K}|Y\right)$, and the estimation of the hyperparameters is done including the possibility of shocks to the observable variable $\{y_t\}_{t=1}^T$. The posterior distribution of the intervention variable is used to decide whether shocks are present or not. For each t we look at $P(K_t = \delta|Y)$ for the additive formulation and $P(K_t = \tilde{K}|Y)$ for the mixture of distributions. Our decision about the presence of a shock at time t' depends on the proportion of times $K_{t'}$ was sampled different from zero, for the additive case, or different from one, for the mixture formulation.

The choice between the inclusion of K_t in an additive or multiplicative way is not straightforward, and it is directly related to the sampling scheme used for estimating the model, as well as with the prior knowledge assumed about the data.

Using the additive formulation allows for more general specifications for the intervention prior, namely the use of continuous priors. Additionally, given the classification we are using for the different type of interventions, inclusion as location parameter is preferred; we will see later that the magnitude of the shock can be assumed to have a more general distribution. This allows us to make inference about the magnitude, based on samples from its posterior distribution. Furthermore, the additive formulation permits the definition of more general models. For example, suppose we formulate a model with a non-normal distribution for the measurement innovation process, which can be approximated by a mixture of normals (Carlin,

Polson, and Stoffer, 1992, Shephard, 1994). If in addition we want to include intervention variables, this is not feasible using a mixture of normal distributions, as that device is already being used to approximate the innovations process distribution.

Intervention variables are included as a location parameter by McCulloch and Tsay (1993), in the context of ARMA models, where a normal prior is used for detecting additive outliers and level shifts (considering these two situations separately). In Chib, Nardari, and Shephard (2002), in the context of volatility models, a normal distribution is assumed as a prior distribution for a variable that approximates the magnitude of an intervention in the mean process. Examples of the use of mixture distributions for modeling interventions are found in Barnett, Kohn, and Sheather (1997), for ARMA models, and Carter and Kohn (1996) for a general SSM. In both these papers a discrete prior is used for the scale parameter. This type of formulation, has the drawback of restricting the choice for the magnitude of the interventions to a set of values that is chosen *a priori*. It has the advantage of yielding a discrete posterior for the intervention variable.

In what follows the formulation of intervention models is done using a location parameter, that is an intervention variable that affects the observations or the state space vector in an additive way.

3.4.1 Gibbs sampling assuming a discrete prior for K_t

The Gibbs sampler is the MCMC method we use for estimating the parameters of SSM with interventions. Two type of interventions are considered: outliers and shifts in the transition equation. As before, for simplicity of exposition, the methods are applied to a local level model.

Let $\{y_t\}_{t=1}^T$ be an univariate time series, which is modeled according to a state space model, with an univariate state vector $\{\alpha_t\}_{t=1}^T$, where the possibility of the existence of shocks is taken in account, by the use of the intervention vector $\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2)$, where $\mathbf{K}_i = (K_{i,1}, \ldots, K_{i,T})$, for $i = 1, 2$:

$$y_t = \alpha_t + K_{1,t} + \mathbf{G}_t \mathbf{u}_t, \tag{3.16}$$

$$\alpha_{t+1} = \alpha_t + K_{2,t} + \mathbf{H}_t \mathbf{u}_t. \tag{3.17}$$

Initially we consider the coefficients in $\mathbf{H}_t$ and $\mathbf{G}_t$ (represented by $\boldsymbol{\theta}$) to be known and we focus on the properties of the joint posterior distribution $f(\boldsymbol{\alpha}, \mathbf{K} | Y, \boldsymbol{\theta})$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$. We assume that $\mathbf{H}_t' \mathbf{G}_t = 0$, $rank(\mathbf{H}_t) \neq 0$, $rank(\mathbf{G}_t) \neq 0$, together with the diffuse initial conditions for the state vector given in expression (3.11).

The intervention variables $K_{i,s}$ and $K_{j,t}$ are assumed to be a priori independent for every $s \neq t$ and $i, j = 1, 2$. For each of the components of the intervention variables we assume a Bernoulli prior distribution:

$$K_{i,t} = \begin{cases} 0, & \text{with probability} \quad 1 - p_i \\ k_i, & \text{with probability} \quad p_i, \end{cases} \tag{3.18}$$

for $i = 1, 2$, with k_i and $0 < p_i < 1$ chosen *a priori*. This assumption implies that if there is an intervention in one of the equations defining the state space model, its size is restricted to the value defined by the prior distribution.

For now, the parameters p_i and k_i of the prior distribution of $K_{i,t}$ are specified according to each particular case. We will see later that the prior probability of having an intervention is easily included in the parameters to be sampled.

In order to sample from $(\boldsymbol{\alpha}, \mathbf{K} | Y, \boldsymbol{\theta})$ the Gibbs sampler is used. Suppose we start the sampler with a set of initial values $\left(\boldsymbol{\alpha}^{(0)}, \mathbf{K}^{(0)}\right)$ and, after i iterations the current state of the vector is $\left(\boldsymbol{\alpha}^{(i)}, \mathbf{K}^{(i)}\right)$. Then for iteration $i+1$, $\left(\boldsymbol{\alpha}^{(i+1)}, \mathbf{K}^{(i+1)}\right)$ is obtained through the updating sampling scheme:

$$\boldsymbol{\alpha}^{(i+1)} \sim f\left(\boldsymbol{\alpha} | Y, \boldsymbol{\theta}, \mathbf{K}^{(i)}\right); \tag{3.19}$$

$$\mathbf{K}^{(i+1)} \sim f\left(\mathbf{K} | Y, \boldsymbol{\theta}, \boldsymbol{\alpha}^{(i+1)}\right). \tag{3.20}$$

After convergence is achieved, this allows us to sample from the target posterior for $(\boldsymbol{\alpha}, \mathbf{K})$:

$$(\boldsymbol{\alpha}, \mathbf{K}) \sim f(\boldsymbol{\alpha}, \mathbf{K} | Y, \boldsymbol{\theta}).$$

For simplicity, in the next derivations we omit the conditioning parameter vector $\boldsymbol{\theta}$. Conditional on the intervention variables, equations (3.16) and (3.17) define a Gaussian state space model. Thus sampling from (3.19) is done in the same way

as in Chapter 2, using a simulation smoother (De Jong and Shephard, 1995). We focus our attention in sampling from (3.20). We start by noticing that:

$$
\begin{aligned}
f\left(\mathbf{K} \mid Y, \boldsymbol{\alpha}\right) &= f\left(\mathbf{K}_1, \mathbf{K}_2 \mid Y, \boldsymbol{\alpha}\right) = f\left(\mathbf{K}_1 \mid Y, \boldsymbol{\alpha}\right) f\left(\mathbf{K}_2 \mid Y, \boldsymbol{\alpha}\right) 11 \quad (3.21) \\
&= \prod_{t=1}^{n} f\left(K_{1,t} \mid Y, \boldsymbol{\alpha}\right) f\left(K_{2,t} \mid Y, \boldsymbol{\alpha}\right). \quad (3.22)
\end{aligned}
$$

The assumption of prior independence between the different types of intervention implies the factorization in (3.21). On the other hand, given the assumption of serial independence for the intervention variables we obtain factorization in (3.22). Furthermore, given the temporal dynamics of the model, we have that expression (3.22) can be written as:

$$
\prod_{t=1}^{n} f\left(K_{1,t} \mid y_t, \alpha_t\right) f\left(K_{2,t} \mid \alpha_{t+1}, \alpha_t\right). \quad (3.23)
$$

A similar factorization to (3.23) is used in Shephard (1993). Expression (3.23) is obtained by the independence assumptions on $\mathbf{K}$ and the formulation of the measurement and transition equations and does not depend on the specific prior distribution assumed for that variable. Therefore sampling from the full conditional of the intervention variables is done by sampling from

$$
f\left(K_{1,t} \mid y_t, \alpha_t\right), \quad (3.24)
$$

$$
f\left(K_{2,t} \mid \alpha_{t+1}, \alpha_t\right), \quad (3.25)
$$

for $t = 1, \ldots, T$.

For $K_{1,t}$'s distribution in (3.24) we have that:

$$
\begin{aligned}
f\left(K_{1,t} \mid y_t, \alpha_t\right) &\propto f\left(y_t \mid \alpha_t, K_{1,t}\right) f\left(K_{1,t} \mid \alpha_t\right) \quad (3.26) \\
&= f\left(y_t \mid \alpha_t, K_{1,t}\right) f\left(K_{1,t}\right). \quad (3.27)
\end{aligned}
$$

With the assumption of a discrete prior for $K_{1,t}$, its full conditional distribution is also discrete. To sample from this distribution, the right hand side of expression (3.27) has to be evaluated for the different values of $K_{1,t}$ defined *a priori*. The likelihood function in (3.27) is a normal density:

$$
y_t \mid \alpha_t, K_{1,t} \sim N\left(\alpha_t + K_{1,t}, \mathbf{G}_t \mathbf{G}_t'\right).
$$

Having the values of this likelihood for the current state vector and for the different values that $K_{1,t}$ assumes, together with its prior probabilities, we get the posterior probabilities of $K_{1,t}$, up to a constant factor, and from that its posterior distribution.

Doing the same type of derivation for intervention $K_{2,t}$'s distribution in (3.25) we get:

$$
\begin{aligned}
f\left(K_{2,t} | \alpha_{t+1}, \alpha_t\right) & \propto f\left(\alpha_{t+1}, \alpha_t | K_{2,t}\right) f\left(K_{2,t}\right) \\
& = f\left(\alpha_{t+1} | \alpha_t, K_{2,t}\right) f\left(K_{2,t}\right) f\left(\alpha_t | K_{2,t}\right) \qquad (3.28) \\
& \propto f\left(\alpha_{t+1} | \alpha_t, K_{2,t}\right) f\left(K_{2,t}\right) . \qquad (3.29)
\end{aligned}
$$

The intervention variable $K_{2,t}$ affects the state space vector in the following period, which justifies the passage from expressions (3.28) to (3.29). The one step-ahead state prediction is normally distributed as:

$$
\alpha_{t+1} | \alpha_t, K_{2,t} \sim N\left(\alpha_t + K_{2,t}, \mathbf{H}_t \mathbf{H}_t'\right)
$$

Evaluating this density for each value of $K_{2,t}$, together with its prior probabilities we obtain the full conditional probabilities for the intervention variable $K_{2,t}$.

After convergence is achieved, we have samples from the posterior distribution $(\boldsymbol{\alpha}, \mathbf{K} | Y)$. The posterior samples from the interventions are used to establish the existence and location of shocks. To illustrate how this is accomplished, suppose we have a sample of size N of draws from the posterior distribution of the intervention variable for the presence of an outlier, $\left\{K_{1,t}^{(i)}\right\}_{t=1}^{T}$ for $i = 1, \ldots, N$. Let $\bar{p}_{1,t}$ be the posterior probability of having an outlier at any instant t, obtained, for a posterior sample of size N, as the proportion of times $K_{1,t}$ was sampled with a value different from zero. The position of the outliers is established by analyzing the values of the posterior probabilities of having a shock, over time.

There are two distinct approaches to determining the location of the shocks. We may set a threshold for the value of the posterior probability of a shock, for example 0.5. Any point with posterior probability greater than this value is classified as the location of a shock. An alternative is the use of informative plots to identify outlying values for the posterior probability. In McCulloch and Tsay (1994), the posterior probabilities are plotted and the shocks are identified as the observations

with outstanding values for that variable. Using this procedure, for the detection of level shifts, several points are identified as level shifts, with posterior probabilities ranging from 0.40 to 0.55, as they are outstanding values when compared with the posterior probabilities for the remaining observations. Examples of this informative plots approach are found in Ray and Tsay (2001) for the detection of change points to a long memory parameter, where change points are detected with posterior probability ranging from 0.40 to 0.77. In So (1999), for the detection of outliers, seven observations are classified as possible shocks, with posterior probabilities with values greater than 0.15. The other decision method is classifying as shocks observations such that the correspondent posterior probability has a value greater than c, for a certain $0 < c < 1$. In Justel, Pena, and Tsay (2001), the threshold value c is taken to be 0.5, for detection of patches of outliers in autoregressive models. Using a threshold value for determining the position of the shocks has the advantage of being an automatic procedure. It is the natural procedure to use when doing Monte Carlo studies. In that case is not feasible to inspect the set of posterior probabilities for all the simulated data sets. When working with real data set the threshold method can also be used. However, it should be complemented with the analysis of the posterior probabilities set, as a whole. Using only the inspection of this set of probabilities for the detection of outlying values might induce in an over fitting of the data, and the detection of too many shocks. Using only a threshold value does not take in account the specific behaviour of the data set in study. Therefore, in empirical applications, the procedure of detecting the position of the shocks should be a compromise between an informative and an automatic method.

In the sampling scheme described above the intervention variables are sampled conditioned on the observations and the state space vector. This approach is used in Shephard (1994), where outliers are dealt with by the use of either a t-distribution for the innovation processes or by a mixture of normals. In Gerlach, Carter, and Kohn (2000), a different scheme is used, where the interventions, included as scale parameters, are sampled conditional on the observations instead of the observations and the states. Although this approach is more efficient, as the interventions are conditioned on fewer variables, (Liu, Wong, and Kong, 1994) the resulting poste-

rior distribution is tractable only in the case of discrete priors for the intervention variables. The limitations of this approach are discussed in the next section.

3.5 Generalizing the interventions prior distribution

In the previous section we described how to sample from the posterior distribution of the intervention variables, when a Bernoulli prior distribution is assumed. This distribution can be easily generalized to a multinomial distribution. However, this approach demands a considerable amount of prior information to define the set of values the shocks can take, together with its probabilities.

To overcome this problem, we assume a continuous prior distribution for the size of the intervention variables. Consider as before a local level model, formulated for allowing the detection of outliers. The additive intervention term in the measurement equation is now composed by two factors, $k_{1,t}$ and $p_{1,t}$, the first one the magnitude of the shock and the second one an indicator variable for the existence of a shock at that point in time. With these definitions the corresponding SSM is formulated as:

$$y_t = \alpha_t + k_{1,t}p_{1,t} + \varepsilon_t, \tag{3.30}$$

$$\alpha_{t+1} = \alpha_t + \eta_t. \tag{3.31}$$

A possible choice for the prior distribution of the outlier size $k_{1,t}$ is a normal distribution:

$$k_{1,t} \sim N(\mu, \sigma^2) \tag{3.32}$$

This prior is used, with $\mu = 0$ in McCulloch and Tsay (1993) for detection of additive outliers, and in Justel, Pena, and Tsay (2001), for the detection of patches of outliers, both for autoregressive models. In Chib, Nardari, and Shephard (2002), for a stochastic volatility model with jumps in the mean equation, the jump is assumed to follow a normal distribution $N(-0.5\delta^2, \delta^2)$ with δ sampled together with the other parameters of the model. We consider μ and σ^2 as set *a priori*.

The indicator variable $p_{1,t}$ prior is a Bernoulli distribution

$$p_{1,t} \sim Bernoulli(q_1) \tag{3.33}$$

The probability q_1 is sampled together with the other variables. We assume a beta prior for q_1:

$$q_1 \sim Beta(a_1, b_1), \tag{3.34}$$

where a_1 and b_1 are set *a priori*, and their values are function of the prior expectations of number of shocks. This prior assumption for the prior probability of a shock is commonly used in the literature (McCulloch and Tsay, 1993 and Chib, Nardari, and Shephard, 2002, for example).

The prior assumption for the state vector is given in expression (3.11). The variances of the model are *a priori* distributed as inverse gammas, as assumed in Chapter 2, for the SSM without intervention variables,

$$\sigma_\varepsilon^2 \sim IG\left(\frac{c_1}{2}, \frac{s_1}{2}\right),$$
$$\sigma_\varepsilon^2 \sim IG\left(\frac{c_2}{2}, \frac{s_2}{2}\right).$$

Using this setup, and defining $\mathbf{k}_1 = (k_{1,1}, \ldots, k_{1,T})$, and $\mathbf{p}_1 = (p_{1,1}, \ldots, p_{1,T})$, we are interested in sampling from the posterior distribution:

$$\left(\alpha, \sigma_\varepsilon^2, \sigma_\eta^2, \mathbf{k}_1, \mathbf{p}_1, q_1 | Y\right),$$

which, using the Gibbs sampler, is done by sequentially sampling from the conditional distributions:

$$\alpha \sim f\left(\alpha | Y, \sigma_\varepsilon^2, \sigma_\eta^2, \mathbf{k}_1, \mathbf{p}_1, q_1\right) \tag{3.35}$$

$$\sigma_\varepsilon^2 \sim f\left(\sigma_\varepsilon^2 | Y, \alpha, \mathbf{k}_1, \mathbf{p}_1\right) \tag{3.36}$$

$$\sigma_\eta^2 \sim f\left(\sigma_\eta^2 | \alpha\right) \tag{3.37}$$

$$\mathbf{k}_1 \sim f\left(\mathbf{k}_1 | Y, \alpha, \sigma_\varepsilon^2, \mathbf{p}_1\right) \tag{3.38}$$

$$\mathbf{p}_1 \sim f\left(\mathbf{p}_1 | Y, \alpha, \sigma_\varepsilon^2, \mathbf{k}_1, q_1\right) \tag{3.39}$$

$$q_1 \sim f\left(q_1 | \mathbf{p}_1\right). \tag{3.40}$$

Conditionally on the intervention term the measurement equation is still Gaussian and, therefore, draws from (3.35) are obtained as in Chapter 2, using a simulation smoother (De Jong and Shephard, 1995).

The conditional distributions in (3.36) and (3.37) are derived using Lemma 2.4.1, and are inverse gamma distributions, respectively:

$$\sigma_\varepsilon^2 | Y, \boldsymbol{\alpha}, \mathbf{k}_1, \mathbf{p}_1 \sim IG \left(\frac{c_1 + T}{2}, \frac{s_1 + \sum_{t=1}^{T} (y_t - \alpha_t - k_{1,t} p_{1,t})^2}{2} \right)$$

and

$$\sigma_\eta^2 | \boldsymbol{\alpha} \sim IG \left(\frac{c_2 + T - 1}{2}, \frac{s_2 + \sum_{t=1}^{T-1} (\alpha_{t+1} - \alpha_t)^2}{2} \right)$$

To sample from the intervention variable distribution in (3.38) we assume that the occurrence of an outlier does not depend of past or future shocks, that is, $k_{1,t}$ and $k_{1,s}$ are independent for every $t \neq s$, which implies:

$$f \left(\mathbf{k}_1 | Y, \boldsymbol{\alpha}, \sigma_\varepsilon^2, \mathbf{p}_1 \right) = \prod_{t=1}^{T} f \left(k_{1,t} | y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} \right) \tag{3.41}$$

Let $f_N (x | \mu, \sigma^2)$ be the $N(\mu, \sigma^2)$ corresponding density evaluated at x. For every $t = 1, \ldots, T$

$$\begin{aligned} f \left(k_{1,t} | y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} \right) &\propto f \left(y_t | \alpha_t, k_{1,t}, \sigma_\varepsilon^2, p_{1,t} \right) f (k_{1,t}) \\ &= f_N \left(y_t | \alpha_t + k_{1,t} p_{1,t}, \sigma_\varepsilon^2 \right) f_N \left(k_{1,t} | \mu, \sigma^2 \right). \end{aligned} \tag{3.42}$$

If $p_{1,t} = 0$ the posterior for $k_{1,t}$ is proportional to its normal prior, and for $t = 1, \ldots, T$ we draw:

$$k_{1,t} | y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} \sim N \left(\mu, \sigma^2 \right).$$

If $p_{1,t} = 1$ the product of normal densities in (3.42) is given by:

$$\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left[-\frac{(y_t - (\alpha_t + k_{1,t}))^2}{2\sigma_\varepsilon^2} \right] \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(k_{1,t} - \mu)^2}{2\sigma^2} \right]$$

Conditional on $(y_t, \alpha_t, \sigma_\varepsilon^2)$ the product in (3.42) is proportional to

$$\exp \left[-\frac{k_{1,t}^2 - 2 \left(\frac{1}{\sigma_\varepsilon^2} + \frac{1}{\sigma^2} \right)^{-1} \left(\frac{y_t - \alpha_t}{2\sigma_\varepsilon^2} + \frac{\mu}{\sigma^2} \right) k_{1,t}}{2 \left(\frac{1}{\sigma_\varepsilon^2} + \frac{1}{\sigma^2} \right)^{-1}} \right]$$

Thus, for the case of $p_{1,t} = 1$, sampling from the posterior distribution in (3.38) is achieved by sampling from a normal distribution

$$N\left(\mu^*, \sigma^{*2}\right),\tag{3.43}$$

where:

$$\sigma^{*2} = \left(\frac{1}{\sigma_\varepsilon^2} + \frac{1}{\sigma^2}\right)^{-1}$$
$$\mu^* = \sigma^{*2}\left[(y_t - \alpha_t)\sigma_\varepsilon^{-2} + \mu\sigma^{-2}\right].$$

Using a similar argument to the one used for the size of intervention variable, we have that the conditional distribution for the indicator variable, in (3.39), can be written as:

$$f\left(\mathbf{p_1}|Y, \boldsymbol{\alpha}, \sigma_\varepsilon^2, \mathbf{k_1}, q_1\right) = \prod_{t=1}^{T} f\left(p_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, k_{1,t}, q_1\right)\tag{3.44}$$

Thus, sampling from the full conditional distribution of $p_{1,t}$ is done by sampling for $t = 1\ldots, T$ from the distributions on the righthand side of (3.44). Given the prior Bernoulli assumption for p_{1t}, the full conditional is a Bernoulli. To obtain the probabilities of this distribution, we note that:

$$f\left(p_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, k_{1,t}, q_1\right) \propto f\left(y_t|\alpha_t, \sigma_\varepsilon^2, k_{1,t}, p_{1,t}, q_1\right)f\left(p_{1,t}|q_1\right)$$

This implies that the probabilities for the full conditional Bernoulli distribution are obtained by the following expressions:

$$P\left(p_{1,t} = i|y_t, \alpha_t, \sigma_\varepsilon^2, k_{1,t}, q_1\right) \propto f_N\left(y_t|\alpha_t + k_{1,t}i, \sigma_\varepsilon^2\right)P\left(p_{1,t} = i\right),\tag{3.45}$$

for $i = 0, 1$, where

$$P\left(p_{1,t} = i\right) = iq_1 + (1-i)(1-q_1).$$

Sampling from the conditional distribution of q_1, in (3.40), given the prior $Beta\,(a_0, a_1)$, is done by using a standard result on the conjugate property of the beta and Bernoulli distributions (DeGroot, 1970).

Theorem 3.5.1 *Suppose that $X_1, \ldots, X_n$ is a random sample from a Bernoulli distribution with an unknown parameter q. Suppose also that the prior distribution of q is a beta distribution with parameters a and b such that $a > 0$ and $b > 0$. Then the posterior distribution of q when $X_i = x_i$ $(i = 1, \ldots, n)$ is a beta distribution with parameters $a + \sum_{i=1}^{n} x_i$ and $b + n - \sum_{i=1}^{n} x_i$.*

Using this result we get that

$$q_1 | \mathbf{p}_1 \sim Beta\left(a_1 + \sum_{t=1}^{T} p_{1,t}, b_1 + T - \sum_{t=1}^{T} p_{1,t}\right),$$

that is, we draw from the full conditional of q_1 by sampling from

$$Beta\left(a_1 + n_1, b_1 + n_0\right), \tag{3.46}$$

where n_1 the number of indicators sampled equal to one and $n_0 = T - n_1$.

3.5.1 Uninformative prior for the intervention magnitude

The Gaussian prior assumption in (3.38) has some drawbacks. Suppose that after i iterations of the Gibbs sampler, and for a certain t^*, we have $p_{1,t^*}^{(i)} = 0$. Additionally, assume that for the true data generating process there is an outlier at t^*. In next iteration we are interested in sampling an absolute value for $k_{1,t^*}^{(i+1)}$ large enough for the indicator variable $p_{1,t^*}^{(i+1)}$ to be next sampled with value one. As we are assuming a normal prior $N\left(\mu, \sigma^2\right)$ for the size $k_{1,t}$, either we take μ different from zero or set a high value for σ^2. If these parameters are fixed *a priori*, setting a value non zero for μ requires some previous information about the data. An alternative is fixing it equal to zero, as it is done in McCulloch and Tsay (1993) and McCulloch and Tsay (1994). However, it is not desirable that the prior distribution is concentrated around zero. In this, case we have to set σ^2 large enough to model outlying points. Finally we note that, when working with a real data set, there might not be enough prior information for a reasonable choice of the prior normal distribution parameters.

To provide an example of the sensitivity of the estimates of the intervention variables to the choice of the parameters of the prior distribution for the size variable,

we conduct a small Monte Carlo study. We simulate 100 data sets of size $T = 100$, from the local level model in equations (3.30) to (3.31). The variances are set to $\sigma_\epsilon^2 = 2$ and $\sigma_\eta^2 = 1$. A shock is introduced at $t = 25$ with size 10, to produce an outlier. The mean parameter μ is fixed to the value zero. For each simulated data set, we run 2000 Gibbs iteration, discarding the first 1000, for different values of the normal prior variance: $\sigma^2 = 10, 20, 50, 100, 200, 500$. The results obtained are summarized in Table 3.1

Table 3.1: Summary of output from posterior sample means, for the local level model with intervention variables in the measurement equation, across 100 simulated replications, for the normal prior parameter $\sigma^2 = 10, 20, 50, 100, 200, 500$.

	Truth	σ^2	10	20	50	100	200	500
$k_{1,25}$	10	Mean	7.805	8.741	9.634	9.721	9.542	10.115
		SD	1.400	1.511	1.358	1.536	1.602	1.850
$p_{1,25}$	1	Mean	0.953	0.974	0.953	0.980	0.955	0.965
		SD	0.140	0.078	0.140	0.0781	0.147	0.118
σ_ϵ^2	2	Mean	1.919	1.940	1.963	1.948	1.897	1.969
		SD	0.358	0.360	0.334	0.359	0.368	0.346
σ_η^2	1	Mean	1.136	1.057	1.097	1.104	1.111	1.143
		SD	0.290	0.280	0.266	0.286	0.360	0.320

The values reported are the mean and standard deviation of the posterior mean estimates across 100 samples, for different values of the variance of the normal prior distribution for the size of intervention variable. Based on 2,000 Gibbs draws, discarding the first 1,000.

We can see that for a variance with values of 10 or 20, the size of the outlier is underestimated. For the values considered for σ^2 over 50, the mean values of the estimates obtained across the replications are close to the true value. The posterior sample averages, obtained across the simulated replications, for the other parameters are not so sensitive to the different choices of the normal prior variance. The sensitivity, to the different values of σ^2, of the size of outlier estimated across the 100 simulated replications can be seen from the box-plots in Figure 3.2.

From all the values for the prior parameter the one with the best overall estimation results is $\sigma^2 = 100$. When using real data there is not enough prior information to decide which value of the intervention magnitude variance parameter will yield the best results. From the previous experience we conclude that in that case the

Figure 3.2: Box-plots of posterior mean size of outlier, across 100 simulated replications, for normal prior parameter $\sigma^2 = 10, 20, 50, 100, 200, 500$, for a local level model.
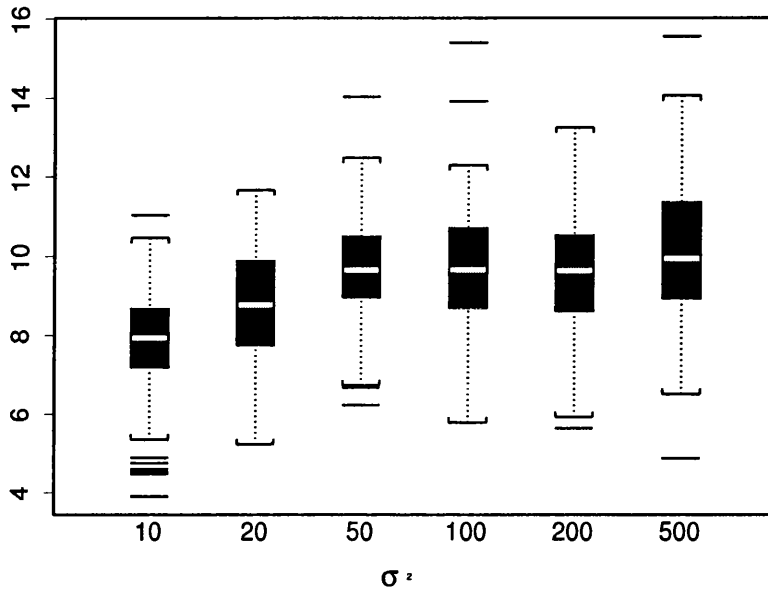
wiser choice is to take $\sigma^2 = M$, for a high value of M. In the limit this corresponds to having a diffuse prior, uninformative about the prior behaviour of $k_{1,t}$. An alternative is sampling μ and σ^2, in the Gibbs sampling process. In Chib, Nardari, and Shephard (2002), μ and σ^2 are defined as function of a parameter δ, which is sampled together with the other parameters of the model. The sampling from the full conditional of δ is done by a Metropolis-Hasting algorithm. This procedure adds more complexity to the sampling algorithm, particularly if there is not enough prior information to establish a functional relation between μ and σ^2, and we have to sample from both parameters posterior separately.

Given the above considerations, we propose a simpler approach; assuming an uninformative prior for the magnitude of intervention variable:

$$k_{1,t} \propto 1, \quad t = 1, \ldots, T. \tag{3.47}$$

A description of sampling from the size of intervention full conditional when assuming the flat prior in (3.47) follows. The factorization in (3.41), holds independently of the prior distributional assumption for k_1, and it implies that sampling

from the full conditional for this variable is done by sampling from the conditional distributions for $k_{1,t}$, for $t = 1, \ldots, T$. With the flat prior assumption we have that

$$f\left(k_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t}\right) \propto f_N(y_t|\alpha_t + k_{1,t}p_{1,t}, \sigma_\varepsilon^2),\tag{3.48}$$

for $t = 1, \ldots, T$.

Using the normal density definition and the fact that we are conditioning on everything except $k_{1,t}$, expression (3.48) can be expressed as:

$$f\left(k_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} = 1\right) \propto \exp\left[-\frac{k_{1,t}^2 p_{1,t}^2 - 2\left(y_t - \alpha_t\right)k_{1,t}p_{1,t}}{2\sigma_\varepsilon^2}\right].\tag{3.49}$$

If $p_{1,t} = 1$, to sample from the size of the intervention full conditional we sample from a normal distribution

$$f\left(k_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} = 1\right) \sim f_N\left(k_{1,t}|y_t - \alpha_t, \sigma_\varepsilon^2\right).\tag{3.50}$$

The full conditional distribution in (3.50), reflects the uninformative character of the prior distribution for the size of intervention variable. If the indicator variable has value one, to update the size variable, no prior information is incorporated in the full conditional distribution. We note that, although the prior distribution for the size of intervention variable is improper, when $p_{1,t} = 1$ the full conditional for the size of intervention at that instant is a proper posterior.

When $p_{1,t} = 0$, the full conditional distribution in expression (3.49) is proportional to a constant, and therefore has an improper distribution .

$$f\left(k_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} = 0\right) \propto 1.\tag{3.51}$$

To overcome the problem of sampling from the full conditional in this case, we approximate the flat distribution in (3.51) by a flat uniform distribution on a bounded range of definition:

$$f\left(k_{1,t}|y_t, \alpha_t, \sigma_\varepsilon^2, p_{1,t} = 0\right) \sim U\left[u_1, v_1\right],\tag{3.52}$$

with $u_1 < 0 < v_1$. The choice of these parameters reflects the prior knowledge of what should be the range of the size of the shocks for the data set, and does not

require an intensive prior analysis of the data. For example, setting

$u_1 = \min\{y_t : t = 1, \ldots, T\} - \max\{y_t : t = 1, \ldots, T\}$ and

$v_1 = \max\{y_t : t = 1, \ldots, T\} - \min\{y_t : t = 1, \ldots, T\}$, does not require much knowledge on the data set in analysis and, as we will see later, does not present the problems that arise for the normal prior.

The approximation in (3.52) has a negligible effect on the posterior distribution obtained for the size of the intervention variable. Consider an instant $t = t'$. Suppose there is not an outlier at $t = t'$, and that the current value for the indicator variable is $p_{1,t'} = 0$. The size variable is updated by sampling from the uniform ranging from u_1 to v_1. As long as the value sampled is not approximately zero, when updating the indicator variable, it will be updated as $p_{1,t'} = 0$. Therefore $k_{1,t'}p_{1,t'} = 0$ and for that iteration the size of intervention variable will not contribute to the average size of intervention for $t = t'$. Consider now that there is an outlier at $t = t'$ and that the current value for the indicator variable is $p_{1,t'} = 0$. By limiting the range from which to update the value of the intervention variable, we are increasing the probability of sampling a value of $k_{1,t'}$ close to the true value of the shock, and therefore increasing the probability that in the next updating of the intervention variable it will be sampled as $p_{1,t'} = 1$. Finally we notice that, when obtaining the posterior mean of the size of a shock, the product $k_{1,t}p_{1,t}$ is averaged over all the iterations where $p_{1,t}$ was sampled with value one. Consequently the major impact of the choice of the bounded uniform will be for those instances with a high posterior mean for the indicator of intervention variable. We show later in a sensitivity study, that the choice of the uniform boundaries does not affect the posterior samples of the intervention variables in these cases.

The assumption of an improper prior for a size of a shock has also been considered in Abraham and Box (1979) and Franses, Hoek, and Paap (1997). In Abraham and Box (1979) in the context of the detection of outliers for autoregressive models. In this paper the posterior distributions are derived analytically. A flat prior is assumed for the size of an outlier, occurring at any instant in time. The assumption of a flat prior, implies an approximation in the posterior distribution obtained for some of the parameters, since the flat prior is improper. In Franses, Hoek, and Paap

73

(1997), a flat prior is used for the size of a seasonal structural break, in the context of autoregressive models. The structural shift is assumed to occur only once in time, and the size of the shift is sampled conditioned on that shift having occurred. The full conditional distribution for the size of the seasonal shift is a normal distribution, and it corresponds to the normal full conditional from which we sample when the indicator variable is equal to one.

An alternative approach could have been taken: assumption of a uniform prior on a bounded range for the size of intervention variables, instead of the improper prior in (3.51). In this case, the full conditional to sample from, when $p_{1,t} = 1$ would be a truncated normal, with the prior uniform bounds defining the truncating points. The requirements of prior information about the data would be increased, to insure that the truncated normal would be wide enough to include the true sizes of the shocks. We have concluded that, when assuming a bounded uniform prior wide enough, the results obtained are very similar to the results we get when assuming an improper flat prior.

Sampling from the conditional densities of $\mathbf{p}_1$ and q_1 is done using expressions in (3.45) and (3.46), as these expressions do not depend on the prior assumption for the magnitude variable.

We propose a Gibbs sampling scheme, using the priors given by expressions (3.47), (3.33), (3.34), inverse gamma priors for the hyperparameters, $\sigma_\varepsilon^2 \sim IG\left(\frac{c_1}{2}, \frac{s_1}{2}\right)$ and $\sigma_\eta^2 \sim IG\left(\frac{c_2}{2}, \frac{s_2}{2}\right)$, and the normal diffuse assumption for α_1. Starting from a vector of initial values $\left(\boldsymbol{\alpha}^{(0)}, \sigma_\varepsilon^{2(0)}, \sigma_\eta^{2(0)}, \mathbf{k}_1^{(0)}, \mathbf{p}_1^{(0)}, q_1^{(0)}\right)$, suppose that after i iterations we have the sampled vector $\left(\boldsymbol{\alpha}^{(i)}, \sigma_\varepsilon^{2(i)}, \sigma_\eta^{2(i)}, \mathbf{k}_1^{(i)}, \mathbf{p}_1^{(i)}, q_1^{(i)}\right)$. Then on iteration $i + 1$ we sample from the full conditional distributions using the following sampling scheme:

Sampling scheme 3.1:

1. Draw from $\boldsymbol{\alpha}^{(i+1)}|Y, \sigma_\varepsilon^{2(i)}, \sigma_\eta^{2(i)}, \mathbf{k}_1^{(i)}, \mathbf{p}_1^{(i)}$, using the simulation smoother (De Jong and Shephard, 1995);

2. $\sigma_\varepsilon^{2(i+1)}|Y, \boldsymbol{\alpha}^{(i+1)}, \mathbf{k}_1^{(i)}, \mathbf{p}_1^{(i)} \sim IG\left(\frac{c_1+T}{2}, \frac{s_1 + \sum_{t=1}^{T}\left(y_t - \alpha_t^{(i+1)} - k_{1,t}^{(i)}p_{1,t}^{(i)}\right)^2}{2}\right)$;

3. $\sigma_\eta^{2(i+1)} | \alpha^{(i+1)} \sim IG \left(\frac{c_2 + T - 1}{2}, \frac{s_2 + \sum_{t=1}^{T-1} \left(\alpha_{t+1}^{(i+1)} - \alpha_t^{(i+1)} \right)^2}{2} \right);$

4. For $t = 1, \ldots, T$, if $p_{1,t}^{(i)} = 1$, then:

$k_{1,t}^{(i+1)} | y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, p_{1,t}^{(i)} \sim N \left(y_t - \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)} \right);$

Otherwise:

$k_{1,t}^{(i+1)} | y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, p_{1,t}^{(i)} \sim U[u_1, v_1];$

5. For $t = 1, \ldots, T$, draw $p_{1,t}^{(i+1)} | y_t, \alpha^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)}$ from the discrete two points distribution, with probabilities given by:

$P \left(p_{1,t}^{(i+1)} = 1 | y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)} \right) \propto f_N \left(y_t | \alpha_t^{(i+1)} + k_{1,t}^{(i+1)}, \sigma_\varepsilon^{2(i)} \right) q_1^{(i)},$

$P \left(p_{1,t}^{(i+1)} = 0 | y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)} \right) \propto f_N \left(y_t | \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)} \right) (1 - q_1^{(i)});$

6. $q_1^{(i+1)} | \mathbf{p}_1^{(i+1)} \sim Beta \left(a_1 + \sum_{t=1}^{T} p_{1,t}^{(i+1)}, b_1 + T - \sum_{t=1}^{T} p_{1,t}^{(i+1)} \right).$

If $\sigma_\varepsilon^2 \neq 0$ the later scheme defines an irreducible and an aperiodic Markov chain. In this case, independently from the initial conditions, we can reach any point of the sample space with positive probability (see §2.2.3, equation (2.8)) The case of $\sigma_\varepsilon^2 = 0$ will be dealt with in Chapter 5.

Comparative sensitivity analysis between the normal and uniform prior assumptions

To establish whether by using a flat prior for the size of the intervention variable, there is an improvement of the results obtained for the analysis of sensitivity with a normal prior, we conduct a similar Monte Carlo study, now assuming a uniform prior. We use the same 100 simulated data sets. The different values used for the parameters of the approximating bounded uniform distribution $U[u_1, v_1]$, are such that the two first moments of this distribution correspond to the moments of the normal prior $N(\mu, \sigma^2)$, with $\mu = 0$, and

$$\sigma^2 = 10, 20, 50, 100, 200, 500. \tag{3.53}$$

Let $X \sim U[u_1, v_1]$. Then, $E[X] = \frac{v_1 + u_1}{2}$ and $var[X] = \frac{(v_1 - u_1)^2}{12}$. Comparing these expressions with $\mu = 0$ and σ^2 we obtain that, given a normal distribution $N(0, \sigma^2)$,

the correspondent uniform distribution $U[u_1, v_1]$, with the same two moments, verifies $u_1 = -v_1$ and $v_1 = \sqrt{3\sigma^2}$.

