Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives

Durbin, J.; Koopman, S.J.M.

Publication date:
1998

Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives

J. DURBIN
Department of Statistics,
London School of Economics and Political Science,
Houghton Street, London WC2A 2AE, UK.

S.J. KOOPMAN
CentER for Economic Research,
Tilburg University,
PO Box 90153, 5000 LE Tilburg, Netherlands.

December, 1998

Summary

The analysis of non-Gaussian time series using state space models is considered from both classical and Bayesian perspectives. The treatment in both cases is based on simulation using importance sampling and antithetic variables; Monte Carlo Markov chain methods are not employed. Non-Gaussian disturbances for the state equation as well as for the observation equation are considered. Methods for estimating conditional and posterior means of functions of the state vector given the observations, and the mean square errors of their estimates, are developed. These methods are extended to cover the estimation of conditional and posterior densities and distribution functions. Choice of importance sampling densities and antithetic variables is discussed. The techniques work well in practice and are computationally efficient. Their use is illustrated by applying to a univariate discrete time series, a series with outliers and a volatility series.

Keywords: Antithetic variables; Conditional and posterior statistics; Exponential family distributions; Heavy-tailed distributions; Importance sampling; Kalman filtering and smoothing; Monte Carlo simulation; Non-Gaussian time series models; Posterior distributions.

1 Introduction

This paper discusses the analysis of non-Gaussian time series using state space models from both classical and Bayesian points of view. A major advantage of the state space approach is that we can model the behaviour of different components of the series separately and then put the sub-models together to form an overall model for the series. State space models are very general and can handle a remarkably wide range of applications ranging from ARIMA models and unobserved components time series models to smoothing models with roughness penalties.

An example of the application of state space methods to a problem in applied time series analysis was the assessment for the Department of Transport of the effects of seat belt legislation on road traffic accidents in the United Kingdom described by Harvey and Durbin (1986). Although the observations were count data and hence non-Gaussian, the analysis was based on linear Gaussian methods since these were the only appropriate state space methods available at the time. The realisation that no exact treatment of count data existed at the time led to the work in this paper.

State space models contain two classes of variables, the unobserved state variables which describe the development over time of the underlying system, and the observations. We consider departures from normality both for the state variables and for the conditional distributions of the observations given the state. For the state, our primary interest is in heavy-tailed densities which enable us to model structural shifts. For the conditional densities of the observations, we consider general classes of distributions which include both exponential family distributions and heavy-tailed densities. The exponential family densities allow us to model count data such as Poisson, binomial and multinomial observations as well as to model skewed data by, for example, Gamma densities. The heavy-tailed densities allow us to model outliers. For a classical analysis we calculate maximum likelihood estimates of model parameters and then estimate conditional means of functions of the state given the observations, together with the mean square errors of the estimates. We also show how to estimate conditional distribution functions and conditional densities. For Bayesian analysis we estimate posterior means and variances, posterior distribution functions and densities and show how to draw random samples from the estimated posterior distributions of functions of the state. The methods are simple, practical and computationally efficient. For the most part we present a general theory for multivariate observations.

The techniques used are based on the Kalman filter and smoother and on Monte Carlo simulation using Gaussian importance sampling and antithetic variables. Using these techniques we develop methods that are new, elegant and efficient for problems in time series analysis, and we provide estimates that are as accurate as is desired. Our simulation techniques are based on independent samples and not on Markov chains, thus enabling us to avoid convergence problems and also to obtain simple and accurate estimates of sampling variances due to simulation.

Some early work on state space modelling with non-Gaussian data is reviewed in Chapter 8 of Anderson and Moore's (1979) text book. A further review of early work is given by Kitagawa (1987) and in the accompanying published discussion, particularly in the extensive comments of Martin and Raftery (1987). Gaussian mixtures were used by Harrison and Stevens (1971, 1976) under the name multi-process models for problems involving non-Gaussian data. Most of this work deals only with filtering. However, a comprehensive treatment of both filtering and smoothing was given by Kitagawa (1989, 1990) based on approximating non-Gaussian densities by Gaussian mixtures. At each update he collapses the conditional density into a smaller number of components to prevent the number of components in the mixtures becoming unmanageable, so the method is essentially approximative.

State space models for exponential family observations with Gaussian state were introduced by West, Harrison and Migon (1985). They used a Bayesian approach using conjugate priors

and at each update the posterior density was approximated in order to retain the conjugate structure. Their model was considered further by Fahrmeir (1992) who estimated the state variables by approximating their conditional modes given the observations. Frühwirth-Schnatter (1994) developed an approximate Bayesian technique by approximating the prior of the state density at each step of the filtering process by a Gaussian density and then performing the update using the new observation by means of a numerical integration of dimensionality equal to the dimensionality of the observation vector.

The disadvantage of all these methods is that they involve approximation errors of unknown magnitude whereas with our techniques, errors are due only to simulation and their extent can be measured and made as small as desired. Smith (1979, 1981) and Harvey and Fernandes (1989) gave an exact solution for a special case; they based their methods on conjugate distributions and they developed them for specific count data models for which the state equation is a univariate random walk. However, this approach does not lend itself to generalisation.

Using full Bayesian inference models, simulation techniques based on Monte Carlo Markov chain (MCMC) for non-Gaussian state space models have been developed by Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996, 1997), Shephard (1994), Shephard and Pitt (1997) and Cargnoni, Müller and West (1997). General accounts of Bayesian methodology and computation are given by Gelman *et.al.* (1995), Bernardo and Smith (1994) and Gelfand and Smith (1999). New developments in this paper are based on earlier work of Durbin and Koopman (1992, 1997). In the first paper we considered conditional mode estimation based on Kalman filtering and smoothing methods for exponential family models; in the second paper we considered the special case where the observations given the state are non-Gaussian while the state is Gaussian and the objective was to calculate maximum likelihood estimates of model parameters by simulation. The simulation methods were highly efficient computationally in the sense that accurate results were obtained using small simulation sample sizes in the low hundreds. Shephard and Pitt (1997) also considered maximum likelihood estimation of parameters of non-Gaussian state space models by simulation. Geyer and Thompson (1992) have developed simulation methods of estimation for specific autologistic models and other exponential family models without dynamic structures.

The structure of the paper is as follows. In section 2 we present the state space models that we shall consider. Section 3 develops some basic formulae that underly the simulation techniques that we shall describe in detail later. In section 4 we obtain a linear Gaussian model that approximates the non-Gaussian model in the neighbourhood of the conditional mode of the stacked state vector given the observations; this is used to provide the Gaussian densities that we use for importance sampling. Section 5 develops the computational techniques that are required for practical applications. These are based on importance sampling using two types of antithetic variables, one for location and one for scale. We obtain computationally efficient estimates of the means and variances of arbitrary functions of the stacked state vector given the observations; these enable us to estimate conditional distribution and density functions and to draw random samples from conditional distributions. We also obtain simple estimates of the variances of errors due to simulation. The results are extended in a straightforward manner to analogous problems in Bayesian inference.

Section 6 applies the techniques to three real data sets. The first refers to deaths in road accidents, the second is a series of UK gas consumption and the third is an exchange rate volatility series. The results demonstrate the feasibility of the techniques for different models and show the differences between results based on the classical methods and results using a Bayesian approach. Section 7 discusses our approach. We conclude that our methods for time series analysis of non-Gaussian observations based on state space models are elegant, practical and computationally efficient for both classical and Bayesian inference.

2 Models

2.1 The linear Gaussian model

In this section we present the state space models that will be considered in the paper. We begin with the linear Gaussian model. Although our main concern is with non-Gaussian models, the linear Gaussian model provides the basis from which all our methods will be developed. The model can be formulated in a variety of ways; we shall take the form

$$y_t = Z_t\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H_t), \tag{1}$$

$$\alpha_t = T_t\alpha_{t-1} + R_t\eta_t, \qquad \eta_t \sim N(0, Q_t), \tag{2}$$

for $t = 1, \ldots, n$. Here, y_t is a $(p \times 1)$ vector of observations, α_t is an unobserved $(m \times 1)$ state vector, R_t is a selection matrix composed of r columns of the identity matrix I_m, which need not be adjacent, and the variance matrices H_t and Q_t are nonsingular. The disturbance vectors ε_t and η_t are serially independent and independent of each other. Matrices H_t, Q_t, Z_t and T_t are assumed known apart from possible dependence on a parameter vector ψ which in classical inference is assumed fixed and unknown, and in Bayesian inference is assumed to be random. Equations (1) and (2) are called respectively the observation equation and the state equation of the state space model. It is worth noting that (1) can be regarded as a multiple regression model whose coefficient vector α_t is determined by the first order vector autoregression (2). The state space model (1) and (2) is essentially equivalent to model (16) and (17) of the seminal Kalman (1960) paper.

2.2 Non-Gaussian models

We shall use the generic notation $p(\cdot)$, $p(\cdot,\cdot)$ and $p(\cdot|\cdot)$ for marginal, joint and conditional densities. The general non-Gaussian model that we shall consider has a similar state space structure to (1) and (2) in the sense that observations are determined by a relation of the form

$$p\left(y_t|\alpha_1, \ldots, \alpha_t, y_1, \ldots, y_{t-1}\right) = p\left(y_t|Z_t\alpha_t\right), \tag{3}$$

while the state vectors are determined independently of previous observations by the relation

$$\alpha_t = T_t\alpha_{t-1} + R_t\eta_t, \qquad \eta_t \sim p\left(\eta_t\right), \tag{4}$$

for $t = 1, \ldots, n$, where the η_t's are serially independent. Here, either $p\left(y_t|Z_t\alpha_t\right)$ or $p\left(\eta_t\right)$ or both can be non-Gaussian. We denote $Z_t\alpha_t$ by θ_t and refer to it as the signal. While we begin by considering a general form for $p\left(y_t|\theta_t\right)$, we shall pay particular attention to two special cases: (i) observations which come from exponential family distributions with densities of the form

$$p\left(y_t|\theta_t\right) = \exp\left[y_t'\theta_t - b_t(\theta_t) + c_t(y_t)\right], \tag{5}$$

where $b_t\left(\theta_t\right)$ is twice differentiable and $c_t\left(y_t\right)$ is a function of y_t only; (ii) observations generated by the relation

$$y_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim p\left(\varepsilon_t\right), \tag{6}$$

where the ε_t's are non-Gaussian and serially independent.

In the next section we will develop estimation formulae which provide the basis for our simulation methodology. We will do this for both classical and Bayesian inference. In the terminology of Bayesian analysis, all the models in this section are hierarchical models, in which the elements of $\alpha_1, \ldots, \alpha_n$ are the parameters and the elements of ψ are the hyperparameters; see, for example, Bernardo and Smith (1994, p.371).

