Partial Non-Gaussian State Space
Author(s): Neil Shephard
Reviewed work(s):
Source: *Biometrika,* Vol. 81, No. 1 (Mar., 1994), pp. 115-131
Published by: Biometrika Trust
Stable URL: http://www.jstor.org/stable/2337055
Accessed: 21/04/2012 15:44

Partial non-Gaussian state space

By NEIL SHEPHARD

Nuffield College, Oxford OX1 1NF, U.K.

Summary

In this paper we suggest the use of simulation techniques to extend the applicability of the usual Gaussian state space filtering and smoothing techniques to a class of non-Gaussian time series models. This allows a fully Bayesian or maximum likelihood analysis of some interesting models, including outlier models, discrete Markov chain components, multiplicative models and stochastic variance models. Finally we discuss at some length the use of a non-Gaussian model to seasonally adjust the published money supply figures.

Some key words: EM algorithm; Filtering; Metropolis algorithm; Seasonal adjustment; Smoothing; Unobserved components.

1. Motivation

The Gaussian state space form

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t),$$

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

$$\alpha_0 | Y_0 \sim N(a_{0|0}, P_{0|0}) \quad (t = 1, \ldots, n),$$

where $\alpha_0 | Y_0$, ε_t and η_s are independent for all t and s and Y_t denotes the information available at time t, has had an important role in time series in recent years (West & Harrison, 1989). The columns of G_t will be subsets of the columns of I and are present to ensure that Q_t can be assumed to be positive definite even in cases where there are identities in the transition equation. We will also always assume that H_t is positive definite.

Any small modification to the above form, such as removing Gaussianity from one element of η_t, means the Kalman filter and the corresponding smoothing algorithms lose their optimality properties, although occasionally some of the best linear qualities still hold. This leaves the modeller with the difficulty of working with approximate filters, which in turn tend to yield inconsistent parameter estimators.

In this paper we suggest ways of modifying the usual Gaussian filtering and smoothing algorithms to produce exact results. We progress by example, only suggesting a new framework, called a partial non-Gaussian state space, towards the end of the paper. The examples we discuss include outlier models, multiplicative models, discrete Markov chain components, and stochastic variance models. Finally we explore at some length the use of a non-Gaussian model to seasonally adjust the published money supply figures.

2. Outlier models

2·1. *Student's t distribution*

We focus on the local level model discussed by Harvey (1989, Ch. 2) and West & Harrison (1989, Ch. 2), although similar types of results hold for more general unobserved

component models. It takes on the form

$$y_t = \alpha_t + \varepsilon_t, \quad \alpha_t = \alpha_{t-1} + \eta_t, \quad H_t = \sigma_\varepsilon^2, \quad Q_t = \sigma_\eta^2. \tag{2}$$

Suppose we wished to modify this model to allow for outliers. There are two main ways in which this could be achieved: by using a Student's t distribution or a mixture distribution. The techniques we use to tackle these two models will be similar. In this section we deal with the former, in the next the latter.

In the Student's t case, we take $\zeta_t = \varepsilon_t/\omega_{1t}^{\frac{1}{2}}$ and $\theta_t = \eta_t/\omega_{2t}^{\frac{1}{2}}$ instead of ε_t and η_t respectively. Here $p_1\omega_{1t}$ and $p_2\omega_{2t}$ are independent $\chi_{p_1}^2$ and $\chi_{p_2}^2$ random variables respectively. As ζ_t and θ_s are uncorrelated for all t and s, the Kalman filter and smoother will still be best linear estimators; however, they will be poor if p_1 or p_2 are small.

Initially we assume all the parameters in our models are known. We focus on estimating $\alpha = (\alpha_1, \ldots, \alpha_n)'$ and $\omega = (\omega_{11}, \omega_{21}, \ldots, \omega_{1n}, \omega_{2n})'$, by reporting the location and scale of $f\{(\alpha', \omega')' | Y_n\} = f(\delta | Y_n)$ say.

General solutions to the problem of non-Gaussian state space filtering and smoothing have been put forward by a number of authors. Kitagawa (1987) suggests sequential approximation using numerical integration rules. This is only feasible if the dimension of the state is very small, 1, 2 or perhaps 3. To cover more complicated situations Kitagawa (1989) has suggested using an approximate Gaussian sum filter (Alspach & Sorenson, 1972; Anderson & Moore, 1979).

More recently there has been a very active interest in methods which allow the production of simulations from densities like $\delta | Y_n$: see Carlin, Polson & Stoffer (1992), Fruhwirth-Schnatter (1994), Shephard (1993) and unpublished papers by M. West and P. Müller. Most work is within the Metropolis–Hastings framework, more particularly the single move Gibbs sampler. Ripley (1987, p. 113) and Smith & Roberts (1993) elucidate the ideas behind these techniques.