Using the values in (3.53) we get as values for the upper boundary of the uniform distribution: $v_1 = 5.5, 7.7, 12.2, 17.3, 24.5, 38.7$. A summary of the Monte Carlo experiment output is given in Table 3.2.

Table 3.2: Summary of output from posterior sample means, for the local level model with intervention variables in the measurement equation, across 100 simulated replications, for the approximating uniform distribution parameters $v_1 = 5.5, 7.7, 12.2, 17.3, 24.5, 38.7$ and $u_1 = -v_1$.

	Truth	v_1	5.5	7.7	12.2	17.3	24.5	38.7
$k_{1,25}$	10	Mean	9.834	9.844	9.854	9.851	9.863	9.870
		SD	1.833	1.836	1.819	1.815	1.833	1.814
$p_{1,25}$	1	Mean	0.979	0.979	0.971	0.959	0.950	0.938
		SD	0.071	0.071	0.102	0.125	0.136	0.164
σ_ε^2	2	Mean	1.815	1.837	1.899	1.929	1.952	1.971
		SD	0.352	0.343	0.359	0.370	0.385	0.379
σ_η^2	1	Mean	1.076	1.091	1.094	1.096	1.095	1.095
		SD	0.268	0.266	0.278	0.277	0.284	0.269

The values reported are the mean and standard deviation of the posterior mean estimates across 100 samples, for different values of the parameters of the uniform prior for the size of intervention. Based on 2,000 Gibbs draws, discarding the first 1,000.

The mean estimates of the size of the outlier, across the replications, are much less sensitive to the change of the uniform parameters, when compared with the results obtained using the normal distribution, presenting mean values estimates very close to the true value. The remaining parameters also present a small sensitivity to the choice of the uniform prior parameters. Comparatively to the results in Table 3.1, the level component hyperparameter, presents less sensitivity to the choice of the uniform prior parameters than to the choice of the normal prior variance. The average estimates of this hyperparameter, obtained using the uniform prior, present overall values closer to the true one. The irregular component hyperparameter presents a slightly higher sensitivity to the choice of the uniform parameters than to the choice of the normal variance. Still, using the uniform prior, the posterior sample means for this hyperparameter are on average accurate estimates of the true

value. In Figure 3.3 we plot the box-plots for the estimates of the size variable, across the simulated replications, and for the different values of parameters of the approximating uniform prior distribution.

Figure 3.3: Box-plot of posterior mean size of outlier, across 100 simulated replications, for approximating uniform prior parameter v_1 =5.5, 7.7, 12.2, 17.3, 24.5, 38.7, for a local level model.
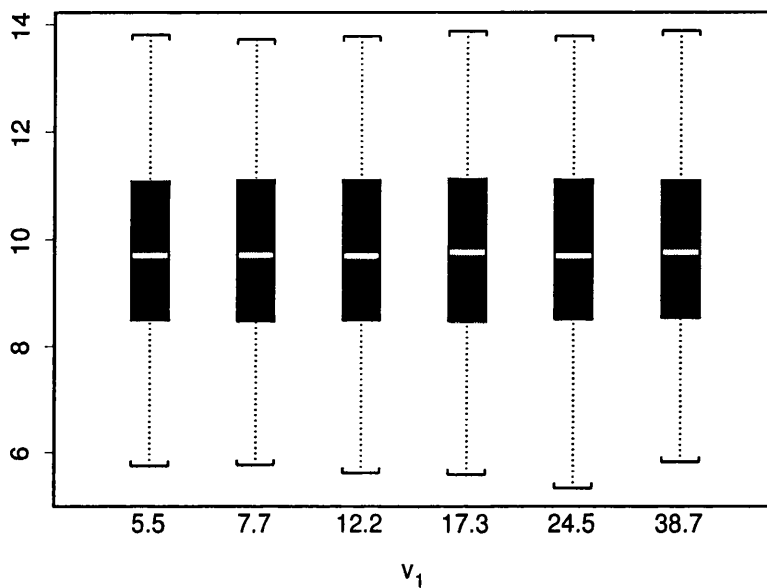
The insensitivity of the magnitude estimates to prior assumptions using a uniform distribution are clear from Figure 3.3. The estimates, across the simulated replications, obtained using this prior are overall better than the ones obtained using the normal prior. Only for the higher value of σ^2 considered, $\sigma^2 = 500$, we obtain an average estimate closer to the true value of the size of the outlier, when using the normal prior. Comparing with the average estimate of the magnitude, for the correspondent choice of parameters for the approximating uniform prior, we conclude that, even in this case, the accuracy of the average estimates, for the two prior distributions, is very similar. The use of a flat prior for the size variable presents evident advantages. The parameters' choice for the bounded uniform approximation does not have a significant impact on the posterior samples obtained for this variable.

77

3.5.2 Detection of outliers and level shifts with an uninformative prior for the shocks magnitude

We consider the local level model, allowing for the presence of outliers and level shifts, and present a Gibbs sampler algorithm using the uninformative prior for the shocks magnitude proposed in §3.5.1. The model is specified as:

$$y_t = \alpha_t + k_{1,t} p_{1,t} + \varepsilon_t, \tag{3.54}$$

$$\alpha_{t+1} = \alpha_t + k_{2,t} p_{2,t} + \eta_t. \tag{3.55}$$

There is a level shift at $t = t^*$ of size k_{2,t^*-1} if $p_{2,t^*-1} = 1$, and in a similar way we have an outlier at $t = t'$ of size $k_{1,t'}$ if $p_{1,t'} = 1$.

We assume the usual conditions of independence between the measurement and transition noise processes, together with the distribution assumptions in (3.9) and (3.10), and the diffuse initial condition for the state space vector given in (3.11). For the remaining parameters of the model the prior distributions assumed are:

$$\sigma_\varepsilon^2 \sim IG\left(\frac{c_1}{2}, \frac{s_1}{2}\right),$$

$$\sigma_\eta^2 \sim IG\left(\frac{c_2}{2}, \frac{s_2}{2}\right),$$

$$k_{1,t} \propto 1, \text{for } t = 1, \ldots, T,$$

$$k_{2,t} \propto 1, \text{for } t = 1, \ldots, T,$$

$$p_{1,t} \sim Bernoulli(q_1), \text{for } t = 1, \ldots, T,$$

$$p_{2,t} \sim Bernoulli(q_2), \text{for } t = 1, \ldots, T,$$

$$q_1 \sim Beta(a_1, b_1),$$

$$q_2 \sim Beta(a_2, b_2).$$

The parameters $c_1, c_2, s_1, s_2, a_1, a_2, b_1, b_2$ are set *a priori*. We assume that the occurrence of a shock is independent across equations and for different instances in time. With these assumptions the derivation of the posterior distribution for the variables $\mathbf{k}_1$, $\mathbf{p}_1$ and q_1 is identical to that given in §3.5.1. In a similar way, we obtain the expressions for sampling from the full conditionals for the variables $\mathbf{k}_2 = (k_{2,1}, \ldots, k_{2,T})$, $\mathbf{p}_2 = (p_{2,1}, \ldots, p_{2,T})$ and q_2.

For the variable representing the size of a level shift we have that:

$$f\left(\mathbf{k_2}|Y, \boldsymbol{\alpha}, \sigma_\eta^2, \sigma_\eta^2, \mathbf{k_1}, \mathbf{p_1}, \mathbf{p_2}, q_1, q_2\right) = \prod_{t=1}^{T} f\left(k_{2,t}|\alpha_{t+1}, \alpha_t, \sigma_\eta^2, p_{2,t}\right).$$

On the other hand,

$$f\left(k_{2,t}|\alpha_{t+1}, \alpha_t, \sigma_\eta^2, p_{2,t}\right) \propto f_N\left(\alpha_{t+1}|\alpha_t + k_{2,t}p_{2,t}, \sigma_\eta^2\right). \tag{3.56}$$

Hence if the actual value of $p_{2,t}$ is one we update the value of $k_{2,t}$ by sampling from the normal distribution

$$k_{2,t}|\alpha_{t+1}, \alpha_t, \sigma_\eta^2, p_{2,t} \sim N\left(\alpha_{t+1} - \alpha_t, \sigma_\eta^2\right).$$

Otherwise, $k_{2,t}$ is updated by sampling from the uniform distribution $U[u_2, v_2]$. This uniform distribution is an approximation of the improper full conditional of $k_{2,t}$. The parameters u_2, v_2 are set *a priori*.

The updating of the indicator variable $p_{2,t}$ is done by noting that:

$$f\left(\mathbf{p_2}|Y, \boldsymbol{\alpha}, \sigma_\varepsilon^2, \sigma_\eta^2, \mathbf{k_1}, \mathbf{k_2}, \mathbf{p_1}, q_1, q_2\right) = \prod_{t=1}^{T} f\left(p_{2,t}|\alpha_{t+1}, \alpha_t, \sigma_\eta^2, k_{2,t}, q_2\right).$$

Consequently, $p_{2,t}$ is sampled from a Bernoulli distribution with probabilities obtained through:

$$P\left(p_{2,t} = l|\alpha_{t+1}, \alpha_t, \sigma_\eta^2, k_{2,t}, q_2\right) \propto f_N\left(\alpha_{t+1}|\alpha_t + lk_{2,t}, \sigma_\eta^2\right) P\left(p_{2,t} = l\right),$$

for $l = 0, 1$, where $P\left(p_{2,t} = 1\right) = q_2$. We set $p_{2,T} = 0$ corresponding to having no level shift at time $T+1$, that is, no level shifts outside the time range considered for the observations. Hence, the updating of the size and indicator variables, for level shifts, is done for $t = 1, \ldots, T-1$.

With the beta prior assumption for q_2, this variable full conditional distribution is also a beta distribution, given by

$$q_2|\mathbf{p_2} \sim Beta\left(a_2 + \sum_{t=1}^{T} p_{2,t}, b_2 + T - \sum_{t=1}^{T} p_{2,t}\right).$$

We now propose a Gibbs sampler scheme for estimation of SSM in the presence of outliers and level shifts.

Starting from a vector of initial values

$$\left(\boldsymbol{\alpha}^{(0)}, \sigma_\varepsilon^{2(0)}, \sigma_\eta^{2(0)}, \mathbf{k}_1^{(0)}, \mathbf{k}_2^{(0)}, \mathbf{p}_1^{(0)}, \mathbf{p}_2^{(0)}, q_1^{(0)}, q_2^{(0)}\right),$$

suppose that after i iterations we have the sampled vector

$$\left(\boldsymbol{\alpha}^{(i)}, \sigma_\varepsilon^{2(i)}, \sigma_\eta^{2(i)}, \mathbf{k}_1^{(i)}, \mathbf{k}_2^{(i)}, \mathbf{p}_1^{(i)}, \mathbf{p}_2^{(i)}, q_1^{(i)}, q_2^{(i)}\right).$$

Then on iteration $i + 1$ we sample from the full conditional distributions using the following sampling scheme:

Sampling scheme 3.2

1. Draw from $\boldsymbol{\alpha}^{(i+1)}|Y, \sigma_\varepsilon^{2(i)}, \sigma_\eta^{2(i)}, \mathbf{k}_1^{(i)}, \mathbf{k}_2^{(i)}, \mathbf{p}_1^{(i)}, \mathbf{p}_2^{(i)}$, using the simulation smoother (De Jong and Shephard, 1995);

2. $\sigma_\varepsilon^{(i+1)}|Y, \boldsymbol{\alpha}^{(i+1)}, \mathbf{k}_1^{(i)}, \mathbf{p}_1^{(i)} \sim IG\left(\frac{c_1+T}{2}, \frac{s_1+\sum_{t=1}^{T}\left(y_t - \alpha_t^{(i+1)} - k_{1,t}^{(i)} p_{1,t}^{(i)}\right)^2}{2}\right);$

3. $\sigma_\eta^{(i+1)}|\boldsymbol{\alpha}^{(i+1)}, \mathbf{k}_2^{(i)}, \mathbf{p}_2^{(i)} \sim IG\left(\frac{c_2+T-1}{2}, \frac{s_2+\sum_{t=1}^{T-1}\left(\alpha_{t+1}^{(i+1)} - \alpha_t^{(i+1)} - k_{2,t}^{(i)} p_{2,t}^{(i)}\right)^2}{2}\right);$

4. For $t = 1, \ldots, T$, if $p_{1,t}^{(i)} = 1$, then:

   $k_{1,t}^{(i+1)}|y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, p_{1,t}^{(i)} \sim N\left(y_t - \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}\right);$

   Otherwise:

   $k_{1,t}^{(i+1)}|y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, p_{1,t}^{(i)} \sim U[u_1, v_1];$

5. For $t = 1, \ldots, T - 1$, if $p_{2,t}^{(i)} = 1$, then:

   $k_{2,t}^{(i+1)}|\alpha_{t+1}^{(i+1)}, \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}, p_{2,t}^{(i)} \sim N\left(\alpha_{t+1}^{(i+1)} - \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}\right);$

   Otherwise:

   $k_{2,t}^{(i+1)}|\alpha_{t+1}^{(i+1)}, \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}, p_{2,t}^{(i)} \sim U[u_2, v_2];$

6. For $t = 1, \ldots, T$, draw $p_{1,t}^{(i+1)}|Y, \alpha^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)}$ from the discrete two points distribution, with probabilities given by:

   $P\left(p_{1,t}^{(i+1)} = 1|y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)}\right) \propto f_N\left(y_t|\alpha_t^{(i+1)} + k_{1,t}^{(i+1)}, \sigma_\varepsilon^{2(i+1)}\right) q_1^{(i)},$

   $P\left(p_{1,t}^{(i+1)} = 0|y_t, \alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)}\right) \propto f_N\left(y_t|\alpha_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}\right) (1 - q_1^{(i)});$

7. For $t = 1, \ldots, T - 1$, draw $p_{2,t}^{(i+1)} | \alpha_{t+1}^{(i+1)}, \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}, k_{2,t}^{(i+1)}, q_2^{(i)}$ from the discrete two points distribution, with probabilities given by:

$$P\left(p_{2,t}^{(i+1)} = 1 | \alpha_{t+1}^{(i+1)}, \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}, k_{2,t}^{(i+1)}, q_2^{(i)}\right) \propto$$
$$f_N\left(\alpha_{t+1}^{(i+1)} | \alpha_t^{(i+1)} + k_{2,t}^{(i+1)}, \sigma_\eta^{2(i+1)}\right) q_2^{(i)},$$

$$P\left(p_{2,t}^{(i+1)} = 0 | \alpha_{t+1}^{(i+1)}, \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}, k_{2,t}^{(i+1)}, q_2^{(i)}\right) \propto$$
$$f_N\left(\alpha_{t+1}^{(i+1)} | \alpha_t^{(i+1)}, \sigma_\eta^{2(i+1)}\right) (1 - q_2^{(i)});$$

8. $q_1^{(i+1)} | \mathbf{p}_1^{(i+1)} \sim Beta\left(a_1 + \sum_{t=1}^{T} p_{1,t}^{(i+1)}, b_1 + T - \sum_{t=1}^{T} p_{1,t}^{(i+1)}\right).$

9. $q_2^{(i+1)} | \mathbf{p}_2^{(i+1)} \sim Beta\left(a_2 + \sum_{t=1}^{T} p_{2,t}^{(i+1)}, b_2 + T - \sum_{t=1}^{T} p_{2,t}^{(i+1)}\right).$

This sampling scheme converges to the target posterior distribution as long as $\sigma_\varepsilon^2 \neq 0$ and $\sigma_\eta^2 \neq 0$. The case of one of these variances being equal to zero is dealt with in Chapter 5.

3.5.3 A Monte Carlo study

In what follows we conduct a Monte Carlo experiment for assessing the performance of *sampling scheme 3.2*. We simulate artificial data sets of size $T = 100$, using as data generating process:

$$y_t = \alpha_t + k_{1,t} p_{1,t} + \varepsilon_t,$$
$$\alpha_{t+1} = \alpha_t + k_{2,t} p_{2,t} + \eta_t,$$

with $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$. Two type of shocks are introduced in the data; outliers at time $t = 20$ and $t = 50$ with sizes $k_{1,20} = 7$ and $k_{1,50} = -6$; level shifts at time $t = 40$ and $t = 75$ with sizes $k_{2,39} = 7$ and $k_{2,74} = -7$. This process is repeated 10000 times. In Figure 3.4, we plot one of the artificial data sets used.

Sampling scheme 3.2 is used to estimate the parameters of each data set. The Gibbs sampler is run for 10,000 iterations, with the first 5,000 discarded. The prior distributions used are:

$$\sigma_x^2 \sim IG\left(\frac{5}{2}, \frac{5}{2}\right), \text{ for } x = \varepsilon, \eta;$$
$$k_{i,t} \propto 1, \text{ for } t = 1, \ldots, T;$$
$$q_i \sim Beta\left(2, 100\right), \text{ for } i = 1, 2.$$

Figure 3.4: Artificial data set generated from a local level model, with T=100, with outliers at time $t = 20$ and $t = 50$, of size 7 and -6, and level shifts at time $t = 40$ and $t = 75$ of size 7 and -7.

The approximating uniform distributions are taken as $U[-10,10]$. In Table 3.3 we present a summary of the output obtained for the posterior samples of the parameters across the 10,000 replications of this simulation experiment.

The mean posterior probability of having an outlier or a level shift, in the instances where they were input to the simulate data sets, presents an average value, across the simulate replications, above 0.75, presenting higher values in the case of the outliers. The average posterior mean of the size of the shocks is also accurate and close to the their true values. The estimated values for the hyperpameters agree in average with their true values, which is also a consequence of the fact that the shocks are correctly accounted for in the model. Taking as threshold value 0.5, for the detection of a shock, we can observe that in more than 78% of the simulated replications, each shock is correctly detected. Analyzing the values for the median of the posterior mean samples we can see that this central measure presents, for the majority of the parameters, values closer to the true ones.

In Figure 3.5, we plot the box-plots of the samples, obtained over the 10,000 replications, for the posterior mean of the indicator variables, for the instances where their true value is one.

Figure 3.5: Box-plot of posterior sample means of probability of outliers (a)-(b) and level shifts (c)-(d), across 10,000 simulated replications, for a local level model.
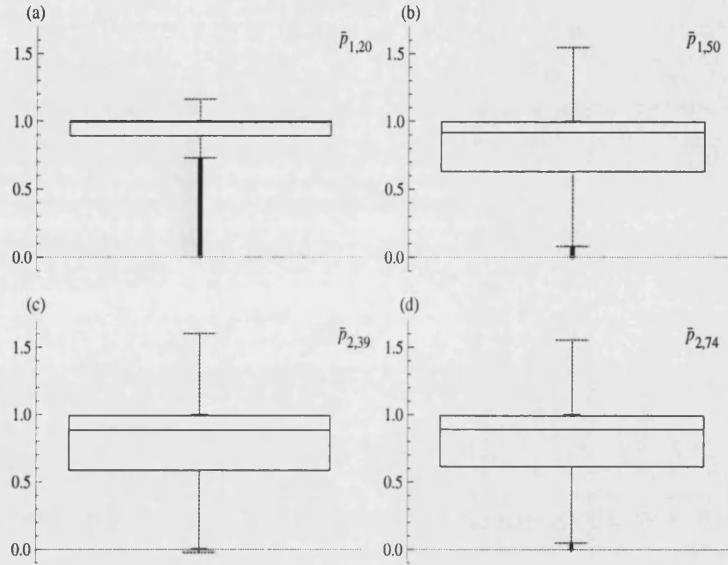
Figure 3.6: Histograms (a)-(b) and and box-plots (c)-(d) of posterior sample means of hyperparameters, across 10,000 replications, for a local level model.
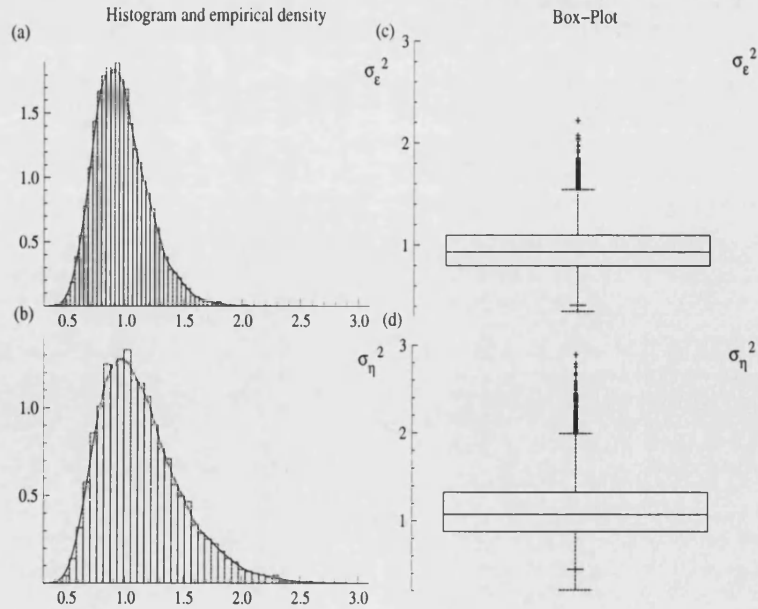
Table 3.3: Summary of output from posterior sample averages, across 10,000 simulated replications, for a local level model.

		Truth	Posterior mean sample					
			Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$	≥ 0.5
	σ_ε^2	1	0.964	0.936	0.227	0.599	1.483	-
	σ_η^2	1	1.134	1.075	0.345	0.631	1.956	-
Outlier	$p_{1,20}$	1	0.891	0.992	0.208	0.193	1	0.924
	$k_{1,20}$	7	6.978	6.999	2.767	4.227	9.778	-
	$p_{1,50}$	1	0.770	0.917	0.293	0.0538	1	0.806
	$k_{1,50}$	-6	-6.091	-5.990	2.808	-8.755	-3.485	-
Level shift	$p_{2,39}$	1	0.750	0.883	0.296	0.0518	1	0.790
	$k_{2,39}$	7	6.868	6.965	3.755	3.762	9.939	-
	$p_{2,74}$	1	0.758	0.889	0.290	0.0588	1	0.801
	$k_{2,74}$	-7	-6.954	-7.000	3.122	-9.976	-3.883	-
	q_1	0.02	0.0242	0.0243	0.0033	0.0175	0.0311	-
	q_2	0.02	0.0249	0.0251	0.0056	0.0141	0.0374	-

The values reported are the mean, standard deviation, 2.5th and 97.5th percentiles of the posterior mean estimates across 10,000 samples. " ≥ 0.5" is the proportion of times the shocks were correctly detected, for a threshold of 0.5, across the 10,000 samples. Based on 10000 Gibbs draws, discarding the first 5,000.

For all the four shocks in the data sets, the first quartile is greater than the threshold of 0.5. For the outlier at $t = 20$ the estimates for the indicator variable are less spread, whereas for the remaining shocks the distributions are similar.

In Figure 3.6, we plot the histogram and box- plots of the posterior sample means obtained for σ_ε^2 and σ_η^2, across the simulated replications. We note that the sample of posterior means for σ_η^2 is slightly right skewed, and has a distribution more spread than the one for σ_ε^2. This is related to the fact, that in average, the posterior sample mean for the indicator variables is higher for the outliers than for the level shifts present in the data sets.

In summary, the results obtained indicate that *sampling scheme 3.2* works efficiently, and overall the parameters of the model are estimated satisfactorily.

3.5.4 Stationary model: AR(1) plus noise

To illustrate the performance of the method proposed for detection of shocks for a stationary time series model, we present a numerical application for the AR(1) plus noise model. An artificial data set, of size $T = 100$ is generated from the model:

$$y_t = \alpha_t + k_{1,t}p_{1,t} + \varepsilon_t,$$
$$\alpha_{t+1} = \phi\alpha_t + +k_{2,t}p_{2,t} + \eta_t,$$

with $\phi = 0.5$ and $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$. An outlier was input at $t = 50$, and a shock to the transition equation at $t = 50$, with sizes -7 and 7, respectively. The purpose of this numerical illustration is to assess how the stationary nature of the model affects the performance of the methodology in terms of shocks diagnostics. For this reason, sampling of the autoregressive parameter is not performed, and it is taken as equal to its true value. With this assumption adapting *sampling scheme 3.2* for the stationary model is straightforward. The Gibbs sampler was run for $200,000$ iterations with the first $100,000$ discarded. In Figure 3.7 we present the posterior sample averages of the intervention variables. The true location of the shocks is detected, with both type of shocks being detected with posterior mean probabilities above 0.7. The posterior averages of the intervention variables for the instances where shocks are detected are:

$$\bar{p}_{1,50} = 0.774 \quad \bar{k}_{1,50} = -6.215,$$
$$\bar{p}_{2,74} = 0.770 \quad \bar{k}_{2,74} = 6.084,$$

A summary of the properties of the posterior samples obtained for the hyperparameters is presented in Table 3.4. We can see that the posterior sample mean estimates for both hyperparameters present values close to the true ones.

Further study of the hyperparameters posterior samples properties is be done by analyzing the plots in Figure 3.8. From Figure 3.8 (a) we see that the hyperparameters present a negative correlation. The cross-correlations have a rapid rate of converge towards zero, as it can be observed from Figure 3.8 (b). The autocorrelations of the posterior sample of the transition equation hyperparameter, presents a slighter slower rate of convergence to zero, as it is seen in Figure 3.8 (c) and (d).

Figure 3.7: Plots of posterior means of indicator and size of intervention variables for outliers (a)-(b) and level shifts (c)-(d), for AR(1) plus noise model.
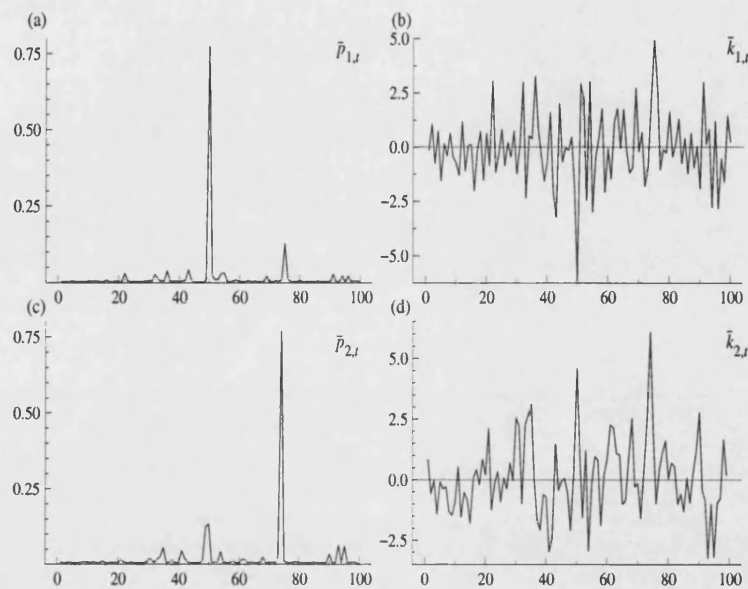
Figure 3.8: Scatter plot of posterior samples for $\left(\sigma_\varepsilon^2, \sigma_\eta^2\right)$ (a), cross correlations (b), autocorrelation functions (c)-(d) for posterior samples of hyperparameters until lag 100, for AR(1) plus noise model.
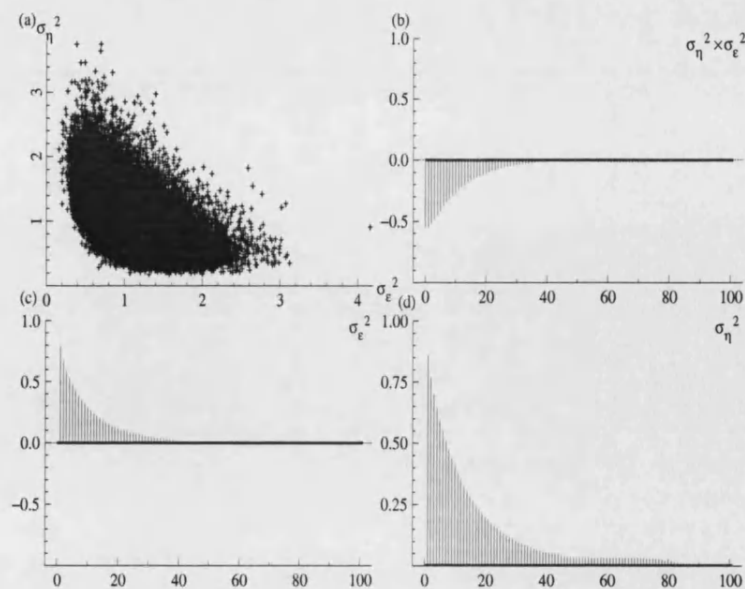
Table 3.4: Summary of output from posterior sample means of hyperparameters, for artificial data set from AR(1) plus noise model.

| | Truth | Posterior mean sample | | | | |
		Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$
σ_ε^2	1	1.117	1.088	0.00645	0.508	1.898
σ_η^2	1	0.956	0.892	0.00776	0.396	1.869

The values reported are the mean, median, 2.5th and 97.5th percentile of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated with a bandwidth of 5,000 for the Parzen window). Based on 200,000 Gibbs draws, discarding the first 100,000.

The results obtained for the application of the sampling scheme proposed allow us to conclude that the method proposed performs satisfactorily when applied to this stationary model.

3.5.5 Empirical application: coal consumption data set

We illustrate the detection of outliers and level shifts using the method in §3.5.2 and apply it to the logarithm of the quarterly coal consumption in the UK, from 1960 quarter 1 to 1986 quarter 4 (note Harvey (1989) uses the data set up to 1983Q4) in total 108 observations. This data set is presented in Figure 3.9. In an

Figure 3.9: Quarterly coal consumption (logarithm), in the UK, from 1960 quarter 1 to 1986 quarter 4.
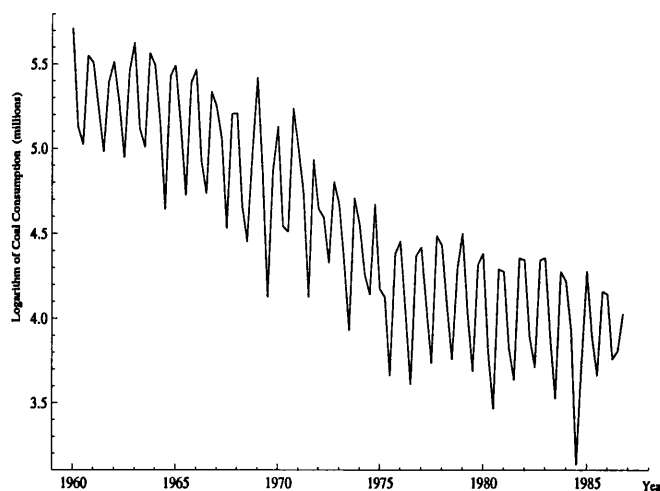
initial analysis of the data, we can see that there is a seasonal pattern, as well as a trend, with a slope component. Also there seems to be some shocks to the data. We started by fitting an unobserved component model with trend, with stochastic slope, and quarterly seasonal dummy component. Our results showed that the slope and seasonal variances were approximately zero. Hence the model we estimated, allowing for the presence of outliers and level shifts is formulated as:

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + k_{1,t} p_{1,t} + \varepsilon_t, \\
\mu_{t+1} &= \mu_t + \beta_t + k_{2,t} p_{2,t} + \eta_t, \\
\beta_{t+1} &= \beta_t, \\
\gamma_{t+1} &= -\gamma_t - \gamma_{t-1} - \gamma_{t-2}.
\end{aligned}
$$

Although *sampling scheme 3.2* was presented for the case where there is only a level component, it is easily extended to this case as the additional components are non-stochastic, and therefore do not imply any additional parameter estimation. The type of shocks considered are outliers and level shifts.

The prior distributions used are:

$$
\begin{aligned}
\sigma_\varepsilon^2 &\sim IG\left(\frac{5}{2}, \frac{0.05}{2}\right); \\
\sigma_\eta^2 &\sim IG\left(\frac{5}{2}, \frac{0.01}{2}\right); \\
k_{i,t} &\propto 1, \text{ for } i = 1, 2, \ t = 1, \ldots, T; \\
q_1 &\sim Beta\,(8, 100); \\
q_2 &\sim Beta\,(2, 100).
\end{aligned}
$$

The parameters u_i and v_i are taken to be -0.5 and 0.5, respectively, for $i = 1, 2$.

The Gibbs sampler was run for 100,000 iterations, with the first 50,000 discarded. With less runs the results obtained do not differ substantially from the ones we present. However, as the algorithm with this number of iterations took only 157 seconds to run, we have opted for this high number of iterations. The Ox program implemented is presented in Appendix A.

Figure 3.10 is a plot of the posterior sample mean of probability of an outlier and a level shift, and the posterior sample average size of a shock. There are

Figure 3.10: Plots of posterior means of indicator and size of intervention variables for outliers (a)-(b) and level shifts (c)-(d), for the coal data set, from 1960Q1 to 1986Q4.
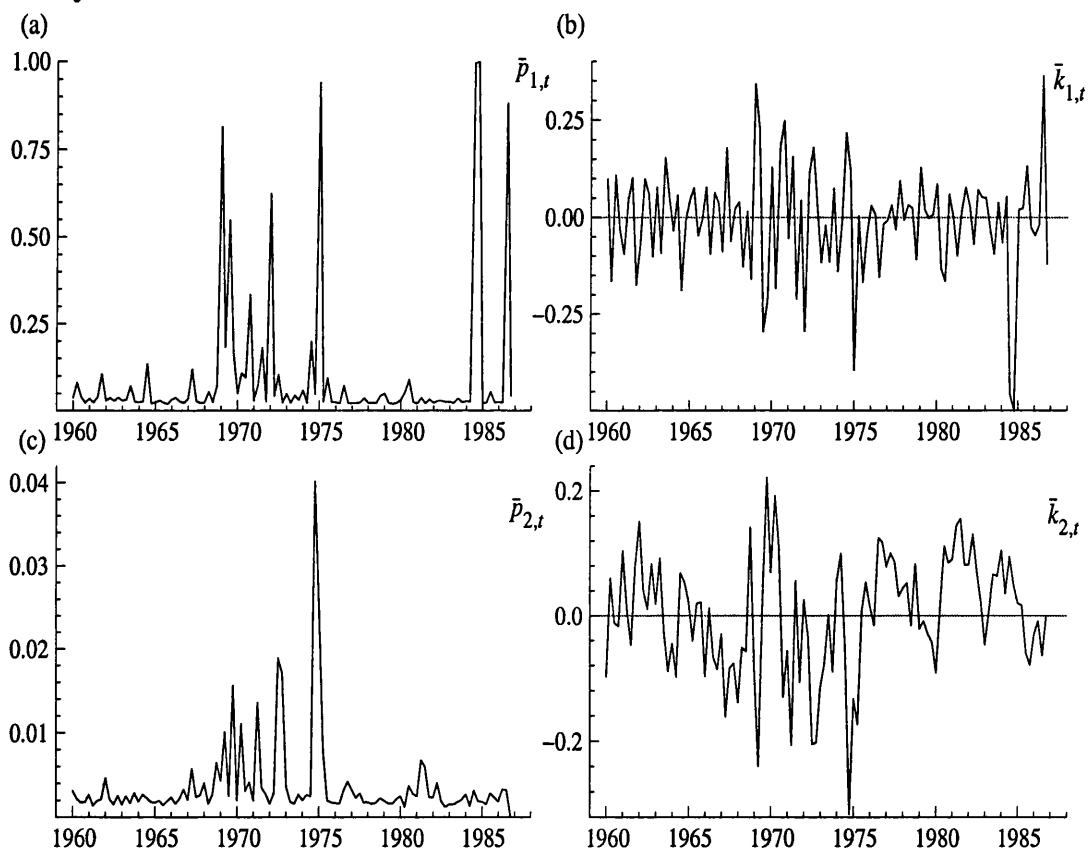
seven points with outlying mean posterior probability of being an outlier. For all these observations the posterior probability presents a value greater than 0.5. and are therefore classified as outliers. No level shift is detected. The highest posterior probability obtained for this type of shock occurs in 1974Q4, with a posterior average probability of 0.04, and would correspond to a level shift in 1975Q1. An outlier is detected in 1975Q1, with a mean posterior probability of 0.94.

The average size of a shock is obtained as the weighted average of the posterior sample of the size variable, the weights being the indicator variable. This amounts to including for the calculation of the mean size, only the sampled values for which the corresponding indicator variable is non zero. Using this method of calculation for estimating the size variable, the ergodic result for the Gibbs sampler in §2.2.3 still holds. If convergence has been achieved, we are using a subset of draws from the posterior distribution of this variable to calculate the size estimate. In Table 3.5, the shock detection results are reported.

In Atkinson, Koopman, and Shephard (1997), for the subset of observations until 1983Q4, using a non-Bayesian method, two outliers and a level shift are identified. The outliers in 1969 quarters 1 and 3 and the level shift in 1974Q4. The level shift corresponds to the the peak we noted in Figure 3.10, with an average posterior probability of 0.04. To consider the possibility of the existence of a level shift at that point, that is being masked by the outlier detected in 1975Q1 (observation 61), we run the sampler with the restriction of $p_{1,61} = 0$. The results obtained point in the direction of a level shift at that instant, but with a posterior probability much lower than the value reported in Table 3.5 for the presence of an outlier in observation 61. Overall, we conclude there is a stronger evidence of an outlier in 1975Q1, than a level shift at 1974Q4. The results presented at the end of this empirical study support these findings. Finally, we note that Atkinson, Koopman, and Shephard (1997) mention the possibility of outliers around 1975Q1. In Table 3.6 we summarize the results obtained for the hyperparamters and prior probabilities of shocks.

Although the posterior samples, for any of the parameters of the model, are not uncorrelated, if convergence as been achieved, the autocorrelation function should quickly converge towards zero (see §2.2.4). In Figure 3.11 we plot the autocorrelation

Table 3.5: Posterior average probability and size of outliers detected for the coal data set, from 1960Q1 to 1984Q4.

		Posterior sample mean						
	Year-Quarter	1969-1	1969-3	1972-1	1975-1	1984-3	1984-4	1986-3
Outlier	Probability	0.815	0.550	0.623	0.940	0.995	0.999	0.881
	Size	0.342	-0.295	-0.294	-0.395	-0.458	-0.498	0.362

The values reported are the posterior mean of the probability and size of the shocks detected, using a threshold of 0.5 for the indicator variable. Based on 100,000 Gibbs draws, discarding the first 50,000.

Table 3.6: Summary of output from posterior samples of the hyperparameters, and prior probabilities of an outlier and a level shift, for the coal data set, from 1960Q1 to 1984Q4.