3 Basic simulation formulae

3.1 Introduction

In this section we develop the basic formulae underlying our simulation methods; details for practical calculation will be given in section 5. Denote the stacked vectors $(\alpha_1', \ldots, \alpha_n')'$ and $(y_1', \ldots, y_n')'$ by α and y. Most of the problems considered in this paper are essentially the estimation of the conditional mean

$$\bar{x} = \mathrm{E}[x(\alpha)|y] \tag{7}$$

of an arbitrary function $x(\alpha)$ of α given the observation vector y. This formulation includes estimates of quantities of interest such as the mean $\mathrm{E}(\alpha_t|y)$ of the state vector α_t given y and its conditional variance matrix $\mathrm{Var}(\alpha_t|y)$; it also includes estimates of the conditional density and distribution function of $x(\alpha)$ given y in the classical case and the posterior density and distribution function of $x(\alpha)$ in the Bayesian case. We shall estimate $\bar{x}$ by simulation methods that are similar to those used in Shephard and Pitt (1997) and Durbin and Koopman (1997) for estimating the likelihood in non-Gaussian state space models. The methods are based on standard ideas in simulation methodology, namely importance sampling and antithetic variables, as described, for example, in Ripley (1987); in particular, we make no use of Markov chain Monte Carlo (MCMC) methods. As a result, our simulation samples are independent so we can easily calculate variances of errors due to simulation, and we avoid the convergence problems associated with MCMC techniques. Nevertheless, our methods are computationally very efficient as we shall demonstrate. The techniques we shall describe will be based on Gaussian importance densities. We shall use the generic notation $g(\cdot)$, $g(\cdot,\cdot)$ and $g(\cdot|\cdot)$ for Gaussian marginal, joint and conditional densities.

3.2 Formulae for classical inference

Let us first consider the classical inference case where the parameter vector ψ is assumed to be fixed and unknown and is estimated by its maximum likelihood estimate $\hat{\psi}$ obtained by numerically maximising the Monte Carlo likelihood function as discussed in section 5.4. For given ψ, let $g(\alpha|y)$ be a Gaussian importance density which is chosen to resemble $p(\alpha|y)$ as closely as is reasonably possible; we have from (7),

$$\bar{x} = \int x(\alpha)p(\alpha|y)d\alpha = \int x(\alpha)\frac{p(\alpha|y)}{g(\alpha|y)}g(\alpha|y)d\alpha = \mathrm{E}_g\left[x(\alpha)\frac{p(\alpha|y)}{g(\alpha|y)}\right], \tag{8}$$

where E_g denotes expectation with respect to the importance density $g(\alpha|y)$. For the models of section 2, $p(\alpha|y)$ and $g(\alpha|y)$ are complicated algebraically, whereas the corresponding joint densities $p(\alpha, y)$ and $g(\alpha, y)$ are straightforward. We therefore put $p(\alpha|y) = p(\alpha, y)/p(y)$ and $g(\alpha|y) = g(\alpha, y)/g(y)$ in (8), giving

$$\bar{x} = \frac{g(y)}{p(y)}\mathrm{E}_g\left[x(\alpha)\frac{p(\alpha, y)}{g(\alpha, y)}\right]. \tag{9}$$

Putting $x(\alpha) = 1$ in (7) and (9) we have

$$1 = \frac{g(y)}{p(y)}\mathrm{E}_g\left[\frac{p(\alpha, y)}{g(\alpha, y)}\right]. \tag{10}$$

Taking the ratios of these gives

$$\bar{x} = \frac{\mathrm{E}_g\left[x(\alpha)w(\alpha, y)\right]}{\mathrm{E}_g\left[w(\alpha, y)\right]}, \quad \text{where} \quad w(\alpha, y) = \frac{p(\alpha, y)}{g(\alpha, y)}. \tag{11}$$

This formula provides the basis for the bulk of the work in this paper. For example, it can be used to estimate conditional variances of quantities of interest as well as conditional densities and distribution functions. We could in principle obtain a Monte Carlo estimate $\hat{x}$ of $\bar{x}$ in the following way. Choose a series of independent draws $\alpha^{(1)}, \ldots, \alpha^{(N)}$ from the distribution with density $g(\alpha|y)$ and take

$$\hat{x} = \frac{\sum_{i=1}^{N} x_i w_i}{\sum_{i=1}^{N} w_i}, \quad \text{where} \quad x_i = x(\alpha^{(i)}) \quad \text{and} \quad w_i = w(\alpha^{(i)}, y). \tag{12}$$

Since the draws are independent, and under assumptions which are satisfied in practical cases, $\hat{x}$ converges to $\bar{x}$ probabilistically as $N \to \infty$. However, this simple estimate is numerically inefficient and we shall refine it considerably in section 5.

An important special case is where the observations are non-Gaussian but the state vector is generated by the linear Gaussian model (2). We then have $p(\alpha) = g(\alpha)$ so

$$\frac{p(\alpha, y)}{g(\alpha, y)} = \frac{p(\alpha)\, p(y|\alpha)}{g(\alpha)\, g(y|\alpha)} = \frac{p(y|\alpha)}{g(y|\alpha)} = \frac{p(y|\theta)}{g(y|\theta)}.$$

Thus (11) becomes the simpler formulae

$$\bar{x} = \frac{\mathrm{E}_g\left[x(\alpha) w^*(\theta, y)\right]}{\mathrm{E}_g\left[w^*(\theta, y)\right]} \quad \text{where} \quad w^*(\theta, y) = \frac{p(y|\theta)}{g(y|\theta)}; \tag{13}$$

its estimate $\hat{x}$ is given by an obvious analogue of (12).

3.3 Formulae for Bayesian inference

Now let us consider the problem from a Bayesian point of view. The parameter vector ψ is regarded as random with prior density $p(\psi)$ which to begin with we take as a proper prior. As before, suppose we wish to calculate $\bar{x} = \mathrm{E}[x(\alpha)|y]$. This now takes the form

$$\bar{x} = \int x(\alpha) p(\psi, \alpha|y) d\psi d\alpha.$$

We have

$$p(\psi, \alpha|y) = p(\psi|y) p(\alpha|\psi, y)$$

where by Bayes' theorem

$$p(\psi|y) = K p(\psi) p(y|\psi)$$

in which K is a normalising constant. Thus

$$\bar{x} = K \int x(\alpha) p(\psi) p(y|\psi) p(\alpha|\psi, y) d\psi d\alpha. \tag{14}$$

Consider the approximation of the posterior density $p(\psi|y)$ by its large sample normal approximation

$$g(\psi|y) = \mathrm{N}(\hat{\psi}, \hat{V}),$$

where $\hat{\psi}$ is the solution of the equation

$$\frac{\partial \log p(\psi|y)}{\partial \psi} = \frac{\partial \log p(\psi)}{\partial \psi} + \frac{\partial \log p(y|\psi)}{\partial \psi} = 0, \tag{15}$$

and

$$\hat{V}^{-1} = -\frac{\partial^2 \log p(\psi)}{\partial \psi \partial \psi'} - \frac{\partial^2 \log p(y|\psi)}{\partial \psi \partial \psi'}\bigg|_{\psi = \hat{\psi}}. \tag{16}$$

The value $\hat{\psi}$ is computed iteratively by an obvious extension of the techniques of Durbin and Koopman (1997) using a linearisation at a trial value $\tilde{\psi}$ of ψ, while the second derivatives can be calculated numerically. For discussion of large sample approximations to $p(\psi|y)$ see Gelman *et.al.* (1995, Chapter 4) and Bernardo and Smith (1994, section 5.3).

We shall use $g(\psi|y)$ as an importance density for $p(\psi|y)$. Let $g(\alpha|\psi, y)$ be an appropriate Gaussian importance density for $p(\alpha|\psi, y)$ analogous to $g(\alpha|y)$ in (8). We can then rewrite (14) as

$$\begin{aligned}
\bar{x} &= K \int x(\alpha) \frac{p(\psi)p(y|\psi)}{g(\psi|y)} \frac{p(\alpha|\psi, y)}{g(\alpha|\psi, y)} g(\psi|y)g(\alpha|\psi, y)d\psi d\alpha \\
&= K \int x(\alpha) \frac{p(\psi)g(y|\psi)}{g(\psi|y)} \frac{p(\alpha, y|\psi)}{g(\alpha, y|\psi)} g(\psi, \alpha|y)d\psi d\alpha \\
&= K \, \mathrm{E}_g \left[x(\alpha) \frac{p(\psi)g(y|\psi)}{g(\psi|y)} \frac{p(\alpha, y|\psi)}{g(\alpha, y|\psi)} \right], \qquad (17)
\end{aligned}$$

where E_g now denotes expectation with respect to the importance joint density $g(\psi, \alpha|y) = g(\psi|y)g(\alpha|\psi, y)$. It is very fortunate that the quantity $p(y|\psi)$, which is hard to compute, conveniently drops out of this expression. Taking the ratio of this expression for $\bar{x}$ to the same expression with $x(\alpha)$ equal to one, the term K disappears, giving analogously to (11),

$$\bar{x} = \frac{\mathrm{E}_g \left[x(\alpha)z(\psi, \alpha, y) \right]}{\mathrm{E}_g \left[z(\psi, \alpha, y) \right]}, \quad \text{where} \quad z(\psi, \alpha, y) = \frac{p(\psi)g(y|\psi)}{g(\psi|y)} \frac{p(\alpha, y|\psi)}{g(\alpha, y|\psi)}. \qquad (18)$$

This formula provides the basis for our results in the Bayesian case. It can be used to obtain estimates of posterior means, variances, densities and distribution functions. In principle we could compute a Monte Carlo point estimate $\hat{x}$ of $\bar{x}$ as follows. Let $\psi^{(i)}$ be a random draw from $g(\psi|y)$ and let $\alpha^{(i)}$ be a random draw from $g(\alpha|\psi^{(i)}, y)$ for $i = 1, \ldots, N$; we assume here that we only draw one $\alpha^{(i)}$ for each $\psi^{(i)}$ though, of course, more could be drawn if desired. Then take

$$\hat{x} = \frac{\sum_{i=1}^N x_i z_i}{\sum_{i=1}^N z_i}, \quad \text{where} \quad x_i = x(\alpha^{(i)}) \quad \text{and} \quad z_i = z(\psi^{(i)}, \alpha^{(i)}, y). \qquad (19)$$

We see that the only difference between (19) and (12) is the replacement of w_i by z_i which allows for the effect of drawing values of ψ from $g(\psi|y)$. This simple form of the simulation will be improved later. The term $g(y|\psi^{(i)})$ in z_i is easily calculated by the Kalman filter.

For cases where a proper prior is not available we may wish to use a non-informative prior in which we assume that the prior density is proportional to a specified function $p(\psi)$ in a domain of ψ of interest even though the integral $\int p(\psi)d\psi$ does not exist. For a discussion of non-informative priors see, for example, in Chapters 2 and 3 of Gelman *et.al.* (1995). Where it exists, the posterior density is $p(\psi|y) = Kp(\psi)p(y|\psi)$ as in the proper prior case so all the previous formulae apply without change. This is why we use the same symbol $p(\psi)$ for both cases even though in the non-informative case $p(\psi)$ is not a density. An important special case is the diffuse prior for which $p(\psi) = 1$ for all ψ.

3.4 Bayesian analysis for the linear Gaussian model

Although this paper is directed at non-Gaussian models, let us digress briefly to consider the application of the above Bayesian treatment to the linear Gaussian model (1) and (2), since this model is important in practical applications and our methodology is new. Let

$$\bar{x}(\psi) = \mathrm{E}[x(\alpha)|\psi, y] = \int x(\alpha)p(\alpha|\psi, y)d\alpha,$$

and assume that for given ψ, $\bar{x}(\psi)$ is obtainable by a routine Kalman filtering and smoothing operation; for example, $x(\psi)$ could be an estimate of the trend at time t or it could be a forecast of y_t at time $t > n$. Then

$$\begin{aligned}\bar{x} &= \int \bar{x}(\psi)p(\psi|y)d\psi = K\int \bar{x}(\psi)p(\psi)p(y|\psi)d\psi = K\int \bar{x}(\psi)z^g(\psi,y)g(\psi|y)d\psi \\ &= K\ \mathrm{E}_g[\bar{x}(\psi)z^g(\psi,y)], \qquad \text{where} \quad z^g(\psi,y) = \frac{p(\psi)g(y|\psi)}{g(\psi|y)}\end{aligned}$$

and E_g is expectation with respect to the importance density $g(\psi|y)$; note that we write $g(y|\psi)$ in place of $p(y|\psi)$ since $p(y|\psi)$ is Gaussian. Analogously to (18) we therefore have

$$\bar{x} = \frac{\mathrm{E}_g[\bar{x}(\alpha)z^g(\psi,y)]}{\mathrm{E}_g[z^g(\psi,y)]}, \tag{20}$$

while for practical calculation there is an obvious analogue of (19). In (20), $z^g(\psi,y)$ depends on the likelihood $g(y|\psi)$ which can be computed by a routine Kalman filtering operation for the linear Gaussian model.