Unfortunately if $\delta | Y_n$ has elements which are highly correlated then the desired asymptotic distribution of the sampler will only be approached very slowly. To illustrate this point it is useful to work with the Gaussian version of (2). Figure 1 displays the results of a Gibbs sampler for some simulated data with $n = 23$, $\sigma_\varepsilon^2 = 1$ and $\sigma_\eta^2 = 5 \times 10^{-5}$, using 1000 parallel samplers all initialised with $\alpha_t = y_t/2$, recording values after 5, 10, 50, 250, 500 and 1000 iterations. The results show how slowly the Gibbs sampler converges; indeed it is possible to show that the sampler slows as σ_η^2 falls, eventually ceasing to converge when $\sigma_\eta^2 = 0$. This should not be surprising for the nature of time series modelling is that some of the smoothed estimates of the components, especially estimates of the trend, will be highly correlated through time.

In this paper we suggest that in partial non-Gaussian models it is more useful, as well as simpler, to make many moves simultaneously. The model's block structure, of high correlation within the elements of $\alpha | Y_n$ and $\omega | Y_n$ but low correlation between these two vectors, suggests the following multimove Gibbs sampler, which should be iterated until convergence.

Multimove sampler
1. Generate a candidate point $\delta = (\alpha', \omega')'$.
2. Draw α^* from the density $f(\alpha | \omega, Y_n)$.
3. Draw ω^* from the density $f(\omega | \alpha^*, Y_n)$.
4. Write $\delta = (\alpha^{*'}, \omega^{*'})'$ and go to 2.

Convergence will be guaranteed if the joint density of the $\omega_1, \ldots, \omega_n$ is everywhere
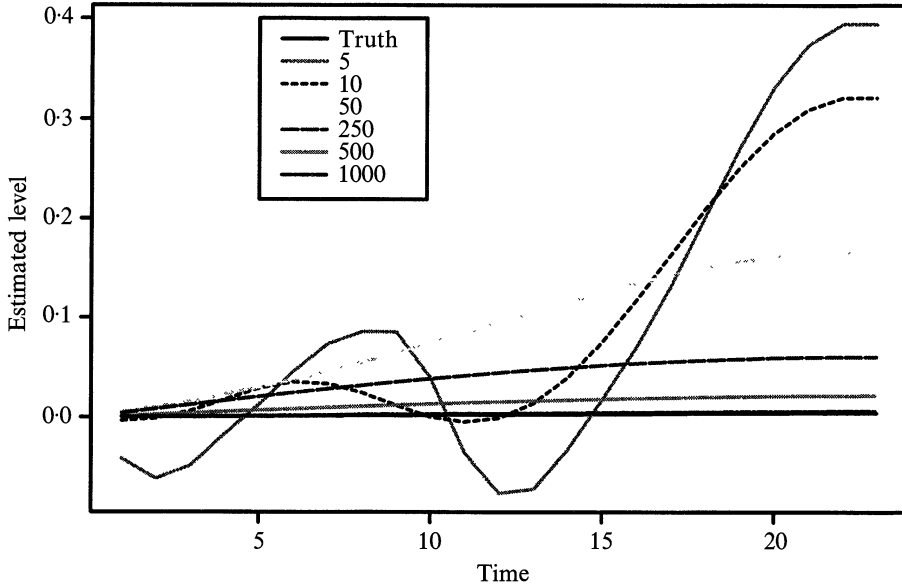
Fig. 1. Gibbs estimate, after displayed iterations; 'truth' line almost coincides with line
for 1000 iterations.

positive (Tierney, 1994). To initialise this sampler, it is only necessary to deal with ω, which can often be selected from its unconditional joint distribution, $f(\omega)$.

For the outlier model, $f(\alpha|\omega, Y_n)$ and $f(\omega|\alpha^*, Y_n)$ take on particularly attractive forms. We exploit a result which will recur throughout the paper:

$$f(\alpha|\omega, Y_n) = f(\alpha_n|\omega, Y_n) \prod_{t=1}^{n-1} f(\alpha_t|\alpha_{t+1}, \alpha_{t+2}, \ldots, \alpha_n, \omega, Y_n)$$

$$= f(\alpha_n|\omega, Y_n) \prod_{t=1}^{n-1} f(\alpha_t|\alpha_{t+1}, \omega, Y_t). \qquad (3)$$

This is a slight modification of results given recently by Fruhwirth-Schnatter (1994) and Carter & Kohn (1994). In our context this decomposition of the joint density is useful because each expression on the right-hand side involves a Gaussian density, as we have conditioned out all the nonnormality, or in a different context, the nonlinearity, in the system. It turns out that we can solve out each Gaussian density and so, by simulating from each conditional density, we can simulate efficiently from $f(\alpha|\omega, Y_n)$.