	Prior		Posterior sample			
	Mean	SD	Mean	SD	$Q_{0.025}$	$Q_{0.975}$
σ_ε^2	0.017	0.024	0.00723	3.6e-05	0.00441	0.01113
σ_η^2	0.0033	0.0047	0.00191	2.2e-05	0.00085	0.00384
q_1	0.074	0.025	0.0862	2.4e-04	0.0461	0.1370
q_2	0.0196	0.014	0.0067	1.6e-04	0.0002	0.0245

The values reported are the mean, 2.5th and 97.5th percentile of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated with a bandwidth of 5,000 for the Parzen window). Based on 100,000 Gibbs draws, discarding the first 50,000.

functions for σ_ε^2 and σ_η^2 until lag 500. For both parameters, the behaviour of the

Figure 3.11: Autocorrelation functions for posterior samples for σ_ε^2 and σ_η^2, for 50 lags, for the coal data set, from 1960Q1 to 1986Q4.
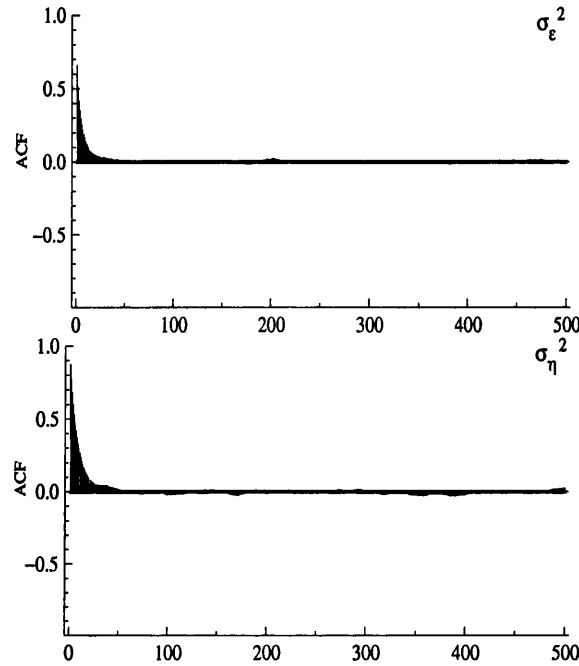
autocorrelation functions does not contradict the assumption of having achieved convergence to the posterior distribution.

A more detailed study of the statistical properties of the posterior samples of the size of intervention variables is now presented. The next figures are based on the full posterior samples of the size of an outlier, for the instances where this type of shock was detected. In Figure 3.12 the scatter plots of the size of outliers are presented. From these plots there is no apparent relation between the sizes of shocks, for the different instances in time, across the iterations recorded.

In Figure 3.13, we present the autocorrelation functions for the size of the outliers detected. It is visible that the convergence towards zero is much slower for those instances where the posterior probability of having a shock is higher.

As means of checking the quality of the model fitted, we use the methodology

Figure 3.12: Scatter plots of posterior samples for size of intervention variables of outliers detected, for the coal data set, from 1960Q1 to 1986Q4.
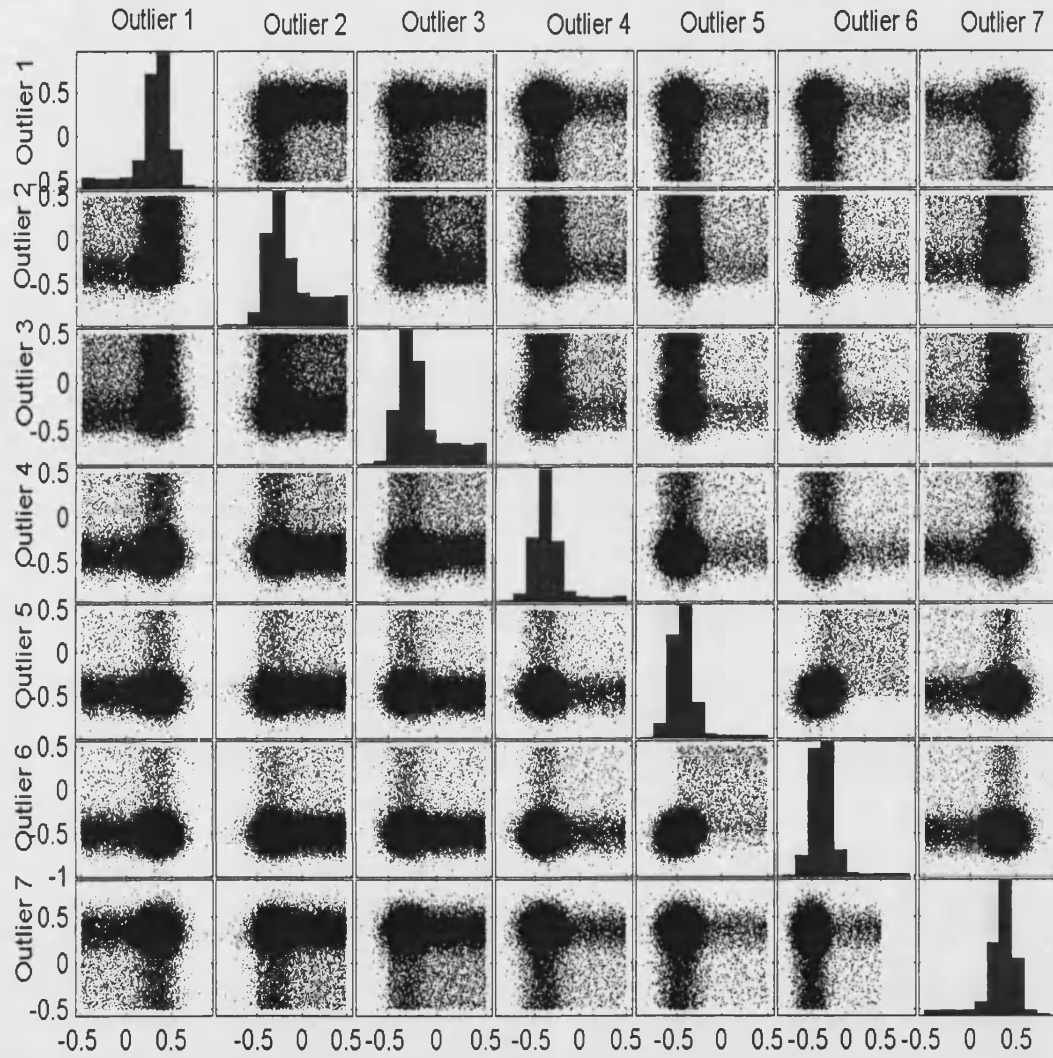
Outlier 1: 1969-1, Outlier 2: 1969-3, Outlier 3: 1972-1, Outlier 4: 1975-1, Outlier 5: 1984-3, Outlier 6: 1984-4, Outlier 7: 1986-3.

Figure 3.13: Autocorrelation functions for posterior samples for size of intervention variables of outliers detected , for 500 lags, for the coal data set, from 1960Q1 to 1986Q4.
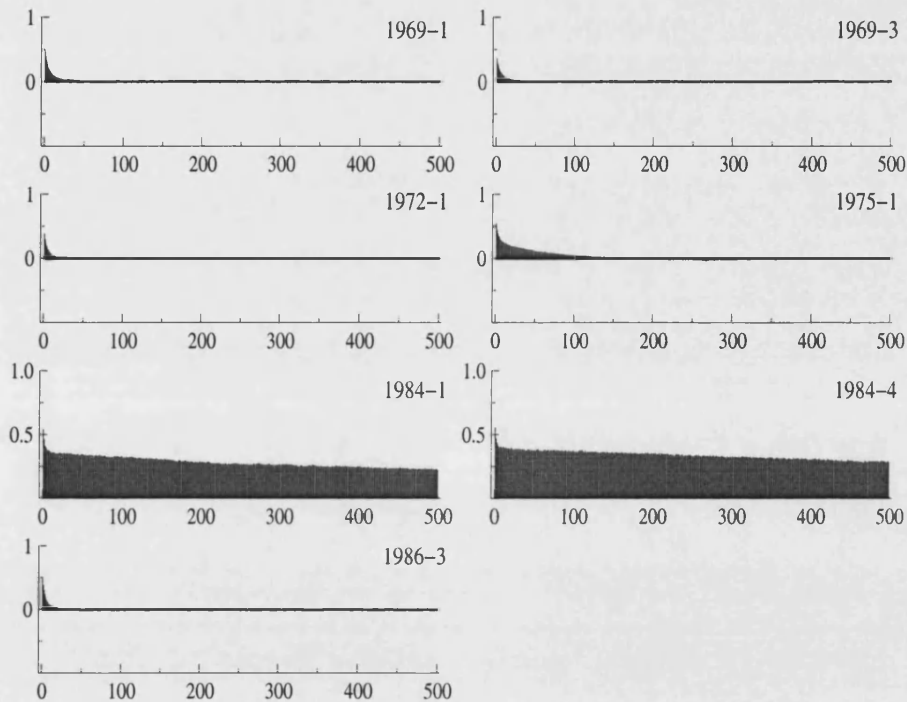
presented in Harvey and Koopman (1992), and discussed in more detail in §2.3. It is based on the analysis of the innovations and auxiliary residuals. The innovations are the one step-ahead prediction error, and are obtained from running the Kalman filter (see §2.3, equations (2.17) to (2.21)):

$$v_t = \boldsymbol{E}\left(\varepsilon_t | Y_{t-1}\right),$$
$$F_t = \operatorname{var}\left(v_t\right).$$

If the model is well specified they should be uncorrelated and normally distributed.

The auxiliary residuals are useful to detect features of the data, such as outliers and structural shifts, that the original model did not account for. They are smoothed estimates of the disturbances of the unobserved components. For a BSM, we can obtain the auxiliary residuals corresponding to the irregular, level, slope and seasonal components. They are obtained from running the Kalman filter and smoother. For

94

the level component the auxiliary residual is:

$$\hat{\eta}_t = \boldsymbol{E}\left(\eta_t | Y\right).$$

Using the outliers detected, with the estimated sizes, as reported in Table 3.5, and the estimated hyperparameters in Table 3.6, we run the Kalman filter and smoother and obtain the standardized innovations, and irregular and level standardized auxiliary residuals: $v_t/\sqrt{F_t}$, $\hat{\varepsilon}_t/\sigma_{\hat{\varepsilon}_t}$ and $\hat{\eta}/\sigma_{\hat{\eta}_t}$, for $t = 1, \ldots, T$ (see §2.3).

Figure 3.14: Descriptive plots of standardized innovations, for the coal data set, from 1960Q1 to 1986Q4.
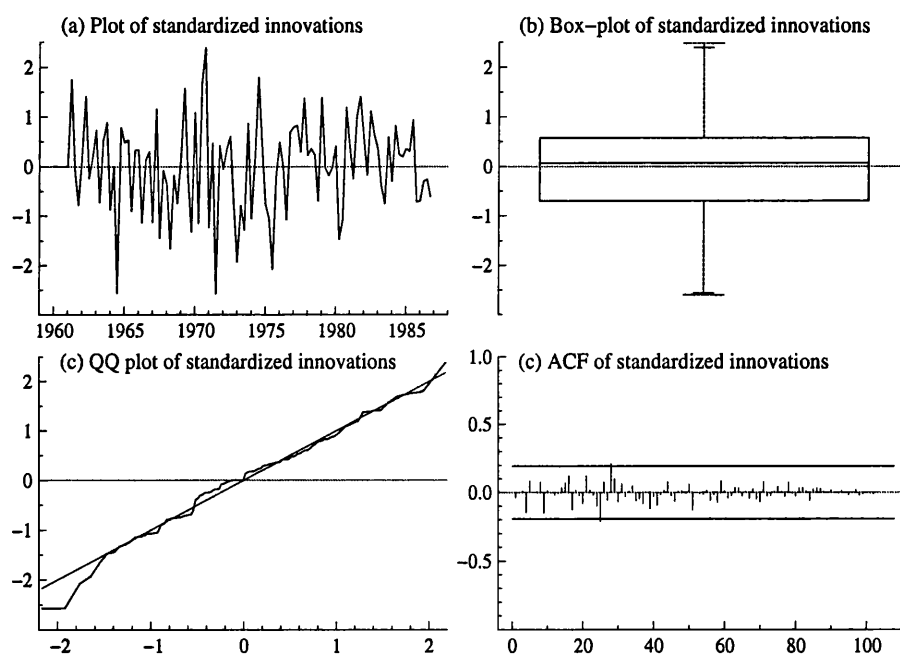
In Figure 3.14 we present plots for the standardized innovations. The standardized innovations present absolute values less than 2.6, so there is no outstanding value. This observation is corroborated from the inspection of the plot and box-plot for this series, in Figure 3.14 (a) and (b), respectively. From analysis of the plots in Figure 3.14 (c) and (d), there is no indication of departure from the hypothesis that the standardized innovations are distributed as a standard normal and present no serial correlation. In summary, from the analysis of the standardized innovations, there is no indication of misspecification of the model or undetected shocks.

Figure 3.15: Plots (a)-(b) and box-plots (c)-(d) of standardized auxiliary residuals, for the coal data set, from 1960Q1 to 1986Q4.
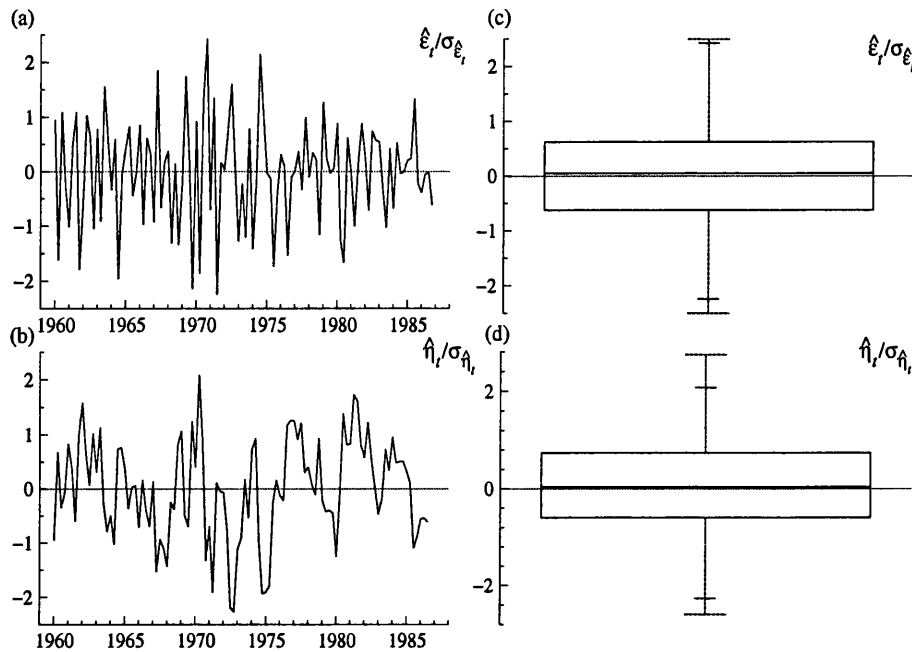
In Figure 3.15 we present plots for the standardized auxiliary residuals. The analysis of these residuals is done with the purpose of detecting shocks that were unaccounted for, and were not identified through the analysis of the standardized innovations. Following the exposition in §2.3, standardized auxiliary residuals in absolute value greater than 3.5, would indicate the presence of shocks unaccounted by the model. The standardized auxiliary residuals for the irregular and level component present absolute values bellow 2.5, which indicates that all the shocks present in the data set were corrected detected in terms of their position and estimated size.

3.6 Conclusions

We propose a Gibbs sampler algorithm for estimation of SSM in the presence of outliers and level shifts, where the prior for the intervention size variable is an uninformative flat distribution. This assumption is less restrictive than other continuous priors in the literature. To use the uninformative prior, the parameters of an approx-

imating bounded uniform have to be defined *a priori*. We show that the posterior samples of the variables of the model present a low sensitivity to their choice. On the other hand, it does not have the constrains of considering a discrete prior, where the set of values of the intervention variable size have to be chosen *a priori*. The Gibbs sampler presented converges to the target distribution as long as both the measurement and transition variances are different from zero. This restriction is often overlooked in the literature and we shall deal with it in Chapter 5.

Several numerical illustrations are given, in order to describe and assess the efficiency of the method. The flat and normal prior assumptions were compared, in terms of the sensitivity of the estimates to the choice of the prior parameters. The results obtained show that the parameter estimates are much less sensitive to the choice of the uniform distribution. In particular, the variable representing the size of an outlier, presents very consistent estimates across different values for the parameters of the uniform prior. This represents a key advantage of the use of a flat prior in relation to a normal prior.

The sampler we propose for detection of outliers and level shifts was applied in a Monte Carlo study, where we considered artificial data sets with two outliers and two level shifts. The average estimates of the hyperparameters, obtained across the simulated replications, give accurate estimates of the their true values. The shocks input to the data were detected in most replications, with estimated sizes in average very close to the true ones. The application of the sampler is also exemplified for shocks diagnostics in the context of stationary models, namely an AR(1) plus noise model.

Finally, the methodology proposed was applied to a real data set, consisting of the logarithm of quarterly coal consumption in the UK (in millions) from 1960Q1 to 1986Q4. A BSM was fit, with non-stochastic slope and dummy seasonal component, allowing for the existence of outliers and level shifts. Seven outliers were identified and no level shift. We conclude that the model fit the data satisfactory, and the shocks were correctly identified.

Chapter 4

Estimation of unobserved component models in the presence of outliers and structural shifts

4.1 Introduction

Unobserved component models (UCM), (Harvey, 1989) are a special case of state space models. The temporal dynamics of the time series are described as the sum of different effects (components) with stochastic or deterministic evolution over time. The basic structural model (BSM), is an unobserved component model with irregular, trend and dummy seasonal component, formulated as:

$$
\begin{aligned}
y_t &= \mu_t + \gamma_t + \varepsilon_t, \\
\mu_{t+1} &= \mu_t + \beta_t + \eta_t, \\
\beta_{t+1} &= \beta_t + \zeta_t, \\
\sum_{i=0}^{s-1} \gamma_{t+1-i} &= \omega_t,
\end{aligned}
\tag{4.1}
$$

with the innovation processes defined as Gaussian white noise, and the assumption of independence between the different components.

We consider two types of shocks to the time series process $\{y_t\}_{t=1}^T$; outliers and structural shifts. An outlier at t^* is generated by a shock to the irregular component equation at that point in time, affecting only observation y_{t^*}. A structural shift at t', is generated by a shock to any of the components level, slope or seasonal, and it will affect the observed series for all instances $t \geq t' + 1$, as a consequence of the non-stationarity of these equations. Shocks to the observations which are not accounted

for in the original model will result in biased estimates of the model variances.

In Chapter 3 several works on the detection of shocks using the state space model framework, were cited. Examples of methods for detecting outliers and structural shifts in the specific context of the BSM, using non-Bayesian techniques, are Atkinson, Koopman, and Shephard (1997) and Penzer (1998), for the detection of seasonal shifts. In a Bayesian context, for BSM, in West and Harrison (1997), the shocks to the data are included in the model by feed-forward intervention. Anticipating an abnormal event at instant t^*, the model specifications for that point in time is changed. Several ways of changing the model specification for the instant when an abnormal event is expected are proposed. Examples are, inclusion of an intervention variable, with a normal prior distribution and parameters set by the user; modification of the prior specifications of the model. An example of a method for modeling seasonal shifts, is presented in Franses, Hoek, and Paap (1997), for unit roots processes and using a non-stochastic dummy seasonal component.

We propose a Bayesian method for estimating the BSM, with the inclusion of intervention variables, to account for the presence of general transition shocks in the data. This method is a generalization of the methodology presented in Chapter 3, where intervention models were consider for the detection of outliers and level shifts.

Our main contribution is the introduction of a continuous distribution for the magnitude of the intervention variables; namely a uniform distribution. Additionally, we consider a BSM with interventions on all the components, estimated using the Gibbs sampler, which to our knowledge has not been considered in the literature. By running the Gibbs sampler, we obtain posterior samples for the hyperparameters and intervention variables. We obtain estimates for the four hyperparameters and simultaneously detect the position and estimate the magnitude of the outliers and structural shift that might be present in the data.

We start by describing the formulation of the intervention model, the set of parameters to be estimated, and their prior distributions. Next we derive the Gibbs sampler scheme used for sampling from the posterior distribution of the parameters. To establish the effectiveness of the method we present two numerical illustrations.

Firstly, a Monte Carlo experiment. The data sets are generated from a BSM; the four types of shocks are added to the data: outlier, level, slope and seasonal shifts. Secondly, the methodology proposed is applied to a real data set: the quarterly figures of marriages in the UK, from 1958 quarter 1 to 1982 quarter 4.

4.2 Unobserved components model with intervention variables

Following the formulation in Chapter 3, we include interventions in the model formulated by equations (4.1), through an additive term $k_{i,t}p_{i,t}$, for t=1,...,T, and $i = 1,2,3,4$, corresponding to the irregular, level, slope and seasonal component equations, respectively.

$$y_t = \mu_t + \gamma_t + k_{1,t}p_{1,t} + \varepsilon_t \qquad (4.2)$$

$$\mu_{t+1} = \mu_t + \beta_t + k_{2,t}p_{2,t} + \eta_t \qquad (4.3)$$

$$\beta_{t+1} = \beta_t + k_{3,t}p_{3,t} + \zeta_t \qquad (4.4)$$

$$\sum_{i=0}^{s-1} \gamma_{t+1-i} = k_{4,t}p_{4,t} + \omega_t, \qquad (4.5)$$

with diffuse initial conditions:

$$(\mu_1, \beta_1, \gamma_1, \ldots, \gamma_{-s+3})' \sim N(0, I_s), \quad \kappa \to \infty . \qquad (4.6)$$

The noise processes are mutually independent and normal distributed. For all t,

$$\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right),$$
$$\eta_t \sim N\left(0, \sigma_\eta^2\right),$$
$$\zeta_t \sim N\left(0, \sigma_\zeta^2\right),$$
$$\omega_t \sim N\left(0, \sigma_\omega^2\right).$$

There is a shock to component i at instant t^*, with size k_{i,t^*}, if $p_{i,t^*} = 1$. If $i = 1$ and $p_{1,t^*} = 1$, the shock is an outlier; it has a direct and instant impact on y_t. For $i = 2,3,4$, a shock to any of the components implies a structural shift. Consider for example $i = 4$. There is a seasonal shift at $t^* + 1$ of size k_{4,t^*} if $p_{4,t^*} = 1$. The term

$p_{4,t^*} k_{4,t^*}$ implies a shift in the seasonal component at time $t^* + 1$, which will persist for all the following t and will affect y_t, for $t \geq t^* + 1$.

Some notation is necessary at this point to simplify the explanation of the methodology that follows. Let x_t represent any of the time indexed variables in the model and define $\mathbf{x} = (x_1, \ldots, x_T)$. With this notation we define:

$$\boldsymbol{\alpha} = (\mu, \beta, \gamma)', \tag{4.7}$$

$$\mathbf{k} = (\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)', \tag{4.8}$$

$$\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4)'. \tag{4.9}$$

Using a Bayesian approach, our aim is to sample from the joint posterior distribution of the variables in expressions (4.7) to (4.9). The priors are chosen using an approach very similar to that given in Chapter 3. The prior assumptions for the level, slope and seasonal state variables are defined by the diffuse initial conditions in (4.6). For the hyperpameters, we assume an inverse gamma prior distribution:

$$\sigma_\varepsilon^2 \sim IG\left(\frac{c_1}{2}, \frac{s_1}{2}\right), \tag{4.10}$$

$$\sigma_\eta^2 \sim IG\left(\frac{c_2}{2}, \frac{s_2}{2}\right), \tag{4.11}$$

$$\sigma_\zeta^2 \sim IG\left(\frac{c_3}{2}, \frac{s_3}{2}\right), \tag{4.12}$$

$$\sigma_\omega^2 \sim IG\left(\frac{c_4}{2}, \frac{s_4}{2}\right). \tag{4.13}$$

$$\tag{4.14}$$

The size of intervention variables are assumed to be *a priori* mutually and serially independent with a flat prior distributions. For $t = 1, \ldots, T$ and $i = 1, \ldots, 4$

$$k_{i,t} \propto 1. \tag{4.15}$$

The indicator variables are also assumed to be *a priori* mutually and serially independent, distributed according to a prior Bernoulli distribution. For $t = 1, \ldots, T$ and $i = 1, \ldots, 4$

$$p_{i,t} \sim Bernoulli\,(q_i), \tag{4.16}$$

with $q_i = P\,[p_{i,t} = 1]$.

The prior probabilities of having any of the different type of interventions are sampled together with the other parameters of the model. The prior assumptions for this set of variables are:

$$q_i \sim Beta\left(a_i, b_i\right), \tag{4.17}$$

for $i = 1, \ldots, 4$.

For the vector of parameters not time dependent, the hyperparameters and prior probabilities, we use the following notation:

$$\sigma^2 = \left(\sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2\right),$$
$$q = \left(q_1, q_2, q_3, q_4\right).$$

We assume that intervention variables, for different components, are independent between them. With the previous assumptions and definitions, the estimation of the intervention model will be accomplished via a Bayesian method, by obtaining samples from the posterior distribution:

$$f\left(\boldsymbol{\alpha}, \sigma^2, \mathbf{k}, \mathbf{p}, q | Y\right), \tag{4.18}$$

given the prior assumptions in expressions (4.6) and (4.10) to (4.17), for the set of *a priori* fixed parameters $\{c_i, s_i, u_i, v_i, a_i, b_i, \text{ for } i = 1, 2, 3, 4\}$. We use the Gibbs sampler to draw from the target posterior distribution. Iteratively draws are obtained from the full conditional distributions:

$$\boldsymbol{\alpha} \sim f\left(\boldsymbol{\alpha} | Y, \sigma^2, \mathbf{k}, \mathbf{p}, q\right), \tag{4.19}$$

$$\sigma^2 \sim f\left(\sigma^2 | Y, \boldsymbol{\alpha}, \mathbf{k}, \mathbf{p}, q\right), \tag{4.20}$$

$$\mathbf{k} \sim f\left(\mathbf{k} | Y, \boldsymbol{\alpha}, \sigma^2, \mathbf{p}, q\right), \tag{4.21}$$

$$\mathbf{p} \sim f\left(\mathbf{p} | Y, \boldsymbol{\alpha}, \sigma^2, \mathbf{k}, q\right), \tag{4.22}$$

$$q \sim f\left(q | Y, \boldsymbol{\alpha}, \sigma^2, \mathbf{k}, \mathbf{p}\right), \tag{4.23}$$

$$\tag{4.24}$$

After convergence is achieved this algorithm will generate samples from the distribution in (4.18). The method for sampling from the several full conditionals is

a generalization of the techniques presented in Chapter 3. Conditional on the intervention variables, the state space model defined by equations (4.2) to (4.5) is Gaussian. Sampling from the full conditional distribution of the states in (4.19) is done using the simulation smoother in De Jong and Shephard (1995).

The full conditional distributions for the variance parameters are inverse gammas, a result obtained by applying Lemma 2.4.1. For example, consider the seasonal component variance,

$$
\begin{aligned}
f\left(\sigma_\omega^2|Y, \boldsymbol{\alpha}, \mathbf{k}, \mathbf{p}, q\right) &= f\left(\sigma_\omega^2|\boldsymbol{\gamma}, \mathbf{k_4}, \mathbf{p_4}\right) \\
&\propto f\left(\boldsymbol{\gamma}|\sigma_\omega^2, \mathbf{k_4}, \mathbf{p_4}\right) f\left(\sigma_\omega^2\right).
\end{aligned}
\tag{4.25}
$$

On the other hand the first conditional density on the right hand side of (4.25), satisfies

$$
f\left(\boldsymbol{\gamma}|\sigma_\omega^2, \mathbf{k_4}, \mathbf{p_4}\right) \propto \prod_{t=s}^{T} f\left(\gamma_t|\gamma_{t-1}, \ldots, \gamma_{t-s+1}, k_{4,t-1}, p_{4,t-1}, \sigma_\omega^2\right) f\left(\sigma_\omega^2\right).
$$

Using the same type of derivation used in Lemma 2.4.1 , we obtain

$$
\left(\sigma_\omega^2|\boldsymbol{\gamma}, \mathbf{k_4}, \mathbf{p_4}\right) \sim IG\left(\frac{c_4 + T - s + 1}{2}, \frac{s_4 + \sum_{t=s}^{T}(\gamma_t - \ldots - \gamma_{t-s+1} - k_{4,t-1}p_{4,t-1})^2}{2}\right)
\tag{4.26}
$$

As for the intervention variables, size, indicator, and probability of intervention, we describe how to sample from their full conditional for the case of the seasonal component. Given that the prior assumptions for the interventions in different components are of the same nature, the remaining full conditional distributions are obtained in a similar way. We start by noticing that, given the mutual independence of the noise processes, and the mutual prior independence between the different type of interventions, we have

$$
f\left(\mathbf{k_4}|Y, \boldsymbol{\alpha}, \sigma^2, \mathbf{k_1}, \mathbf{k_2}, \mathbf{k_3}, \mathbf{p}, q\right) = f\left(\mathbf{k_4}|\boldsymbol{\gamma}, \sigma_\omega^2, \mathbf{p_4}, q_4\right).
$$

On the other hand, the size of the seasonal shifts are mutually independent. This property together with the dynamics of the seasonal component equation implies that

$$
f\left(\mathbf{k_4}|\boldsymbol{\gamma}, \sigma_\omega^2, \mathbf{p_4}, q_4\right) = \prod_{t=1}^{T} f\left(k_{4,t}|\gamma_{t+1}, \gamma_t, \ldots, \gamma_{t-s+2}, , \sigma_\omega^2, p_{4,t}\right).
\tag{4.27}
$$

Therefore, sampling from the full conditional of $\mathbf{k}_4$ is done by sampling, for $t = 1, \ldots, T$ from the distributions on the right hand side of expression (4.27). With the flat prior assumption for $k_{4,t}$, for $t = s - 1, \ldots, T - 1$, the full prior conditional becomes

$$
f\left(k_{4,t}|\gamma_{t+1}, \gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, p_{4,t}\right) \propto
$$
$$
f\left(\gamma_{t+1}|\gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, k_{4,t}, p_{4,t}\right). \tag{4.28}
$$

If $p_{4,t} = 1$, then,

$$
k_{4,t}|\gamma_{t+1}, \gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, p_{4,t} \sim N\left(\gamma_{t+1} - \gamma_t - \ldots - \gamma_{t-s+2}, \sigma_\omega^2\right).
$$

If $p_{4,t} = 0$, the full conditional for the size variable is an improper flat distribution. Following the exposed in Chapter 3, sampling from the full conditional for $k_{4,t}$ is achieved by approximating the improper prior by a bounded uniform distribution.

$$
k_{4,t}|\gamma_{t+1}, \gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, p_{4,t} \sim U\left[u_4, v_4\right]
$$

With the range of seasonal components simulated available at each iteration of the sampler, γ_t for $t = 1, \ldots, T$, the expression in (4.28), can be evaluated only for $t = s - 1, \ldots, T - 1$. We assume that there is not enough information prior to $t = s - 1$, to detect seasonal shifts. Therefore, we rule out those type of interventions by setting the corresponding indicators to zero. At the other extreme $p_{4,T}$, indicates whether there is a seasonal shift at instant $T + 1$, which is outside our range of observations. For this reason, it is set fixed to zero.

Using a similar argument, for the full conditional distribution of $\mathbf{p}_4$, we have that

$$
f\left(\mathbf{p}_4|\boldsymbol{\gamma}, \sigma_\omega^2, \mathbf{k}_4, q_4\right) = \prod_{t=1}^{T} f\left(p_{4,t}|\gamma_{t+1}, \gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, k_{4,t}, q_4\right). \tag{4.29}
$$

On the other hand, for each of the densities on the right hand side of expression (4.29), we have

$$
f\left(p_{4,t}|\gamma_{t+1}, \gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, k_{4,t}, q_4\right) \propto
$$
$$
f\left(\gamma_{t+1}|\gamma_t, \ldots, \gamma_{t-s+2}, \sigma_\omega^2, k_{4,t}, p_{4,t}, q_4\right) f\left(p_{4,t}|q_4\right). \tag{4.30}
$$

Given the prior Bernoulli assumed for the indicator variables, for $t = s-1, \ldots, T-1$ the distributions on the left hand side of expression (4.30), are also going to be Bernoulli distributions with probabilities given by:

$$P\left[p_{4,t} = 1\right] \propto f_N\left(\gamma_{t+1}|\gamma_t + \ldots + \gamma_{t-s+2} + k_{4,t}; \sigma_\omega^2\right) q_4,$$

$$P\left[p_{4,t} = 0\right] \propto f_N\left(\gamma_{t+1}|\gamma_t + \ldots + \gamma_{t-s+2}; \sigma_\omega^2\right) (1 - q_4).$$

For the prior probability of having a seasonal shift, q_4 we have

$$f\left(q_4|Y, \alpha, \sigma^2, \mathbf{k}, \mathbf{p}, q_1, q_2, q_3\right) \propto f\left(q_4|\mathbf{p}_4\right),$$

and therefore we sample from a beta distribution

$$q_4|\mathbf{p}_4 \sim Beta\left(a_4 + \sum_{t=1}^T p_{4,t}, b_4 + T - \sum_{t=1}^T p_{4,t}\right).$$

We now present the Gibbs sampler scheme used from sampling from the joint posterior distribution of the parameters.

Sampling scheme 4.1 Suppose we have run i iterations of the Gibbs sampler, initialized with the vector $\left(\mu_1, \beta_1, \gamma_{3-s}, \ldots, \gamma_1, \sigma^{(0)}, \mathbf{k}^{(0)}, \mathbf{p}^{(0)}, q^{(0)}\right)$. After i iterations the current vector is $\left(\alpha^{(i)}, \sigma^{2(i)}, \mathbf{k}^{(i)}, \mathbf{p}^{(i)}, q^{(i)}\right)$. Then, on iteration $i+1$ the sampled vector is updated according to the following scheme :

1. Sample $\alpha^{(i+1)}|Y, \sigma^{(i)}, \mathbf{k}^{(i)}, \mathbf{p}^{(i)}, q^{(i)}$ using the simulation smoother in De Jong and Shephard (1995).

2. Sample $\sigma^{2(i+1)}|Y, \alpha^{(i+1)}, \mathbf{k}^{(i)}, \mathbf{p}^{(i)}, q^{(i)}$ by drawing from the distributions:

$$\sigma_\varepsilon^{2(i+1)}|Y, \alpha^{(i+1)}, \mathbf{k}_1^{(i)}, \mathbf{p}_1^{(i)} \sim IG\left(\frac{c_1 + T}{2}, \frac{s_1 + \tilde{s}_1}{2}\right),$$

with $\tilde{s}_1 = \sum_{t=1}^T \left(y_t - \mu_t^{(i+1)} - \gamma_t^{(i+1)} - k_{1,t}^{(i)}p_{1,t}^{(i)}\right)^2$;

$$\sigma_\eta^{2(i+1)}|\mu^{(i+1)}, \beta^{(i+1)}, \mathbf{k}_2^{(i)}, \mathbf{p}_2^{(i)} \sim IG\left(\frac{c_2 + T - 1}{2}, \frac{s_2 + \tilde{s}_2}{2}\right),$$

with $\tilde{s}_2 = \sum_{t=1}^{T-1} \left(\mu_{t+1}^{(i+1)} - \mu_t^{(i+1)} - \beta_t^{(i+1)} - k_{2,t}^{(i)}p_{2,t}^{(i)}\right)^2$;

$$\sigma_\zeta^{2(i+1)}|\beta^{(i+1)}, \mathbf{k}_3^{(i)}, \mathbf{p}_3^{(i)} \sim IG\left(\frac{c_3 + T - 1}{2}, \frac{s_3 + \tilde{s}_3}{2}\right),$$

with $\tilde{s}_3 = \sum\limits_{t=1}^{T-1} \left(\beta_{t+1}^{(i+1)} - \beta_t^{(i+1)} - k_{3,t}^{(i)} p_{3,t}^{(i)} \right)^2$;

$$\sigma_\omega^{2(i+1)} | \gamma^{(i+1)}, \mathbf{k}_4^{(i)}, \mathbf{p}_4^{(i)} \sim IG \left(\frac{c_4 + T - s + 1}{2}, \frac{s_4 + \tilde{s}_4}{2} \right).$$

with $\tilde{s}_4 = \sum\limits_{t=s-1}^{T-1} \left(\gamma_{t+1}^{(i+1)} - \ldots - \gamma_{t-s+2}^{(i+1)} - k_{4,t}^{(i)} p_{4,t}^{(i)} \right)^2$.