4 Approximating linear Gaussian models

4.1 Introduction

In this section we obtain the Gaussian importance densities that we need for simulation by constructing linear Gaussian models which approximate the non-Gaussian model in the neighbourhood of the conditional mode of α given y. Let $g(\alpha|y)$ and $g(\alpha,y)$ be the conditional and joint densities generated by model (1) and (2) and let $p(\alpha|y)$ and $p(\alpha,y)$ be the corresponding densities generated by model (3) and (4). We will determine the approximating model by choosing H_t and Q_t so that densities $g(\alpha|y)$ and $p(\alpha|y)$ have the same mode $\hat{\alpha}$. The possibility that $p(\alpha,y)$ might be multimodal will be considered in section 4.6. Taking the Gaussian model first, $\hat{\alpha}$ is the solution of the vector equation $\partial \log g(\alpha|y)/\partial\alpha = 0$. Now $\log g(\alpha|y) = \log g(\alpha,y) - \log g(y)$. Thus, the mode is also the solution of the vector equation $\partial \log g(\alpha,y)/\partial\alpha = 0$. This version of the equation is easier to manage since $g(\alpha,y)$ has a simple form whereas $g(\alpha|y)$ does not. Since R_t consists of columns of I_m, $\eta_t = R_t'(\alpha_t - T_t\alpha_{t-1})$. We therefore have

$$\begin{aligned}\log g(\alpha,y) &= \text{constant} - \frac{1}{2}\sum_{t=1}^{n}(\alpha_t - T_t\alpha_{t-1})' R_t Q_t^{-1} R_t'(\alpha_t - T_t\alpha_{t-1}) - \\ &\quad \frac{1}{2}\sum_{t=1}^{n}(y_t - Z_t\alpha_t)' H_t^{-1}(y_t - Z_t\alpha_t).\end{aligned}$$

Differentiating with respect to α_t and equating to zero gives the equations

$$\begin{aligned}-R_tQ_t^{-1}R_t'(\alpha_t - T_t\alpha_{t-1}) \;+\; d_tT_{t+1}'R_{t+1}Q_{t+1}^{-1}R_{t+1}'(\alpha_{t+1} - T_{t+1}\alpha_t) \;+\; \\ Z_t'H_t^{-1}(y_t - Z_t\alpha_t) \;=\; 0,\end{aligned} \tag{21}$$

for $t = 1,\ldots,n$, where $d_t = 1$ for $t < n$ and $d_n = 0$. The solution to these equations is the conditional mode $\widehat{\alpha}$. Since $g(\alpha|y)$ is Gaussian the mode is equal to the mean so $\hat{\alpha}$ can be routinely calculated by the Kalman filter and smoother (KFS); for details of the KFS see Harvey (1989, Chapter 3). It follows that linear equations of the form (21) can be solved by the KFS which is known to be very efficient computationally.

Assuming that the non-Gaussian model (3) and (4) is sufficiently well behaved, the mode $\hat{\alpha}$ of $p(\alpha|y)$ is the solution of the vector equation

$$\frac{\partial \log p(\alpha|y)}{\partial\alpha} = 0$$

and hence of the equation

$$\frac{\partial \log p(\alpha, y)}{\partial \alpha} = 0.$$

Let $q_t(\eta_t) = -\log p(\eta_t)$ and let $h_t(y_t|\theta_t) = -\log p(y_t|\theta_t)$. Then,

$$\log p(\alpha, y) = \text{constant} - [q_t(\eta_t) + h_t(y_t|\theta_t)], \tag{22}$$

with $\eta_t = R_t'(\alpha_t - T_t\alpha_{t-1})$ so $\widehat{\alpha}$ is a solution of the equations

$$\begin{aligned} \frac{\partial \log p(\alpha, y)}{\partial \alpha_t} &= -R_t \frac{\partial q_t(\eta_t)}{\partial \eta_t} + d_t T_{t+1}' R_{t+1} \frac{\partial q_{t+1}(\eta_{t+1})}{\partial \eta_{t+1}} - Z_t' \frac{\partial h_t(y_t|\theta_t)}{\partial \theta_t} \\ &= 0, \end{aligned} \tag{23}$$

for $t = 1, \ldots, n$, where, as before, $d_t = 1$ for $t = 1, \ldots, n-1$ and $d_n = 0$. We solve these equations by iteration, where at each step we linearise, put the result in the form (21) and solve by the KFS. Convergence is fast and normally only around ten iterations or less are needed. A different method of solving these equations was given by Fahrmeir and Kaufmann (1991) but it is more cumbersome than our method.

4.2 Linearisation for non-Gaussian observation densities: Method 1

We shall consider two methods of linearising the observation component of (23). The first method enables exponential family observations, such as Poisson distributed observations, to be handled; the second method is given in section 4.4 and deals with observations having the form (6) when $p(\varepsilon_t)$ is a function of ε_t^2; this is suitable for distributions with heavy tails such as the t-distribution.

Suppose that $\widetilde{\alpha} = [\widetilde{\alpha}_1', \ldots, \widetilde{\alpha}_n']'$ is a trial value of α, let $\widetilde{\theta}_t = Z_t\widetilde{\alpha}_t$ and define

$$\dot{h}_t = \left.\frac{\partial h_t(y_t|\theta_t)}{\partial \theta_t}\right|_{\theta_t = \widetilde{\theta}_t}, \qquad \ddot{h}_t = \left.\frac{\partial^2 h_t(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'}\right|_{\theta_t = \widetilde{\theta}_t}. \tag{24}$$

Expanding about $\widetilde{\theta}_t$ gives approximately

$$\frac{\partial h_t(y_t|\theta_t)}{\partial \theta_t} = \dot{h}_t + \ddot{h}_t\left(\theta_t - \widetilde{\theta}_t\right). \tag{25}$$

Substituting in the final term of (23) gives the linearised form

$$-Z_t'\left(\dot{h}_t + \ddot{h}_t\theta_t - \ddot{h}_t\widetilde{\theta}_t\right). \tag{26}$$

To put this in the same format as the final term of (21) put

$$\widetilde{H}_t = \ddot{h}_t^{-1}, \qquad \widetilde{y}_t = \widetilde{\theta}_t - \ddot{h}_t^{-1}\dot{h}_t. \tag{27}$$

Then the final term becomes $Z_t\widetilde{H}_t^{-1}(\widetilde{y}_t - \theta_t)$ as required.

Consider, for example, the important special case in which the state equation retains the original linear Gaussian form (2). Equations (23) then have the linearised form

$$\begin{aligned} -R_tQ_t^{-1}R_t'(\alpha_t - T_t\alpha_{t-1}) + d_tT_{t+1}'R_{t+1}Q_{t+1}^{-1}R_{t+1}'(\alpha_{t+1} - T_{t+1}\alpha_t) + \\ Z_t'\widetilde{H}_t^{-1}(\widetilde{y}_t - Z_t\alpha_t) = 0, \end{aligned} \tag{28}$$

analogous to (21), which can be solved for α by the KFS to give a new trial value and the process is repeated until convergence. The values of α and θ after convergence to the mode are denoted by $\hat{\alpha}$ and $\hat{\theta}$, respectively.

It is evident from (27) that Method 1 only works when $\ddot{h}_t$ is positive definite. When $\ddot{h}_t$ is negative-definite or semi-definite, Method 2 should normally be used. Finally, it is important to note that the first and second derivatives of the log Gaussian density of y_t given θ_t, as implied by (27), are the same as the first and second derivatives of $-h_t(y_t|\theta_t)$ at the mode $\theta = \hat{\theta}$, for $t = 1, \ldots, n$. This means that not only does the approximating linear Gaussian model have the same conditional mode as model (3) and (4), it has the same curvature at the mode also.

4.3 Exponential family observations

The most important application of these results is to time series of observations from exponential family distributions, such as Poisson, binomial and multinomial observations. The model with observational density (5) together with linear Gaussian state equation (2) was introduced by West, Harrison and Migon (1985). They called it the dynamic generalised linear model and they fitted it by an approximate Bayesian technique based on conjugate priors.

For density (5),

$$h_t(y_t|\theta_t) = -\log p(y_t|\theta_t) = -\left[y_t'\theta_t - b_t(\theta_t) + c_t(y_t)\right]. \tag{29}$$

Define

$$\begin{aligned} \dot{b}_t &= \left.\frac{\partial b_t(\theta_t)}{\partial \theta_t}\right|_{\theta_t = \widetilde{\theta}_t}, \\ \ddot{b}_t &= \left.\frac{\partial^2 b_t(\theta_t)}{\partial \theta_t \partial \theta_t'}\right|_{\theta_t = \widetilde{\theta}_t}. \end{aligned}$$

Then $\dot{h}_t = \dot{b}_t - y_t$ and $\ddot{h}_t = \ddot{b}_t$ so using (27) we take $\widetilde{H}_t = \ddot{b}_t^{-1}$ and $\widetilde{y}_t = \widetilde{\theta}_t - \ddot{b}_t^{-1}(\dot{b}_t - y_t)$. These values can be substituted in (28) to obtain a solution for the case where the state equation is linear and Gaussian. Since, as is well known, $\ddot{b}_t = \mathrm{Var}(y_t|\theta_t)$, it is positive definite in non-degenerate cases, so for the exponential family, Method 1 can always be used.

4.4 Linearisation for non-Gaussian observation densities: Method 2

We now consider the case where the observations are generated by model (6). We shall assume that y_t is univariate and that $p(\varepsilon_t)$ is a function of ε_t^2; this case is important for heavy-tailed densities such as the t distribution, and for Gaussian mixtures with zero means.

Let $\log p(\varepsilon_t) = -\frac{1}{2}h_t^*(\varepsilon_t^2)$. Then the contribution of the observation component to the equation $\partial \log p(\alpha, y)/\partial\alpha$ is

$$-\frac{1}{2}\frac{\partial h_t^*(\varepsilon_t^2)}{\partial \varepsilon_t^2}\frac{\partial \varepsilon_t^2}{\partial \alpha_t} = Z_t'\frac{\partial h_t^*(\varepsilon_t^2)}{\partial \varepsilon_t^2}(y_t - \theta_t). \tag{30}$$

Let

$$\dot{h}_t^* = \left.\frac{\partial h_t^*(\varepsilon_t^2)}{\partial \varepsilon_t^2}\right|_{\varepsilon_t = y_t - \widetilde{\theta}_t}. \tag{31}$$

Then take $Z_t'\dot{h}_t^*(y_t - \theta_t)$ as the linearised form of (30). By taking $\widetilde{H}_t^{-1} = \dot{h}_t^*$ we have the observation component in the correct form (21) so we can use the KFS at each step of the solution of the equation $\partial \log p(\alpha, y)/\partial\alpha = 0$. We emphasise that now only the first derivative of the log of the implied Gaussian density of ε_t is equal to that of $p(\varepsilon_t)$, compared to method 1 which equalised the first and second derivatives.

It is of course necessary for this method to work that $\dot{h}_t^*$ is positive with probability one; however, this condition is satisfied for the applications we consider below. Strictly speaking it is not essential that $p(\varepsilon_t)$ is a function of ε_t^2. In other cases we could define

$$\dot{h}_t^* = -\frac{1}{\varepsilon_t}\frac{\partial \log p(\varepsilon_t)}{\partial \varepsilon_t}\Bigg|_{\varepsilon_t = y_t - \widetilde{\theta}_t} , \tag{32}$$

and proceed in the same way. Again, the method only works when $\dot{h}_t^*$ is positive with probability one.

4.5 Linearisation when the state errors are non-Gaussian

We now consider the linearisation of the state component in equations (23) when the state errors η_t are non-Gaussian. Suppose that $\widetilde{\eta} = [\widetilde{\eta}_1', \ldots, \widetilde{\eta}_n']'$ is a trial value of $\eta = [\eta_1', \ldots, \eta_n']'$ where $\widetilde{\eta}_t = R_t'(\widetilde{\alpha}_t - T_t\widetilde{\alpha}_{t-1})$. In this paper we shall confine ourselves to the situation where the elements η_{it} of η_t are mutually independent and where the density $p(\eta_{it})$ of η_{it} is a function of η_{it}^2. These assumptions are not very restrictive since they enable us to deal relatively easily with two cases of particular interest in practice, namely heavy-tailed errors and models with structural shifts using method 2 of subsection 4.4.

Let $q_{it}^*(\eta_{it}^2) = -2\log p(\eta_{it})$ and denote the i-th column of R_t by R_{it}. Then the state contribution to the conditional mode equations (23) is

$$-\frac{1}{2}\sum_{i=1}^{r}\left[R_{it}\frac{\partial q_{it}^*(\eta_{it}^2)}{\partial \eta_{it}} - d_t T_{t+1}' R_{i,t+1}\frac{\partial q_{i,t+1}^*\left(\eta_{i,t+1}^2\right)}{\partial \eta_{i,t+1}}\right], \qquad t = 1, \ldots, n. \tag{33}$$

The linearised form of (33) is

$$-\sum_{i=1}^{r}\left[R_{it}\dot{q}_{it}^*\eta_{it} - d_t T_{t+1}' R_{i,t+1}\dot{q}_{i,t+1}^*\eta_{i,t+1}\right], \tag{34}$$

where

$$\dot{q}_{it}^* = \frac{\partial q_{it}^*(\eta_{it}^2)}{\partial \eta_{it}^2}\Bigg|_{\eta_t = \widetilde{\eta}_t} . \tag{35}$$

Putting $\widetilde{Q}_t^{-1} = diag[\dot{q}_{1t}^*, \ldots, \dot{q}_{rt}^*]$, $\eta_t = R_t'(\alpha_t - T_t\alpha_{t-1})$, and similarly for $\widetilde{Q}_{t+1}$ and η_{t+1}, we see that (34) has the same form as the state component of (21). Consequently, in the iterative estimation of $\widehat{\alpha}$ the KFS can be used to update the trial value $\widetilde{\alpha}$.

4.6 Discussion

So far in this section we have emphasised the use of the mode $\hat{\alpha}$ of $p(\alpha|y)$ to obtain a linear approximating model which we use to calculate the Gaussian densities for simulation using the techniques of the next section. If, however, the sole object of the investigation was to estimate α and if economy in computation was desired, then $\hat{\alpha}$ could be used for the purpose; indeed, this was the estimator used by Durbin and Koopman (1992) and an approximation to it was used by Fahrmeir (1992). Our experience has been that there is very little difference in the examples we have examined between the mode and the mean $\mathrm{E}(\alpha|y)$. A disadvantage of this use of the mode, however, is that there is no accompanying estimate of its error variance matrix.

We have assumed above that there is a single mode and the question arises whether multimodality will create complications. If multimodality is suspected it can be investigated by using different starting points and checking whether iterations from them converge to the same mode. In none of the cases we have examined has multimodality of $p(\alpha|y)$ caused any difficulties. For

this reason we do not believe that this will give rise to problems in routine time series analysis. If, however, multimodality were to occur in a particular case, we would suggest fitting a linear Gaussian model to the data at the outset and using this to define the first importance density $g_1(\eta|y)$ and conditional joint density $g_1(\eta, y)$. Simulation based on these using the methods of the next section is employed to obtain a first estimate $\tilde{\eta}^{(1)}$ of $\mathrm{E}(\eta|y)$ and from this a first estimate $\tilde{\theta}_t^{(1)}$ of θ_t is calculated for $t = 1, \ldots, n$. Now linearise $\log p(y_t|\theta_t)$ at $\tilde{\theta}_t^{(1)}$ as in section 4.2 or 4.4 and $\log p(\eta_t)$ at $\tilde{\eta}^{(1)}$ as in section 4.5. These linearisations give a new approximating linear Gaussian model which defines a new $g(\eta|y)$, $g_2(\eta|y)$, and a new $g(\eta, y)$, $g_2(\eta, y)$. Simulation using these gives a new estimate $\tilde{\eta}^{(2)}$ of $\mathrm{E}(\eta|y)$. This iterative process is continued until adequate convergence is achieved. However, we emphasise that it is not necessary for the value of α at which the model is linearised to be a precisely accurate estimate of either the mode or the mean of $p(\alpha|y)$. The only way that the choice of the value of α used as the basis for the simulation affects the final estimate $\hat{x}$ is in the variances due to simulation which, as we shall show below, are accurately estimated as a routine part of the simulation procedure. Where necessary, the simulation sample size can be increased to reduce these error variances to any required extent. It will be noted that we are basing the iterations on the mean, not the mode. Since the mean, when it exists, is unique, no question of 'multimeanality' can arise.

5 Computational methods

5.1 Introduction

In this section we discuss suitable computational methods for estimating $\bar{x}$ given by (11) when classical inference is used and $\bar{x}$ given by (18) when Bayesian inference is used. We begin with (11). The starting point for classical analysis is that we take $\psi = \hat{\psi}$ where $\hat{\psi}$ is the maximum likelihood estimate of ψ determined as described in section 5.4. During the simulations it is important to work with variables in their simplest forms. Thus for the observation equation (3) we work with the signal $\theta_t = Z_t\alpha_t$ and for the state equation (4) we work with the state disturbance η_t. Substituting for α in terms of η in (11) gives

$$\bar{x} = \frac{\mathrm{E}_g[x(\eta)w(\eta, y)]}{\mathrm{E}_g[w(\eta, y)]} \qquad \text{where} \qquad w(\eta, y) = \frac{p(\eta, y)}{g(\eta, y)} \tag{36}$$

and α is obtained from η using the relations $\alpha_t = T_t\alpha_{t-1} + R_t\eta_t$ for $t = 1, \ldots, n$. We take the symbols $x(\eta)$ and $w(\eta, y)$ as denoting here the functions of η that we obtain by substituting for α in terms of η in $x(\alpha)$ and $w(\alpha, y)$ in (11). Also we take E_g as denoting expectation with respect to the importance density $g(\eta|y)$.

5.2 Simulation smoother and antithetic variables

The simulations are based on random draws of η from the importance density $g(\eta|y)$ using the simulation smoother of de Jong and Shephard (1995); this computes efficiently a draw for η as a linear function of rn independent standard normal deviates where r is the dimension of vector η_t and n is the number of observations. Efficiency is increased by the use of antithetic variables. We shall employ two types of antithetic variables. The first is the standard one given by $\check{\eta} = 2\hat{\eta} - \eta$ where $\hat{\eta} = \mathrm{E}_g(\eta)$ can be obtained via the disturbance smoother; for details of the disturbance smoother see Koopman (1993). Since $\check{\eta} - \hat{\eta} = -(\eta - \hat{\eta})$ and η is normal, the two vectors η and $\check{\eta}$ are equi-probable. Thus we obtain two simulation samples from each draw of the simulation smoother; moreover, values of conditional mean calculated from the two samples are negatively correlated, giving further efficiency gains.

The second antithetic variable was developed by Durbin and Koopman (1997). Let u be the vector of rn $N(0, 1)$ variables that is used in the simulation smoother to generate η and let

$c = u'u$; then $c \sim \chi^2_{rn}$. For a given value of c let $q = \Pr(\chi^2_{rn} < c) = F(c)$ and let $\acute{c} = F^{-1}(1-q)$. Then as c varies, c and $\acute{c}$ have the same distribution. Now take, $\acute{\eta} = \hat{\eta} + \sqrt{\acute{c}/c}(\eta - \hat{\eta})$. Then $\acute{\eta}$ has the same distribution as η. This follows because c and $(\eta - \hat{\eta})/\sqrt{c}$ are independently distributed. Finally, take $\grave{\eta} = \hat{\eta} + \sqrt{\acute{c}/c}(\check{\eta} - \hat{\eta})$. Thus we obtain a balanced set of four equi-probable values of η for each run of the simulation smoother.

The number of antithetics can be increased without difficulty. For example, take c and q as above. Then q is uniform on $(0,1)$ and we write $q \sim U(0,1)$. Let $q_1 = q + 0.5$ modulo 1; then $q_1 \sim U(0,1)$ and we have a balanced set of four $U(0,1)$ variables, q, q_1, $1-q$ and $1-q_1$. Take $\acute{c} = F^{-1}(1-q)$ as before and similarly $c_1 = F^{-1}(q_1)$ and $\acute{c}_1 = F^{-1}(1-q_1)$. Then each of c_1 and $\acute{c}_1$ can be combined with η and $\check{\eta}$ as was $\acute{c}$ previously and we emerge with a balanced set of eight equi-probable values of η for each simulation. In principle this process could be extended indefinitely by taking $q_1 = q$ and $q_{j+1} = q_j + 2^{-k}$ modulo 1, for $j = 1, \ldots, 2^{k-1}$ and $k = 2, 3, \ldots$; however, four values of q are probably enough in practice. By using the standard normal distribution function applied to elements of u, the same idea could be used to obtain a new balanced value η_1 from η so by taking $\check{\eta}_1 = 2\hat{\eta} - \eta_1$ we would have four values of η to combine with the four values of c. In the following we will assume that we have generated N draws of η using the simulation smoother and the antithetic variables; in practice, we will work with the two basic antithetics so N will be a multiple of 4.

In theory, importance sampling could give an inaccurate result on a particular occasion if in the basic formulae (36) very high values of $w(\eta, y)$ are associated with very small values of the importance density $g(\eta|y)$ in such a way that together they make a significant contribution to $\bar{x}$, and if also, on this particular occasion, these values happen to be over- or under-represented; for further discussion of this point see Gelman *et.al.* (1995, p.307). In practice we have not experienced difficulties from this source in any of the examples we have considered. Nevertheless we recognise that difficulties could occur if the tail densities of $p(\eta|y)$ were substantially thicker than those of $g(\eta|y)$. We have developed a way of simulating values of η with thicker tails than those of the Gaussian but the methods are not used in this paper and there is not the space here to discuss details; we refer to the technical report of Durbin and Koopman (1999).