3. Sample $\mathbf{k}^{(i+1)} | Y, \boldsymbol{\alpha}^{(i+1)}, \sigma^{2(i+1)}, \mathbf{p}^{(i)}, q^{(i)}$ by drawing from the distributions:

3.1 For $t = 1, \ldots, T$, if $p_{1,t}^{(i)} = 1$ then

$$k_{1,t}^{(i+1)} | y_t, \mu_t^{(i+1)}, \gamma_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, p_{1,t}^{(i)} \sim N \left(y_t - \mu_t^{(i+1)} - \gamma_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)} \right);$$

If $p_{1,t}^{(i)} = 0$ then

$$k_{1,t}^{(i+1)} | y_t, \mu_t^{(i+1)}, \gamma_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, p_{1,t}^{(i)} \sim U \left[u_1, v_1 \right].$$

3.2 For $t = 1, \ldots, T - 1$, if $p_{2,t}^{(i)} = 1$ then

$$k_{2,t}^{(i+1)} | \mu_{t+1}^{(i+1)}, \mu_t^{(i+1)}, \beta_t^{(i+1)}, \sigma_\eta^{2(i+1)}, p_{2,t}^{(i)} \sim N \left(\mu_{t+1}^{(i+1)} - \mu_t^{(i+1)} - \beta_t^{(i+1)}, \sigma_\eta^{2(i+1)} \right);$$

If $p_{2,t}^{(i)} = 0$ then

$$k_{2,t}^{(i+1)} | \mu_{t+1}^{(i+1)}, \mu_t^{(i+1)}, \beta_t^{(i+1)}, \sigma_\eta^{2(i+1)}, p_{2,t}^{(i)} \sim U \left[u_2, v_2 \right].$$

3.3 For $t = 1, \ldots, T - 1$, if $p_{3,t}^{(i)} = 1$ then

$$k_{3,t}^{(i+1)} | \beta_{t+1}^{(i+1)}, \beta_t^{(i+1)}, \sigma_\zeta^{2(i+1)}, p_{3,t}^{(i)} \sim N \left(\beta_{t+1}^{(i+1)} - \beta_t^{(i+1)}, \sigma_\zeta^{2(i+1)} \right);$$

If $p_{3,t}^{(i)} = 0$ then

$$k_{3,t}^{(i+1)} | \beta_{t+1}^{(i+1)}, \beta_t^{(i+1)}, \sigma_\zeta^{2(i+1)}, p_{3,t}^{(i)} \sim U \left[u_3, v_3 \right].$$

3.4 For $t = s - 2, \ldots, T - 1$, if $p_{4,t}^{(i)} = 1$ then

$$k_{4,t}^{(i+1)} | \gamma_{t+1}^{(i+1)}, \gamma_t^{(i+1)}, \ldots, \gamma_{t-s+2}^{(i+1)}, \sigma_\omega^{2(i+1)}, p_{4,t}^{(i)} \sim$$
$$N \left(\gamma_{t+1}^{(i+1)} - \gamma_t^{(i+1)} - \ldots - \gamma_{t-s+2}^{(i+1)}, \sigma_\omega^{2(i+1)} \right);$$

If $p_{4,t}^{(i)} = 0$ then

$$k_{4,t}^{(i+1)} | \gamma_{t+1}^{(i+1)}, \gamma_t^{(i+1)}, \ldots, \gamma_{t-s+2}^{(i+1)}, \sigma_\omega^{2(i+1)}, p_{4,t}^{(i)} \sim U \left[u_4, v_4 \right].$$

4. Sample $\mathbf{p}^{(i+1)}|Y, \boldsymbol{\alpha}^{(i+1)}, \sigma^{2(i+1)}, \mathbf{k}^{(i+1)}, q^{(i)}$ by drawing each $p_{j,t}^{(i+1)}$, for each t, for $j = 1, \ldots, 4$, from a Bernoulli distribution with probabilities defined by:

4.1 For $t = 1, \ldots, T$ and $l = 0, 1$

$$P\left[p_{1,t}^{(i+1)} = l | y_t, \mu_t^{(i+1)}, \gamma_t^{(i+1)}, \sigma_\varepsilon^{2(i+1)}, k_{1,t}^{(i+1)}, q_1^{(i)}\right] \propto$$
$$f_N\left(y_t | \mu_t^{(i+1)} + \gamma_t^{(i+1)} + l k_{1,t}^{(i+1)}; \sigma_\varepsilon^{2(i+1)}\right)\left(l q_1^{(i)} + (1-l)(1-q_1^{(i)})\right);$$

4.2 For $t = 1, \ldots, T-1$ and $l = 0, 1$

$$P\left[p_{2,t}^{(i+1)} = l | \mu_{t+1}^{(i+1)}, \mu_t^{(i+1)}, \beta_t^{(i+1)}, \sigma_\eta^{2(i+1)}, k_{2,t}^{(i+1)}, q_2^{(i)}\right] \propto$$
$$f_N\left(\mu_{t+1}^{(i+1)} | \mu_t^{(i+1)} + \beta_t^{(i+1)} + l k_{2,t}^{(i+1)}; \sigma_\eta^{2(i+1)}\right)\left(l q_2^{(i)} + (1-l)(1-q_2^{(i)})\right);$$

4.3 For $t = 1, \ldots, T-1$ and $l = 0, 1$

$$P\left[p_{3,t}^{(i+1)} = l | \beta_{t+1}^{(i+1)}, \beta_t^{(i+1)}, \sigma_\zeta^{2(i+1)}, k_{3,t}^{(i+1)}, q_3^{(i)}\right] \propto$$
$$f_N\left(\beta_{t+1}^{(i+1)} | \beta_t^{(i+1)} + l k_{3,t}^{(i+1)}; \sigma_\zeta^{2(i+1)}\right)\left(l q_3^{(i)} + (1-l)(1-q_3^{(i)})\right);$$

4.4 For $t = s-1, \ldots, T-1$ and $l = 0, 1$

$$P\left[p_{4,t}^{(i+1)} = l | \gamma_{t+1}^{(i+1)}, \ldots, \gamma_{t-s+2}^{(i+1)}, \sigma_\omega^{2(i+1)}, k_{4,t}^{(i+1)}, q_4^{(i)}\right] \propto$$
$$f_N\left(\gamma_{t+1}^{(i+1)} | \gamma_t^{(i+1)} + \ldots + \gamma_{t-s+2}^{(i+1)} + l k_{4,t}^{(i+1)}; \sigma_\omega^{2(i+1)}\right)\left(l q_4^{(i)} + (1-l)(1-q_4^{(i)})\right);$$

5. For $j = 1, 2, 3, 4$ sample $q_j^{(i+1)}|\mathbf{p}_j^{(i+1)}$ from the beta distributions:

$$q_j^{(i+1)}|\mathbf{p}_j^{(i+1)} \sim Beta\left(a_j + \sum_{t=1}^{T} p_{j,t}^{(i+1)}, b_j + T - \sum_{t=1}^{T} p_{jt}^{(i+1)}\right).$$

By the convergence results in §2.2.3, the Gibbs sampler scheme define by *sampling scheme 4.1* is irreducible and aperiodic, and converges to the target posterior distribution, if all the hyperparameters are different from zero. After convergence is achieved, the sample means of the posterior samples for each of the parameters of the model, provide consistent estimates of the true parameters.

Establishing the existence of shocks to the observed variables is done through the posterior samples of the size and indicator variables. As an illustration, consider

the detection of seasonal shifts. Take a sample, of dimension N from the posterior distribution of the seasonal indicator variable, obtained after convergence is achieved $\left\{p_{4,t}^{(i)}\right\}_{t=1,\ldots,T}^{i=1,\ldots,N}$, and obtain the corresponding sample mean, $\bar{p}_{4,t} = \frac{\sum_{i=1}^{N} p_{4,t}^{(i)}}{N}$, for $t = 1,\ldots,T$. Using a threshold method there is a seasonal shift at instant t^* if $\bar{p}_{4,t^*-1} \geq c$, where c is a benchmark proportion for the detection of a structure shift across the replications, after convergence is achieved. The value of c is chosen empirically and following the discussion in §3.4.1, is often chosen as 0.5.

Having established the existence of a seasonal shift at instant t^* we are interested in estimating the size of that shift, by considering the sample $\left\{k_{4,t^*-1}^{(i)}\right\}^{i=1,\ldots,N}$. The estimation of the size is obtained by a weighted average of this sample, taking as weights the correspondent sample of the indicator $\left\{p_{4,t^*-1}^{(i)}\right\}^{i=1,\ldots,N}$:

$$\bar{k}_{4,t^*-1} = \frac{\sum_{i=1}^{N} k_{4,t^*-1}^{(i)} p_{4,t^*-1}^{(i)}}{\sum_{i=1}^{N} p_{4,t^*-1}^{(i)}}. \tag{4.31}$$

4.3 A Monte Carlo study

In order to assess the performance of the sampling scheme described above we perform a Monte Carlo experiment. Using as the data generating process the BSM defined by equations (4.2) to (4.5), we obtained 5000 data sets of size $T = 100$. The true hyperparameters used for generating the data sets are: $\sigma_\varepsilon^2 = \sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\omega^2 = 0.1$. An outlier is introduced at $t = 20$ with magnitude 3; a level shift at instant $t = 50$ with magnitude -5; a slope shift at $t = 70$ with magnitude -3; a seasonal shift at instant $t = 60$ with magnitude 3. An example of the data set generated by the model with these parameters is in Figure 4.1.

For each data set, the Gibbs sampler defined in *sampling scheme 4.1* is run with a burn in period of 5000 iterations, and the following 5000 sampled values are recorded. The following prior distributions are used:

$$\begin{aligned}
\sigma_x^2 &\sim IG\left(\tfrac{5}{2}, \tfrac{0.5}{2}\right) &&\text{for } x = \varepsilon, \eta, \zeta, \omega, \\
k_{i,t} &\propto 1 &&\text{for } i = 1,2,3,4, \ t = 1,\ldots,T, \\
q_i &\sim Beta\left(2,100\right) &&\text{for } i = 1,2,3,4.
\end{aligned}$$

The parameters for the approximating uniform distributions are $u_i = -3$, $v_i = 3$, for $i = 1,3,4$ and $u_2 = -5$, $v_2 = 5$.

Figure 4.1: Artificial data set generated from BSM with outlier, level, slope and seasonal shifts.
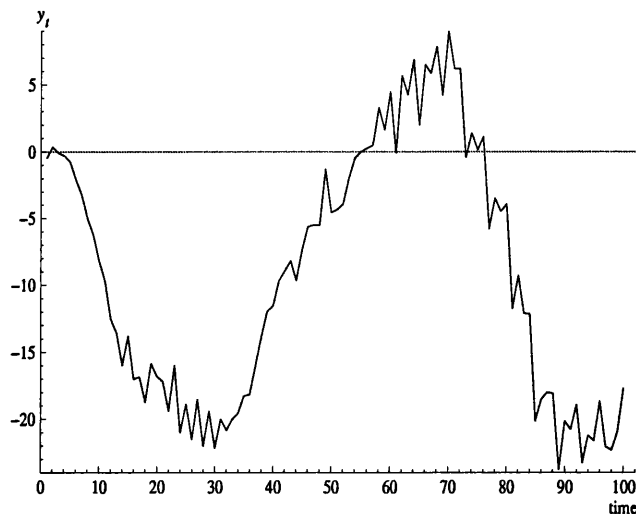
In Table 4.1 we present a summary of the estimates of the hyperparameters across the 5000 simulated replications.

The hyperparameters for all the components with exception of the irregular, present a sample mean across the values estimated close to the true value of 0.1. For σ_ε^2 we get an empirical 95% confidence interval that does not include the true value of the parameter. In Figure 4.2 we present the histograms and box-plots for the posterior mean of the hyperparameters estimated across the replications. The empirical distributions for the trend, slope and seasonal hyperparameters estimated, across the simulated replications, present similar shapes. They are slightly right skewed, and present a comparable degree of dispersion. The sample of estimates for the irregular hyperparameter, is also somewhat right skewed, but comparatively less dispersed. In Table 4.2, we present a summary of the results obtained for the intervention variables: indicator of the presence of shock $p_{i,t}$ and size of shock $k_{i,t}$, across the 5,000 simulations. We focus on these estimates for the instances where a shock exists. Given the model formulation, a component shift should be detected through the intervention variables for the instant before its effect is felt. For example, given that there is a seasonal shift at $t = 70$, the indicator of its presence is $p_{4,69}$ and $k_{4,69}$ its size.

109

Table 4.1: Summary of output from posterior sample means of hyperparameters, across 5,000 simulated replications, for BSM.

	Truth	Posterior mean sample				
		Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$
σ_ε^2	0.10	0.0605	0.0592	0.00823	0.0478	0.0801
σ_η^2	0.10	0.1230	0.118	0.0279	0.0839	0.1903
σ_ζ^2	0.10	0.0970	0.0937	0.0242	0.0597	0.1539
σ_ω^2	0.10	0.0997	0.0971	0.0235	0.0621	0.1535

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimates across 5,000 samples. Based on 10,000 Gibbs draws, discarding the first 5,000.

Figure 4.2: Histograms (a)-(d) and box-plots (e)-(f) of posterior mean estimates of hyperparameters, across 5,000 simulated replications, for BSM.
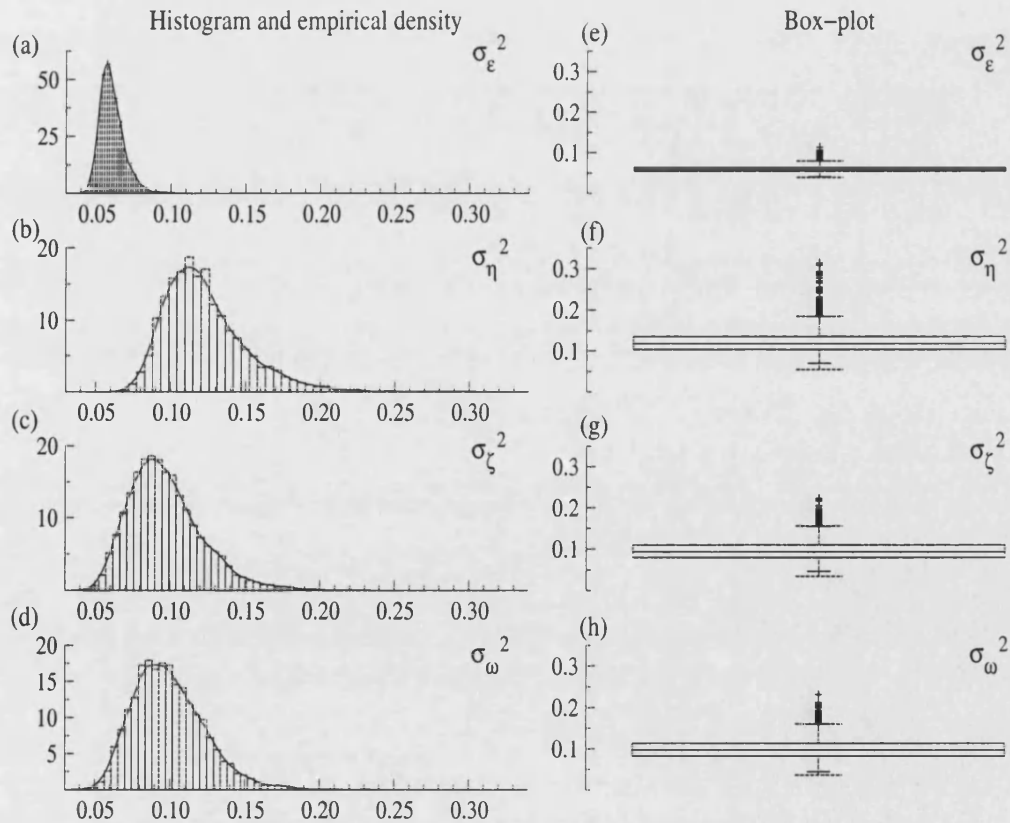
Table 4.2: Summary of output from posterior sample means for intervention variables, across 5,000 simulated replications, for BSM.

Truth								Posterior sample mean				
		Indicator $\bar{p}_{i,t}$						Size of shock $\bar{k}_{i,t}$				
t	$k_{i,t}$	Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$	≥ 0.5	Mean	Median	SD	Q_1	Q_3
Outlier	20	3	0.697	0.849	0.340	0.0134	1	0.718	2.898	2.951	0.973	0.888	4.349
Level s.	49	-5	0.803	1.000	0.328	0.0046	1	0.816	-4.522	-4.859	2.089	-6.738	0.391
Slope s.	69	-3	0.747	0.876	0.296	0.0482	1	0.788	-2.952	-2.976	1.140	-4.273	-1.425
Seas. s.	59	3	0.949	1.000	0.143	0.452	1	0.970	2.998	3.001	0.527	1.987	4.018

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 5,000 samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5, across the 5,000 samples. Based on 10,000 Gibbs draws, discarding the first 5,000.

For all the type of shocks the average, across the simulated replications, of the posterior sample mean probability of correctly detected shock, presents values above 50%, which we take as a threshold value for considering there is a shock. Comparing the different type of shifts, the seasonal shift is detected more often, with a percentage very close to 100%. The average percentage of detection of the outlier presents the smallest value. All types of shocks, are correctly detected more than 70% of the time, when we use a critical value of 0.5 for detection .

The mean values of the estimated shocks magnitude across the simulations, are for all shocks very close to the true ones. The distribution of the posterior mean of the shocks size, for the 5,000 replications is plotted in Figure 4.3. We can see that

Figure 4.3: Histogram and box-plot of estimates of magnitude of outlier (a)-(b), level shift (c)-(d), slope shift (e)-(f), seasonal shift (g)-(h), across 5,000 simulated replications, for BSM.

112
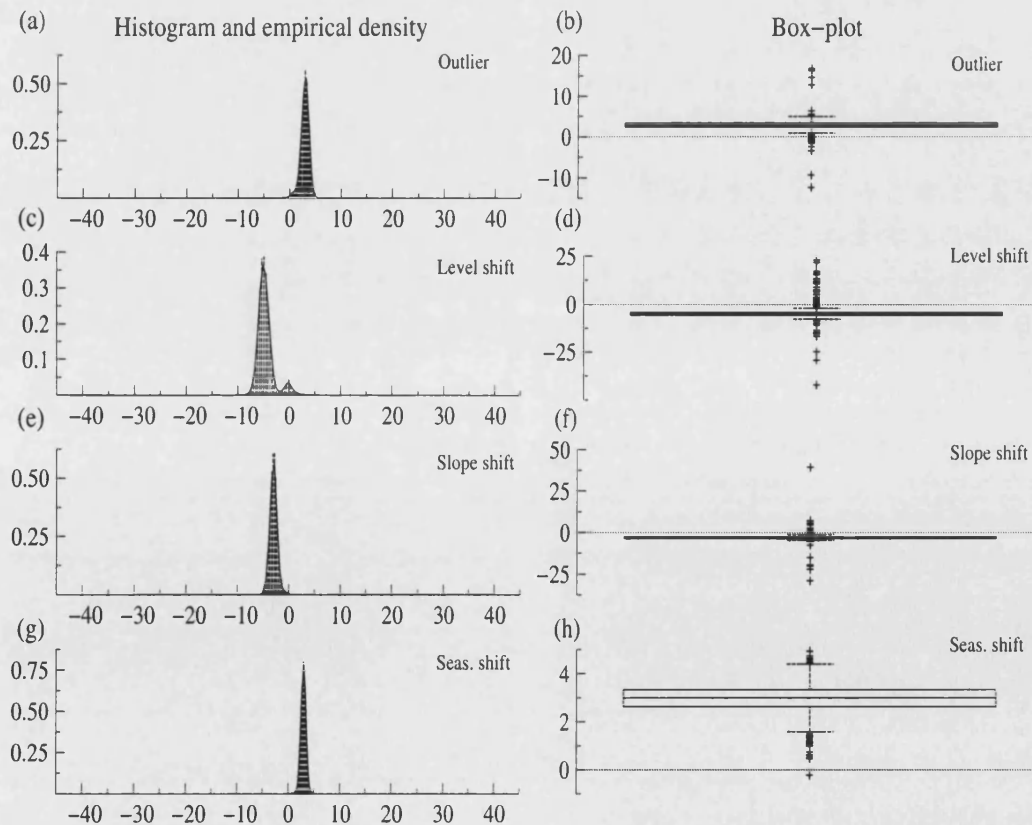
the seasonal shift presents the least dispersed distribution, approximately symmetric around the true value of the magnitude of the shift. For the level shift, we can see that for a subset of the simulated replications, the size of the shift is distributed around zero. Excluding this minority of cases, the simulated replications of the posterior magnitude are approximately symmetrically distributed around the true value.

Together with the parameters given above, the prior probability of a shock was also sampled. The following values are the average of the posterior means across the replications, for each of the four different types of shocks:

$$q_1 = 0.0219, \quad q_2 = 0.0207, \quad q_3 = 0.0242, \quad q_4 = 0.0208.$$

From this Monte Carlo experiment we conclude that the method we have proposed performs satisfactorily. Without much prior knowledge of the behaviour of the data, the shocks present in the simulated data sets were detected in the majority of the replications and the shock magnitude estimated in an accurate way. As a consequence of correctly accounting for the shocks, the hyperparameters of the model were also correctly estimated. The irregular hyperparameter is an exception, presenting in average an underestimated value. This is a problem inherent to the formulation of the model. Given the amount of parameters to be estimated, the results are satisfactory.

4.4 Empirical application: marriages data set

Sampling scheme 4.1 is applied to modeling the data set consisting of the quarterly number of marriages (in thousands) in the UK from the first quarter of 1958 (1958Q1) to the fourth quarter of 1984 (1984Q4), in a total of 100 observations. The data is plotted in Figure 4.4. This data set has been previously analyzed in West and Harrison (1997) and Penzer (1998), for the subset of observations between 1965Q1 and 1970Q4.

We started by fitting a BSM, with $s = 4$, with intervention variables for all the components, given by equations (4.2) to (4.5). From this initial analysis we have

Figure 4.4: Quarterly number of marriages in the UK from 1958Q1 to 1984Q4
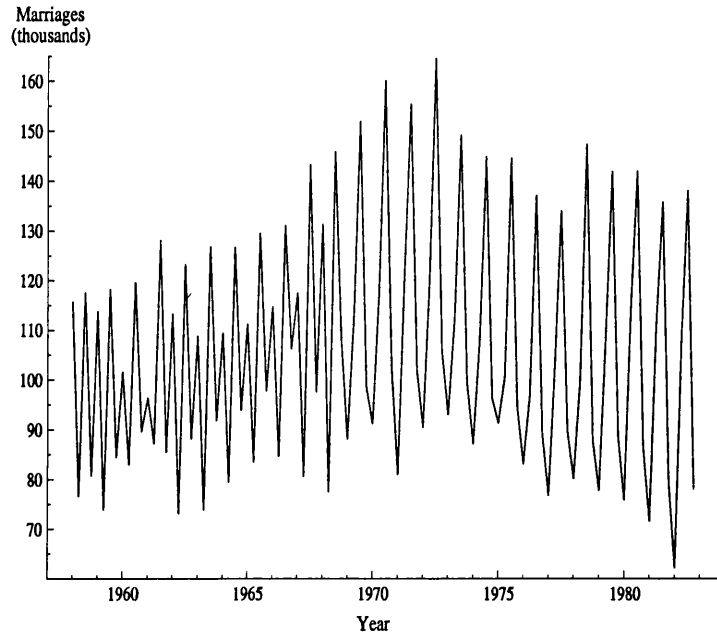
concluded there are no outliers or slope shifts in the data and so we reduced the intervention variables to level and seasonal. The model we are fitting is then:

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

$$\mu_{t+1} = \mu_t + \beta_t + k_{2,t}p_{2,t} + \eta_t$$

$$\beta_{t+1} = \beta_t + \zeta_t$$

$$\sum_{i=0}^{3} \gamma_{t+1-i} = k_{4,t}p_{4,t} + \omega_t,$$

We use a particular case of *sampling scheme 4.1*, by taking the intervention variables corresponding to outliers and slope shifts to be zero and not running steps *3.1*, *3.3*, *4.1* and *4.3*.

As prior distributions we use:

$$\sigma_\varepsilon^2 \sim IG\left(\tfrac{5}{2}, \tfrac{0.0001}{2}\right),$$

$$\sigma_\eta^2 \sim IG\left(\tfrac{5}{2}, \tfrac{5}{2}\right),$$

$$\sigma_\zeta^2 \sim IG\left(\tfrac{5}{2}, \tfrac{0.01}{2}\right),$$

$$\sigma_\omega^2 \sim IG\left(\tfrac{5}{2}, \tfrac{50}{2}\right),$$

$$k_{i,t} \propto 1, \qquad \text{for } i = 2,4, \ t = 1,\ldots,T,$$

$$q_i \sim Beta\,(2,100), \qquad \text{for } i = 2,4.$$

114

For the uniform approximation of the flat prior, when the indicator variable is equal to zero, we use:

$$k_{2,t} \sim U[-10, 10],$$
$$k_{4,t} \sim U[-40, 20].$$

The Gibbs sampler defined by *sampling scheme 4.1* is run for 100,000 iterations, from which the first 50,000 are discard. We start the analysis of the results obtained by considering the intervention variables. The plots (a) and (c) in Figure 4.5 are obtained by averaging the values of the indicator variable, for each t, across the draws from the sampler. They give the posterior detected proportion in each point in time. For the indicator of a level shift we have a peak in 1973Q1, with a mean posterior value of 0.60. Taking the threshold of 0.5 this is indicative of a level shift in 1973Q2. For the indicator of a seasonal shift we have two outstanding values in 1961Q4 and 1968Q4. They correspond to average values above 0.5. In fact for 1968Q4, in every recorded draw a seasonal shift is detected. Therefore, two seasonal shifts are detected, at 1962Q1 and 1969Q1. Having established the existence and position of the shocks, we turn our attention to their size, by analyzing plots (b) and (d) in Figure 4.5. These plots are obtained as the weighted averages given by expression (4.31). We have a downward level shift at 1973Q2, an upward seasonal shift at 1962Q1, followed by a downward seasonal shift at 1969Q1. A summary of the detected shocks and their sizes is given in Table 4.3.

The hyperparameters estimated, and some descriptive statistics of the posterior samples are given in Table 4.4. In Figure 4.6, the empirical distributions of these samples are plotted. The empirical distribution of the posterior sample of σ_ω^2 is approximately symmetric in relation to its estimated value of 10.796, whereas the distribution for σ_η^2 is slightly right skewed. The empirical distributions of σ_ε^2 and σ_ζ^2 are concentrated close to zero. Although they present very small values when compared with the values of the two other variances, given the the posterior standard errors, in Table 4.4, they are statistical significant at 5%.

In Figure 4.7, we plot the autocorrelation functions up to lag 1,000. We can see that there is a rapid convergence towards zero, confirming that convergence have been achieved after the burn in period used. We note that we have used a large

Figure 4.5: Plots of posterior means of indicator and size of intervention variables for level shifts (a)-(b) and seasonal shifts (c)-(d), for the marriages data set from 1958Q1 to 1984Q4.
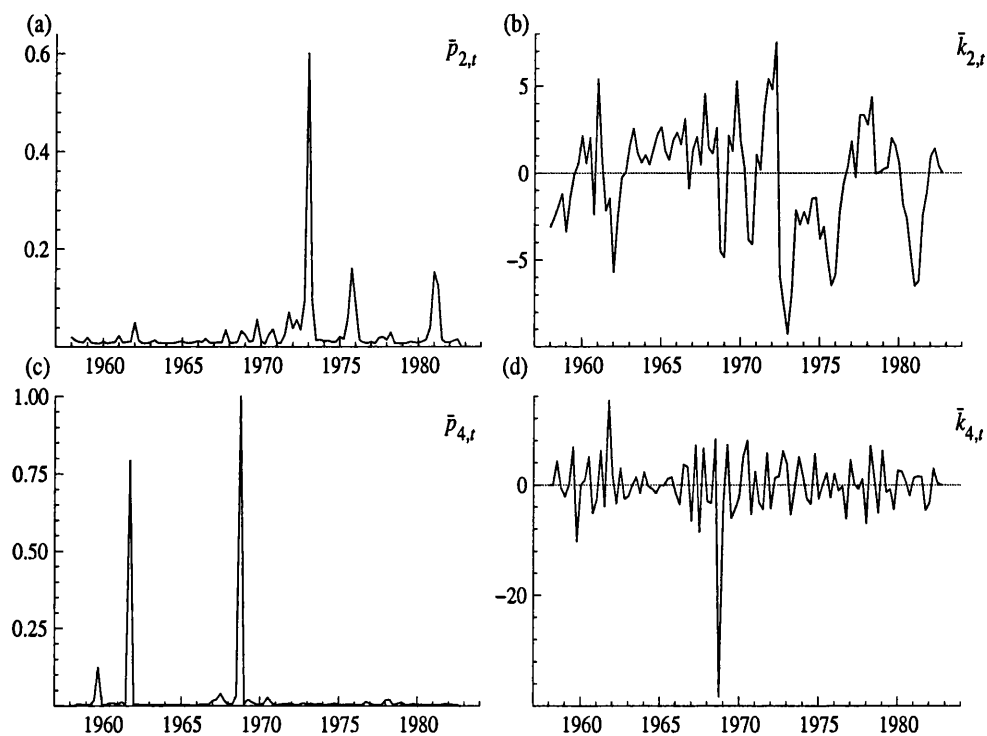
Table 4.3: Summary of output from posterior sample of interventions variables for structural shifts detected for the marriages data set, from 1958Q1 to 1984Q4.

		Posterior sample	
Shift	t (year-q)	$\bar{k}_{i,t}$	$\bar{p}_{i,t}$
Level	1973-1	-9.26	0.60
Seasonal	1961-4	15.21	0.79
	1968-4	-38.30	1

The values reported are the weighted mean of the size of intervention and the mean of the indicator of intervention, of posterior samples, for t corresponding to shocks detected. Based on 100,000 Gibbs draws, discarding the first 50,000.

Table 4.4: Summary of output from posterior sample of hyperparameters, for the marriages data set, from 1958Q1 to 1984Q4.

Posterior sample	σ_ε^2	σ_η^2	σ_ζ^2	σ_ω^2
Mean	0.0157	1.777	0.00384	10.796
SD	0.00094	0.0326	0.00021	0.0288

The values reported are the mean of the posterior sample and standard deviation of estimates (SD estimated with a bandwidth of 5,000 for the Parzen window) of the hyperparameters. Based on 100,000 Gibbs draws, discarding the first 50,000.

Figure 4.6: Histograms (a)-(d) and box-plots (e)-(f) of posterior samples of hyperparameters, for the marriages data set, from 1958Q1 to 1984Q4.
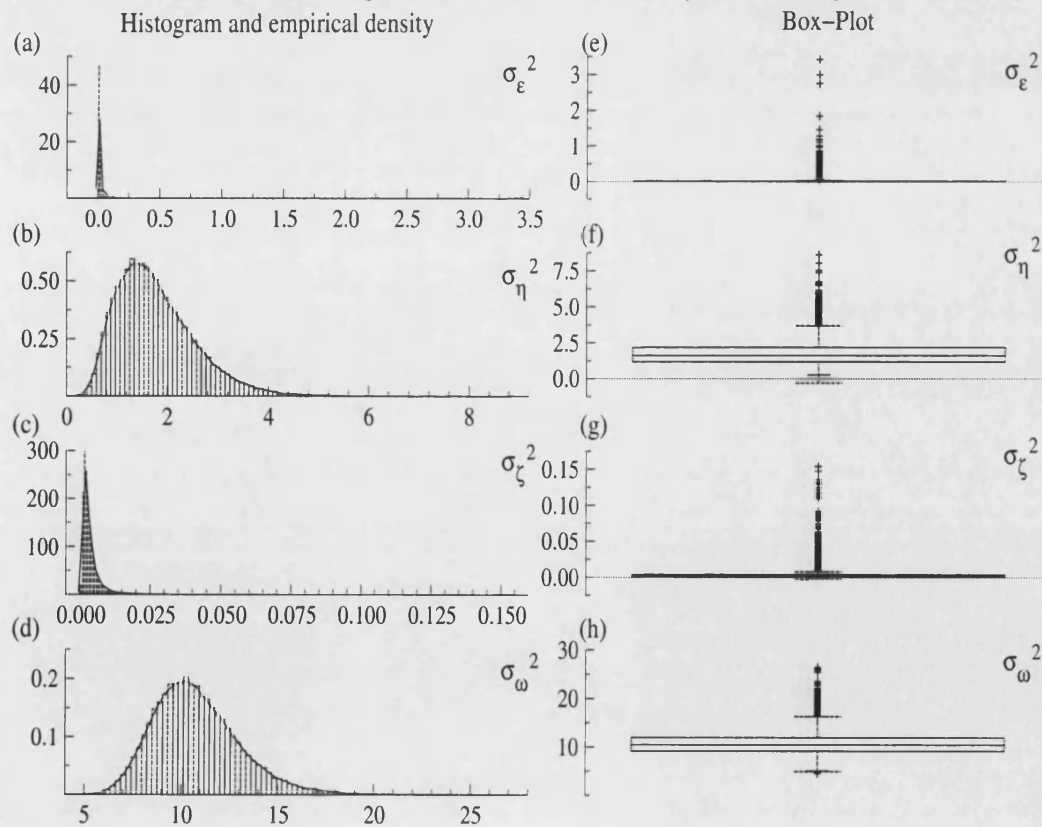
value for the burn-in-period, to ensure that convergence had been achieved. Smaller burn in periods produced similar results. As it was not computational expensive to use a burn in period of 50,000 iterations, we have opted for this value.

Figure 4.7: Autocorrelation functions for the posterior samples of hyperparameters, for 1,000 lags, for the marriages data set, from 1958Q1 to 1984Q4.
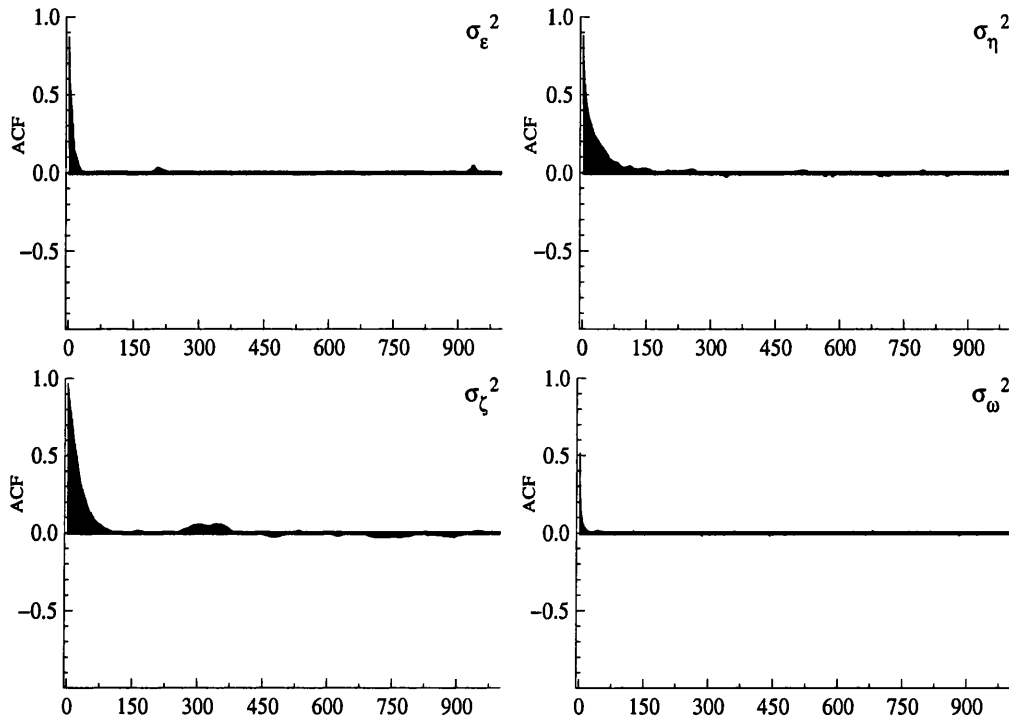
In Figure 4.8 we present scatter plots for the posterior samples of the hyperparameters. The variables that seem to present an higher degree of correlation are the level and seasonal variance. This finding is confirmed by the values obtained for the variance-covariance matrix for the hyperparameters posterior samples, presented in Table 4.5. The covariance between the posterior samples for the level and seasonal variances, presents the higher absolute value for the covariances, corresponding to a negative correlation between these two variables.

The mean values of the posterior samples of the prior probabilities of a structural shift are $q_2 = 0.025$ and $q_4 = 0.022$, respectively for the level and seasonal shifts. These parameters estimates do not reflect accurately the proportion of shifts

118

Figure 4.8: Scatter plots of the posterior samples of hyperparameters, for the marriages data set, from 1958Q1 to 1984Q4.
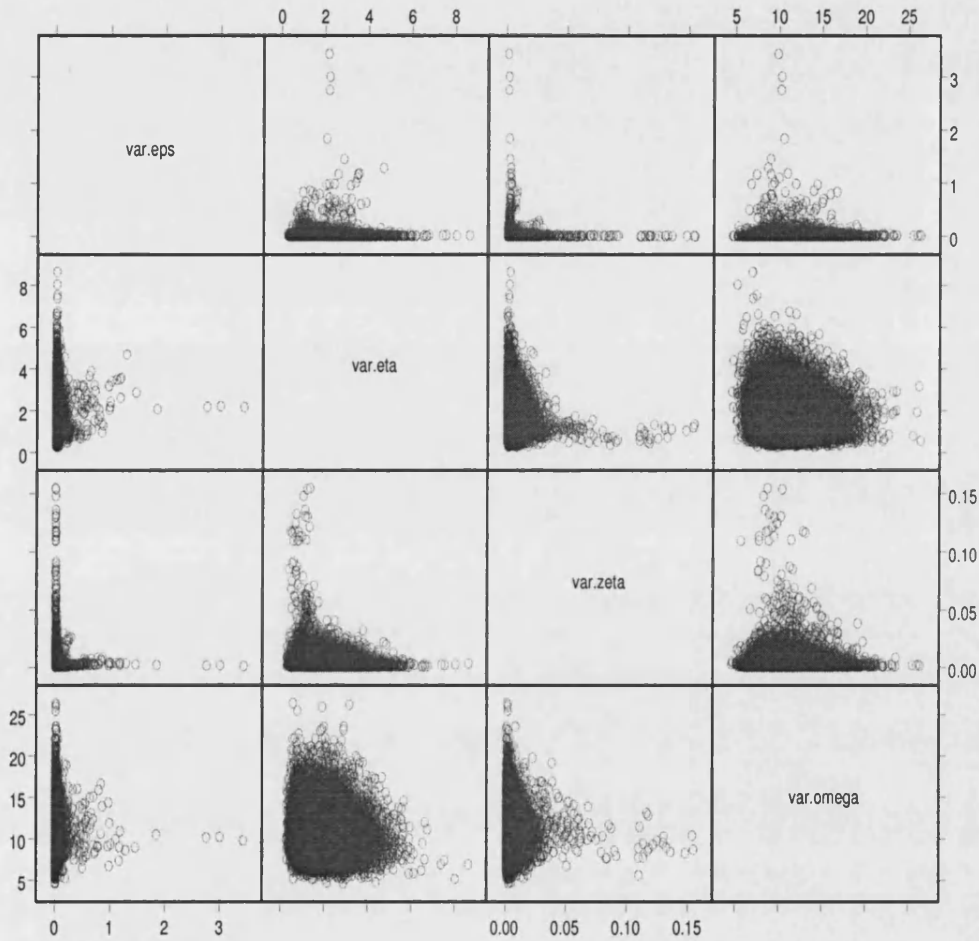
Table 4.5: Variance-covariance matrix of posterior samples of hyperparameters, for the marriages data set, from 1958Q1 to 1984Q4.

	σ_ε^2	σ_η^2	σ_ζ^2	σ_ω^2
σ_ε^2	0.00156	0.000347	-2.45E-06	0.00131
σ_η^2	0.000347	0.600	-0.000288	-0.174
σ_ζ^2	-2.45E-06	-0.000288	2.20E-05	-9.63E-05
σ_ω^2	0.00131	-0.174	-9.63E-05	4.851

detected in the data. For any individual draw the indicator variable may have the value one for observations other than those corresponding to the location of the shocks. Therefore, we do not necessarily obtain $q_4 > q_2$, in spite of having detected more seasonal shifts than level shifts.

In Figure 4.9, we plot the standardized innovations. This set is obtained by running the Kalman filter, inputting the estimates of the hyperparameters and location of structural shifts previously reported (in §2.3 these diagnostic tools are described). If the model fits the data well, the standardized innovations should be a standard

Figure 4.9: Descriptive plots of standardized innovations, for the marriages data set, from 1958Q1 to 1984Q4.

(a) Plot of standardized innovations

(b) Box-plot of standardized innovations

(c) ACF for standardized innnovations

(d) QQ plot of standardized innovations
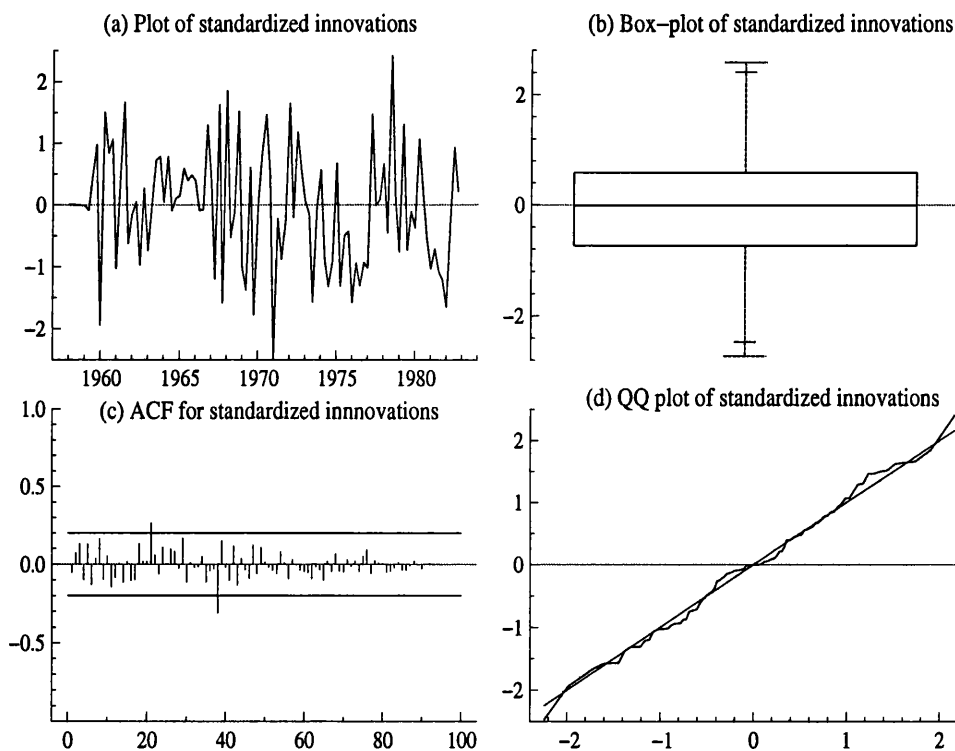
Gaussian white noise. From the plots in Figure 4.9 we can see that they behave approximately as expected.