5.3 Estimating means, variances, densities and distribution functions

We first consider estimation of conditional means and error variances of our estimates. Let $w(\eta) = p(\eta, y)/g(\eta, y)$, taking the dependence on y as implicit since y is constant from now on. Then (36) gives

$$\bar{x} = \mathrm{E}_g\left[x(\eta)w(\eta)\right] \ / \ \mathrm{E}_g\left[w(\eta)\right]. \tag{37}$$

which is estimated by

$$\hat{x} = \sum_{i=1}^{N} x_i w_i \ / \ \sum_{i=1}^{N} w_i \qquad \text{where} \quad x_i = x(\eta^{(i)}), \quad w_i = w(\eta^{(i)}) = \frac{p(\eta^{(i)}, y)}{g(\eta^{(i)}, y)}, \tag{38}$$

and $\eta^{(i)}$ is the i-th draw from the importance density $g(\eta|y)$ for $i = 1, \ldots, N$. For the case where $x(\eta)$ is a vector we could at this point present formulae for estimating the matrix $\mathrm{Var}[x(\eta)|y]$ and also the variance matrix due to simulation of $\hat{x} - \bar{x}$. However, from a practical point of view the covariance terms are of little interest so it seems sensible to focus on variance terms by taking $x(\eta)$ as a scalar; extension to include covariance terms is straightforward. We estimate $\mathrm{Var}[x(\eta)|y]$ by

$$\widehat{\mathrm{V}}\mathrm{ar}[x(\eta)|y] = \left(\sum_{i=1}^{N} x_i^2 w_i \ / \ \sum_{i=1}^{N} w_i\right) - \hat{x}^2. \tag{39}$$

The estimation error from the simulation is

$$\hat{x} - \bar{x} = \sum_{i=1}^{N} w_i(x_i - \bar{x}) \ / \ \sum_{i=1}^{N} w_i.$$

Denote the sum of the four values of $w_i(x_i - \bar{x})$ that come from the j-th run of the simulation smoother by v_j and the sum of the corresponding values of $w_i(x_i - \hat{x})$ by $\hat{v}_j$. For N large enough, since the draws from the simulation smoother are independent, the variance due to simulation is, to a good approximation,

$$\mathrm{Var}_s(\hat{x}) = \frac{N}{4}\mathrm{Var}(v_j) \ / \ \left(\sum_{i=1}^{N} w_i\right)^2, \tag{40}$$

which we estimate by

$$\widehat{\mathrm{Var}}_s(\hat{x}) = \sum_{j=1}^{N/4} \hat{v}_j^2 \ / \ \left(\sum_{i=1}^{N} w_i\right)^2. \tag{41}$$

The fact that we can estimate simulation variances so easily is one of the advantages of our methods over Markov chain Monte Carlo methods.

When $x(\eta)$ is a scalar the above technique can be used to estimate the conditional distribution function and the conditional density function of x. Let $G[x|y] = \Pr[x(\eta) \leq x|y]$ and let $I_x(\eta)$ be an indicator which is unity if $x(\eta) \leq x$ and is zero if $x(\eta) > x$. Then $G(x|y) = \mathrm{E}(I_x(\eta)|y)$. Since $I_x(\eta)$ is a function of η we can treat it in the same way as $x(\eta)$. Let S_x be the sum of the values of w_i for which $x_i \leq x$, for $i = 1, \ldots, N$. Then estimate $G(x|y)$ by

$$\hat{G}(x|y) = S_x \ / \ \sum_{i=1}^{N} w_i. \tag{42}$$

This can be used to estimate quantiles. Similarly, if δ is the interval $\left(x - \frac{1}{2}d, x + \frac{1}{2}d\right)$ where d is suitably small and positive, let S^δ be the sum of the values of w_i for which $x(\eta) \in \delta$. Then the estimator of the conditional density $p(x|y)$ of x given y is

$$\hat{p}(x|y) = d^{-1}S^\delta \ / \ \sum_{i=1}^{N} w_i. \tag{43}$$

This estimator can be used to construct a histogram.

We now show how to generate a sample of M independent values from the estimated conditional distribution of $x(\eta)$ using importance resampling; for further details of the method see Gelfand and Smith (1999) and Gelman *et.al.* (1995). Take $x^{[k]} = x_j$ with probability $w_j / \sum_{i=1}^{N} w_i$ for $j = 1, \ldots, N$. Then

$$\Pr(x^{[k]} \leq x) = \frac{\sum_{x_j \leq x} w_j}{\sum_{i=1}^{N} w_i} = \hat{G}(x|y).$$

Thus $x^{[k]}$ is a random draw from the distribution function given by (42). Doing this M times with replacement gives a sample of $M \leq N$ independent draws. The sampling can also be done without replacement but the values are not then independent.

A weakness of the classical approach is that it does not automatically allow for the effect on estimates of variance of estimation errors in $\hat{\psi}$. For the present problem the effect is usually $O(n^{-1})$ relative to the variance under estimate so the investigator could decide just to ignore it. If an allowance for the effect is desired, we suggest that an easy way to achieve it is to perform a Bayesian analysis as described in section 5.5 with a diffuse prior for ψ. Estimates of posterior variances in this analysis automatically contain an allowance for the effect and can be used in a classical analysis to provide estimates of conditional variance that are unbiased to $O(n^{-1})$.

5.4 Maximum likelihood estimation of parameter vector

Estimation of the parameter vector ψ by maximum likelihood using importance sampling was considered by Shephard and Pitt (1997) and in more detail by Durbin and Koopman (1997) for the case where α_t is generated by the linear Gaussian model (2). We now extend the treatment to models with non-Gaussian state errors under the assumptions made in section 2 about the density of η. Denote the likelihood for models (6) and (3) by $L(\psi)$ and the likelihood for the linear Gaussian approximating model by $L_g(\psi)$. In terms of the notation of section 2, $L(\psi) = p(y)$ and $L_g(\psi) = g(y)$, so it follows from (10) that

$$L(\psi) = L_g(\psi)\mathrm{E}_g[w(\eta)]$$

where E_g and $w(\eta)$ are defined in section 5.1. We estimate this by

$$\hat{L}(\psi) = L_g(\psi)\bar{w}, \tag{44}$$

where $\bar{w} = \frac{1}{N}\sum_{i=1}^{N} w_i$. We note that $\hat{L}(\psi)$ is obtained as an adjustment to $L_g(\psi)$; thus the closer the underlying model is to a linear Gaussian model the smaller the value of N is needed to attain preassigned accuracy. In practice we work with $\log \hat{L}(\psi)$ which has a bias of $O(N^{-1})$; if desired, a correction can be made as in Durbin and Koopman (1997, equation 16), but for most cases in practice the bias will be small enough to be neglected.

To estimate ψ, $\log \hat{L}(\psi)$ is maximised by any convenient and effective numerical technique. In order to ensure stability in the iterative process, it is important to use the same random numbers from the simulation smoother for each value of ψ. Initial parameter values for ψ are obtained by maximising the approximate loglikelihood

$$\log L(\psi) \simeq \log L_g(\psi) + \log w\left(\hat{\eta}\right); \tag{45}$$

this does not require simulation. Alternatively, the more accurate non-simulated approximation given in (21) of Durbin and Koopman (1997) may be used.

Denote the resulting maximum likelihood estimate of ψ by $\hat{\psi}$, and denote the 'true' estimate that would be obtained by maximising $\log L\left(\psi\right)$, if this could be done exactly, by $\tilde{\psi}$. We estimate the mean square error (MSE) matrix of errors due to simulation, $\mathrm{MSE}_g(\hat{\psi}) = \mathrm{E}_g\left\{(\hat{\psi}-\tilde{\psi})(\hat{\psi}-\tilde{\psi})'\right\}$ as in Durbin and Koopman (1997) by

$$\widehat{\mathrm{MSE}}_g\left(\tilde{\psi}\right) = \hat{V}\left\{\frac{1}{N}\sum_{i=1}^{N}(q^{(i)}-\bar{q})(q^{(i)}-\bar{q})'\right\}\hat{V}, \tag{46}$$

where $q^{(i)} = \partial w_i/\partial\psi$, $\bar{q} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}$ and

$$\hat{V} = \left\{-\frac{\partial^2 \log L(\hat{\psi})}{\partial\psi\partial\psi'}\right\}^{-1} \tag{47}$$

is the large-sample estimate of the variance matrix of $\hat{\psi}$. The derivatives $q^{(i)}$ and $-\hat{V}^{-1}$ are calculated numerically from neighbouring values of ψ in the neighbourhood of $\hat{\psi}$. Square roots of diagonal elements of (46) can be compared with square roots of diagonal elements of (47) to give relative standard errors due to simulation. These methods are very efficient computationally. For the examples considered in Durbin and Koopman (1997) it was shown that simulation sample sizes of N equal to around 800 based on 200 draws from the simulation smoother were sufficient for accurate estimation of ψ.

5.5 Bayesian inference

To perform a Bayesian analysis we begin by implementing formula (18) by simulation. When the prior density of ψ is diffuse the approximate density of ψ given y is $g(\psi|y) = \mathrm{N}(\hat{\psi}, \hat{V})$ where $\hat{\psi}$ and $\hat{V}$ are obtained as described in the previous section. When the prior is not diffuse there is a straightforward modification based on (15) and (16). Usually, $\hat{V}$ is $O(n^{-1})$ while $\mathrm{Var}[x(\alpha)|\psi, y]$ is $O(1)$ so it is reasonable to expect that the coefficients of variation of elements of ψ given y will be significantly smaller than those of $x(\alpha)$ given ψ and y. Let us therefore assume to begin with that antithetics are not needed in simulation from density $g(\psi|y)$ whereas they are definitely needed in simulation from density $g(\eta|\psi, y)$. Substitute for α in terms of η in (18) giving, analogously to (36),

$$\bar{x} = \frac{\mathrm{E}_g[x(\eta)z(\psi, \eta, y)]}{\mathrm{E}_g[z(\psi, \eta, y)]} \quad \text{where} \quad z(\psi, \eta, y) = \frac{p(\psi)g(y|\psi)}{g(\psi|y)} \frac{p(\eta, y|\psi)}{g(\eta, y|\psi)} \tag{48}$$

and where E_g denotes expectation with respect to density $g(\psi, \eta|y)$. Let $\psi^{(i)}$ be a random draw from $g(\psi|y)$, which is obtainable in a routine way from a sample of independent $\mathrm{N}(0, 1)$ variables, and let $\eta^{(i)}$ be a random draw from density $g(\eta|\psi^{(i)}, y)$ for $i = 1, \ldots, N$. To do this we need an approximation to the mode $\hat{\eta}^{(i)}$ of density $g(\eta|\psi^{(i)}, y)$ but this is rapidly obtained in a few iterations starting from the mode of $g(\eta|\hat{\psi}, y)$. Let

$$z_i = \frac{p(\psi^{(i)})g(y|\psi^{(i)})}{g(\psi^{(i)}|y)} \frac{p(\eta^{(i)}, y|\psi^{(i)})}{g(\eta^{(i)}, y|\psi^{(i)})} = \frac{p(\psi^{(i)})g(y|\psi^{(i)})}{g(\psi^{(i)}|y)} w_i, \tag{49}$$

and estimate $\bar{x}$ in (18) by

$$\hat{x} = \frac{\sum_{i=1}^N x_i z_i}{\sum_{i=1}^N z_i}. \tag{50}$$

Estimates of the posterior distribution function and density of $x(\eta)$ can be obtained in the same way as for the conditional distribution function and density in section 5.2. Similarly, the posterior variance and simulation variance are obtained from formulae that are analogous to (39) and (41) except that w_i are replaced by z_i. Formula (50) for $\hat{x}$ has been written on the assumption that no antithetics are used for the draws from density $g(\psi|y)$; however, the formula is easily extended to the case where antithetics are used.