The standardized auxiliary residuals, obtained from running the disturbance smoother, are a more powerful diagnostic tool, to detect the existence of shocks not accounted for by the model fitted. They are plotted in Figure 4.10. We can

Figure 4.10: Plots of standardized auxiliary residuals, for marriages data set, from 1958Q1 to 1984Q4.
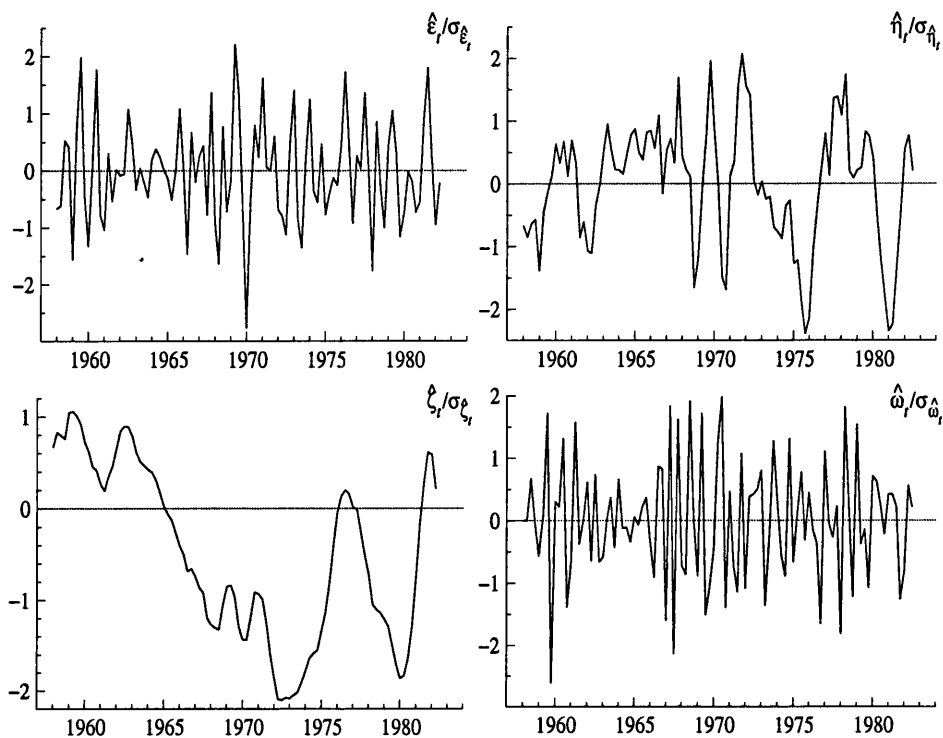
observe that for any of the standardized auxiliary residuals the values plotted are in general between -2 and 2, or present values very close to these boundaries. For the irregular standardized auxiliary residuals there is an outstanding value at 1971Q1, of approximately -2.7. If the series of auxiliary residuals was independent, for testing simultaneously the significance of these series, the critical value for a test of size 0.05 would be approximately 3.5 (see Penzer, 2001 and discussion in §2.3). Given that, for any of the series of residuals, the values are in absolute value inferior to 2.8, we conclude that all the shocks in the data are correctly accounted for,

As mentioned before, this data set, for the sub-period of 1965Q1 to 1970Q4, as been analyzed in West and Harrison (1997) and Penzer (1998). In both papers a seasonal shift as been considered in 1969Q1. When comparing our method with the one in Penzer (1998), we conclude that it has the advantage of performing the estimation of the hyperparameters and detection of the shocks simultaneously, in an

automatic way, not requiring much prior information about the data set in study.

4.5   Conclusions

We have proposed a Gibbs sampler for detecting the existence of outliers, level, slope and seasonal shifts, simultaneously with the estimation of the hyperparamters, for the basic structural model. The detection of shocks is accomplished by the inclusion of two intervention variables, for each of the equations of the model, for the location and size of the shocks. Using a flat distribution for the prior of the shocks magnitude does not require much prior information about the behaviour of the data. Therefore it has the advantage of allowing greater automation of the process of estimation of the intervention model.

The Monte Carlo study we have conducted shows that the method works efficiently when applied to data sets with outliers and the three different types of structural shifts. When applying the methodology to a real data set we conclude that the methodology allow us to estimate the model in a satisfactory way. The requirement of little prior information is clearly an advantage in modeling real series.

Chapter 5

Detection of outliers and level shifts in unobserved component models with a null noise variance

5.1 Introduction

Chapter 3 presents a method for estimating the hyperparameters of a state space model in the presence of outliers and level shifts. The model is formulated as:

$$y_t = \alpha_t + k_{1,t} p_{1,t} + \varepsilon_t, \tag{5.1}$$

$$\alpha_{t+1} = \alpha_t + k_{2,t} p_{2,t} + \eta_t. \tag{5.2}$$

The noise processes are assumed to be normally distributed:

$$\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \tag{5.3}$$

$$\eta_t \sim N\left(0, \sigma_\eta^2\right), \tag{5.4}$$

$$\tag{5.5}$$

for all $t = 1, \ldots, T$.

From the algorithm given in *sampling scheme 3.2*, it is clear that the method proposed will be degenerate if one of the hyperparamters is null. To see how that might happen suppose we know, *a priori*, that $\sigma_\varepsilon^2 = 0$ and $\sigma_\eta^2 \neq 0$. If for a certain iteration i, $p_{1,t}^{(i)}$ is sampled as one, then on iteration $i + 1$ we sample the magnitude $k_{1,t}^{(i+1)}$ from a degenerate normal distribution with zero variance. This problem is also present in the scheme proposed in Chib, Nardari, and Shephard (2002), and is

referred to in Gerlach, Carter, and Kohn (2000) as being a consequence of sampling the magnitude conditioning on the states.

Sampling the magnitude conditioning only on the observations is, in fact, a more efficient procedure, as we are conditioning on fewer variables; see Liu, Wong, and Kong (1994). However it is not a feasible approach when we assume a continuous prior for the magnitude. In Gerlach, Carter, and Kohn (2000) a method is presented for sampling the magnitude of the shocks without conditioning on the state variables. Consider a time series process $\{y_t\}_{t=1}^T$, with locally constant level component μ_t, where the existence of a level shift at time t is accounted for through a scale parameter K_t, for $t = 1, \ldots, T$. The state space model formulation is:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \\
\mu_{t+1} &= \mu_t + K_t \eta_t,
\end{aligned}
$$

with the noise processes distributed as in (5.3) and (5.4). The level changes are modeled using a mixture of Gaussian distributions for $K_t \eta_t$. There is not a change in the level of the series at $t+1$, with a prior probability of p, if $K_t = 0$. With a prior probability of $1 - p$, $K_t = k$, with k set *a priori* to a value big enough to account for a level shift at time $t + 1$. More generally K_t is assumed to have a multinomial prior distribution. Sampling from its posterior is based on the relation:

$$
p\left(K_t | Y\right) \propto p\left(K_t | K_{s \neq t}\right) p\left(y^t | y^{1, t-1}, K^{1, t}\right) p\left(y^{t+1, T} | y^{1, t}, K^{1, T}\right), \tag{5.6}
$$

where $y^{i,j} = (y_i, \ldots, y_j)$ and $K^{i,j} = (K_i, \ldots, K_j)$. Gerlach, Carter, and Kohn (2000) present an efficient way of evaluating these probabilities. Expression (5.6) is evaluated for the finite set of values of the prior multinomial for K_t. This approach is feasible given the assumption of a discrete prior distribution for the magnitude of a shock K_t. The drawback of a multinomial prior assumption is that we have to define *a priori* a reasonable set of values for the magnitude of the shocks. This is quite restrictive and demands considerable prior knowledge of the time series in question.

With one of the noise variances equal to zero, a problem arises with the simulation smoother (De Jong and Shephard, 1995) which we have used for sampling from the

full conditional of the states vector. This method involves inverting a variance-covariance matrix that will be singular if any of the individual variances is equal to zero.

To overcome these difficulties, in §5.2 we describe an alternative simulation smoother, put forward by Durbin and Koopman (2002), and its use for modeling interventions. This method overcomes the question of sampling from the states full conditional, when one of the variances is set equal to zero. They formulate a model without intervention variables. Therefore, the question of sampling from the interventions full conditionals is not tackled by their methodology. The method presented in Durbin and Koopman (2002) is a simulation smoother method, different from the simulation smoother in De Jong and Shephard (1995), in the techniques used for obtaining the states posterior samples. It was not the method used throughout the entire thesis as it has only recently been published.

Our main contribution is a method for sampling from the intervention variables full conditionals, when $\sigma_\varepsilon^2 = 0$, $\sigma_\eta^2 \neq 0$, and $\sigma_\eta^2 = 0$, $\sigma_\varepsilon^2 \neq 0$, assuming a continuous prior distribution for the size of intervention variables.

We propose a sampling scheme run in two stages. A first stage works on an auxiliary data set. This auxiliary data is constructed in such a way that it follows a local level model, with both hyperparameters different from zero. In the case of $\sigma_\varepsilon^2 = 0$, this data set is constructed so that the position and location of the outliers are the same as for the original data set. The purpose of this auxiliary data set is to obtain posterior samples of the intervention variables for the presence of an outlier, and the prior probability of an outlier parameter. When $\sigma_\eta^2 = 0$, the auxiliary data set constructed so that the level shifts in the original data are preserved in terms of their location and size. We obtain posterior samples for the variables related to the presence of level shifts, by estimating a local level model with interventions for the auxiliary data set. On a second stage, the original data set used for estimation and we obtain posterior samples for the remaining variables of the model. These posterior samples are obtained by running a second Gibbs sampler. The shocks detected on the first stage, are input as dummy variables with fixed positions and sizes.

We describe this two stage method, considering separately the cases of $\sigma_\varepsilon^2 = 0$ and $\sigma_\eta^2 = 0$. In §5.3, we explain the sampling scheme when the irregular hyperparameter is equal to zero. Two numerical examples are presented. Firstly, a Monte carlo study. The methodology is applied for 5000 artificial data sets, generated from a local level model, with $\sigma_\varepsilon^2 = 0$, and where an outlier and a level shift are input. Secondly, we perform an empirical application of our methodology, to a real data set. The data considered consists of monthly quotes of bonds issued by the Greek government, from August 1916 to June 1930 (Christodoulaki and Penzer, 2002). §5.4 has the same structure. We present a sampling scheme for the case when the level hyperparameter is zero. The effectiveness of the method proposed is assessed by a Monte Carlo study. Finally, an empirical application is done using the flow of the Nile at Aswan data set, from 1871 to 1970 (Cobb, 1978).

In §5.5, an overall analysis of the methodologies proposed and results obtained is done.

The method we propose requires less prior information about the type of shocks present in the data, than the method presented in Gerlach, Carter, and Kohn (2000). For each of the stages of our sampling scheme, a Gibbs sampler similar to the one in *sampling scheme 3.2* is run. The prior distribution assumed for the size of shocks is a flat distribution. As we have discussed in Chapter 3, the sensitivity of the estimates of the shocks sizes to the choice of the prior's parameters is small. The method we propose, implies an *a priori* choice of a variance parameter, when generating the auxiliary data set. An analysis of sensitivity of the method to the choice of this parameter is conducted in §5.3.1 for the $\sigma_\varepsilon = 0$ case and in §5.4.1 for the $\sigma_\eta = 0$ case. From this analysis we conclude that, even when that choice is not optimal, the results obtained for the size of the shocks detected using the auxiliary data set provide a good indicator of the true size of the shocks. Suppose we have chosen a variance for generating the auxiliary data set that masks the shocks present in the data, In that case, the posterior probability of having a shock will be underestimated. However, the posterior mean size of the shock, is accurately estimated. This estimate gives an indication on how much the variance parameter should be shrunk, in order not to mask the presence of the shocks. Thus, the auxiliary data set is generated without

126

requiring much prior knowledge about the size of the shocks present in the data set.

When using a multinomial distribution, as in Gerlach, Carter, and Kohn (2000), the values from which to sample the size of the shocks have to been chosen *a priori*. This choice implies a considerable *prior* knowledge about the magnitude of the shocks present in the data. This set of values have a direct implication on the posterior probability of having a shock at a certain instance. Additionally, the estimate of the size of shocks detected is predetermined by the prior definition of the multinomial distribution.

5.2 Sampling from the state space variable full conditional distribution

When using the simulation smoother (De Jong and Shephard, 1995) the variance - covariance matrix $\mathbf{C}_t$, defined in (2.38), has to be inverted, for evaluating expression (2.41). As noted in De Jong and Shephard (1995), a necessary condition for $\mathbf{C}_t$ to be non-singular is that the rows of the choice matrix $\mathbf{S}_t$, in (2.37), that defines which of the noise processes we are interested in sampling from, do not span the row space of $(\mathbf{G}_t, \mathbf{H}_t)$. If we consider the model defined by equations (5.1) and (5.2), with univariate observations and state space vector, the case of one of the variances equal to zero is ruled out when using this simulation method.

Recently, a new method for sampling from the states full conditional has been proposed by Durbin and Koopman (2002). Following their notation suppose we have a state space model

$$y_t = \mathbf{Z}_t \alpha_t + \varepsilon_t, \tag{5.7}$$

$$\alpha_{t+1} = \mathbf{T}_t \alpha_t + \mathbf{R}_t \eta_t, \tag{5.8}$$

with $\varepsilon_t \sim N(0, \mathbf{G}_t)$ and $\eta_t \sim N(0, \mathbf{H}_t)$, for $t = 1, \ldots, T$. For simplicity of exposition we consider univariate observations $\{y_t\}_{t=1}^T$. The state vector α_t is $m \times 1$ and we take $\mathbf{R}_t = I_p$, for $t = 1, \ldots, T$, where p is the dimension of the vector column η_t. The matrices $\mathbf{Z}_t$, $\mathbf{T}_t$, $\mathbf{G}_t$, and $\mathbf{H}_t$ are assumed to be known, for all t. Note that the difference between this formulation and the one we have been using is the fact that

the constant terms are not present in either equations. We will see later how they can be included. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)$, $\eta = (\eta_1', \ldots \eta_T')$ and $\alpha = (\alpha_1', \ldots, \alpha_T')$. Our aim is to sample from the full conditional,

$$\alpha \sim P\left(\alpha | Y, \{Z_t, T_t, G_t, H_t\}_{t=1}^T\right),\tag{5.9}$$

given the prior $\alpha_1 \sim (a_1, P_1)$. Given the assumption that the coefficient matrices are known, expression (5.9) can be simplified to

$$\alpha \sim P\left(\alpha | Y\right).\tag{5.10}$$

Sampling from (5.10) is accomplished by simulating from the full conditional of the noise processes,

$$(\varepsilon, \eta) \sim P\left(\varepsilon, \eta | Y\right),\tag{5.11}$$

followed by a recursive process for sampling from the full conditional distribution of the states.

Let $\mathbf{u} = (\varepsilon_1, \eta_1', \ldots, \varepsilon_T, \eta_T')'$. As we are working in the context of linear Gaussian state space models,

$$\mathbf{u} | Y \sim P\left(\mathbf{u} | Y\right) = f_N\left(E\left(\mathbf{u} | Y\right), \mathrm{var}\left(\mathbf{u} | Y\right)\right).\tag{5.12}$$

We are then interested in getting the quantities

$$\hat{\mathbf{u}} = E\left(\mathbf{u} | Y\right),\tag{5.13}$$

$$\mathbf{W} = \mathrm{var}\left(\mathbf{u} | Y\right).\tag{5.14}$$

Obtaining the quantity in expression (5.13) is done by applying the Kalman filter and smoother to the model in (5.1) and (5.2), using the recursions in (2.17) to (2.21), and (2.22) to (2.26), respectively.

The method in Durbin and Koopman (2002) for obtaining the quantity in (5.14) relies on the following result, which can be found, together with proof, in Andersen (1984), pp. 37:

Theorem 5.2.1 *Let the components of* $\mathbf{X}$ *be divided into two groups composing the sub-vectors* $\mathbf{X}^{(1)}$ *and* $\mathbf{X}^{(2)}$. *Suppose the mean of* $\mathbf{X}$, $\boldsymbol{\mu}$ *is similarly divided into* $\boldsymbol{\mu}^{(1)}$ *and* $\boldsymbol{\mu}^{(2)}$, *and suppose the covariance matrix* $\boldsymbol{\Sigma}$ *of* $\mathbf{X}$ *is divided into* $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{22}$, *the covariance matrices of* $\mathbf{X}^{(1)}$, *of* $\mathbf{X}^{(1)}$ *and* $\mathbf{X}^{(2)}$, *and of* $\mathbf{X}^{(2)}$, *respectively. Then if the distribution of* $\mathbf{X}$ *is normal, the conditional distribution of* $\mathbf{X}^{(1)}$ *given* $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ *is normal with mean* $\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}\right)$ *and covariance matrix* $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Using the notation of Theorem 5.2.1 let $\mathbf{X}^{(1)} = \mathbf{u}$ and $\mathbf{X}^{(2)} = Y$. We have that $Y = (y_1, \ldots, y_T)$ is distributed as a multivariate normal, and $\mathbf{u}$ is unconditional distributed also as a multivariate normal. By Theorem 5.2.1 $\text{var}(\mathbf{u}|Y) = \mathbf{W}$ does not depend on Y. The draws from the full conditional $\mathbf{u}|Y \sim N(\hat{\mathbf{u}}, \mathbf{W})$ are then obtained by adding a draw from $N(0, \mathbf{W})$, independent of Y, to the mean correction term $\hat{\mathbf{u}} = E(\mathbf{u}|Y)$.

5.2.1 Sampling from $N(0, \mathbf{W})$

Let $\mathbf{u}^+$ be a draw from the unconditional distribution of the noise process:

$$\mathbf{u}^+ \sim N(0, \boldsymbol{\Omega}), \tag{5.15}$$

where $\boldsymbol{\Omega} = \text{diag}(\mathbf{G}_1, \mathbf{H}_1, \ldots, \mathbf{G}_T, \mathbf{H}_T)$. Using equations (5.1) and (5.2) and replacing $\mathbf{u}$ by $\mathbf{u}^+$ we recursively generate Y^+, using as initial value for the state vector $\boldsymbol{\alpha}_1$ a draw from its prior distribution. The case of diffuse initial conditions is considered later.

Applying the Kalman filter and smoother for the state space model with Y replaced by Y^+ we obtain :

$$\hat{\mathbf{u}}^+ = E\left(\mathbf{u}^+|Y^+\right). \tag{5.16}$$

It can be shown that:

$$\mathbf{u}^+ - \hat{\mathbf{u}}^+ \sim N(0, \mathbf{W}) \tag{5.17}$$

To get the result in (5.17) we start by showing that $E(\mathbf{u}^+ - \hat{\mathbf{u}}^+) = 0$.

$$
\begin{aligned}
E\left(\mathbf{u}^+ - \hat{\mathbf{u}}^+\right) &= E_{Y^+}\left[E\left(\mathbf{u}^+ - \hat{\mathbf{u}}^+|Y^+\right)\right] \\
&= E_{Y^+}(0) = 0,
\end{aligned}
$$

where we have used expression (5.16). On the other hand

$$\text{var}\left(\mathbf{u}^+ - \hat{\mathbf{u}}^+\right) = E_{Y^+}\left[\text{var}\left(\mathbf{u}^+ - \hat{\mathbf{u}}^+|Y^+\right)\right].$$

From Theorem 5.2.1 $\text{var}\left(\mathbf{u}^+|Y^+\right)$ does not depend on the conditioning variable and we get

$$\begin{aligned}
\text{var}\left(\mathbf{u}^+ - \hat{\mathbf{u}}^+\right) &= E_{Y^+}\left(\mathbf{W}\right) \\
&= \mathbf{W},
\end{aligned}$$

which concludes the derivation of the distribution in (5.17) , which can also be written as

$$\mathbf{u}^+ - \hat{\mathbf{u}}^+ \sim N\left(0, \text{var}\left(\mathbf{u}|Y\right)\right) \tag{5.18}$$

5.2.2 Sampling from $N\left(E\left(\mathbf{u}|Y\right), \text{var}\left(\mathbf{u}|Y\right)\right)$

Suppose we have sampled $\mathbf{u}^+ - \hat{\mathbf{u}}^+$ from a normal distribution as in (5.17), using the method described in the previous section, and that we have also $\hat{\mathbf{u}} = E\left(\mathbf{u}|Y\right)$. Then if we define

$$\tilde{\mathbf{u}} = \hat{\mathbf{u}} + \mathbf{u}^+ - \hat{\mathbf{u}}^+, \tag{5.19}$$

it will have a posterior distribution given by

$$\tilde{\mathbf{u}}|Y = \left(\hat{\mathbf{u}} + \mathbf{u}^+ - \hat{\mathbf{u}}^+\right)|Y \sim N\left(\hat{\mathbf{u}}, \mathbf{W}\right), \tag{5.20}$$

and therefore $\tilde{\mathbf{u}}$ is a draw from the posterior in (5.12).

To show the correctness of the method described we have to show that:

$$E\left(\tilde{\mathbf{u}}|Y\right) = \hat{\mathbf{u}}, \tag{5.21}$$

and

$$\text{var}\left(\tilde{\mathbf{u}}|Y\right) = \mathbf{W}. \tag{5.22}$$

Starting by expression (5.21):

$$\begin{aligned}
E\left(\tilde{\mathbf{u}}|Y\right) &= E\left(\hat{\mathbf{u}} + \mathbf{u}^+ - \hat{\mathbf{u}}^+|Y\right) \\
&= E\left[E\left(\mathbf{u}|Y\right)|Y\right] + E\left(\mathbf{u}^+ - \hat{\mathbf{u}}^+|Y\right), \tag{5.23}
\end{aligned}$$

130

and using that $u^+ - \hat{u}^+$ is independent from Y together with the expected value of the distribution in (5.17) we get that (5.23) can be written as:

$$\hat{u} + E\left(u^+ - \hat{u}^+\right) = \hat{u},$$

which proves (5.21). As for the conditional variance in (5.22):

$$
\begin{aligned}
\text{var}\left(\tilde{u}|Y\right) &= E\left[\left(\tilde{u} - \hat{u}\right)'\left(\tilde{u} - \hat{u}\right)|Y\right] \\
&= E\left[\left(u^+ - \hat{u}^+\right)'\left(u^+ - \hat{u}^+\right)|Y\right] \\
&= E\left[\left(u^+ - \hat{u}^+\right)'\left(u^+ - \hat{u}^+\right)\right] \\
&= W
\end{aligned}
$$

where we have used the fact that the distribution of $u^+ - \hat{u}^+$ is independent from Y and given by (5.17).

The case of diffuse initial conditions for the state space vector is discussed in Durbin and Koopman (2002). Suppose that

$$\alpha_1 \sim N\left(a_1, \kappa I\right), \quad \text{with } \kappa \to \infty. \tag{5.24}$$

To obtain Y^+, recursively from equations (5.7) and (5.8) we need an initial state vector α_1, distributed as in (5.24). In Durbin and Koopman (2002) is proved that α_1 can be taken as any vector, including a vector of zeros, as long as an exact diffuse Kalman filter and smoother is used afterwards (see Koopman, 1997).

5.2.3 Generalization for including intervention variables

The formulation presented in Durbin and Koopman (2002) for the simulation smoother, can be generalized to include the intervention terms $k_{1,t}p_{1,t}$ and $k_{2,t}p_{2,t}$, in the model defined by equations (5.1) and (5.2). Let

$$c_t = k_{1,t}p_{1,t}, \tag{5.25}$$

$$d_t = k_{2,t}p_{2,t}, \tag{5.26}$$

for $t = 1, \ldots, T$. We are interested in sampling from

$$\left(\alpha|Y, \sigma_\varepsilon^2, \sigma_\eta^2, k_1, k_2, p_1, p_2\right), \tag{5.27}$$

where $\mathbf{k}_i = (\mathbf{k}_{i,1}, \ldots, \mathbf{k}_{i,T})$ and $\mathbf{p}_i = (\mathbf{p}_{i,1}, \ldots, \mathbf{p}_{i,T})$, for $i = 1, 2$.

Hence with the definitions in (5.25) and (5.26), sampling from the full conditional in (5.27) is equivalent to sample from the posterior distribution of the state space vector $(\boldsymbol{\alpha}|Y)$, for the following SSM,

$$y_t = c_t + \alpha_t + \varepsilon_t, \tag{5.28}$$

$$\alpha_{t+1} = d_t + \alpha_t + \eta_t, \tag{5.29}$$

given that all the parameters of the model are known.

Define

$$m_t = c_t + \varepsilon_t, \tag{5.30}$$

$$n_t = d_t + \eta_t, \tag{5.31}$$

and $\mathbf{v} = (m_1, n_1, \ldots, m_T, n_T,)'$. The SSM in (5.28) and (5.29) is equivalent to:

$$y_t = \alpha_t + m_t, \tag{5.32}$$

$$\alpha_{t+1} = \alpha_t + n_t, \tag{5.33}$$

with the innovations distributed *a priori* as:

$$m_t \sim N\left(c_t, \sigma_\varepsilon^2\right),$$

$$n_t \sim N\left(d_t, \sigma_\eta^2\right),$$

for $t = 1, \ldots, T$, and the diffuse initial condition:

$$\alpha_1 \sim N(0, \kappa), \quad \kappa \to \infty. \tag{5.34}$$

The posterior distribution of innovation vector $\mathbf{v}$ is Gaussian with parameters:

$$\boldsymbol{E}(\mathbf{v}|Y) = \boldsymbol{E}(\mathbf{u}|Y) + (c_1, d_1, \ldots, c_T, d_T)',$$

$$\text{var}(\mathbf{v}|Y) = \text{var}(\mathbf{u}|Y).$$

The arguments used for obtaining the sample from the posterior of the innovation processes in (5.7) and (5.8) still hold. We have that $\mathbf{v}|Y \sim N(\hat{\mathbf{v}}, \mathbf{W})$, where, $\hat{\mathbf{v}} = \boldsymbol{E}(\mathbf{v}|Y)$ is obtained from running the exact diffuse Kalman filter and smoother

(Koopman, 1997); using the result in Theorem 5.2.1, $\text{var}\,(\mathbf{v}|Y) = \mathbf{W}$ is independent of Y, and is obtained as in §5.2.1.

Next we explain in more detail how to sample from the full conditional of the innovations for the model in (5.28) and (5.29). We assume that all the parameters of the model are known. The state space variable is diffuse.

A draw $\tilde{\mathbf{v}} \sim \mathbf{v}|Y$ is obtained in the following way.

1. Draw $\mathbf{v}^+ \sim P\,(\mathbf{v})$, by sampling, for $t = 1, \ldots, N$

$$\mathbf{v}_t^+ \sim N\,(\boldsymbol{\mu}_t, \boldsymbol{\Omega})\,,$$

with

$$
\begin{aligned}
\boldsymbol{\mu}_t &= (c_t, d_t)'\,, \\
\boldsymbol{\Omega} &= \text{diag}\left(\sigma_\varepsilon^2, \sigma_\eta^2\right);
\end{aligned}
$$

2. Starting with $\alpha_1 = 0$, use equations (5.32), (5.33) and $\mathbf{v}^+$ to generate Y^+;

3. Run the exact diffuse Kalman filter and smoother for Y and Y^+, for the model defined by (5.32) and (5.33). From the output of the smoother algorithm we obtain:

$$
\begin{aligned}
\hat{\mathbf{v}} &= \boldsymbol{E}\,(\mathbf{v}|Y)\,, \\
\hat{\mathbf{v}}^+ &= \boldsymbol{E}\left(\mathbf{v}|Y^+\right)
\end{aligned}
$$

A draw from $\tilde{\mathbf{u}} \sim \mathbf{u}|Y$ is then obtained as

$$\tilde{\mathbf{u}} = \hat{\mathbf{v}} + \mathbf{v}^+ - \hat{\mathbf{v}}^+ - (c_1, d_1, \ldots, c_T, d_T)'\,. \tag{5.35}$$

Note that the step at which a matrix had to be inverted was in the Kalman filter, for the variance of $v_t = \boldsymbol{E}\,(\varepsilon_t|Y_t)$, F_t. Consequently the case $\sigma_\varepsilon^2 = 0$ or $\sigma_\eta^2 = 0$ can be handled by this simulation smoother without any modification.

In order to obtain a draw from the full conditional of the states for the original model, in equations (5.1) and (5.2), we use the posterior draw from the innovation vector in (5.35). Starting with an initial value α_1, draw from the distribution in

(5.34), we use equations (5.1) and (5.2) and recursively generate a draw from the states full conditional $\alpha|Y, \sigma_\varepsilon^2, \sigma_\eta^2, \mathbf{k}_1, \mathbf{k}_2, \mathbf{p}_1, \mathbf{p}_2$. The simulation smoother in Durbin and Koopman (2002) gives a method for sampling from the states full conditional, even when one of the hyperparameters is equal to zero, but it does not consider the issue of sampling from the intervention variables full conditionals. To solve the question of sampling from the intervention variables in this case, we propose a new sampling scheme in §5.3 and §5.4.

5.3 Detection of outliers and level shifts when $\sigma_\varepsilon^2 = 0$

Suppose that $\sigma_\varepsilon^2 = 0$. The method used in Chapter 3 to derive the full conditional distribution of the intervention variables for the presence of outliers relies on having noise in the measurement equation, which does not happen when $\sigma_\varepsilon^2 = 0$. In this case we propose an alternative method for detecting the presence of this type of shock. The model we are considering is defined by

$$y_t = \alpha_t + k_{1,t} p_{1,t}, \tag{5.36}$$

$$\alpha_{t+1} = \alpha_t + k_{2,t} p_{2,t} + \eta_t, \tag{5.37}$$

with

$$\eta \sim N\left(0, \sigma_\eta^2\right), \tag{5.38}$$

$$\alpha_1 \sim N\left(0, \kappa\right), \text{ with } \kappa \to \infty, \tag{5.39}$$

The prior assumptions for the variables we are sampling are:

$$\sigma_\eta^2 \sim IG\left(\frac{c_2}{2}, \frac{s_2}{2}\right); \tag{5.40}$$

$$k_{i,t} \propto 1, \text{ for } i = 1, 2, \ t = 1, \ldots, T; \tag{5.41}$$

$$p_{i,t} \sim Bernoulli\left(q_i\right), \text{ for } i = 1, 2, \ t = 1, \ldots, T; \tag{5.42}$$

$$q_i \sim Beta\left(a_i, b_i\right), \text{ for } i = 1, 2. \tag{5.43}$$

We propose a two stages algorithm for detecting the presence of outliers and level shifts. The first stage is based on running the Gibbs sampler for an artificial data

set, for which the irregular variance is different from zero. The posterior samples of the parameters of the model for this artificial data set allow us to detect and characterize the outliers in the data. With the outliers detected, the Gibbs sampler is run in a second stage, for the original data set, to obtain posterior samples for the intervention variables for level shifts and the transition equation variance.

From the first stage we obtain posterior samples for the set of random variables related to the detection of outliers, namely $(\mathbf{k}_1, \mathbf{p}_1, q_1)$. The auxiliary data set $\{\tilde{y}_t\}_{t=1}^{T}$ is constructed in a such a way that it follows a local level model, with both hyperparameters different from zero. Hence, to detect the shocks present, namely the outliers, we use *sampling scheme 3.2*. From the posterior samples obtained from running the Gibbs sampler for $\{\tilde{y}_t\}_{t=1}^{T}$, we focus our attention on the intervention variables for the presence of outliers. As this auxiliary data set is constructed in such a way that the position and size of outliers of the original data set are the same, running this first Gibbs sampler enable us to detect and characterize the set of outliers present in $\{y_t\}_{t=1}^{T}$. Using this information we run a second Gibbs sampler, as defined in *sampling scheme 3.2*, not sampling from the intervention variables for the presence of outliers. These variables are fixed and set to values that reflect the set of outliers detected in the first stage. With this second run of the sampler, we obtain posterior samples for the remaining variables, $\left(\boldsymbol{\alpha}, \sigma_\eta^2, \mathbf{k}_2, \mathbf{p}_2, q_2 \right)$.

Suppose we have a data set $\{y_t\}_{t=1}^{T}$ that follows the model defined by equations (5.36) and (5.37). We construct an auxiliary data set $\{\tilde{y}_t\}_{t=1}^{T}$, by adding an innovation process to the data set in study:

$$\tilde{y}_t = y_t + \tilde{\varepsilon}_t, \tag{5.44}$$

for $t = 1, \ldots, T$, where

$$\tilde{\varepsilon}_t \sim N\left(0, \sigma_{\tilde{\varepsilon}}^2\right), \tag{5.45}$$

are independent and identically distributed for $t = 1, \ldots, T$. From its construction in (5.44) and the model formulation for y_t in (5.36) and (5.37), the auxiliary data set follows also a local level model:

$$\tilde{y}_t = \alpha_t + k_{1,t} p_{1,t} + \tilde{\varepsilon}_t, \tag{5.46}$$

$$\alpha_{t+1} = \alpha_t + k_{2,t} p_{2,t} + \eta_t, \tag{5.47}$$

with a non-zero variance for the measurement equation innovations. The original data set is replaced by $\{\tilde{y}_t\}_{t=1}^{T}$, and $\sigma_\varepsilon^2 = 0$ by $\sigma_{\tilde{\varepsilon}}^2$. All the other variables of the model remain unchanged. The model defined by equations (5.46) and (5.47) fits in the framework of *sampling scheme 3.2*. The choice of the perturbation $\tilde{\varepsilon}_t$, namely of its variance, should be so that if any outliers are present in the initial data set, their position and size is preserved by adding the noise $\tilde{\varepsilon}_t$.

Suppose there is an outlier at instant t', of size $k_{1,t'}$. Assume for simplicity of explanation that all the parameters of the model are known *a priori* with exception of the outlier position. We sample the indicator variables for the presence of outliers using the auxiliary model defined by equations (5.46) and (5.47). The full conditional distribution of the indicator variable for the presence of an outlier at t' is a Bernoulli distribution with probabilities given by:

$$P\left(p_{1,t'} = 1 | \tilde{y}_{t'}, \alpha_{t'}, \sigma_{\tilde{\varepsilon}}^2, k_{1,t'}, q_1\right) \quad \propto \quad f_N\left(\tilde{y}_{t'} | \alpha_{t'} + k_{1,t'}, \sigma_{\tilde{\varepsilon}}^2\right) q_1, \tag{5.48}$$

$$\propto \quad \exp\left[-\frac{\left(\tilde{y}_{t'} - \alpha_{t'} - k_{1,t'}\right)^2}{2\sigma_{\tilde{\varepsilon}}^2}\right] q_1,$$

and

$$P\left(p_{1,t'} = 0 | \tilde{y}_{t'}, \alpha_{t'}, \sigma_{\tilde{\varepsilon}}^2, k_{1,t'}, q_1\right) \quad \propto \quad f_N\left(\tilde{y}_{t'} | \alpha_{t'}, \sigma_{\tilde{\varepsilon}}^2\right) (1 - q_1), \tag{5.49}$$

$$\propto \quad \exp\left[-\frac{(\tilde{y}_{t'} - \alpha_{t'})^2}{2\sigma_{\tilde{\varepsilon}}^2}\right] (1 - q_1).$$

Given the construction of the auxiliary data set as in (5.44), the probabilities in expressions (5.48) and (5.49) are proportional to

$$\exp\left[-\frac{\left(y_{t'} - \alpha_{t'} - k_{1,t'} + \tilde{\varepsilon}_{t'}\right)^2}{2\sigma_{\tilde{\varepsilon}}^2}\right] q_1, \tag{5.50}$$

and

$$\exp\left[-\frac{(y_{t'} - \alpha_{t'} + \tilde{\varepsilon}_{t'})^2}{2\sigma_{\tilde{\varepsilon}}^2}\right] (1 - q_1), \tag{5.51}$$

respectively.

Given that there is an outlier at the instant considered, the value of $y_{t'} - \alpha_t$ should be significantly different from zero. Hence, the value of $\sigma_{\tilde{\varepsilon}}^2$ should be such that

$\tilde{\varepsilon}_t$, which is measured by $\sigma_{\tilde{\varepsilon}}^2$. On the other hand, we do not allow $\sigma_{\tilde{\varepsilon}}^2 \to 0$. For too small values of $\sigma_{\tilde{\varepsilon}}^2$ the Gaussian distributions in (5.48) and (5.49) become degenerate. In practice, the choice of the disturbance variance does not require much prior knowledge of the data. For example running the Gibbs sampler for estimating the model without interventions can provide information on the disturbance to be used. We will go back to the this issue later, when we present an analysis of sensitivity for the choice of $\sigma_{\tilde{\varepsilon}}^2$.

We consider now that we know the position of the outlier, at $t = t'$, and therefore we take $p_{1,t'} = 1$. We assume that all the parameters of the model are known with the exception of $k_{1,t'}$. We sample from the full conditional distribution of the size of the outlier using the auxiliary model. Using the results we have used for deriving *sampling scheme 3.2* , given that $p_{1,t'} = 1$, a draw from the full conditional of the outlier size is obtained by sampling from:

$$\left(k_{1,t'} | \tilde{y}_{t'}, \alpha_{t'}, \sigma_{\tilde{\varepsilon}}^2, p_{1,t'} = 1 \right) \sim N \left(\tilde{y}_{t'} - \alpha_{t'}, \sigma_{\tilde{\varepsilon}}^2 \right). \tag{5.52}$$

As when sampling from the indicator variable, the perturbation variance should be chosen so that the distribution in (5.52) is not too spread, and there is a high probability of sampling values close to $\boldsymbol{E}\left[\tilde{y}_{t'} - \alpha_{t'} | \tilde{y}_{t'}, \alpha_{t'}, p_{1,t'} = 1, \sigma_{\tilde{\varepsilon}}^2 \right] = k_{1,t'}$. On the other hand, it should be big enough to avoid sampling from a degenerate normal distribution.

We now present an algorithm for estimating the model defined by equation (5.36) and (5.37). The transition innovations are normally distributed as in (5.38), and the state space variable has a diffuse initial condition as in (5.39). The prior distributions are presented in expressions (5.40) to (5.43), and

$$\sigma_{\tilde{\varepsilon}}^2 \sim IG \left(\frac{c_1}{2}, \frac{s_1}{2} \right). \tag{5.53}$$

The parameters of the prior distribution are set *a priori*, together with the perturbation variance $\sigma_{\tilde{\varepsilon}}^2$.

Sampling scheme 5.1:

Stage 1 Run *sampling scheme 3.2*, for the model:

$$\tilde{y}_t = \alpha_t + k_{1,t}p_{1,t} + \tilde{\varepsilon}_t,$$

$$\alpha_{t+1} = \alpha_t + k_{2,t}p_{2,t} + \eta_t,$$

where $\tilde{y}_t = y_t + \tilde{\varepsilon}_t$ and $\tilde{\varepsilon}_t \sim N\left(0, \sigma_{\tilde{\varepsilon}}^2\right)$ are serially uncorrelated and identically distributed, for $t = 1, \ldots, T$. Obtain posterior samples for $\mathbf{k}_1$, $\mathbf{p}_1$, and q_1. From the posterior samples of the indicator variables for the presence of an outlier, obtain the set $\{t_1, \ldots, t_m\}$ of the location of the outliers. From the posterior sample of the variables for the size of outliers, obtain the size of the outliers detected $\left\{\bar{k}_{1,t_1}, \ldots, \bar{k}_{1,t_m}\right\}$.

Stage 2 Run *sampling scheme 3.2* for the model:

$$y_t = \alpha_t + k_{1,t}p_{1,t},$$

$$\alpha_{t+1} = \alpha_t + k_{2,t}p_{2,t} + \eta_t,$$

without sampling from the full conditional of $\sigma_{\varepsilon}^2, \mathbf{k}_1$, $\mathbf{p}_1$ and q_1. The measurement variance is set equal to zero; the indicator variables for the presence of an outlier are set equal to zero for all $t \notin \{t_1, \ldots, t_m\}$. For $t \in \{t_1, \ldots, t_m\}$, $p_{1,t} = 1$. The size of outlier variables are set equal to the sizes obtained in stage one, for $t = t_i$, for $i = 1, \ldots, m$ and set to zero otherwise.

In summary, in the first stage we establish if outliers are present, their location and size, by using the posterior samples of the intervention variables for the presence of outliers. This is accomplished by running the Gibbs sampler for an auxiliary data set. From this run we record the posterior samples for the variables $\mathbf{k}_1, \mathbf{p}_1$ and q_1. In the second stage, having detected the outliers, the Gibbs sampler is run for the original data set, with the position of outliers and sizes as input; we obtain posterior samples for the variables $\sigma_{\eta}^2, \mathbf{k}_2, \mathbf{p}_2$ and q_2.

138

5.3.1 Sensitivity analysis to the choice of $\sigma_{\tilde{\varepsilon}}^2$

A Monte Carlo study is performed to assess the impact of the choice of the variance parameter $\sigma_{\tilde{\varepsilon}}^2$ in (5.45), used to generate the auxiliary data set $\{\tilde{y}\}_{t=1}^T$, as in expression (5.44). *Sampling scheme 5.1* is used to model the artificial data sets generated from the model in equations (5.36) and (5.37), with dimension $T = 100$. The transition equation variance is set to $\sigma_\eta^2 = 1$. An outlier is input at $t = 50$ with size $k_{1,50} = -5$. At instant $t = 75$, a level shift is input with size $k_{2,74} = 5$. For obtaining the auxiliary data set, and run the first stage of *sampling scheme 5.1*, we use four different values for the perturbation variance:

$$\sigma_{\tilde{\varepsilon}}^2 = 0.25, 0.5, 1, 2. \tag{5.54}$$

The prior distributions' parameters used, including for the approximating bounded uniform, are:

$$
\begin{aligned}
c_1 &= 5, u_1 = -5, v_1 = 5, a_1 = 1, b_1 = 100, \\
c_2 &= 5, s_2 = 5, u_2 = -5, v_2 = 5, a_2 = 1, b_2 = 100,
\end{aligned}
$$

and $s_1 = 1, 2.5, 5, 7$, respectively for the values in (5.54).

In Table 5.1, we present a summary of the properties of the posterior samples obtained for the variables related to the detection of outliers, which compose the relevant output from stage one.

The value that implies an higher percentage of detection of the outlier, is $\sigma_{\tilde{\varepsilon}}^2 = 0.5$, with the shock being detected in 84% of the simulated data sets. For $\sigma_{\tilde{\varepsilon}}^2 = 2$, only in 34% of the replications an outlier was detected at $t = 50$. This result is not surprising given that in this case the model being estimated has an irregular standard deviation of approximately 1.4. The outlier to be detected is only, in absolute value, 3.5 times this standard deviation. This is the reason for a lower number of detections of the shock. In contrast, for example when $\sigma_{\tilde{\varepsilon}}^2 = 0.5$, the absolute value of the size of the outlier is 7 times the irregular standard deviation. The percentage of detection of an outlier is decreasing with the increase of the value of $\sigma_{\tilde{\varepsilon}}^2$, exception made to the change from $\sigma_{\tilde{\varepsilon}}^2 = 0.25$ to $\sigma_{\tilde{\varepsilon}}^2 = 0.5$. For the size of the outlier, the value that gives in average more accurate estimates is $\sigma_{\tilde{\varepsilon}}^2 = 0.5$. Also for this value, the variability

Table 5.1: Summary of output from posterior sample averages, across 500 simulated replications, obtained from application of stage 1 of *sampling scheme 5.1*, for $\sigma_\varepsilon^2 = 0.25, 0.5, 1, 2$.

σ_ε^2	Posterior sample mean												
	$\bar{k}_{1,50}$				$\bar{p}_{1,50}$					$\bar{q}_1$			
	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	> 0.5	Mean	SD	$Q_{0.025}$	$Q_{0.975}$
0.25	-4.86	1.26	-7.31	-2.21	0.73	0.31	0.040	1	0.76	0.024	0.004	0.016	0.034
0.5	-4.98	1.12	-7.16	-2.62	0.80	0.28	0.061	1	0.84	0.023	0.005	0.016	0.033
1	-4.96	1.31	-7.47	-2.32	0.62	0.33	0.028	0.999	0.64	0.022	0.003	0.016	0.029
2	-4.71	1.86	-8.19	-1.12	0.38	0.33	0.012	0.994	0.34	0.022	0.004	0.016	0.032

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 500 simulated samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5, across the 500 simulated samples. Based on 10,000 Gibbs draws, discarding the first 5,000.

of the estimates across the replications is smaller. An important point to note is that, although the choice of the perturbation variance affects the number of times the outlier is detected, the sensitivity of the size estimated is much smaller. This feature can be used to make a more efficient choice of the value of $\sigma_{\bar{\varepsilon}}^2$. For example, in this case, by analyzing the results obtained for the choice of $\sigma_{\bar{\varepsilon}}^2 = 2$, we have an indication that, if there is an outlier at $t = 50$, its size should be approximately -5. With this information, the choice of a lower value for the variance is indicated. Therefore, we would run again the first stage of *sampling scheme* 5.1, using a lower value for the variance used for generating the auxiliary data set. In summary, if the initial choice of the variance $\sigma_{\bar{\varepsilon}}^2$ is not optimal, by analyzing the posterior samples for the size of an outlier, a better choice of this value can be done. In this case, this will imply only one additional step in *sampling scheme 5.1*, corresponding to running one more time stage one.

Figure 5.1: Box-plots of posterior sample means of size of outlier, across 500 simulated replications, for $\sigma_{\bar{\varepsilon}}^2 = 0.25, 0.5, 1, 2$.

Irregular variance
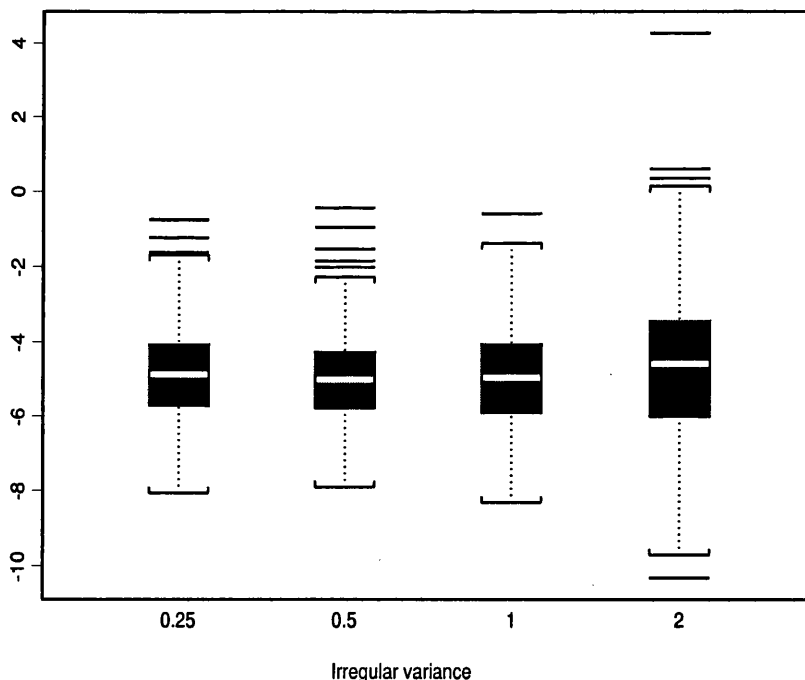
The sensitivity of the size of the outlier to the different choices of the variance can

141

also be observed from the box-plots in Figure 5.1. From these plots we can see that the median of the sizes estimated does not present a high level of sensitivity to the different values of σ_ε^2 used. The variability of the estimates, across the replications, is more sensitive to the choice of this parameter.

The results presented in Table 5.2, are obtained by running the second stage of *sampling scheme 5.1*. The results reported correspond to simulated data sets for which the outlier was detected in the first stage. Suppose we have a data set, and that for certain values of σ_ε^2, the posterior sample mean of the indicator variable $p_{1,50}$ has a value greater than 0.5. Then, the sampling algorithm in stage two is run for this data set, inputting the sizes of outlier estimated for those values of σ_ε^2. As

Table 5.2: Summary of output for posterior sample averages, across simulated replications, obtained from application of stage 2 of *sampling scheme 5.1*, for $\sigma_\varepsilon^2 = 0.25, 0.5, 1, 2$.

σ_ε^2	Posterior sample mean								
	$\bar{k}_{2,74}$				$\bar{p}_{2,74}$				
	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	> 0.5
0.25	5.10	0.98	3.23	6.86	0.94	0.15	0.44	1	0.97
0.5	5.06	0.98	3.20	6.85	0.93	0.16	0.42	1	0.97
1	5.11	0.96	3.27	6.86	0.94	0.14	0.54	1	0.98
2	4.92	0.94	3.28	6.61	0.92	0.17	0.34	1	0.95
	$\bar{\sigma}_\eta^2$				$\bar{q}_2$				
	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	
0.25	0.955	0.13	0.72	1.22	0.022	0.0028	0.017	0.028	-
0.5	0.968	0.13	0.73	1.24	0.022	0.0028	0.017	0.028	-
1	0.973	0.14	0.72	1.24	0.022	0.0027	0.018	0.027	-
2	0.996	0.14	0.74	1.30	0.023	0.0042	0.017	0.035	-

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 500 simulated samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5, across the simulated samples where an outlier was detected at $t = 50$. Based on 10,000 Gibbs draws, discarding the first 5,000.

it might be expected these parameter estimates present very little sensitivity to the choice of the variance σ_ε^2. This result is expected for two reasons. Firstly, we are inputting the estimates of the size of outlier estimated in the first stage only for those cases where the posterior mean probability of having an outlier at $t = 50$ is

superior to 0.5. Secondly, as it was noticed previously, in average the estimate of the size of the outlier presents values very close to the true one, for all values of σ_ε^2 considered.

From this analysis we conclude that although the choice of σ_ε^2 affects the detection of outliers on the first stage, this effect is small regarding the estimated size of the outlier . The choice of the perturbation variance is mainly function of the size of the outliers present in the data. In this way, even when there is not enough prior information for making a suitable choice of this parameter, the results obtained by running the first stage of *sampling scheme 5.1*, can be used for choosing a more appropriate value for σ_ε^2, and rerun stage one of *sampling scheme 5.1* if necessary.

5.3.2   A Monte Carlo study

To illustrate the application of *sampling scheme 5.1* and its properties we present the results obtained with a Monte Carlo study. We generate 5000 simulated data sets, of size $T = 100$, from the model defined by equations (5.36) and (5.37). An outlier and a level shift are input to the data, at instances $t = 50$ and $t = 75$, respectively. The sizes of the shocks are -5, for the outlier, and 5 for the level shift. The true value of the transition equation variance is $\sigma_\eta^2 = 1$. In each of the stages of *sampling scheme 5.1*, the Gibbs sampler is run for 10000 iterations with the first 5000 discarded. For generating the auxiliary data sets, we set the variance parameter to $\sigma_\varepsilon^2 = 0.5$. The parameters for the prior distributions in (5.40) to (5.43) and (5.53) are set to the following values:

$$c_1 = 5, s_1 = 2.5, a_1 = 1, b_1 = 100,$$
$$c_2 = s_2 = 5, a_2 = 1, b_2 = 100.$$

The parameters for the bounded uniform distributions, used when the indicator variable has value one are $u_1 = -5, v_1 = 5, u_2 = -5, v_2 = 5$.

In Table 5.3, we present a summary of the results obtained from running the first Gibbs sampler. We present a summary of the results of all the posterior samples obtained when applying *sampling scheme 5.1* . As was explained previously, from stage 1 we are interested only on the variables related to the detection of outliers.

143

We present the results for the remaining variables only for comparison with the results we obtain for these variables on the second stage.

Table 5.3: Summary of output from posterior samples averages, across 5,000 simulated replications, for a local level model with $\sigma_\varepsilon^2 = 0$.

		Truth	Posterior sample mean					
			Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$	> 0.5
	σ_η^2	1	0.995	0.980	0.210	0.631	1.452	-
	σ_ε^2	0.5	0.498	0.480	0.118	0.318	0.770	-
	$p_{1,50}$	1	0.796	0.921	0.268	0.089	1	0.844
Stage 1	$k_{1,50}$	-5	-4.989	-4.989	1.155	-7.273	-2.826	-
	$p_{2,74}$	1	0.616	0.719	0.336	0.027	1	0.634
	$k_{2,74}$	5	4.946	4.917	2.856	2.250	7.544	-
	q_1	0.01	0.022	0.022	0.003	0.016	0.029	-
	q_2	0.01	0.022	0.022	0.004	0.016	0.032	-
	σ_η^2	1	0.991	0.984	0.139	0.737	1.282	-
	$p_{2,74}$	1	0.913	0.993	0.185	0.263	1	0.942
Stage 2	$k_{2,74}$	5	4.976	4.980	1.009	3.022	6.924	-
	q_2	0.01	0.021	0.021	0.003	0.016	0.028	-

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 5000 simulated samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5. The results referring to stage 2, correspond to the simulated data sets for which an outlier was detected at $t = 50$. Based on 10,000 Gibbs draws, discarding the first 5,000.

For 84% of the simulated data sets, the posterior probability of having an outlier at $t = 50$, is higher than 0.5, and therefore the position of the outlier was correctly estimated. The size estimated for that shock, is on average, very close to the true one. The average value estimated for the irregular parameter of the auxiliary data set is not relevant to the original model estimation. However we note that it presents a very accurate value of 0.498, for a true value of 0.5.

The second stage is run for the subset of simulated data sets for which an outlier was detected at $t = 50$, which amount to 4220 simulated replications. In this stage, the variables for the detection of that type of shocks are not sampled. The size of the outlier is set to the value estimated when running the sampler for the auxiliary data set. For this reason, it would be meaningless to run stage 2 for the data sets

where an outlier was not detected, as the results obtained for stage two would be biased by the non identification of the outlier.

The rate of detection of the level shift presents an higher value than for the detection of outliers. In around 94% of the cases, the level shift was detected. The average estimated value for the size of the level shift gives a very accurate estimate of this parameter.

The results for the posterior sample of the intervention variables for the presence of a level shift obtained when running stage one are not used, for fitting the model. We make a brief comparison of the results obtained for these variables in stage 1 and stage 2. The average size of a level shift is quite similar in both stage. However, the posterior probability of detecting a level shift at $t = 75$, presents higher values for stage 2. This implies that by using the posterior samples from stage 2, we obtain better results in terms of the number of times the level shift input at that instant is detected.

To conclude this analysis in Figure 5.2, we present some plots to characterize the distribution of the posterior average of the size of the shocks, obtained for the simulated data sets. Both type of shocks present a very symmetric empirical distribution, around the parameter's true value. In terms of dispersion, the size of an outlier is slightly more dispersed. That is confirmed by the values of the standard deviations, of 1.16 and 1.01, for the outliers and level shifts, respectively.

5.3.3 Empirical application: bonds data set

Figure 5.3 shows the closing prices of bonds issued by the Greek government in 1910. They were traded on the London Stock Exchange, from August 1916 to June 1930, making a total of 167 observations. The observations presented correspond to a percentage of the par value quoted monthly at the London Stock Exchange. This data set is used by Christodoulaki and Penzer (2002), where a more detailed overview can be found. In Christodoulaki and Penzer (2002), an analysis of this data set is done with the main purpose of identifying possible shocks to the data set. We shall compare our results with those obtained in that study.

Given that the data is monthly, the model we estimate includes a seasonal com-

Figure 5.2: Histograms, and box-plots of estimates of size of outlier (a)-(b) and level shift (c)-(d), across 5,000 simulated replications, for a local level model with $\sigma_\varepsilon^2 = 0$.
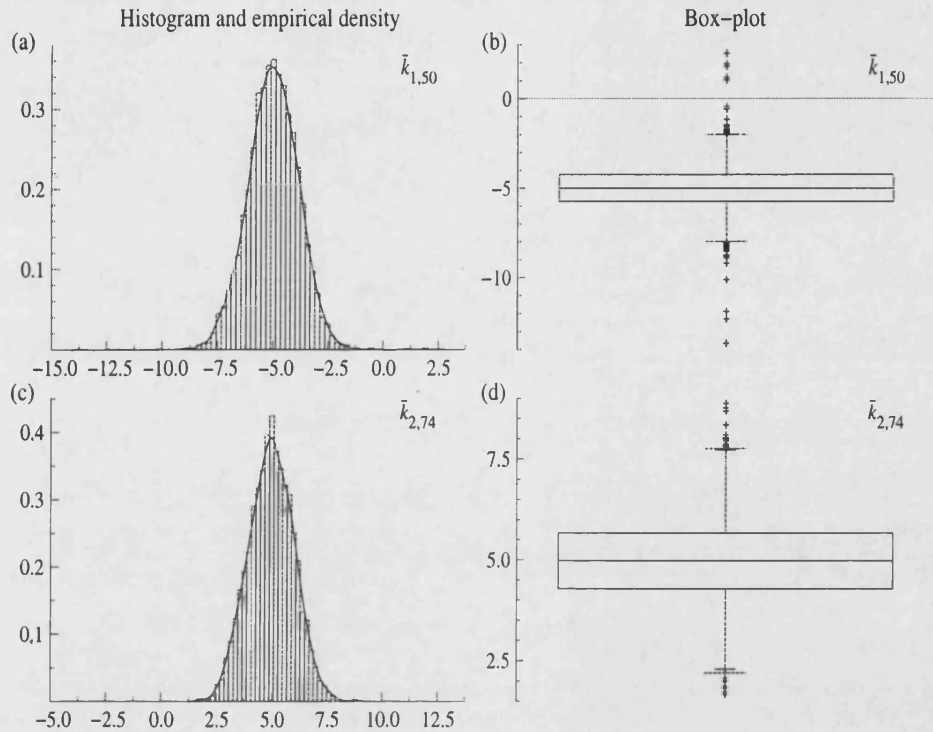
ponent. We start by estimating an unobserved component model with irregular, level and seasonal component. When estimating this model without including intervention variables, we obtain that the irregular variance is estimated as zero. On the other hand, the analysis of the innovations and auxiliary residuals indicate the possible existence of outliers and level shifts. Given this initial analysis, the model we estimate is defined by:

$$y_t = \mu_t + \gamma_t + k_{1,t}p_{1,t}, \tag{5.55}$$

$$\mu_{t+1} = \mu_t + k_{2,t}p_{2,t} + \eta_t, \tag{5.56}$$

$$\sum_{i=0}^{s-1} \gamma_{t+1-1} = \omega_t, \tag{5.57}$$

with

$$\eta_t \sim N\left(0, \sigma_\eta^2\right),$$

146

Figure 5.3: Monthly quotes of Greek Government bonds from August 1916 to June 1930.
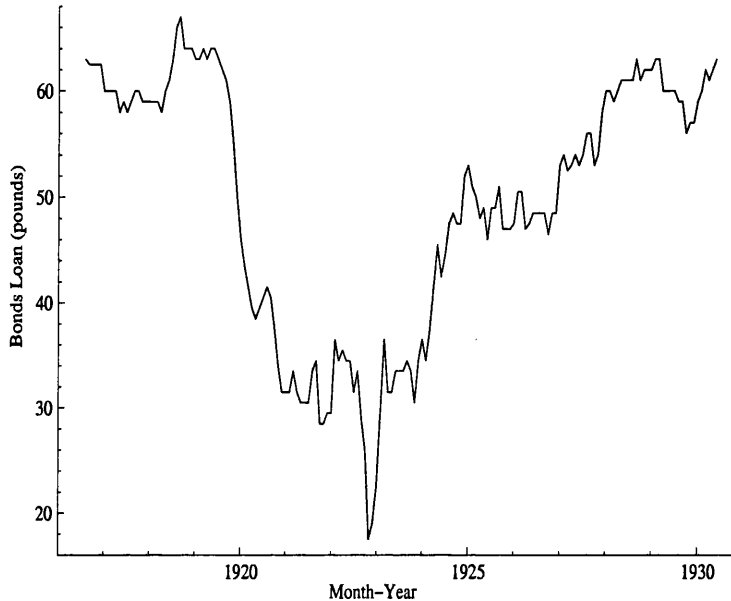
$$\omega_t \sim N\left(0, \sigma_\omega^2\right),$$

for $t = 1, \ldots, T$.

Following the algorithm described in *sampling scheme 5.1*, we construct an auxiliary data set $\{\tilde{y}_t\}_{t=1}^T$. We run the Gibbs sampler for $\{\tilde{y}_t\}_{t=1}^T$, for obtaining posterior samples for the variables for the presence of outliers. The auxiliary data set is obtained as in expression (5.44). From an initial inspection of the data we would expect the presence of an outlier in the year 1923. If we take the difference of two consecutive observations for that year, we get as maximum absolute difference the value 7. We take this value as a rough approximation of what would be the size of an outlier, if present in that period. This gives an indication of the dimensions of the variance σ_ε^2 to be used. We take $\sigma_\varepsilon^2 = 0.5$. If in fact there is an outlier of size approximately 7 present in the data set, its size is 10 times the standard deviation of the irregular component, for the auxiliary data set, and should be correctly detected.

The model we estimate in this first stage is defined as:

$$\tilde{y}_t = \mu_t + k_{1,t}p_{1,t} + \tilde{\varepsilon}_t,$$

$$\mu_{t+1} = \mu_t + k_{2,t}p_{2,t} + \eta_t,$$

$$\sum_{i=0}^{s-1} \gamma_{t+1-1} = \omega_t.$$

The prior distributions are defined as in expressions (5.40) to (5.43), (5.53), and

$$\sigma_\omega^2 \sim IG\left(\frac{c_3}{2}, \frac{s_3}{2}\right).$$

The values for the prior distributions parameters are set to:

$$c_1 = 5, s_1 = 2, a_1 = 1, b_1 = 100,$$

$$c_2 = c_3 = 5, s_2 = 20, s_3 = 0.01, a_2 = 4, b_2 = 100, \qquad (5.58)$$

and for the approximating bounded uniform distributions $u_1 = -5, v_1 = 5, u_2 = -10, v_2 = 10$. The Gibbs sampler defined in *sampling scheme 3.2* is applied, with an additional step for sampling from the seasonal component variance full conditional distribution. The sampler is run for 100,000 iterations, with the first 50,000 discarded. In Figure 5.4 we plot the mean of the posterior samples for the intervention variables, for the presence of outliers.

From this stage we are interested in detecting the position and size of outliers. We have two outstanding values for the posterior mean probability of an outlier, at 1919-12 and 1923-3, with mean posterior probability of an outlier of 0.35 and 0.44, respectively. In 1919-12, the posterior mean size of a shock has a meaningless value of -0.37; the relatively high value of the posterior probability, when compared with the rest of the sample, is related to the possible occurrence of a level shift in the previous instant, as we will see when analyzing the results obtained for the second stage of the estimation. At 1923-3, the correspondent posterior mean size of an outlier presents a value of 5.77. Using a threshold of 0.5 to detect a shock at a certain instant if the mean posterior probability of having a shock is superior to that value, we would not detect any outlier. Following the arguments presented in §3.4.1, it is advantage to analyze the posterior means of probability of shocks for

Figure 5.4: Plots of posterior means of indicator (a) and size (b) of an outlier, for bonds data set, from 1916-8 to 1930-6.
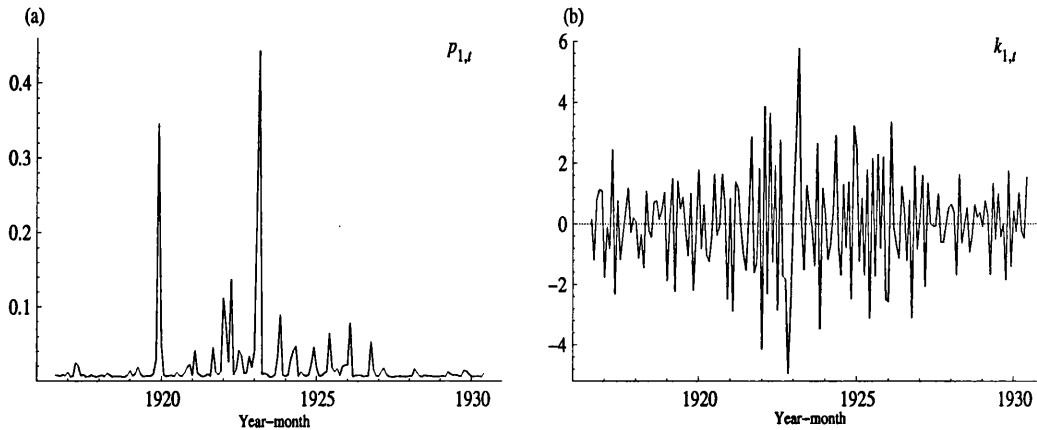
the sample as a whole. From that we conclude that, given its outstanding value and proximity to the boundary value of 0.5, there is an outlier of size 5.77 in March of 1923.

Having detected and characterized the outliers present in the data, we run the second stage of *sampling scheme 5.1*, to obtain the posterior samples of the remaining parameters of the model defined by equations (5.55) to (5.57). This amounts to running the Gibbs sampler to obtain posterior samples of the level and seasonal variances, intervention variables for the presence of level shifts, and the prior probability q_2. The prior distributions parameters are set to the values in (5.58). The Gibbs sampler is run for 100,000 iterations, discarding the first 50,000. In Figure 5.5, we present the posterior mean of the interventions variables for the presence of a level shift.

There are four outlying observations, with posterior mean probability of being a level shift superior to 0.5, corresponding to level shifts in 1919-12, 1922-2, 1922-11 and 1923-2. In Table 5.4 we summarize the results for the instances where shocks were detected. The samples used to obtain the statistics presented in Table 5.4, for the size of the shocks, are sub samples of the posterior sample obtained for those variables. They correspond to those iterations of the Gibbs sampler in which the correspondent indicator variable was sampled with value one. The outlier posterior

149

Figure 5.5: Plots of posterior means of indicator (a) and size (b) of level shift, for bonds data set, from 1916-8 to 1930-6.
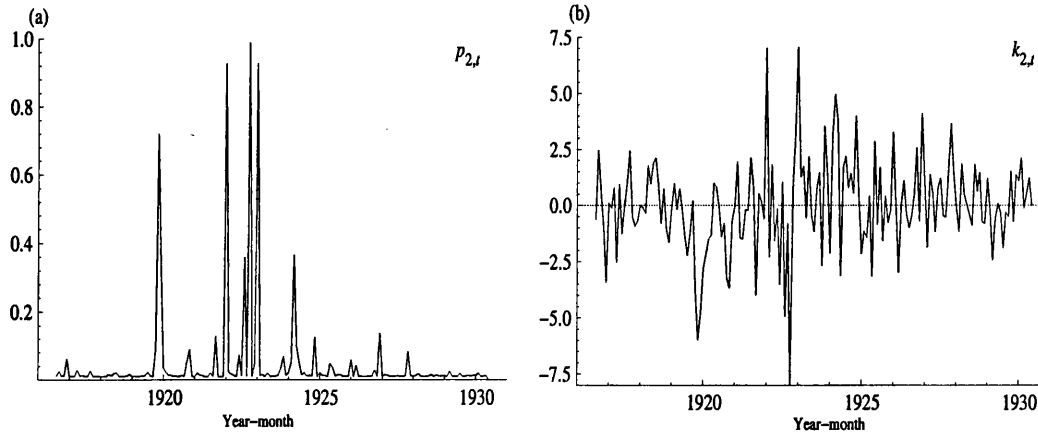
sample presents the higher standard deviation as well as less symmetry. These observations are confirmed by inspection of the plots for these posterior samples, in Figures 5.6 and 5.7. Comparing the posterior distributions of the size of the level shifts detected, they all present an approximately symmetrical empirical distribution.

In Table 5.5 we summarize the properties of the posterior samples of the hyperparameters and the prior probabilities q_1 and q_2. Note that the posterior sample for the prior probability of an outlier was obtained when running the first Gibbs sampler in *sampling scheme 5.1*.

For comparison purpose, in Table 5.6 we present the results obtained by maximum likelihood, given the position of the shocks. These are the results presented in Christodoulaki and Penzer (2002). In Christodoulaki and Penzer (2002), the position of the shocks is established by estimating the model without interventions and using its estimation output to localize the shocks (see De Jong and Penzer, 1998). Comparing the results in Tables 5.4 and 5.5 with the ones in Table 5.6 we can see that the results obtained are very similar. Our method has the advantage of being based on the estimation of the hyperparameters and shocks together.

To check the adequacy of our estimates in Figures 5.8 and 5.9, we present plots for the standardized innovations and standardized auxiliary residuals, respectively.

Table 5.4: Summary of output from posterior samples of intervention variables for outliers and level shifts detected, for the bonds data set, from 1916-8 to 1930-6.

Shock	Year-month	Posterior sample					
		Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$	$\bar{p}_{i,t}$
Outlier	1923-3	5.771	5.241	4.02e-03	2.136	16.053	0.443
Level shift	1919-12	-5.968	-5.985	8.03e-05	-9.304	-2.525	0.722
	1922-2	7.013	7.023	4.98e-05	3.552	10.414	0.928
	1922-11	-7.987	-7.985	4.62e-05	-11.441	-4.518	0.990
	1923-2	7.044	7.065	5.76e-05	3.564	10.413	0.929

The values reported are the mean, median, 2.5th and 97.5th percentile of subsets of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated with a bandwidth of 5,000 for the Parzen window) for the sizes of the shocks detected. The subsets of the posterior samples correspond to iterations where the indicator variable was sampled with value one. "$\bar{p}_{i,t}$" is the posterior sample mean of the indicator variable. Based on 100,000 Gibbs draws, discarding the first 50,000.

From the plots in Figure 5.8, we see that the standardized innovations present a behaviour that indicates they are independently generated from a standard normal distribution. The non existence of outstanding values for the standardized auxiliary residuals implies that all the shocks existent in the data set were correctly accounted for. In conclusion, the model satisfactorily fits the data, in what concerns the hyperparameters estimated and the shocks detected.

5.4 Detection of outliers and level shifts when $\sigma_\eta^2 = 0$

We consider now the problem of estimating a local level model, where outliers and level shifts might be present, when the transition equation variance is equal to zero, $\sigma_\eta^2 = 0$. The SSM formulation is the following:

$$y_t = \alpha_t + k_{1,t}p_{1,t} + \varepsilon_t, \qquad (5.59)$$

$$\alpha_{t+1} = \alpha_t + k_{2,t}p_{2,t}, \qquad (5.60)$$

with

$$\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \qquad (5.61)$$

Figure 5.6: Histogram and box-plot of posterior sample for size of outlier detected in 1923-3, for the bonds data set, from 1916-8 to 1930-6.
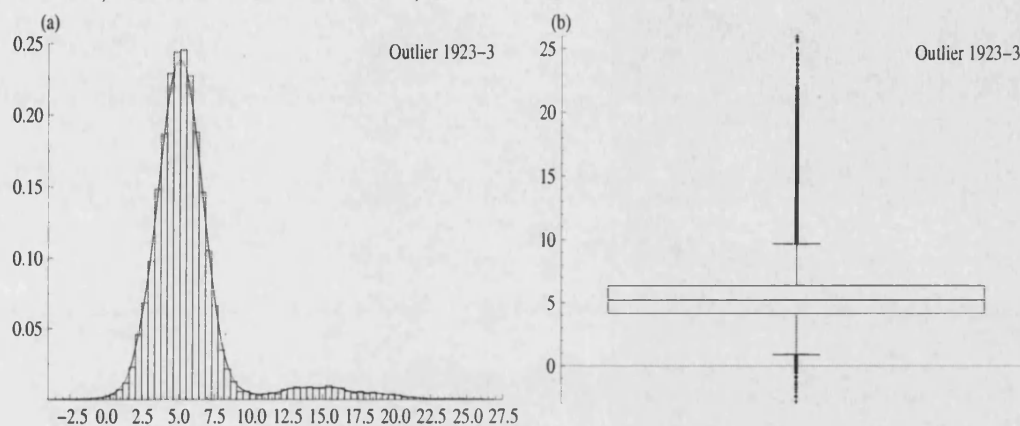
Table 5.5: Summary of output from posterior samples of the hyperparameters, and prior probabilities of an outlier and a level shift, for the bonds data set, from 1916-8 to 1930-6.

| | Prior | | Posterior sample | | | |
	Mean	SD	Mean	SD	$Q_{0.025}$	$Q_{0.975}$
σ_η^2	6.67	9.43	2.868	0.00605	2.146	3.797
σ_ω^2	0.0033	0.0047	0.003062	1.17e-04	7.81e-04	0.0106
q_1	0.0099	0.0098	0.0168	8.27e-04	7.15e-04	0.0493
q_2	0.038	0.019	0.0433	1.61e-04	0.0171	0.0778

The values reported are the mean, 2.5th and 97.5th percentile of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated with a bandwidth of 5,000 for the Parzen window). Based on 100,000 Gibbs draws, discarding the first 50,000.

152

Figure 5.7: Histograms and box-plots of posterior samples for size of level shifts detected in 1919-12 (a)-(b), 1922-2 (c)-(d), 1922-11 (e)-(f) and 1923-2 (g)-(h), for the bonds data set, from 1916-8 to 1930-6.
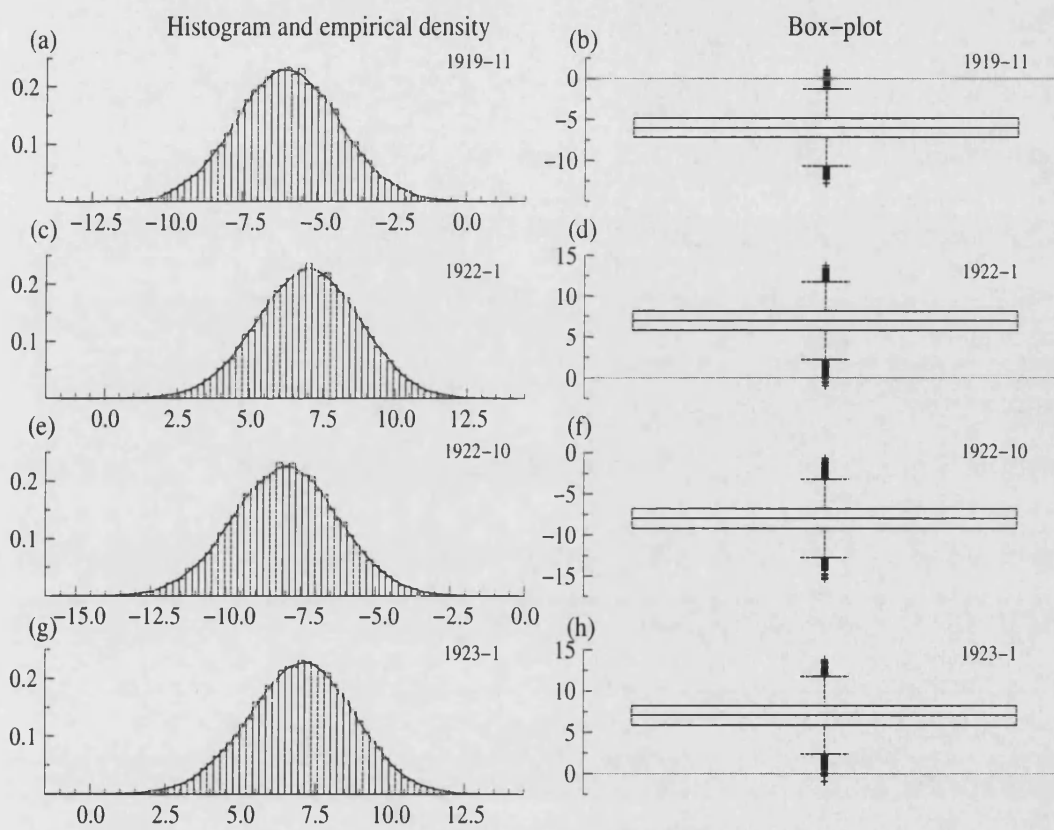
Figure 5.8: Descriptive plots of standardized innovations, for bonds data set, from 1916-8 to 1930-6.