We now consider the estimation of the posterior density of a single element of ψ, which we can take to be ψ_1, the first element of ψ. Denote the vector of the remaining elements of ψ by ψ_2, giving $\psi = (\psi_1, \psi_2')'$. Let $g(\psi_2|\psi_1, y)$ be the conditional density of ψ_2 given ψ_1, which is easily obtained by applying standard regression theory to $g(\psi|y)$. We shall use $g(\psi_2|\psi_1, y)$ as an importance density in place of $g(\psi|y)$. Then

$$p(\psi_1|y) = \int p(\psi|y)d\psi_2 = \int \frac{p(\psi|y)}{g(\psi_2|\psi_1, y)} g(\psi_2|\psi_1, y)d\psi_2 \tag{51}$$

By the methods of sections 3.2, 3.3, 5.4 and 5.5 we have

$$p(\psi|y) = Kp(\psi)p(y|\psi) = Kp(\psi)g(y|\psi) \int \frac{p(\eta, y|\psi)}{g(\eta, y|\psi)} g(\eta|\psi, y)d\eta. \tag{52}$$

Putting (51) and (52) together gives

$$p(\psi_1|y) = K\mathrm{E}_g \left[\frac{p(\psi)g(y|\psi)}{g(\psi_2|\psi_1, y)} \frac{p(\eta, y|\psi)}{g(\eta, y|\psi)} \right], \tag{53}$$

where E_g denotes expectation with respect to the joint importance density $g(\psi_2, \eta|\psi_1, y)$. Let $\psi_2^{(i)}$ be a draw from density $g(\psi_2|\psi_1, y)$, take $\psi^{(i)} = (\psi_1, \psi_2^{(i)\prime})'$, let $\eta^{(i)}$ be a draw from $g(\eta|\psi^{(i)}, y)$ and let

$$z_i^* = \frac{p(\psi^{(i)})g(y|\psi^{(i)})}{g(\psi_2^{(i)}|\psi_1, y)} \frac{p(\eta^{(i)}, y|\psi^{(i)})}{g(\eta^{(i)}, y|\psi^{(i)})}, \qquad i = 1, \ldots, N. \tag{54}$$

Noting that the form of z_i^* differs from the form of z_i in (49) only in the substitution of $g(\psi_2^{(i)}|\psi_1, y)$ for $g(\psi^{(i)}|y)$ and that $K^{-1} = \mathrm{E}_g[z(\psi, \eta, y)]$ as is easily shown, the required estimate of $p(\psi_1, y)$ has the simple and elegant form

$$\hat{p}(\psi_1|y) = \frac{\sum_{i=1}^N z_i^*}{\sum_{i=1}^N z_i}. \tag{55}$$

In the implementation of (55) it is important that the draw of $\psi_2^{(i)}$ from $g(\psi_2|\psi_1, y)$ is obtained directly from the draw of $\psi^{(i)}$ from $g(\psi|y)$ in (49). Details can easily be worked out from elementary regression theory but there is not the space to include them here. A simpler alternative is to calculate the posterior density of ψ_1 while ψ_2 held fixed at its maximum likelihood estimate.

6 Real data illustrations

In this section we discuss the use of the methodology by applying it to three real data sets. The calculations are carried out using the object oriented matrix programming language *Ox 2.0* of Doornik (1998) together with the library of state space functions *SsfPack 2.2* by Koopman, Shephard and Doornik (1998). The data and programs are freely available on the internet at `http://center.kub.nl/stamp/ssfpack`. We will refer to specific *Ox* programs and *SsfPack* functions in the discussion below where appropriate. Documentation of the functions used here and a discussion of computational matters can be found on the internet workpage of *SsfPack*. Because of limitations of space we cannot present in this paper a complete analysis of each of the three examples; instead, we focus on items of particular interest in each of the three cases in such a way that we cover collectively the main features of the output that can be obtained using our approach.

6.1 Van drivers killed in UK: a Poisson application

The data are monthly numbers of light goods vehicle (van) drivers killed in road accidents from 1969 to 1984 in Great Britain. These data led directly to the work presented in this paper. They were part of the data set that Durbin and Harvey (1985) analysed on behalf of the Department of Transport to provide an independent assessment of the effects of the British seat belt law on road casualties. Durbin and Harvey analysed all the data except these van data by an approximating linear Gaussian state space model. However they used an ad-hoc method to analyse the van data because they thought that the number of deaths were too small to justify the use of the linear Gaussian model. The *Ox* program `dkrss_van.ox` is used for calculating all the reported results below.

We model the data by the Poisson density with mean $\exp(\theta_t)$,

$$p(y_t|\theta_t) = \exp\{\theta_t' y_t - \exp(\theta_t) - \log y_t!\}, \qquad t = 1, \ldots, n. \tag{56}$$

with signal θ_t generated by

$$\theta_t = \mu_t + \gamma_t + \lambda x_t,$$

where the trend μ_t is the random walk

$$\mu_t = \mu_{t-1} + \eta_t, \tag{57}$$

λ is the intervention parameter which measures the effects of the seat belt law, x_t is an indicator variable for the post legislation period and the monthly seasonal γ_t is generated by

$$\sum_{j=0}^{11} \gamma_{t-j} = \omega_t, \tag{58}$$

The disturbances η_t and ω_t are mutually independent Gaussian white noise terms with variances $\sigma_\eta^2 = \exp(\psi_\eta)$ and $\sigma_\omega^2 = \exp(\psi_\omega)$, respectively. The parameter estimates are reported by Durbin and Koopman (1997) as $\widehat{\sigma}_\eta = \exp(\hat{\psi}_\eta) = \exp(-3.708) = 0.0245$ and $\widehat{\sigma}_\omega = 0$. The fact that $\widehat{\sigma}_\omega = 0$ implies that the seasonal is constant over time.

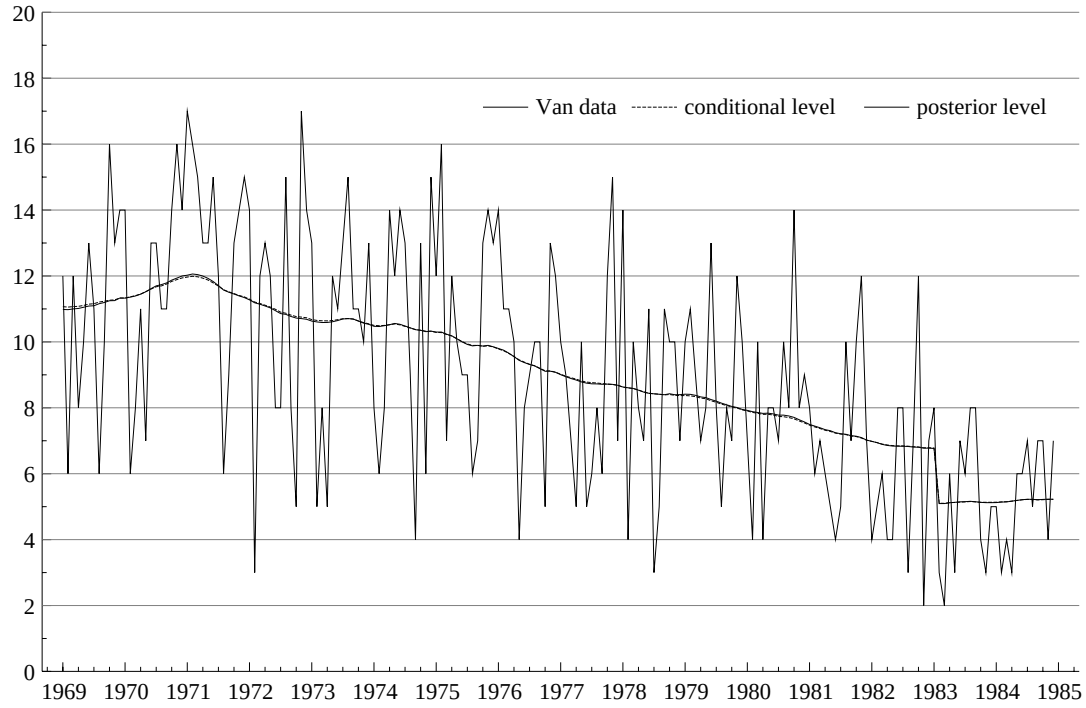

Figure 1: Van data and estimated level including intervention

It follows that $b_t(\theta_t) = \exp(\theta_t)$ in (29), so $\dot{b}_t = \ddot{b}_t = \exp(\tilde{\theta}_t)$ and from section 4.3, $\tilde{H}_t = \exp(-\tilde{\theta}_t)$ and $\tilde{y}_t = \tilde{\theta}_t + \tilde{H}_t y_t - 1$ where $\tilde{\theta}_t$ is some trial value for θ_t $(t = 1, \ldots, n)$. The iterative process of determining the approximating model as described in section 4.2 converges quickly; usually, between three and five iterations are needed. The conditional mean of $\mu_t + \lambda x_t$ for ψ_η fixed at $\hat{\psi}_\eta$ is computed from a classical perspective and exponentiated values of this mean are plotted together with the raw data in Figure 1. The posterior mean from a Bayesian perspective with ψ_η diffuse was also calculated and its exponentiated values are also plotted in Figure 1. The difference between the graphs is almost imperceptible. Conditional and posterior standard deviations of $\mu_t + \lambda x_t$ are plotted in Figure 2. The posterior standard deviations are about 12% larger than the conditional standard deviations; this is due to the fact that in the Bayesian analysis ψ_η is random. The ratios of simulation standard deviations to standard deviations proper never exceeded the 9% level before the break and never exceed the 7% level after the break. The ratios for a Bayesian analysis increases slightly obtaining 10% and 8%, respectively.

In a real analysis, the main objective is the estimation of the effect of the seat-belt law on the number of deaths. Here, this is measured by λ which in the Bayesian analysis has a posterior

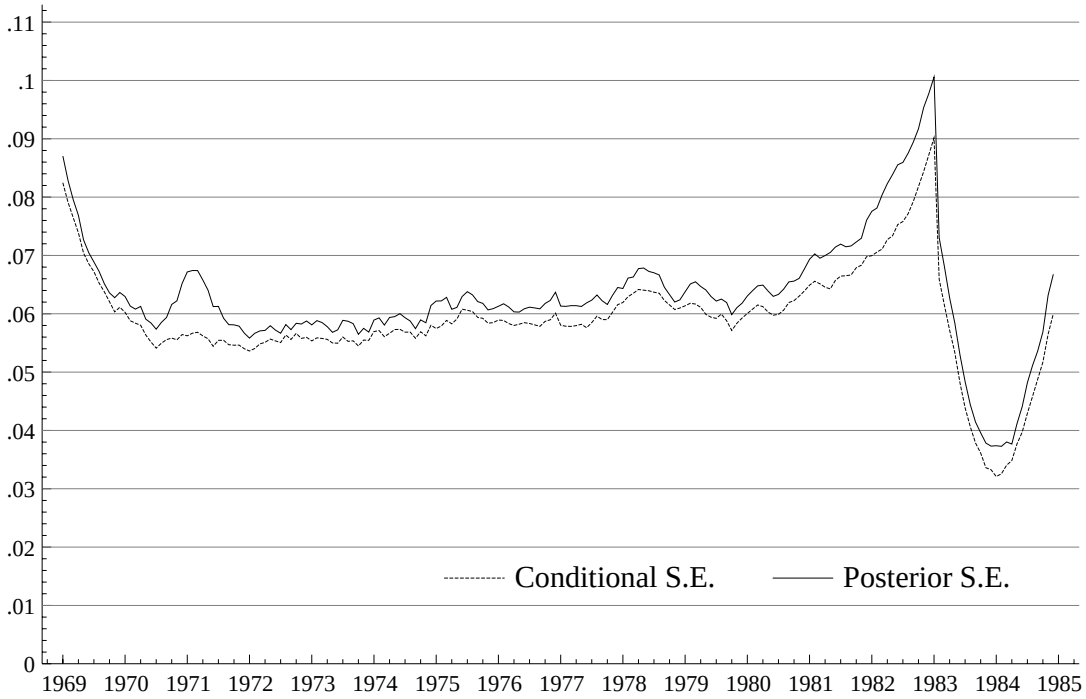

Figure 2: Standard errors for level including intervention

mean of $-.280$; this corresponds to a reduction in the number of deaths of 24.4%. The posterior standard deviation is .126 and the standard error due to simulation is .0040. The corresponding values for the classical analysis are $-.278$, .0114 and .0036, which are not very different. It is clear that the value of λ is significant as is obvious visually from Figure 1. The posterior distribution of λ is presented in Figure 3 in the form of a histogram. This is based on the estimate of the posterior distribution function calculated as indicated in section 5.5. There is a strange dip near the maximum which remains for different simulation sample sizes so we infer that it must be determined by the observations and not the simulation. All the above calculations were based on a sample of 250 draws from the simulation smoother with four antithetics per draw. The reported results show that this relatively small number of samples is adequate for this particular example.