(a) Plot of standardized innovations

(b) Box-plot of standardized innovations

(c) ACF for standardized innnovations

(d) QQ plot of standardized innovations
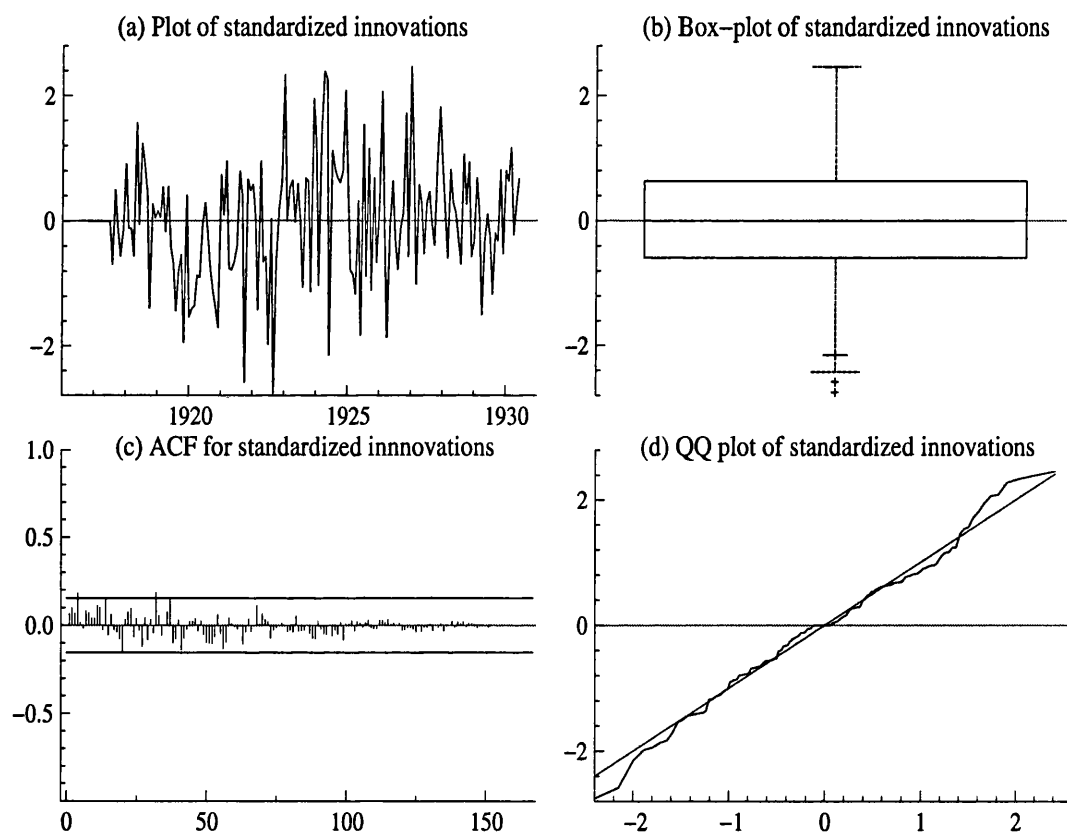
Figure 5.9: Plots (a)-(b) and box-plots (c)-(d) of standardized auxiliary residuals, for bonds data set, from 1916-8 to 1930-6.
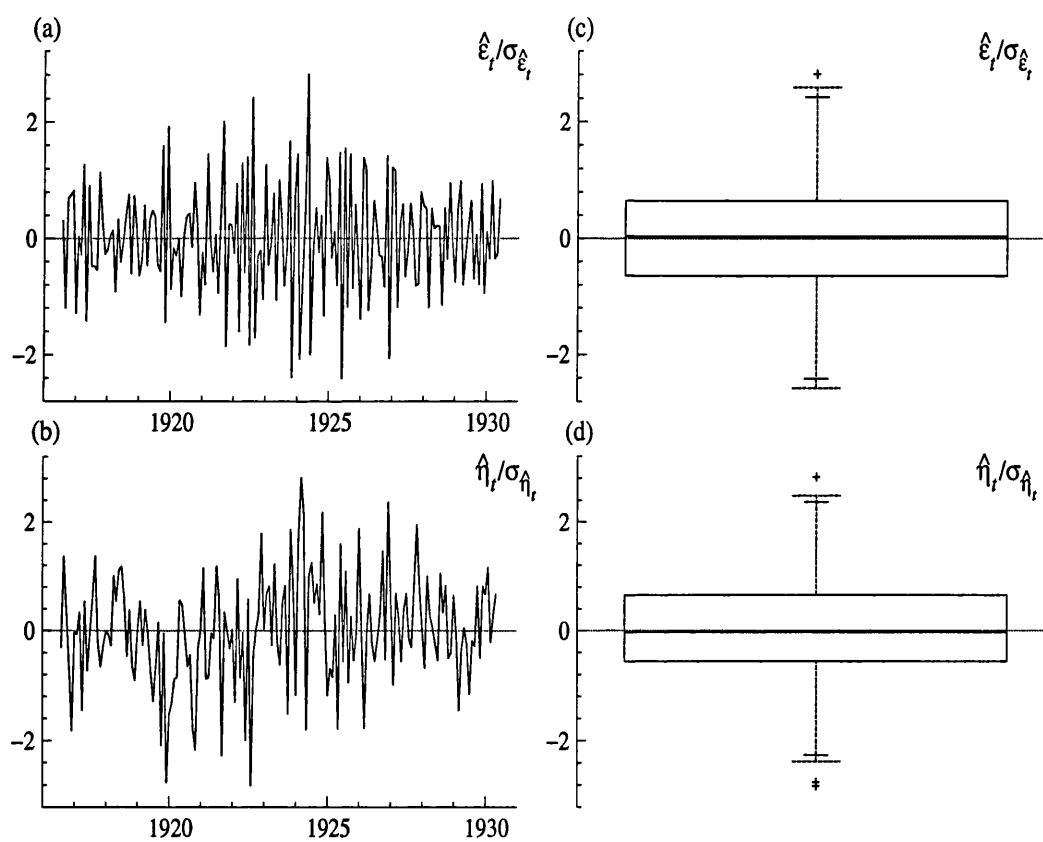
Table 5.6: Estimation results by maximum likelihood, given the position of the shocks, for the bonds data set, from 1916-8 to 1930-6.

σ_ϵ^2	σ_η^2	σ_ω^2	Outlier	Level shift			
			1923-3	1919-12	1922-2	1922-11	1923-2
0.0000	2.915	0.00316	5.641	-6.065	7.062	-7.944	7.09
[0.0000]	[1.0000]	[0.0011]	(4.473)	(−3.413)	(3.970)	(−4.480)	(3.990)

The values in [ ] are the q-ratios for the hyperparameters estimated; the values in ( ) are the t-ratios for the size of shocks estimated.

for $t = 1, \ldots, T$.

A motivation for considering this type of models is the well known flow of the Nile data set (Cobb, 1978). This data set is modeled by a local level model. It has been established by several studies that it presents a level shift (see references in §5.4.3). When this shock is correctly detected, the transition equation variance becomes statistically null. Hence, it fits in the formulation defined by equations (5.59) and (5.60). We shall consider this data set as an empirical application of the method we propose.

The methodology we propose follows the same philosophy as the methodology used for *sampling scheme 5.1*. We estimate this model in two stages. In a first stage, we run the Gibbs sampler for an auxiliary data set $\{y_t^*\}_{t=1}^T$, to establish the location and size of level shifts possibly present in the original data set $\{y_t\}_{t=1}^T$. The auxiliary data set is constructed in such a way that it has the same level shifts, in terms of their location and size, as the original data set. It follows a local level model with the transition variance different from zero. The aim of using this auxiliary data set it to identify the level shifts, their position and size. Having identified and characterized these shocks, the original model is estimated using a second run of the Gibbs sampler; from this second run, we obtain posterior samples for the remaining random variables of the model.

Next we derive the process to obtain $\{y_t^*\}_{t=1}^T$. Let,

$$\eta_t^* \sim N\left(0, \sigma_{\eta^*}^2\right), \tag{5.62}$$

for $t = 0, \ldots, T - 1$, be a process of serially uncorrelated random variables. The

auxiliary data set is obtained in the following manner:

$$y_t^* = y_t + \sum_{i=0}^{t-1} \eta_i^*, \tag{5.63}$$

for $t = 1, \ldots, T - 1$.

If in equations (5.59) and (5.60) we add to both hand sides $\sum_{i=0}^{t-1} \eta_i^*$ and $\sum_{i=0}^{t} \eta_i^*$, respectively, we obtain,

$$y_t + \sum_{i=0}^{t-1} \eta_i^* = \alpha_t + \sum_{i=0}^{t-1} \eta_i^* + k_{1,t}p_{1,t} + \varepsilon_t, \tag{5.64}$$

$$\alpha_{t+1} + \sum_{i=0}^{t} \eta_i^* = \alpha_t + \sum_{i=0}^{t} \eta_i^* + k_{2,t}p_{2,t}. \tag{5.65}$$

Define

$$\alpha_t^* = \alpha_t + \sum_{i=0}^{t-1} \eta_i^*. \tag{5.66}$$

With definitions in expressions (5.63) and (5.66), equations (5.64) and (5.65) are rewritten as:

$$y_t^* = \alpha_t^* + k_{1,t}p_{1,t} + \varepsilon_t, \tag{5.67}$$

$$\alpha_{t+1}^* = \alpha_t^* + k_{2,t}p_{2,t} + \eta_t^*. \tag{5.68}$$

Hence, we obtain a local level model, defined by equations (5.67) and (5.68). The intervention variables in this model are the same as in the original model. The choice of the noise process η_t^*, more specifically of its variance, is such that it does not mask the level shifts that might exist in the original data set.

To assess the impact of the value chosen for $\sigma_{\eta^*}^2$, we next analyze the full conditional distributions used in the first stage, to draw the intervention variables for the presence of level shifts.

Consider the model defined by equations (5.67) and (5.67). Suppose there is a level shift at $t = t'$ of size $k_{2,t'}$, and that the hyperparameters of the model are known. To draw from the full conditional distribution for the indicator variable of a level shift at $t = t'$, we draw from a Bernoulli distribution with probabilities given

by:

$$P\left(p_{2,t'} = 1|\alpha^*_{t'+1}, \alpha^*_{t'}, \sigma^2_{\eta^*}, k_{2,t'}, q_2\right) \quad \propto \quad f_N\left(\alpha^*_{t'+1}|\alpha^*_{t'} + k_{2,t'}, \sigma^2_{\eta^*}\right) q_2, \qquad (5.69)$$

$$\propto \quad \exp\left[-\frac{\left(\alpha^*_{t'+1} - \alpha^*_{t'} - k_{2,t'}\right)^2}{2\sigma^2_{\eta^*}}\right] q_2,$$

and

$$P\left(p_{2,t'} = 0|\alpha^*_{t'+1}, \alpha^*_{t'}, \sigma^2_{\eta^*}, k_{2,t'}, q_2\right) \quad \propto \quad f_N\left(\alpha^*_{t'+1}|\alpha^*_{t'}, \sigma^2_{\eta^*}\right)(1 - q_2), \qquad (5.70)$$

$$\propto \quad \exp\left[-\frac{\left(\alpha^*_{t'+1} - \alpha^*_{t'}\right)^2}{2\sigma^2_{\eta^*}}\right](1 - q_2),$$

Using the definition of the state variables α^*_t in (5.66), the probabilities for the presence of a level shift in expressions (5.69) and (5.70), are proportional to

$$\exp\left[-\frac{\left(\alpha_{t'+1} - \alpha_{t'} - k_{2,t'} + \eta^*_t\right)^2}{2\sigma^2_{\eta^*}}\right] q_2,$$

and

$$\exp\left[-\frac{\left(\alpha_{t'+1} - \alpha_{t'} + \eta^*_t\right)^2}{2\sigma^2_{\eta^*}}\right](1 - q_2),$$

respectively. The argument for the choice of the perturbation variance is similar to the one used when detecting outliers, in §5.3. If there is a level shift in t' the value of $\alpha_{t'+1} - \alpha_{t'}$ should be significantly different from zero. The choice of $\sigma^2_{\eta^*}$ should be such that the value $\alpha_{t'+1} - \alpha_{t'} + \eta^*_{t'}$ has a low probability under a normal distribution $N\left(0, \sigma^2_{\eta^*}\right)$. On the other hand it should not be too small in order to avoid that the full posterior distribution becomes degenerated. As in §5.3, the choice of the perturbation variance can be done by a previous analysis of the data set in study, and does not require an extensive *a priori* knowledge of the data.

We describe now how to sample from the full conditional distribution of the size of a level shift variable. We assume that all the parameters of the model are known, including the position of the level shift at $t = t'$. Hence we have $p_{2,t'} = 1$. The shock size is drawn from the distribution,

$$\left(k_{2,t'}|\sigma^2_{\eta^*}, \alpha^*_{t'+1}, \alpha^*_{t'}, p_{2,t'} = 1\right) \sim N\left(\alpha^*_{t'+1} - \alpha^*_{t'}, \sigma^2_{\eta^*}\right). \qquad (5.71)$$

158

From expression (5.71) we conclude that the choice of $\sigma_{\eta^*}^2$ should be such that avoids degeneracy of the normal distribution in expression (5.71). On the other hand, it should not imply a too high degree of dispersion, so that the values sampled for the level shift size are likely to be close to $E\left[\alpha_{t'+1}^* - \alpha_{t'}^* | \sigma_{\eta^*}^2, \alpha_{t'+1}^*, \alpha_{t'}^*\right] = k_{2,t'}$. We now present a sampling scheme for estimating the model defined by equations (5.59) and (5.60), with the measurement innovation distributed as in (5.61). The prior distributions assumed for the parameters of the model are defined in expressions (5.41) to (5.43), together with

$$\sigma_\varepsilon^2 \sim IG\left(\frac{c_1}{2}\frac{s_1}{2}\right),\tag{5.72}$$

$$\sigma_{\eta^*}^2 \sim IG\left(\frac{c_2}{2}\frac{s_2}{2}\right).\tag{5.73}$$

Sampling scheme 5.2:

Stage 1 Run *sampling scheme 3.2*, for the model:

$$y_t^* = \alpha_t^* + k_{1,t}p_{1,t} + \varepsilon_t,$$

$$\alpha_{t+1}^* = \alpha_t^* + k_{2,t}p_{2,t} + \eta_t^*,$$

where $y_t^* = y_t + \sum_{i=0}^{t-1}\eta_i^*$ and $\eta_t^* \sim N\left(0, \sigma_{\eta^*}^2\right)$ are serially uncorrelated and identically distributed, for $t = 0, \ldots, T - 1$. From the posterior samples of indicator variables for the presence of a level shift, obtain the set $\{t_1, \ldots, t_r\}$ of the location of level shifts detected. From the posterior sample of the variables for the size of level shifts, obtain the size of the level shifts previously detected $\left\{\bar{k}_{2,t_1}, \ldots, \bar{k}_{2,t_r}\right\}$.

Stage 2 Run *sampling scheme 3.2* for the model:

$$y_t = \alpha_t + k_{1,t}p_{1,t} + \varepsilon_t,$$

$$\alpha_{t+1} = \alpha_t + k_{2,t}p_{2,t},$$

without sampling from the full conditionals of σ_η^2, $\mathbf{k}_2$ and $\mathbf{p}_2$. The transition variance is set equal to zero; the indicator variables for the presence of a level shift are set equal to zero for all $t \notin \{t_1, \ldots, t_r\}$. For $t \in \{t_1, \ldots, t_r\}$, $p_{2,t} = 1$. The size of level shift variables are set equal to the sizes previously obtained, for $t = t_i$, for $i = 1, \ldots, r$ and set to zero otherwise.

159

By running the Gibbs sampler in stage 1, we identify the level shifts present in the data, by considering the posterior distributions of the intervention variables for the presence of level shifts. These posterior samples are obtained for the auxiliary model. By construction, the auxiliary data set has the same level shifts as the original data set, in terms of their location and size. From stage 1 we obtain posterior samples for k_2, p_2, and q_2. Having obtained the set of instances where level shifts occur and their sizes, the Gibbs sampler is run once more, applied to the original model and data set. It delivers the posterior samples for the measurement equation variance and the variables related to the presence of an outlier, namely $\sigma_\epsilon^2, k_1, p_1$, and q_1.

5.4.1   Sensitivity analysis to the choice of $\sigma_{\eta^*}^2$

We conduct a Monte Carlo experiment, for 500 simulated data sets, with size $T = 100$, from the model defined by equations (5.59) and (5.60). The variance of the irregular noise in expression (5.61), is set to one. For each data set an outlier of size -5, and a level shift of size 5, are input, at instances $t = 50$ and $t = 75$, respectively.

For each of the simulated data sets, stage one of *sampling scheme 5.2*, is run. Four different values for the variance $\sigma_{\eta^*}^2$ in expression (5.62), are considered, for generating the auxiliary data set, using equation (5.63):

$$\sigma_{\eta^*}^2 = 0.25, 0.5, 1, 2.$$

From stage one, we obtain posterior samples for the size and probability of having a level shift at $t = 75$. Using a threshold of 0.5, the level shift is detected for part of the generated samples, and for each of these, for certain values of $\sigma_{\eta^*}^2$. For this generated samples, stage 2 is run, fixing the size of the level shift, at $t = 75$, to the weighted mean of the posterior sample of $k_{2,74}$ obtained in the previous stage. In stage 2, we obtain posterior samples for the variables related to the detection of an outlier and the measurement equation variance.

The prior assumptions are defined in expressions (5.41) to (5.43), and (5.72) to (5.73). The prior distribution parameters are set to the values

$$c_1 = 5, s_1 = 5, a_1 = 1, b_1 = 100,$$
$$c_2 = 5, a_2 = 1, b_2 = 100,$$

For the approximating bounded uniform distribution we take $u_1 = -5, v_1 = 5, u_2 = -5, v_2 = 5$. The prior parameters for the prior distribution of the auxiliary noise in (5.73) are set to

$$s_2 = 1, 2.5, 5, 7,$$

for $\sigma_{\eta^*}^2 = 0.25, 0.5, 1, 2$, respectively. The Gibbs samplers in stage one and two, are run for 10,000 iterations, discarding the first 5,000.

We start by presenting in Table 5.7, a summary of the properties of the posterior samples obtained from stage 1. The value for $\sigma_{\eta^*}^2$ that presents best results is $\sigma_{\eta^*}^2 = 0.25$. For this value, we obtain the higher percentage of detection of the level shift. For $\sigma_{\eta^*}^2 = 2$, in only 27% of the cases, a level shift is detected at $t = 75$. In this case, we are trying to detect a shock of size 5 to an equation with a hyperparameter of 2. The impact of the shock is not so evident given the dimension of the noise variance, and for that reason is not detected as often as for the other values of $\sigma_{\eta^*}^2$ considered. On average, the posterior mean for the size of the shock, does not exhibit much sensitivity to the choice of the perturbation variance. The standard deviation of this estimate across the simulated replications, presents some sensitivity, increasing with the increase of the perturbation variance.

In Figure 5.10, we plot the posterior sample means of the size of a level shift at $t = 75$, across the simulated replications, for the values of $\sigma_{\eta^*}^2$ used. The median of the sizes estimated, is quite stable, in the same way we have observed for the mean. This lack of sensitivity of these central measures can be used for choosing the value of $\sigma_{\eta^*}^2$. The choice of this parameter can be tuned according to the values obtained for the weighted average of the posterior sample of the size of a level shift, for those instances where the posterior probability of such an event is higher.

Taking the generated data sets for which a level shift is detected in stage 1, we run the second stage of *sampling scheme 5.2*. The size of the level shift is fixed to the estimate obtained in the first stage, accordingly to the value of $\sigma_{\eta^*}^2$ used. In Table 5.8, we present a summary of the posterior samples obtained in this stage. The values of $\sigma_{\eta^*}^2$ that imply more accurate estimates in average, are for $\sigma_{\eta^*}^2 = 0.25, 0.5$. For these values the percentage of outliers detected at $t = 50$ presents its higher

Table 5.7: Summary of output from posterior sample averages, across 500 simulated replications, obtained from application of stage 1 of *sampling scheme 5.2*, for $\sigma^2_{\eta^*} = 0.25, 0.5, 1, 2$.

$\sigma^2_{\eta^*}$	Posterior sample mean												
	$\bar{k}_{2,74}$				$\bar{p}_{2,74}$					$\bar{q}_2$			
	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	> 0.5	Mean	SD	$Q_{0.025}$	$Q_{0.975}$
0.25	4.88	1.18	2.45	6.99	0.76	0.30	0.04	1	0.80	0.022	0.0040	0.016	0.031
0.5	4.89	1.30	2.15	7.16	0.70	0.32	0.03	1	0.73	0.023	0.0044	0.016	0.032
1	4.90	1.54	1.74	7.81	0.50	0.36	0.02	0.999	0.50	0.023	0.0045	0.017	0.032
2	4.75	1.97	1.01	8.58	0.32	0.31	0.01	0.965	0.27	0.024	0.0050	0.018	0.036

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 500 simulated samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5, across the 500 simulated samples. Based on 10,000 Gibbs draws, discarding the first 5,000.

162

Figure 5.10: Box-plot of posterior samples mean of size of level shift, across 500 simulated replications, for $\sigma_{\eta*}^2 = 0.25, 0.5, 1, 2$.
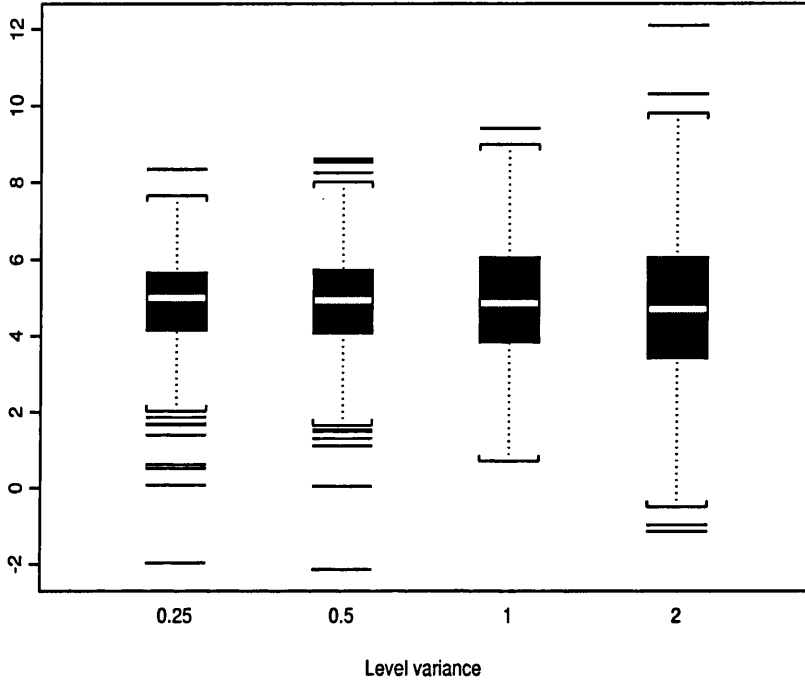
values. Comparing the sensitivity of the results obtained in the second stage, to the choice of $\sigma_{\eta*}^2$, to the results in Table 5.2, for the sensitivity to the choice of σ_ε^2, we conclude that the first ones are more affected by the choice of the perturbation. The justification for this is related to the nature of the shock. A level shift happening at t has impact on all the following observations. If in stage one a level shift is detected at a certain instant t, its estimate is input for running stage 2. The size of the level shift estimated defines the change in the level component of the model for all the following instances.

From the results obtained for this analysis, we conclude that the size of level shift variable presents a satisfactory lack of sensitivity to the choice of $\sigma_{\eta*}^2$. This feature can be used to make an adequate choice of this parameter, without much prior knowledge of the data set being fit. With a sensible choice of the perturbation variance, the estimation results obtained are on average satisfactory.

Table 5.8: Summary of output from posterior sample averages, across simulated replications, obtained from application of stage 2 of *sampling scheme 5.1*, for $\sigma_{\eta*}^2 = 0.25, 0.5, 1, 2$.

$\sigma_{\eta*}^2$	Posterior sample mean								
	$\bar{k}_{1,50}$				$\bar{p}_{1,50}$				
	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	> 0.5
0.25	-4.90	1.08	-6.84	-2.86	0.849	0.25	0.13	1	0.870
0.5	-4.84	1.05	-6.84	-2.88	0.825	0.27	0.08	1	0.857
1	-4.60	1.07	-6.75	-2.65	0.725	0.33	0.03	1	0.770
2	-4.10	1.13	-6.30	-2.07	0.495	0.38	0.02	0.9999	0.471
	$\bar{\sigma}_\varepsilon^2$				$\bar{q}_1$				
	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	Mean	SD	$Q_{0.025}$	$Q_{0.975}$	
0.25	1.123	0.271	0.790	1.844	0.022	0.0027	0.016	0.027	-
0.5	1.190	0.394	0.803	2.358	0.022	0.0029	0.016	0.028	-
1	1.397	0.614	0.804	3.423	0.021	0.0030	0.015	0.027	-
2	1.992	1.181	0.831	4.282	0.021	0.0033	0.016	0.028	-

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 500 simulated samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5, across the simulated samples where a level shift was detected at $t = 75$. Based on 10,000 Gibbs draws, discarding the first 5,000.

5.4.2 A Monte Carlo study

We conduct a Monte Carlo study, where we generate 5,000 simulated data sets, of size $T = 100$, using the model defined by equations (5.59) and (5.60). The measurement equation noise, is generated from a normal distribution as in (5.61), with $\sigma_\varepsilon^2 = 1$. To assess the performance of *sampling scheme 5.2*, an outlier and a level shift are input to the data set. An outlier at instant $t = 50$, of size -5. A level shift at $t = 75$ with size 5. For each data set, we run the first stage of *sampling scheme 5.2*, for 10,000 iterations, discarding the first 5,000. To generate the auxiliary data set, the perturbation variance is set to $\sigma_{\eta*}^2 = 0.25$. From this stage we are interested in storing the posterior samples for the variables related with the detection of level shifts, k_2, p_2 and q_2. A level shift is detected at $t = 75$, if the posterior mean of $p_{2,74}$ is greater than 0.5. In this case, the estimate of the size of the level shift, is the posterior sample weighted average of $k_{2,74}$. This value is input for the second stage of *sampling scheme 5.2*. For those data sets, where a level shift

is detected at $t = 75$ we run the second stage of *sampling scheme 5.2.* We run 10,000 iterations discarding the first 5,000. From this stage, we obtain posterior samples for the remaining variables, $\sigma_\varepsilon^2, \mathbf{k}_1, \mathbf{p}_1$ and q_1. The prior distributions assumed are the ones defined in expressions (5.41) to (5.43), and (5.72) to (5.73). The parameters for these distributions are set to the following values:

$$c_1 = s_1 = 5, a_1 = 1, b_1 = 100,$$
$$c_2 = 5, s_2 = 1, a_2 = 1, b_2 = 100.$$

The boundaries for the approximating uniform distributions are $u_1 = u_2 = -5, v_1 = v_2 = 5$.

In Table 5.9, we present a summary of the statistical properties of all the posterior samples obtained in the two stages composing the sampling scheme. As in the §5.3.2, some of the results presented in Table 5.9, are not used for obtaining estimates of the parameters of the model, and are presented merely for comparison purpose.

The level shift is detected for 76% of the simulated data sets. The average value of the estimated size of a level shift at $t = 75$, is close to the true size of the shock: an average estimated value of 4.9, given a true value of 5. The variance of the perturbation added to the transition equation presents on average an accurate estimate of its true value of 0.25.

The variables of the measurement equation are on average more accurately estimated in the second stage, which is in fact the stage where we obtain their posterior samples.

The outlier at $t = 50$ is detected in 87% of the cases, for which the second stage was run. The mean of the size of the outlier obtained across the simulated data sets, presents an average value of -4.9, and therefore an accurate estimate of its true value of -5. The measurement equation variance is on average estimated with a value close to the true one.

In Figure 5.11, we present some plots of the posterior distributions of the sizes of the two shocks input to the simulated data sets. The empirical distributions presented correspond to the posterior means of the size of a level shift at $t = 75$ and an outlier at $t = 50$. For the level shift, these values are the estimates of the

Table 5.9: Summary of output from posterior samples averages, across 5,000 simulated replications, for a local level model with $\sigma_\eta^2 = 0$.

| | | Truth | Posterior sample mean |
			Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$	> 0.5
	σ_ϵ^2	1	1.423	1.405	0.262	0.954	1.986	-
	$\sigma_{\eta^*}^2$	0.25	0.259	0.243	0.083	0.148	0.470	-
	$p_{1,50}$	1	0.601	0.686	0.346	0.024	1	0.610
Stage 1	$k_{1,50}$	-5	-4.781	-4.899	2.707	-7.602	-2.141	-
	$p_{2,74}$	1	0.735	0.892	0.318	0.037	1	0.756
	$k_{2,74}$	5	4.860	4.894	1.304	2.229	7.192	-
	q_1	0.01	0.023	0.023	0.004	0.016	0.033	-
	q_2	0.01	0.023	0.022	0.004	0.016	0.031	-
	σ_ϵ^2	1	1.131	1.063	0.308	0.767	1.981	-
Stage 2	$p_{1,50}$	1	0.850	0.985	0.256	0.111	1	0.873
	$k_{1,50}$	-5	-4.888	-4.925	1.065	-6.901	-2.818	-
	q_1	0.01	0.022	0.021	0.002913	0.016091	0.028	-

The values reported are the mean, standard deviation, 2.5th and 97.5th percentile of the posterior mean estimate across 5,000 simulated samples. "≥ 0.5" is the proportion of times the shock was correctly detected, for a threshold of 0.5. The results referring to stage 2, correspond to the simulated data sets for which a level shift was detected at $t = 75$. Based on 10,000 Gibbs draws, discarding the first 5,000.

size of the shock, obtained across the 5,000 simulated data sets. For the outlier, across the simulated data sets where the level shift was detected on the first stage, in a total of 3,780 simulated data sets. For the size of the outlier, the estimates are approximately symmetrically distributed around its true value of -5. The size of the level shift presents an higher level of dispersion. However, if we dismiss some of the estimates on the left hand side of the empirical distribution, which are negligible, the estimates are approximately symmetrically distributed around its true value.

In summary, this Monte Carlo study shows that the results obtained across the 5,000 simulated samples are satisfactory. *Sampling scheme 5.2* delivers posterior samples for the variables of the model which allow us to detect on the majority of the cases the correct position of the level shift and outlier input to the data. The size of these shocks and the variance σ_ϵ^2 are on average accurately estimated.

Figure 5.11: Histograms and box-plots of estimates of size of level shift (a)-(b) and outlier (c)-(d), across 5,000 simulated replications, for local level model with $\sigma^2_\eta = 0$.
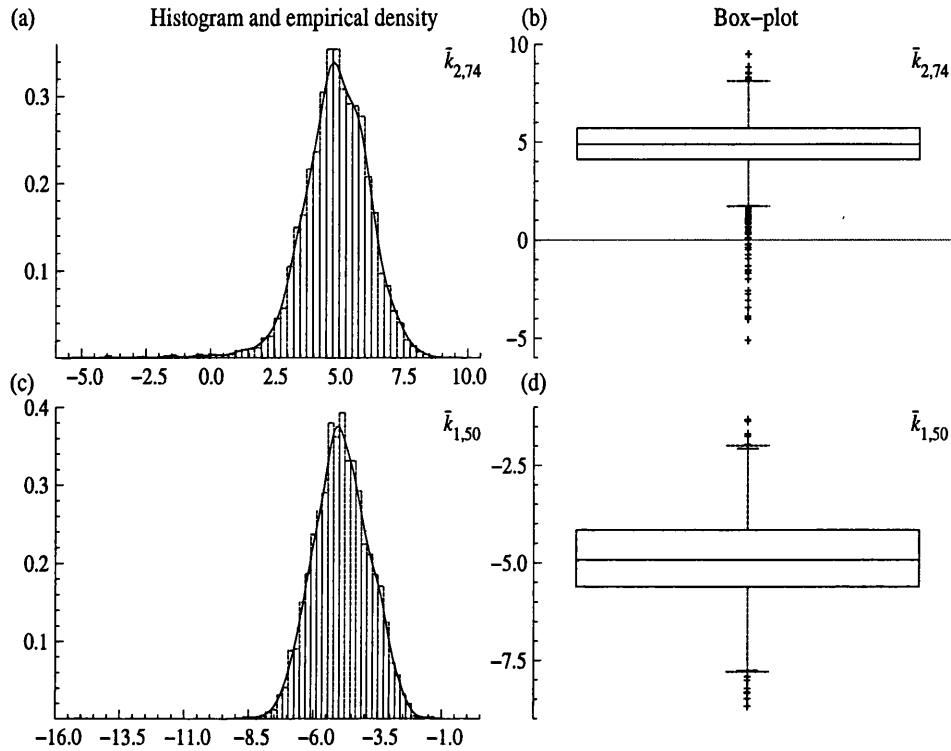
5.4.3 Empirical application: Nile data set

Figure 5.12 shows measurements of annual volume of discharges (in 10^8 m^3), from the river Nile to Aswan, from 1871 to 1970, in a total of 100 annual observations. This data is analyzed in Cobb (1978), where the data set is presented. The study presented in that paper concerns the detection of changing points in the data. Another change point analysis of the Nile data is presented in Carlstein (1988). In these two papers, a change point is detected in 1898. In Balke (1993) an ARMA model is adjusted to the Nile data, with a level shift detected in 1899. In the framework of unobserved component models, Atkinson, Koopman, and Shephard (1997) and De Jong and Penzer (1998), analyze this data set, for the detection of outliers and level shifts. The model fitted is a local level model. Both these papers use a frequentist approach and agree in detecting a level shift in 1899. When including

this shock in the model, the level variance becomes statistically null.

We fit a local level, with intervention variables for the presence of outliers and level shifts, setting the level variance to zero:

$$y_t = \mu_t + k_{1,t}p_{1,t} + \varepsilon_t, \tag{5.74}$$

$$\mu_{t+1} = \mu_t + k_{2,t}p_{2,t}, \tag{5.75}$$

with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, for $T = 1, \ldots, 100$.

Figure 5.12: Annual volume of the river Nile at Aswan, between 1871 and 1970.
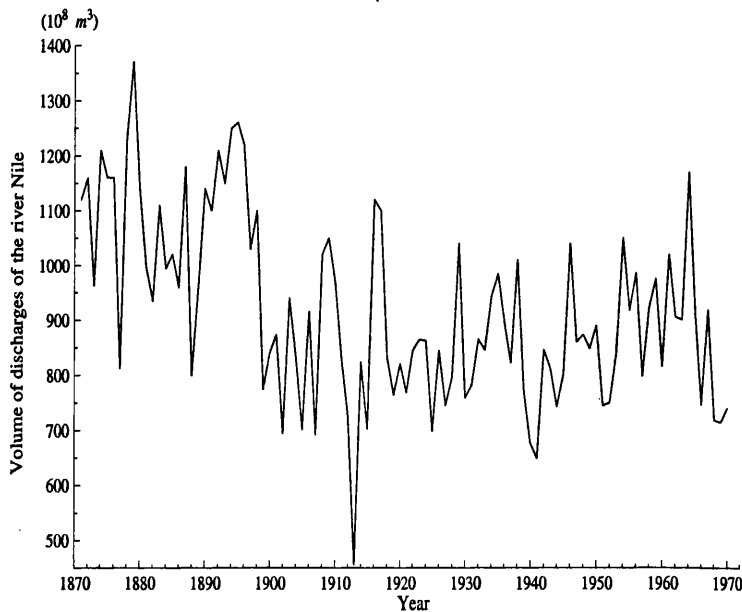
By inspection of the plot in Figure 5.12, there seems to be a change in the level of the series between 1890 and 1900. Analyzing the values of the observations during that decade we can see they range between 774 and 1,260. Given these values, we choose to set $\sigma_{\eta^*}^2 = 20$.

The prior distributions assumed for applying *sampling scheme 5.2* are defined in expressions (5.41) to (5.43) and (5.72) to (5.73). The parameters for these prior distributions are chosen with the values:

$$c_1 = 5, s_1 = 25000, a_1 = 2, b_1 = 100,$$

$$c_2 = 5, s_2 = 60, a_2 = 2, b_2 = 100.$$

The parameters for the bounded uniform are chosen to be $u_1 = -350, v_1 = 350, u_2 = -250, v_2 = 250$. In each stage the Gibbs sampler is run for 100,000 discarding the first 50,000. In Figure 5.13, we plot the posterior samples mean of the intervention variables, for the presence of level shifts and outliers.

Figure 5.13: Plots of posterior means of intervention variables for the presence of a level shift (a)-(b), and for the presence of an outlier (c)-(d), for the Nile data set, from 1871 to 1970.
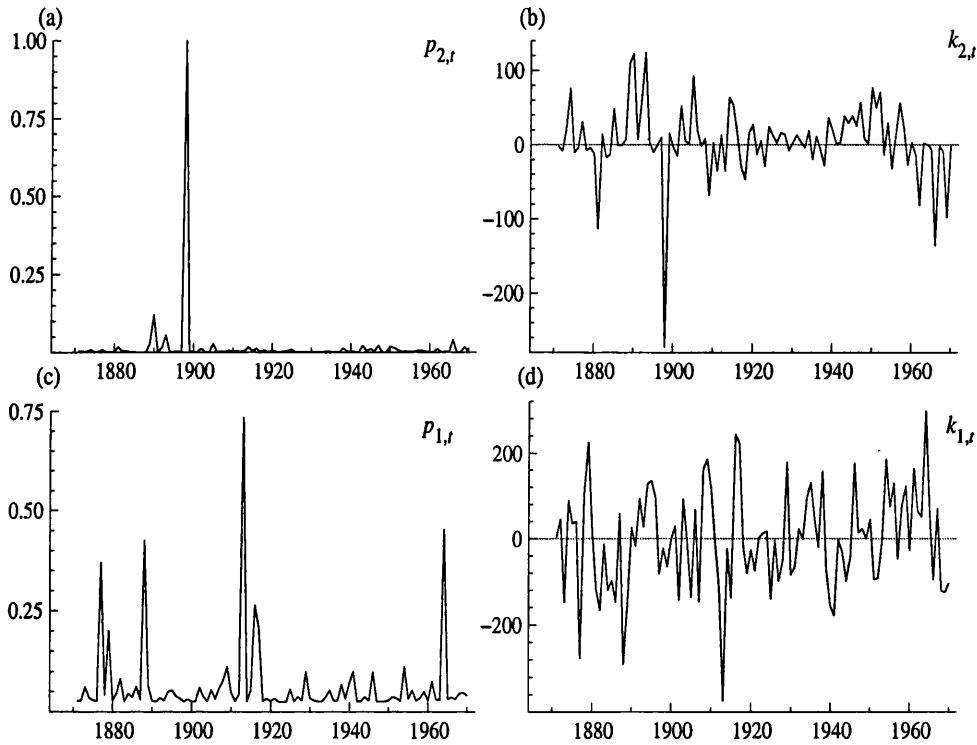
The posterior samples for the variables $p_{2,t}$ and $k_{2,t}$ are obtained by running the Gibbs sampler for the auxiliary data. This auxiliary data set is obtained as in expression (5.63), with $\sigma_{\eta^*}^2 = 20$. The values plotted in Figure 5.13 (a) are the mean of the posterior samples of the indicator variable for the presence of a level shift, for any $t = 1, \ldots, 100$. We observe that, for the observation corresponding to the year 1898, the indicator variable is sampled with value one, for all the iterations stored from the Gibbs sampler. This corresponds to a posterior mean probability of having a level shift in 1899 equal to one. For any of the other instances, the mean

posterior probabilities of having a level shift are meaningless when compared with the threshold of 0.5. In Figure 5.13 (b), we have the weighted mean of the posterior samples of the size of a level shift. Given that, a level shift was detected only for $t = 1899$ we are interested in the value of $\bar{k}_{2,t-1}$, for t corresponding to the year 1899. At 1899 we estimate a downward level shift of size -272.22.

Having detected the level shift present in the data and its size, we input this information for running the second stage of *sampling scheme 5.2*. The posterior samples for the probability of having an outlier are plot in Figure 5.13 (c). We observe that using the threshold of 0.5, one outlier is detected in 1913. From the weighted posterior samples of the size of this type of shock, plotted in Figure 5.13 (d), we have that its size is estimated with value -373.21.

Using an informative approach (see discussion in §3.4.1) we detect three other years with outstanding value for the mean posterior probability of an outlier, in the years 1877, 1888, and 1964. Although the posterior mean probabilities present values below 0.5, with values 0.37, 0.42, and 0.45, respectively, we choose to classify these observations has outliers. We consider that, these values, compared with the mean probabilities for the instances where an outlier is not detected, give a clear indication of the occurrence of an outlier.

In Table 5.10 we summarize the properties of the posterior samples of the variables of interess. We have detected one downward level shift in 1899 and an negative outlier in 1913, which agrees with the results obtained in Atkinson, Koopman, and Shephard (1997) and De Jong and Penzer (1998). The outlier detected in 1877 has also been reported in De Jong and Penzer (1998). In Atkinson, Koopman, and Shephard (1997), the outlier in 1888 is also detected. We detect an outlier in 1964, which, to our knowledge, as not been reported in the literature.

To compare our results with the ones obtained using a different approach, we present is Table 5.11, the results obtained by maximum likelihood. The position of the shocks is input, according to the shocks to have detected.

For all of the shocks we have detected, their estimated sizes by maximum likelihood are statistically significant. The values presented in Table 5.11, for the sizes of the shocks are similar to the ones we have obtained. The t-ratios for the size of

170

Table 5.10: Summary of output from posterior samples, for the Nile data set, from 1871 to 1970.

		Prior		Posterior sample					
		Mean	SD	Mean	Median	SD	$Q_{0.025}$	$Q_{0.975}$	$\bar{p}_{i,t}$
	σ_ϵ^2	8333	11785	13376.25	13193.0	28.76	9146.19	18688.23	–
	q_1	0.02	0.014	0.061	0.058	2.47e-04	0.024	0.11	–
	q_2	0.02	0.014	0.019	0.016	7.91e-04	0.0038	0.045	–
Level shift	1899	0	144	-272.22	-263.21	8.25	-402.87	-189.10	1
Outlier	1877	0	202	-275.42	-279.93	0.57	-471.71	-66.53	0.37
	1888	0	202	-289.82	-293.56	0.72	-489.97	-76.37	0.42
	1913	0	202	-373.21	-375.16	0.46	-588.24	-140.50	0.73
	1964	0	202	298.28	301.21	0.42	83.38	501.73	0.45

The values reported are the mean, median, 2.5th and 97.5th percentile of the posterior samples, and the standard deviation of the sample mean estimates (SD estimated with a bandwidth of 5,000 for the Parzen window) for σ_ϵ^2, q_1, q_2 and the sizes of the shocks detected. "$\bar{p}_{i,t}$" is the posterior sample mean of the indicator variable. Based on 100,000 Gibbs draws, discarding the first 50,000.

Table 5.11: Estimation results by maximum likelihood, given the position of the shocks, for the Nile data set, from 1871 to 1970.

| σ_ε^2	σ_η^2	Level shift 1899	Outlier 1877	1888	1913	1964
12301	0.0000	-269.16	-307.19	-321.19	-395.03	318.97
[1.0000]	[0.0000]	(-10.567)	(-2.7179)	(-2.8418)	(-3.5365)	(2.8556)

The values in [ ] are the q-ratios for the hyperparameters estimated; the values in ( ) are the t-ratios for the size of shocks estimated.

the outliers, present the same order of significance, as the mean posterior probabilities of the occurrence of an outlier. For example, the outlier with an higher mean posterior probability was detected in 1913. For this year, we obtain the estimate of the size, by maximum likelihood, with the higher t-ratio. The year of 1964, an outlier not detected in Atkinson, Koopman, and Shephard (1997) or De Jong and Penzer (1998), presents the second higher posterior mean probability and t-ratio.

We can notice that the estimate for the irregular component variance is lower when estimated by maximum likelihood. One explanation for this difference is that the value obtained by the Gibbs sampler, is the mean of the posterior samples obtained for that variable. Some of the values composing this posterior sample correspond to iterations where not all the indicator variables for the presence of an outlier, for the shocks detected, are sampled with value one.

In Figure 5.14, we plot the empirical densities of the posterior samples for the sizes of the shocks detected. The level shift presents a slightly higher asymmetry, mainly due to a fatter left tail. It implies a posterior sample mean lower than the posterior sample median. In fact, we notice that the posterior mean presents a value closer to the estimate of the size of the level shift obtained by maximum likelihood. For this shock the degree of dispersion is less than for the posterior sizes of the outliers. However, in Table 5.10, the value for the standard deviation of the sizes of the shocks estimated presents an higher value for the level shift detected. This feature is due to the standard deviation being estimated using a Parzen window. The posterior sample for the size of the level shift in 1899 presents an autocorrelation function with a slower convergence towards zero, than the posterior samples for the

Figure 5.14: Histograms and box-plots of posterior samples for size of level shift detected in 1899 (a)-(b) and outliers detected in 1877 (c)-(d), 1888 (e)-(f), 1913 (g)-(h) and 1964(i)-(j), for the Nile data set, from 1871 to 1970.
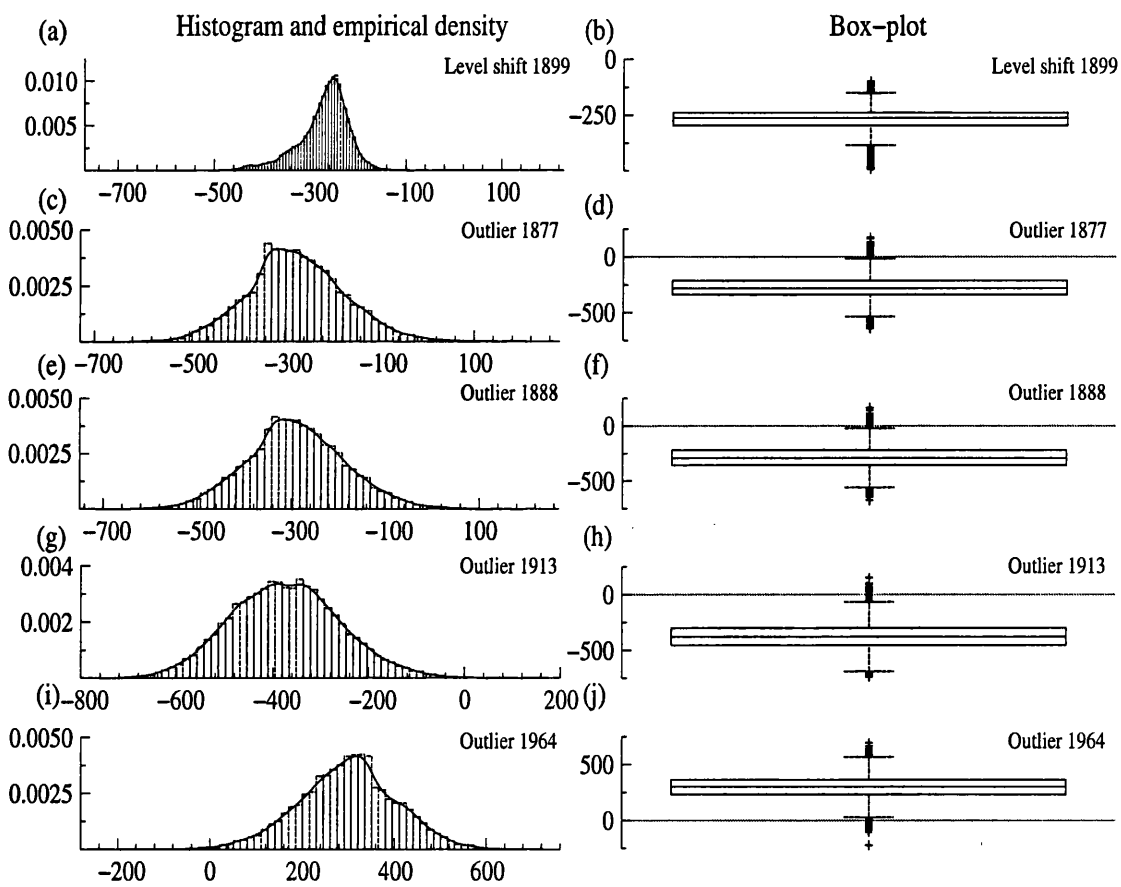
outliers detected.

To check the adequacy of the model fitted in Figures 5.15 and 5.16 we present plots for the standardized innovations and standardized auxiliary residuals, respectively.

The standardized innovations present a behaviour that does not indicate the presence of serial correlation, or departure from the normal distribution. The inspection of the plots for the standardized auxiliary residuals leads to the conclusion that no shock has not been accounted for, and the ones we have detected were estimated with the correct size.

5.5 Conclusions

We have presented two sampling schemes for estimating local level models, in the presence of outliers and level shifts, when one of the hyperparameters is equal to zero. By proposing sampling schemes, with a first stage based on sampling from posterior distributions for an auxiliary data set, we can assume a continuous uniform prior distribution for the size of the shocks. This uniform assumption is chosen with the aim of detecting and estimating the size of the shocks with as few prior knowledge of the data set as possible.

For generating the auxiliary data sets, it is necessary to choose a variance parameter. By the analysis of sensitivity presented for the choice of these parameters, we conclude that it does not imply much assumptions about the characteristics of the shocks that might exist.

The satisfactory performance of the sampling schemes we have proposed has been established by two Monte Carlo studies. As empirical applications we have applied our methodology to two data sets. The results obtained and their comparison with results in previous works confirms the suitable behaviour of our methodology.

Figure 5.15: Descriptive plots of standardized innovations, for the Nile data set, from 1871 to 1970.

(a) Plot of standardized innovations

(b) Box-plot of standardized innovations

(c) ACF for standardized innnovations

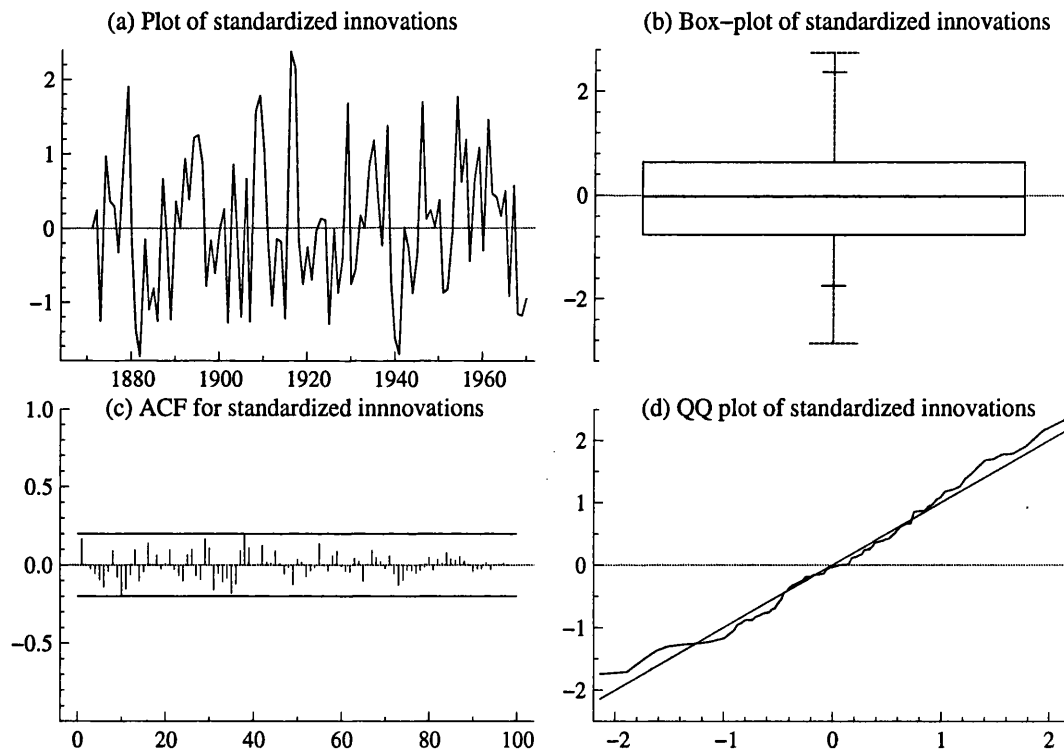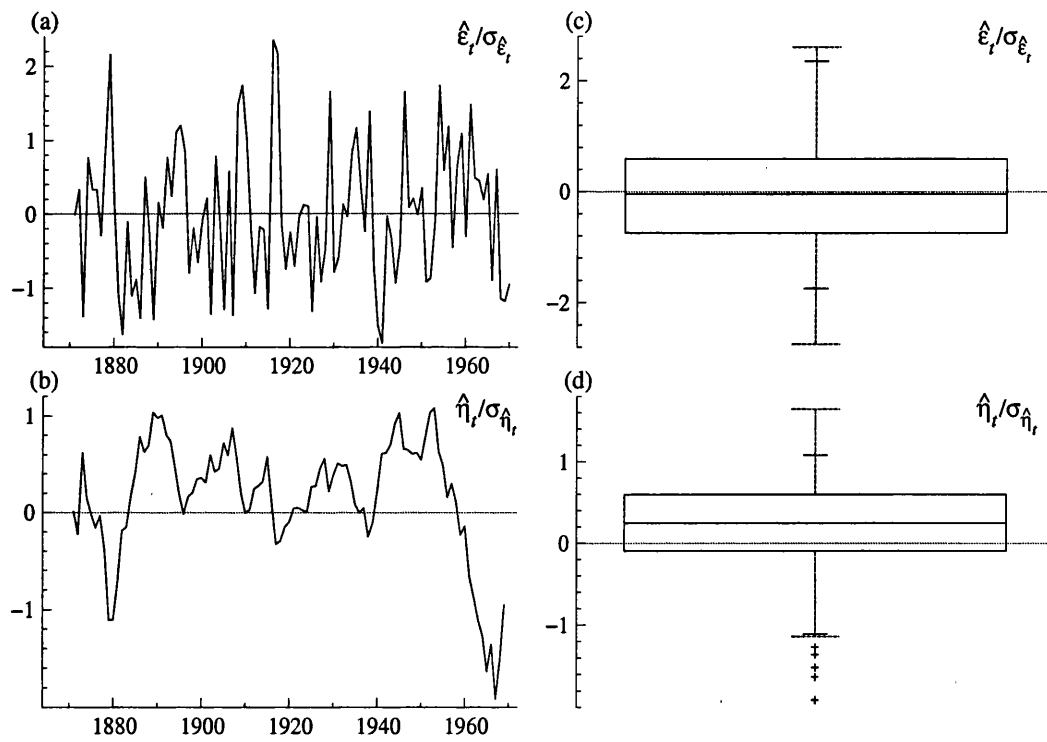(d) QQ plot of standardized innovations

Figure 5.16: Plots (a)-(b) and box-plots (c)-(d) of standardized auxiliary residuals, for the Nile data set, from 1871 to 1970.

Chapter 6

Conclusions

The aim of this work is to present a general methodology for the estimation of structural models, when shocks, namely outliers and structural shifts, are present in the time series data. A Bayesian approach was chosen, because it allows the joint estimation of the hyperparameters and the detection of the shocks' position and size. The Gibbs sampler is the sampling method we use. The detection of shocks is accomplished by introducing a pair of variables, for each type of shock; an indicator variable, for the presence of a shock at an instant in time and a magnitude variable, defining the size of the shock.

Throughout this work, we have considered a flat prior distribution for the size of intervention variable. This prior has not been considered in the literature. It was chosen with the objective of making the estimation process as automatic and free from prior knowledge of the data, as possible. Assuming a flat prior distribution, two parameters have to be chosen *a priori*: the lower and upper bounds of an approximating bounded uniform distribution. We have showed, by numerical illustrations, that the shocks detected, namely their size, present a low sensitivity to the choice of these parameters. Hence, even when there is insufficient prior information to make this choice in the most efficient manner, the shocks are detected satisfactorily.

The advantage of the method we have proposed, when compared to other methodologies is evident. Frequentist approaches have the drawback of not providing a method to simultaneously estimate the hyperparameters and diagnose the shocks. Within Bayesian methodologies, the usual prior assumptions for the intervention

variable are either a multinomial or a normal distribution. When using a discrete prior distribution, the values of possible sizes of the shock are set *a priori*. This requires considerable prior knowledge of the data behaviour. In addition, it prevents the use of information within the data to define the size of the shocks. Using a normal prior, the mean and variance parameters have to be chosen *a priori*. We have showed, in §3.5, that the posterior samples of the size of intervention variable present an undesirable sensitivity to the choice of this distribution's parameters. From the Monte Carlo study conducted in §3.5, we have concluded that, in the absence of relevant prior information about the size of the shocks, the variance parameter should be set to a high value. In the limit this corresponds to using an uninformative prior distribution.

In Chapters 3 and 4, we propose sampling schemes for the estimation of structural models, in the presence of shocks. We start by considering a local level model, in the presence of outliers and level shifts. This formulation was generalized in Chapter 4, where we proposed a method for estimating a basic structural model, in the presence of outliers, and three possible types of structural shifts: level, slope and seasonal shifts. Our main contribution is the assumption of a flat uniform prior distribution for the size of intervention variables. Additionally, the Bayesian techniques of Chapter 4 for estimation of a BSM, allowing for the presence of shocks to any of the model's components, do not appear elsewhere in the literature. The Monte Carlo studies presented show that our methods perform well, when applied to artificial data sets. In particular the shocks input to the simulated data sets have a high proportion of detection, and their sizes are on average accurately estimated. The coal data set considered in Chapter 3, has been previously studied in Atkinson, Koopman, and Shephard (1997), for the subset of observations until 1983Q4, using a frequentist approach for shock detection. In Atkinson, Koopman, and Shephard (1997), outliers are detected in 1969Q1 and 1969Q3, and a level shift in 1975Q1. For the sub-sample until 1983Q4, we have detected the two outliers but, an outlier in 1975Q1, instead of the level shift reported in Atkinson, Koopman, and Shephard (1997). From further investigation of the data, we concluded that a level shift would in fact be detected at that point, if a restriction of not having an outlier at 1975Q1 is

imposed. However, the posterior mean probability of having an outlier at 1975Q1, presented an higher value, than the corresponding posterior mean probability of having a level shift in 1975Q1. We concluded that there is stronger evidence of an outlier at that point in time than a level shift. Moreover, Atkinson, Koopman, and Shephard (1997) mention that there might be an outlier in 1975Q1. In the sub-sample from 1984Q1 to 1984Q4, we detect two other outliers. As an empirical application of the detection of shocks for the BSM, we model the quarterly number of marriages in the UK, from 1958Q1 to 1984Q4. For the period of 1965Q1 to 1970Q4, this data set was analyzed in West and Harrison (1997) and Penzer (1998). This data set is modeled using a BSM, with deterministic slope. We establish the existence of two seasonal shifts; an upward seasonal shift in 1962Q1, and a downward seasonal shift in 1969Q1, which agrees with the findings in West and Harrison (1997) and Penzer (1998). For the period of 1971Q1 to 1984Q4, we detect a downward level shift in 1973Q2. The results we obtained in these empirical applications are overall in agreement with results from using different methodologies. The advantage of our method is that our results were obtained with one run of the Gibbs sampler, which delivered posterior samples for all the variables of the model.

In Chapter 5, we consider the problem of applying the Gibbs sampler for detection of shocks, for the local level model formulation, when one of the hyperparameters is equal to zero. Following our approach of developing a method that does not require substantial prior information, we kept the previous formulation of the intervention model, as well as the prior assumption of a flat distribution for the size of intervention variable. The problem with the sampling schemes presented in Chapters 3 and 4 is that, they will not converge, if we are sampling from intervention variables in equations with a null variance noise. We solved this problem by proposing a two stages sampling scheme. In the first stage, we run the Gibbs sampler for an auxiliary data set, following a local level model, with both hyperparameters different from zero, but with the same shocks as the original data set. In this way we obtain posterior samples for the interventions variables that are included in the model, in an equation with a null hyperparamater. From these posterior samples we detect the position and size of the correspondent shocks. In the second stage, the shocks

estimated in the first stage are input in the model, and a second Gibb sampler is run, for the original data set, for estimating the non null hyperparameter and detection of the other type of shocks. Unlike existing methods our approach assumes a continuous prior distribution for the size of intervention variable for the particular case of one of the hyperparameters being equal to zero. In Gerlach, Carter, and Kohn (2000) the detection of shocks for a state space formulation, allowing for null hyperparameters, is considered. The draw back of the method in Gerlach, Carter, and Kohn (2000), is the assumption of a discrete prior distribution for the size of intervention variable, which as we have argued before, demands a considerable prior knowledge of the data.

The sampling schemes we derive in Chapter 5, are based on a first stage, where an auxiliary data set is generated, for detecting one of the type of shocks. To generate this auxiliary data set, a variance parameter has to be chosen. We presented the results from an analysis of sensitivity of the shocks detected to the choice of this parameter. From the results obtained we conclude that the auxiliary data set can be generated in an efficient way, without extensive prior analysis of the data. Two other Monte Carlo studies were performed. In the first Monte Carlo study the artificial data sets are generated from a local level model, with the irregular variance equal to zero. In the second experiment, the generating process used was local level model with the level variance equal to zero. For both experiments, an outlier and a level shift were added to the data. The results we have obtained and presented show that our methods produce good overall results. The shocks are detected for a high percentage of the simulated replications, with accurate estimates for their size. As an empirical application of the detection of outliers and level shifts, when the irregular variance is equal to zero, we consider the monthly quotes of bonds issued by the Greek government, and transacted in the London Stock Exchange, from August 1916 to June 1930. We fit a unobserved component model, with level and seasonal component, and null irregular variance. The method we propose for the case of a null irregular variance was used including, at stage two, an additional step for sampling from the full conditional for the seasonal component. We detected one outlier and four level shifts. Their positions and sizes agree with

the results obtained in Christodoulaki and Penzer (2002) where this data set was analyzed using a frequentist technique. The method in Christodoulaki and Penzer (2002) for detection of shocks, is based on the output from running the model without interventions. After the position of the shocks is established the model is re-estimated, including dummy variables for the shocks detected, and their size is estimated by maximum likelihood. The estimates for the size of the shocks we obtain using a Bayesian approach, are very similar to the ones obtained by maximum likelihood given the position of the shocks. With our methodology, the shocks are detected by obtaining posterior samples for intervention models. We simultaneously obtain estimates for the hyperparameters, detect the position and estimate the size of the shocks. For the method in Christodoulaki and Penzer (2002) the position of the shocks are obtained on the basis of the null model output.

As a last empirical application, we applied our method for the detection of outliers and level shifts when the level variance is equal to zero, to the data set consisting of the annual volume of the Nile at Aswan, from 1871 to 1970. This data set has been the subject of several works on change-points detection and detection of shocks to state space models. Most of these studies agree in detecting a level shift in 1899. When this shift is accounted for in the model, the level hyperparameter become statistically null. Applying our methodology, we detected a level shift at that instant. This level shift was detected with a posterior sample probability of occurrence of 1. Additionally, we have established the presence of four outliers. The outliers we have detected in 1877, 1888 and 1913 have been reported in other studies. We detected also the presence of an outlier in 1964. The outlier detected with a highest posterior mean of occurrence is in 1913; the outlier detected with a second highest posterior mean probability of occurrence is in 1964. This outlier has not, to our knowledge, been reported before. For comparison purpose we have estimated the intervention model by maximum likelihood, inputting dummy variables for the instances where we have detected shocks. If we order the outliers according to the statistical significance of the estimated sizes using maximum likelihod, it agrees with the order we obtain by ordering them according to the posterior mean probabilities of occurrence of an outlier.

181

Several extensions can be done to the methodologies we have developed in this work. For example, the formulation of a seasonal shift. We defined a seasonal shift has being generated by a shock to the seasonal component at one instant in time. For some time series, this is not the correct formulation. In Penzer (2002), several other formulations are presented, together with examples of real data where they occur. It would be interesting to extend the methodology presented in Chapter 4, for alternative formulations of seasonal shocks. The methodology presented in Chapter 5, for the detection of outliers and level shifts, for the local level model, when one of the hyperparmeters is equal to zero, can be extended to a basic structural model. We considered the local level model for simplicity, but it should be straightforward to consider a more general formulation; a BSM in the presence of shocks to components with a null hyperparameter. Finally, we have have restricted the scope of this work to Gaussian state space models. In Chib, Nardari, and Shephard (2002), a Gibbs sampler is presented for estimation of a non-Gaussian state space model, allowing for shocks in the measurement equation; the model being a stochastic volatility model with jumps in the mean equation. A possible extension of our methodology would be for this type of financial models, or to more general state space processes.

Bibliography

Abraham, B. and G. Box (1979). Bayesian analysis of some outlier problems in time series. *Biometrika 66* (2), 229–236.

Andersen, W. G. (1984). *An Introduction to Multivariate Statistics Analysis* (2nd ed.). New York: John Wiley & Sons.

Anderson, B. D. O. and J. B. Moore (1979). *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall.

Ansley, C. F. and R. Kohn (1985). Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics 13* (4), 1286–1316.

Atkinson, A. C., S. J. Koopman, and N. Shephard (1997). Detecting shocks: outliers and breaks in time series. *Journal of Econometrics* (80), 387–422.

Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business and Economic Statistics 11* (1), 81–92.

Barnett, G., R. Kohn, and S. Sheather (1997). Robust Bayesian estimation of autoregressive-moving-average models. *Journal of time series analysis 18* (1), 11–28.

Barnett, V. and T. Lewis (1984). *Outliers in Statistical Data* (2nd edition ed.). John Wiley & Sons.

Box, G. and G. Jenkins (1970). *Time series analysis - forecasting and control.* Holden-Day.

Box, G. and G. C. Tiao (1968). A Bayesian approach to some outliers problems. *Biometrika 55* (1), 119–129.

Brockwell, P. J. and R. A. Davis (1991). *Time series: theory and methods* (2nd ed.). Springer.

Brooks, S. P. and A. Gelman (1998). Some issues in monitoring convergence of iterative simulations. *Proceedings of the 30th Symposium of the Interface*.

Carlin, B. P., N. G. Polson, and D. S. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state space modelling. *Journal of the American Statistical Association 87*(418), 493–500.

Carlstein, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics 16*(1), 188–197.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Carter, C. K. and R. Kohn (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika 83*(3), 589–601.

Chan, K. S. (1993). Asymptotic behavior of the Gibbs sampler. *Journal of the American Statistical Association 88*(421), 320–326.

Chen, C. and L. Liu (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association 88*(421), 284–297.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* (108), 281–316.

Christodoulaki, O. and J. Penzer (2002). News from London: Greek government bonds on the London Stock Exchange, 1914-1930. mimeo.

Cobb, G. W. (1978). The problem of the Nile, conditional solution to a change point problem. *Biometrika 65*(2), 243–251.

De Jong, P. (1988). A cross-validation filter for time series models. *Biometrika 75*(3), 594–600.

De Jong, P. (1991). The diffuse Kalman filter. *The Annals of Statistics 19*(2), 1073–1083.

De Jong, P. and J. Penzer (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association 93*(442), 796–806.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 22*(2), 339–350.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill. Inc.

Doornik, J. A. (1999). *Ox – An object-oriented matrix programming language* (Third ed.). Timberlake Consultants LTD.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–616.

Durbin, S. and S. J. Koopman (2001). *Time series analysis by state space models.* Oxford University Press.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society, Series B 34*, 229–247.

Franses, P. H., H. Hoek, and R. Paap (1997). Bayesian analysis of seasonal unit roots and saesonal mean shifts. *Journal of Econometrics* (78), 359–380.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of the Time Series Analysis 15*(2), 183–202.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*(410), 398–409.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patter. Anal. Mach. Intell. PAMI-6*(6), 721–741.

Gerlach, R., C. Carter, and R. Kohn (2000). Efficient Bayesian inference for dynamic mixture models. *Journal of the American Statistical Association 95*(451), 819–601.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*(6), 1317–1339.

185

Geyer, C. (1992). Practical Markov chain Monte Carlo. *Statistical Science* 7(4), 473–511.

Green, P. J. and X. L. Han (1992). Metropolis methods, Gaussian proposals and antithetic variable. In P. Barone, P. Frigessi, and M. Piccioni (Eds.), *Stochastic Models, Statistical Methods and Algorithms in Image Analysis*, Volume 74, pp. 142–164. Berlin: Springer-Verlag.

Harvey, A. C. (1989). *Forecasting structural time series models and the Kalman filter*. Cambridge University Press.

Harvey, A. C. and S. J. Koopman (1992). Diagnostic checking of unobserved component time series models. *Journal of Business and Economic Statistics* 10(4), 377–389.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1), 97–109.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Justel, A., D. Pena, and R. S. Tsay (2001). Detection of outliers patches in autoregressive time series. *Statistical Sinica* 11(3), 651–674.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASME Series D*(82), 35–45.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* 80(1), 117–126.

Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Jounal of the American Statistical Association* 92(440), 1630–1638.

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2000). *STAMP 6.0 Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants.

Koopman, S. J. and N. Shephard (1992). Exact score for time series in state space form. *Biometrika* 74(4), 823–826.

186

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal 2*, 113–166.

Liu, J. S., W. H. Wong, and A. Kong (1994). Covariance structure of the Gibbs sampler with application to the comparisons of estimators and augmentation schemes. *Biometrika 81*(1), 27–40.

McCulloch, R. E. and R. S. Tsay (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association 88*(423), 968–978.

McCulloch, R. E. and R. S. Tsay (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis 15*(2), 235–250.

Metropolis, A., M. Rosenbluth, A. Rosenbluth, A. Teller, and E. Teller (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys. 21*, 1087–1092.

Penzer, J. (1998, November). Messy and mistaken seasonal time series. Statistics Research Report LSERR47, London School of Economics.

Penzer, J. (2001, March). Critical values for time series diagnostics. Statistics Research Report LSERR59, London School of Economics.

Penzer, J. (2002). Fast diagnostics for seasonal shifts using structural time series models. mimeo.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Volume 1: Univariate Series. Academic Press.

Ray, B. K. and R. S. Tsay (2001). Bayesian methods for change-point detection in long-range dependent processes. *Unpublished*.

Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. Springer.

Roberts, G. O. and S. K. Sahu (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society B 59*, 291–317.

Roberts, G. O. and A. F. M. Smith (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications 49*, 207–216.

Roberts, G. O. and R. L. Tweedie (1996). Geometric convergence and Central Limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika 83*, 95–110.

Shephard, N. (1993). Fitting nonlinear time-series models with applications to stochastic variance models. *Journal of Applied Econometrics 8*, S135–S152.

Shephard, N. (1994). Partial non-gaussian state space. *Biometrika 81*, 115–131.

So, M. K. P. (1999). Time series with additve noise. *Biometrika 86*(2), 474–482.

Tierney, L. (1994). Markov chains for exploring posterior distribution. *The Annals of Statistics 22*(4), 1701–1762.

Tsay, R. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting 7*, 1–20.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic models.* Springer.

Appendix A

Ox implementation of the Gibbs sampler for empirical application in §3.5.5

```
/* Gibbs sampler for sampling hyperparameters and interventions variables
for the Coal consumption data set in section 3.5.5    */

#include <oxstd.h>
#include <oxfloat.h>
#include <packages/ssfpack/ssfpack.h>
#include <oxprob.h>

const decl sInputFile  = "ofuCOAL1.in7";

/*declaration of functions */
VarSampler(const m);
AmplitudeSampler(const m);
IndicatorSampler(const m);
probSampler(const m);

static decl mY,mStates,mResiduals,Amplitude,
Indicator,IndicProb,prob,v_p;

/*matrices for State Space model*/
static decl Phi,Sigma,Omega;

/*initializations and constants input*/
static decl IrregVar=1;
static decl LevelVar=1;
static decl SeasVar=0;
static decl SlopeVar=0;
const decl crep=50000;
const decl burn=50000;
const decl Irregdegree1=5;
const decl Irregdegree2=0.05;
static decl Leveldegree1=5;
static decl Leveldegree2=0.01;
const decl a0=<8;2>;
const decl b0=<100;100>;
const decl LowUnif=<-0.5;-0.5>;
const decl UppUnif=<0.5;0.5>;

/*state space model formulation*/
static decl mStsm;

/*vectors for residuals sampling*/
static decl mKF,mWgt,mGamma,md,residstate,residY;
```

190

```
/*seasonal period */
static decl c_s=4;

main()
{
        decl time=timer();
        mY=loadmat(sInputFile)[][0]';             /*loads data set*/

        decl m=columns(mY);                       /*Data dimension*/
        decl v_LevelVar=zeros(1,crep+burn);       /*Level var*/
        decl v_IrregVar=zeros(1,crep+burn);       /*Irreg var*/
        decl v_Amplitude=zeros(2,m);              /*posterior magnitude*/
        decl Percentage= zeros(2,m);              /*posterior probability

/*Initialization*/
        mResiduals=zeros(4,m);
        Amplitude=zeros(2,m);
        Indicator= zeros(2,m);
        IndicProb=zeros(4,m);
        prob=<0.01;0.01>;
        v_p=zeros(2,1);
/*construction of state space structure*/
        decl mStsm=(<CMP_LEVEL>~sqrt(LevelVar)~<0,0>)|
                (<CMP_SLOPE>~sqrt(SlopeVar)~<0,0>)|
                (<CMP_SEAS_DUMMY>~sqrt(SeasVar[0][0])~c_s~0)|
                (<CMP_IRREG>~sqrt(IrregVar)~<0,0>);
        GetSsfStsm(mStsm,&Phi, &Omega,&Sigma);
/*Prints input parameters*/
        print("\n","sample dim=",m,"\n","burn=",burn,"\n","rep=",crep,
                "\n","initial irregvar=",IrregVar,
                "\n", "initial levelvar=",LevelVar,
                "\n","Seasvar=",SeasVar,"\n","Slopevar=",SlopeVar,
                "\n", "Prior parameters",
                "\n","Irreg Var  degree1  degree2" ,
                "\n",Irregdegree1~Irregdegree2,
                "\n","Level Var  degree1  degree2" ,
                "\n",Leveldegree1~Leveldegree2,
                "\n","Prob beta prior  a  b" ,"\n",a0~b0,
                "\n","Uniform Prior  Low  Upper" ,"\n",LowUnif~UppUnif);

        decl i,aux;;
```

191

```
            for(i=0;i<crep+burn;++i)
            {
/*sampling Irregular and Level variances*/
            aux=VarSampler(m);
            Omega[0][0]=aux[0][0];
            Omega[5][5]=aux[0][1];
/*sampling intervention's magnitude*/
            Amplitude=AmplitudeSampler(m);
/*sampling intervention's indicator*/
            IndicatorSampler(m);
/*sampling p of intervention*/
            probSampler(m);
/*stores the sampled values for variances*/
            v_LevelVar[][i]=Omega[0][0];
            v_IrregVar[][i]=Omega[5][5];


/*if after burn in period updates percentage and magnitudes stored*/
            if(i>=burn)
            {
            Percentage[0][]=(Percentage[0][]+ (Indicator[0][].!=0));
            Percentage[1][]=(Percentage[1][]+ (Indicator[1][].!=0));
            v_Amplitude=(v_Amplitude+Amplitude.*Indicator);
            v_p=v_p+prob;


            }
            }
/*Gibbs sampler iterations ended here*/
/*formatting output variables*/
            decl l;
            for(l=0;l<m;++l)
            {
                    if(Percentage[0][l]==0)
                    {
                            v_Amplitude[0][l]=0;
                    }
                    else
                    {
                            v_Amplitude[0][l]=
                            v_Amplitude[0][l]/Percentage[0][l];
                    }
                    if(Percentage[1][l]==0)
                    {
```

192

```
                        v_Amplitude[1][l]=0;
                }
                else
                {

v_Amplitude[1][l]=v_Amplitude[1][l]/Percentage[1][l];
                }
        }
/* posterior probabilities of shocks*/
        Percentage=Percentage./crep;
/* averaged estimates of variances*/
        Omega[0][0]=meanr(v_LevelVar[][burn:]);
        Omega[5][5]=meanr(v_IrregVar[][burn:]);
/* posterior p for indicator variables*/
        v_p=v_p/crep;

/*saves some data output*/
        savemat("Inter.in7",(Percentage|v_Amplitude)');
        savemat("Var.in7",(v_LevelVar[][burn:]|v_IrregVar[][burn:])');
/*prints output*/
        print(\n","IrregVariance",Omega[5][5],
        "\n","LevelVariance",Omega[0][0]);
        print("\n","time elapsed", (timer()-time)/100," sec.");

/*end of main*/
}


VarSampler(const m)
{
/* Gamma chooses the disturbances to sample from*/
        mGamma=diag(<1>~zeros(1,c_s)~0);
        mKF=KalmanFil(mY,Phi,Omega,Sigma,zeros(c_s+2,1),<>,<>
                ,1|constant(-1,c_s,1)|0,(Amplitude[][].*Indicator[][]));
        mWgt=SimSmoWgt(mGamma,mKF,Phi,Omega,
        Sigma,zeros(c_s+2,1),<>,<>,1|constant(-1,c_s,1)|0,
        (Amplitude[][].*Indicator[][]));

/* Draws from the disturbances of level equation*/
        residstate=SimSmoDraw(mGamma,rann(1,m),mWgt,mKF,Phi,
        Omega,Sigma,zeros(c_s+2,1),<>,<> ,
        1|constant(-1,c_s,1)|0,(Amplitude[][].*Indicator[][]));
/*mStates: recursion that given the simulated residuals obtains the states
and signals (last row)*/
```

```
        mStates=SsfRecursion(residstate[][],Phi,Omega,<>,zeros(c_s+2,1),<
        >,<>,1|constant(-1,c_s,1)|0,(Amplitude[][].*Indicator[][])~(0|0));
        mResiduals[0][]=residstate[0][1:];      /*transition residuals*/
        mResiduals[3][]=mY-mStates[5][1:];/*measurement residuals*/

        mStates=mStates[:2][:m-1];
        md=mResiduals;
        md=md*md';
/*returns samples from inverse gamma posterior distributions */
        return (1/rangamma(1,1,(Leveldegree1+m-
        1/2,(Leveldegree2+md[0][0])/2))~
        (1/rangamma(1,1,(Irregdegree1+m)/2,(Irregdegree2+md[3][3])/2));
}


AmplitudeSampler(const m)
{
        decl ampliY, ampliStates;
        ampliY=(rann(1,m).*sqrt(Omega[5][5])+mResiduals[3][]+
        Amplitude[0][].*Indicator[0][]).*Indicator[0][]+
        ((ranu(1,m)*(UppUnif[0][]-LowUnif[0][])+
        constant(LowUnif[0],1,m)).*(constant(1,1,m) -Indicator[0][]));

        ampliStates=(rann(1,m).*sqrt(Omega[0][0])+mResiduals[0][]+
        Amplitude[1][].*Indicator[1][]).*Indicator[1][]+((ranu(1,m)
        *(UppUnif[1][]-LowUnif[1][])+constant(LowUnif[1][],1,m))
        .*(constant(1,1,m)-Indicator[1][]));
        return ampliY|ampliStates;
}

IndicatorSampler(const m)
{
        decl u;

        IndicProb[0][]=densn((mY-mStates[0][]-mStates[2][])/
        (sqrt(Omega[5][5])))*(1-prob[0][]);
        IndicProb[1][]=densn((mY-mStates[0][]-mStates[2][]
        -Amplitude[0][])/(sqrt(Omega[5][5])))*prob[0][];
        IndicProb[:1][]=IndicProb[:1][]./(sumc(IndicProb[:1][]));
        u=ranu(1,m);
        Indicator[0][]=ones(1,m).*(u.>(IndicProb[0][]));
```

194

```
IndicProb[2][]=(densn((mStates[0][1:]-mStates[0][:m-2]-
mStates[1][:m-2]) /(sqrt(Omega[0][0])))*(1-prob[1][]))~1;
IndicProb[3][]=(densn(((mStates[0][1:]-mStates[0][:m-2]-
mStates[1][:m-2])
-Amplitude[1][:m-2])/(sqrt(Omega[0][0])))*prob[1][])~0;

IndicProb[2:][]=IndicProb[2:][]./(sumc(IndicProb[2:][]));
u=ranu(1,m);
Indicator[1][]=(ones(1,m)).*(u.>(IndicProb[2][]));
}

probSampler(const m)
{
    decl n1,n0;
    n1=sumr(Indicator[0][]);
    n0=m-n1;
    prob[0]=ranbeta(1,1,a0[0]+n1,b0[0]+n0);

    n1=sumr(Indicator[1][]);
    n0=m-n1;
    prob[1]=ranbeta(1,1,a0[1]+n1,b0[1]+n0);
}
```

195