What we learn from this exercise so far as the underlying real investigation is concerned is that up to the point where the law was introduced there was a slow regular decline in the number of deaths coupled with a constant multiplicative seasonal pattern, while at that point there was an abrupt drop in the trend of around 25%; afterwards, the trend appeared to flatten out, with the seasonal pattern remaining the same. From a methodological point of view we learn that our simulation and estimation procedures work straightforwardly and efficiently. We find that the results of the conditional analysis from a classical perspective and the posterior analysis from a Bayesian perspective are very similar apart from the densities from the posterior densities of the parameters. So far as computing time is concerned, we cannot present a comprehensive study in this paper because of the pressure of space, but to illustrate with one example, the calculation of trend and variance of trend for $t = 1, \ldots, n$ took 78 seconds on a Pentium II computer for the classical analysis and 216 seconds for the Bayesian analysis. While the Bayesian time is greater, the time required is not large by normal standards.

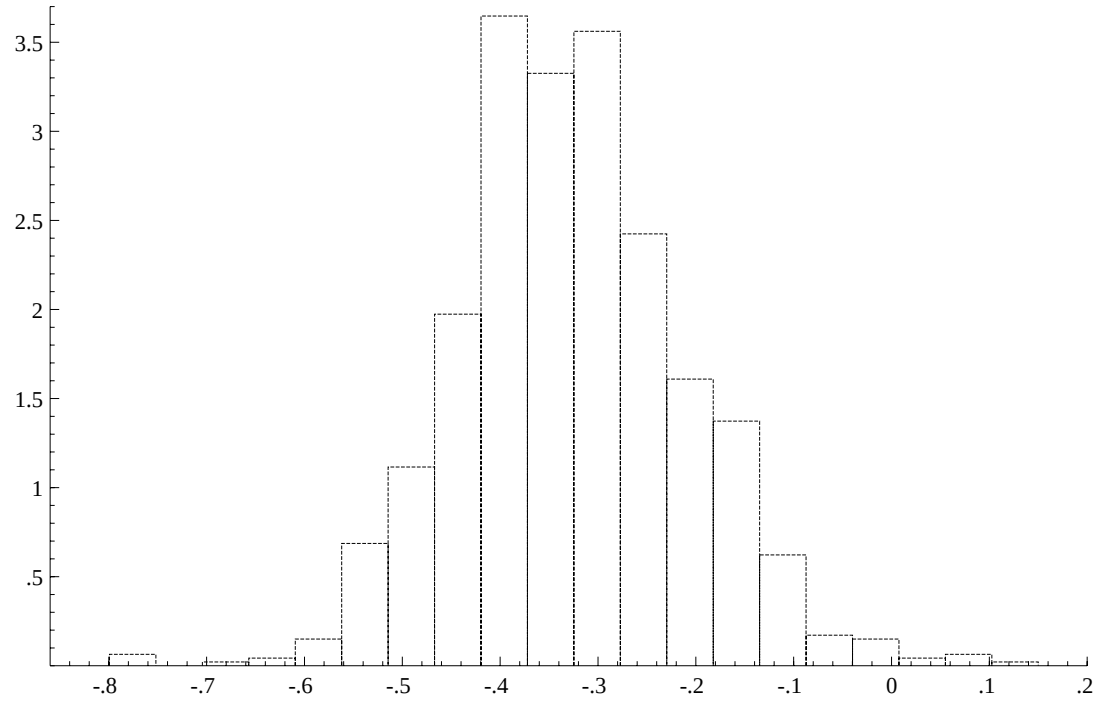

Figure 3: Posterior distribution of intervention effect

6.2 Gas consumption in UK: a heavy-tailed application

In this example we analyse the logged quarterly demand for gas in the UK from 1960 to1986. We use a structural time series model of the basic form

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \tag{59}$$

where μ_t is the trend, γ_t is the seasonal and ε_t is the observation disturbance. Further details of the model are discussed by Harvey (1989, p.172). The purpose of the real investigation underlying the analysis is to study the seasonal pattern in the data with a view to seasonally adjusting the series. It is known that for most of the series the seasonal component changes smoothly over time, but it is also known that there was a disruption in the gas supply in the third and fourth quarters of 1970 which has led to a distortion in the seasonal pattern when a standard analysis based on a Gaussian density for ε_t is employed. The question under investigation is whether the use of a heavy-tailed density for ε_t would improve the estimation of the seasonal in 1970.

To model ε_t we use the t-distribution with log density

$$\log p\left(\varepsilon_t\right) = \text{constant} + \log a\left(\nu\right) + \frac{1}{2}\log k_t - \frac{\nu+1}{2}\log\left(1 + k_t\varepsilon_t^2\right), \tag{60}$$

where

$$a\left(\nu\right) = \frac{, \left(\frac{\nu}{2} + \frac{1}{2}\right)}{, \left(\frac{\nu}{2}\right)}, \qquad k_t^{-1} = \left(\nu - 2\right)\sigma_\varepsilon^2, \qquad \nu > 2, \qquad t = 1, \ldots, n.$$

The mean of ε_t is zero and the variance is σ_ε^2 for any ν degrees of freedom which need not be

an integer. The approximating model is easily obtained by method 2 of section 4.4 with

$$h_t^*\left(\varepsilon_t^2\right) = \ldots + (\nu+1)\log\left(1+k_t\varepsilon_t^2\right), \qquad \dot{h}_t^{*\,-1} = H_t = \frac{1}{\nu+1}\widetilde{\varepsilon}_t^2 + \frac{\nu-2}{\nu+1}\sigma_\varepsilon^2,$$

The iterative scheme is started with $H_t = \sigma_\varepsilon^2$, for $t = 1, \ldots, n$. The number of iterations required for a reasonable level of convergence using the t-distribution is usually higher than for densities from the exponential family; for this example we required around ten iterations. In the classical analysis, the parameters of the model, including the degrees of freedom ν, were estimated by Monte Carlo maximum likelihood as reported in Durbin and Koopman (1997).

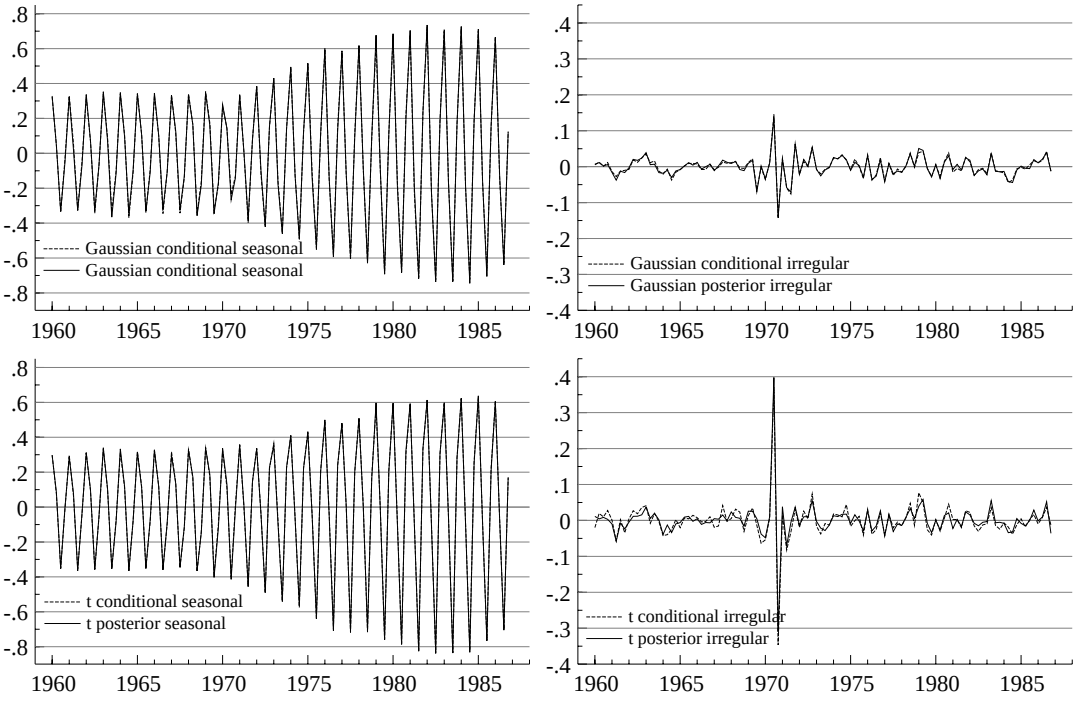

Figure 4: Gaussian and t-model analyses of Gas data

The most interesting feature of this analysis is to compare the estimated seasonal and irregular components based on the Gaussian model and the model with a t-distribution for ε_t. Figure 4 gives the graphs of the estimated seasonal and irregular for both the Gaussian model and the t-model. The most striking feature of those graphs is the greater effectiveness with which the t-model picks and corrects for the outlier relative to the Gaussian model. We observe that in the graph of the seasonal the difference between the classical and Bayesian analyses are imperceptible. Differences are visible in the graphs of the residuals, but they are not large since the residuals themselves are small. The t-model estimates are based on 250 simulation samples from the simulation smoother with four antithetic devices for each sample. The number of simulation samples is sufficient because the ratio of the variance due to simulation to the variance never exceeded 2% for all estimated components in the state vector except at the beginning and end of the series where it never exceeded 4%. The *Ox* program `dkrss_gas.ox` was used for calculating these results.

We learn from the analysis that the change over time of the seasonal pattern in the data is in fact smooth. We also learn that if model (59) is to be used to estimate the seasonal for this

or similar cases with outliers in the observations, then a Gaussian model for ε_t is inappropriate and a heavy-tailed model should be used.

6.3 Pound/Dollar daily exchange rates: a volatility application

The data are the Pound/Dollar daily exchange rates from 1/10/81 to 28/6/85 which have been used by Harvey, Ruiz and Shephard (1994). Denoting the exchange rate by x_t, the daily returns are the series of interest given by $y_t = \triangle \log x_t$, for $t = 1, \ldots, n$. A stochastic volatility (SV) model of the form

$$\begin{aligned} y_t &= \sigma \exp\left(\frac{1}{2}\theta_t\right) u_t, \qquad u_t \sim \mathrm{N}(0,1), \qquad t = 1, \ldots, n, \\ \theta_t &= \phi\theta_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \qquad 0 < \phi < 1, \end{aligned} \tag{61}$$

was used for analysing these data by Harvey, Ruiz and Shephard (1994); for a review of related work and developments of the SV model see Shephard (1996) and Ghysels, Harvey and Renault (1996). Exact treatments of the SV model are developed and they are usually based on MCMC or importance sampling methods; see Jacquier, Polson and Rossi (1994), Danielsson (1994) and Shephard and Pitt (1997). The purpose of the investigations for which this type of analysis is carried out is to study the structure of the volatility of price ratios in the market, which is of considerable interest to financial analysts. The level of θ_t determines the amount of volatility and the value of ϕ measures the autocorrelation present in the logged squared data.

To illustrate our methodology we take the same approach to SV models as Shephard and Pitt (1997) by considering the Gaussian log-density of the SV model,

$$\log p\left(y_t|\theta_t\right) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2}\theta_t - \frac{y_t^2}{2\sigma^2}\exp(-\theta_t). \tag{62}$$

The linear approximating model can be obtained by method 1 of section 4.2 with

$$\tilde{H}_t = 2\sigma^2\frac{\exp(\tilde{\theta}_t)}{y_t^2}, \qquad \tilde{y}_t = \tilde{\theta}_t - \frac{1}{2}\tilde{H}_t + 1,$$

for which $\tilde{H}_t$ is always positive. The iterative process can be started with $\tilde{H}_t = 2$ and $\tilde{y}_t = \log(y_t^2/\sigma^2)$, for $t = 1, \ldots, n$, since it follows from (61) that $y_t^2/\sigma^2 \approx \exp(\theta_t)$. When y_t is zero or very close to zero, it should be replaced by a small constant value to avoid numerical problems; this device is only needed to obtain the approximating model so we do not depart from our exact treatment. The number of iterations required is usually less than ten.

The interest here is usually focussed on the estimates of the parameters or their posterior distributions. For the classical analysis we obtain by the maximum likelihood methods of section 5.4 the following estimates:

$$\begin{array}{lllllllll} \hat{\sigma} &=& .6338, & \hat{\psi}_1 = \log\hat{\sigma} &=& -0.4561, & \mathrm{SE}(\hat{\psi}_1) &=& 0.1033, \\ \hat{\sigma}_\eta &=& .1726, & \hat{\psi}_2 = \log\hat{\sigma}_\eta &=& -1.7569, & \mathrm{SE}(\hat{\psi}_2) &=& 0.2170, \\ \hat{\phi} &=& .9731, & \hat{\psi}_3 = \log\frac{\hat{\phi}}{1-\hat{\phi}} &=& 3.5876, & \mathrm{SE}(\hat{\psi}_3) &=& 0.5007. \end{array}$$

We present the results in this form since we estimate the log transformed parameters, so the standard errors that we calculate apply to them and not to the original parameters of interest. For the Bayesian analysis we present in Figure 5 the posterior densities of each transformed parameter given the other two parameters held fixed at their maximum likelihood values. These results confirm that stochastic volatility models can be handled by our methods from both classical and Bayesian perspectives. The computations are carried out by the *Ox* program `dkrss_sv.ox`.

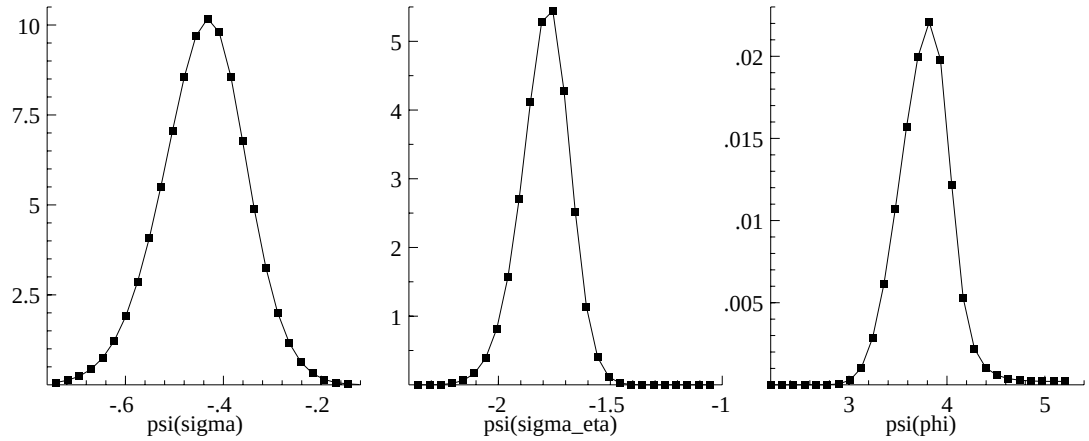

Figure 5: Posterior densities of transformed parameters

7 Discussion

We regard this paper as much more a paper in time series analysis than on simulation. A methodology is developed that can be used by applied researchers for dealing with real non-Gaussian time series data without them having to be time series specialists or enthusiasts for simulation methodology. The ideas underlying the simulation methodology are relatively easy to explain to non-specialists. Also, user-friendly software is freely available on the Internet (`http://center.kub.nl/stamp/ssfpack.htm`) in a relatively straightforward format including documentation.

Methods are developed for classical and Bayesian inference side by side using a common simulation methodology. This widens the choices available for applications. The illustrations provided in the paper show the differences that are found when both approaches are applied to real data. Generally speaking, the differences are small except for the variances of estimates for which the differences are obviously due to the fact that in classical inference the parameters are regarded as fixed whereas in Bayesian inference the parameters are regarded as random variables.

Almost all previous work on non-Gaussian time series analysis by simulation has been done using Monte Carlo Markov chain (MCMC) methodology. In contrast, our approach is based entirely on importance sampling and antithetic variables which have been available for many years but which we have shown to be very efficient for our problem. Because our approach is based on independent samples it has the following advantages relative to MCMC: first, we avoid completely the convergence problems associated with MCMC; second, we can easily compute error variances due to simulation as a routine part of the analysis; thus the investigator can attain any predetermined level of simulation accuracy by increasing the simulation sample size, where necessary, by a specific amount.

Acknowledgement

S.J. Koopman is a Research Fellow of the Royal Netherlands Academy of Arts and Sciences and its financial support is gratefully acknowledged.

References

Anderson, B.D.O. and Moore, J.B. (1979) *Optimal filtering.* Englewood Cliffs, NJ: Prentice–Hall.

Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory.* Chichester: Wiley and Sons.

Cargnoni, C., Müller, P. and West, M. (1997) Bayesian Forecasting of Multinomial Time Series Through Conditionally Gaussian Dynamic Models. *J. Am. Statist. Assoc.,* **92**, 640-647.

Carlin, B.P., Polson, N.G. and Stoffer, D.S. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Am. Statist. Ass.*, **87**, 493-500.

Carter, C.K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika,* **81**, 541-553.

__ (1996) Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika,* **83**, 589-601.

__ (1997) Semiparametric Bayesian inference for time series with mixed spectra. *J.R. Statist. Soc. B*, **59**, 255-268.

de Jong, P. and Shephard, N. (1995) The simulation smoother for time series models. *Biometrika,* **82**, 339-350.

Danielsson, J. (1994) Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *J. Econometrics,* **61**, 375-400.

Doornik, J.A. (1998) *Object-Oriented Matrix Programming using Ox 2.0.* London: Timberlake Consultants Press.

Durbin, J. and Harvey, A.C. (1985) *The Effect of Seat Belt Legislation on Road Casualties.* London: HMSO.

Durbin, J. and Koopman, S.J. (1992) Filtering, smoothing and estimation for time series models when the observations come from exponential family distributions. Unpublished: London School of Economics.

__ (1997) Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, **84**, 669-684.

__ (1999) Thicker-tailed importance densities. Unpublished: London School of Economics.

Fahrmeir, L. (1992) Conditional mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models. *J. Am. Statist. Assoc.,* **87**, 501-509.

Fahrmeir, L. and Kaufmann, H. (1991) On Kalman filtering, conditional mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika*, **38**, 37-60.

Frühwirth-Schnatter, S. (1994) Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. *Statist. and Comput.,* **4**, 259-269.

Gelfand, A.E. and Smith, A.F.M. (1999) *Bayesian Computation.* Chichester: Wiley and Sons.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) *Bayesian Data Analysis.* London: Chapman and Hall.

Geyer, C.J. and Thompson, E.A. (1992) Constrained Monte Carlo Maximum Likelihood for Dependent Data. *J.R. Statist. Soc. B*, **54**, 657-699.

Ghysels, E., Harvey, A.C. and Renault, E. (1996) Stochastic Volatility. In C.R. Rao and G.S. Maddala (eds.) *Statistical Methods in Finance.* Amsterdam: North-Holland.

Harrison, P.J. and Stevens, C.F. (1971) A Bayesian approach to short-term forecasting. *Operational Research Quarterly*, **22**, 341-362.

__ (1976) Bayesian forecasting. *J.R. Statist. Soc. B*, **38**, 205-247.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, A.C. and Durbin, J. (1986) The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *J.R. Statist. Soc. A*, **149**, 187-227.

Harvey, A.C. and Fernandes, C. (1989) Time series models for count or qualitative observations. *J. Bus. Econ. Statist.*, **7**, 407-422.

Harvey, A.C., Ruiz, E. and Shephard, N. (1994) Multivariate stochastic variance models. *Rev. Econ. Studies*, **61**, 247-264.

Jacquier, E., Polson, N.G. and Rossi. P.E. (1994) Bayesian Analysis of Stochastic Volatility Models. *J. Bus. Econ. Statist.*, **12**, 371-417.

Kalman, R.E. (1960) A new approach to linear prediction and filtering problems. *J. Basic Engineering, Trans ASME, Series D*, **82**, 35-45.

Kitagawa, G. (1987) Non–Gaussian state–space modelling of nonstationary time series (with discussion). *J.Am.Statist.Assoc.*, **82**, 1032-1063.

__ (1989) Non-Gaussian seasonal adjustment. *Computers and Mathematics with Applications*, **18**, 503-514.

__ (1990) The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. Technical report, Institute of Statistical Mathematics, Tokyo.

Koopman, S.J. (1993) Disturbance smoother for state space models. *Biometrika*, **80**, 117-126.

Koopman, S.J., Shephard, N. and Doornik (1998) Statistical algorithms for models in state space using SsfPack 2.2. Unpublished: CentER discussion paper.

Martin and Raftery (1987) Discussion of Non–Gaussian state–space modelling of nonstationary time series by G. Kitagawa. *J.Am.Statist.Assoc.*, **82**, 1032-1063.

Ripley, B.D. (1987) *Stochastic Simulation.* New York: Wiley.

Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika*, **81**, 115-131.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R. Cox and D.V. Hinkley *Time Series Models in Econometrics, Finance and Other Fields.* London: Chapman and Hall.

Shephard, N. and Pitt, M.K. (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653-667.

Smith, J.Q. (1979) A generalization of the Bayesian steady forecasting model. *J.R. Statist. Soc. B*, **41**, 375-387.

__ (1981) The multiparameter steady model. *J.R. Statist. Soc. B*, **43**, 256-260.

West, M., Harrison, P.J. and Migon, H.S. (1985) Dynamic generalized linear models and Bayesian forecasting (with discussion). *J.Am.Statist.Assoc.*, **80**, 73-97.