We initially focus on implementing this procedure for the simple outlier model. First, given ω, the model becomes

$$y_t = \alpha_t + \zeta_t, \quad \zeta_t|\omega_{1t} \sim N(0, \sigma_\varepsilon^2/\omega_{1t}),$$
$$\alpha_t = \alpha_{t-1} + \theta_t, \quad \theta_t|\omega_{2t} \sim N(0, \sigma_\eta^2/\omega_{2t}). \qquad (4)$$

The last state α_n^* can be simulated from the density of the conventional Kalman filter $\alpha_n|\omega, Y_n \sim N(a_{n|n}, p_{n|n})$, while

$$\alpha_t|\alpha_{t+1}^*, \omega, Y_t \sim N\left\{a_{t|t} + \frac{p_{t,t+1|t}(\alpha_{t+1}^* - a_{t+1|t})}{p_{t+1|t}}, p_{t|t} - \frac{p_{t,t+1|t}^2}{p_{t+1|t}}\right\}, \qquad (5)$$

suggesting that the rest of the α_t's can be simulated by progressing backwards, simulating

in turn α_{n-1}^*, α_{n-2}^*, ... This is easy to program, involving a run of the Kalman filter, which in general takes on the form

$$a_{t+1|t} = T_{t+1}a_{t|t-1} + K_t v_t, \quad P_{t+1|t} = T_{t+1}P_{t|t-1}L_t' + G_t Q_{t+1}G_t',$$

$$v_t = y_t - Z_t a_{t|t-1}, \quad F_t = Z_t P_{t|t-1}Z_t' + H_t, \quad K_t = T_{t+1}P_{t|t-1}Z_t'F_t^{-1}, \quad L_t = T_{t+1} - K_t Z_t,$$

$$P_{t,t+1|t} = P_{t|t}T_t', \quad a_{t|t} = a_{t|t-1} + P_{t|t-1}Z_t'F_t^{-1}v_t,$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}Z_t'F_t^{-1}Z_t P_{t|t-1}' \quad (t = 1, \ldots, n), \tag{6}$$

where

$$a_{t|s} = E(\alpha_t | \omega, Y_s), \quad P_{t|s} = E\{(\alpha_t - a_{t|s})(\alpha_t - a_{t|s})' | \omega, Y_s\},$$

$$P_{t,t+1|t} = E\{(\alpha_t - a_{t|t})(\alpha_{t+1} - a_{t+1|t})' | \omega, Y_t\}.$$

The remaining problem is to simulate from

$$f(\omega | \alpha^*, Y_n) = \prod_{t=1}^{n} f(\omega_{1t} | y_t, \alpha_t^*) f(\omega_{2t} | \alpha_t^*, \alpha_{t-1}^*). \tag{7}$$

Since $p_1 \omega_{1t} \sim \chi_{p_1}^2$,

$$\left\{ \frac{(y_t - \alpha_t^*)^2}{\sigma_\varepsilon^2} + p_1 \right\} \omega_{1t} | y_t, \alpha_t^* \sim \chi_{p_1+1}^2,$$

$$\left\{ \frac{(\alpha_t^* - \alpha_{t-1}^*)^2}{\sigma_\eta^2} + p_2 \right\} \omega_{2t} | \alpha_t^*, \alpha_{t-1}^* \sim \chi_{p_2+1}^2. \tag{8}$$

Hence the simulations from $f(\omega | \alpha^*, Y_n)$ simply involves the drawing of $2n$ independent χ^2 variables, while the independence of $\omega_{1t} | y_t, \alpha_t^*$ and $\omega_{2t} | \alpha_t^*, \alpha_{t-1}^*$ suggests that this set-up will yield a rapidly converging sampler.

To illustrate the multimove sampler, a series of length 200 was simulated using Cauchy measurement errors and Gaussian transition errors, with $\sigma_\varepsilon^2 = 1$ and $\sigma_\eta^2 = 0\cdot01$. This is displayed in Fig. 2(a). Both Gibbs and multimove samplers were used on these data, running 250 parallel samplers and recording their state estimates after 5, 10, 50, 250, 500, 1000 iterations. The samplers were initialised using the unconditional distribution of the ω and $\alpha_t = y_t/2$. The results are given in Figs 2(b) and (c). After 5 iterations the multimove sampler is almost exact. This rapid convergence compares favourably with the Gibbs sampler, which is slowed by the high correlation amongst the $\alpha | Y_n$.

2·2. Gaussian mixture distributions

The same type of argument can be used to deal with a time series model whose noise is drawn from a Gaussian mixture; Harrison & Stevens (1976) call such a structure a multiprocess model. There is a huge literature on this model (Peña & Guttman, 1988; Kitagawa, 1989). Most existing solutions to this problem are inexact or involve great computational expense.

Consider the local level model, which will be set up in the following way:

$$y_t = \mu_t + \varepsilon_t(\omega_{1t}), \quad \varepsilon_t(\omega_{1t}) | \omega_{1t} \sim N(0, \sigma_{\varepsilon t}^2),$$

$$\mu_t = \mu_{t-1} + \eta_t(\omega_{2t}), \quad \eta_t(\omega_{2t}) | \omega_{2t} \sim N(0, \sigma_{\eta t}^2), \tag{9}$$

where ω_{1t} and ω_{2t} are independent Bernoulli trials, for all values of t, indexed by probability

(a) Observations

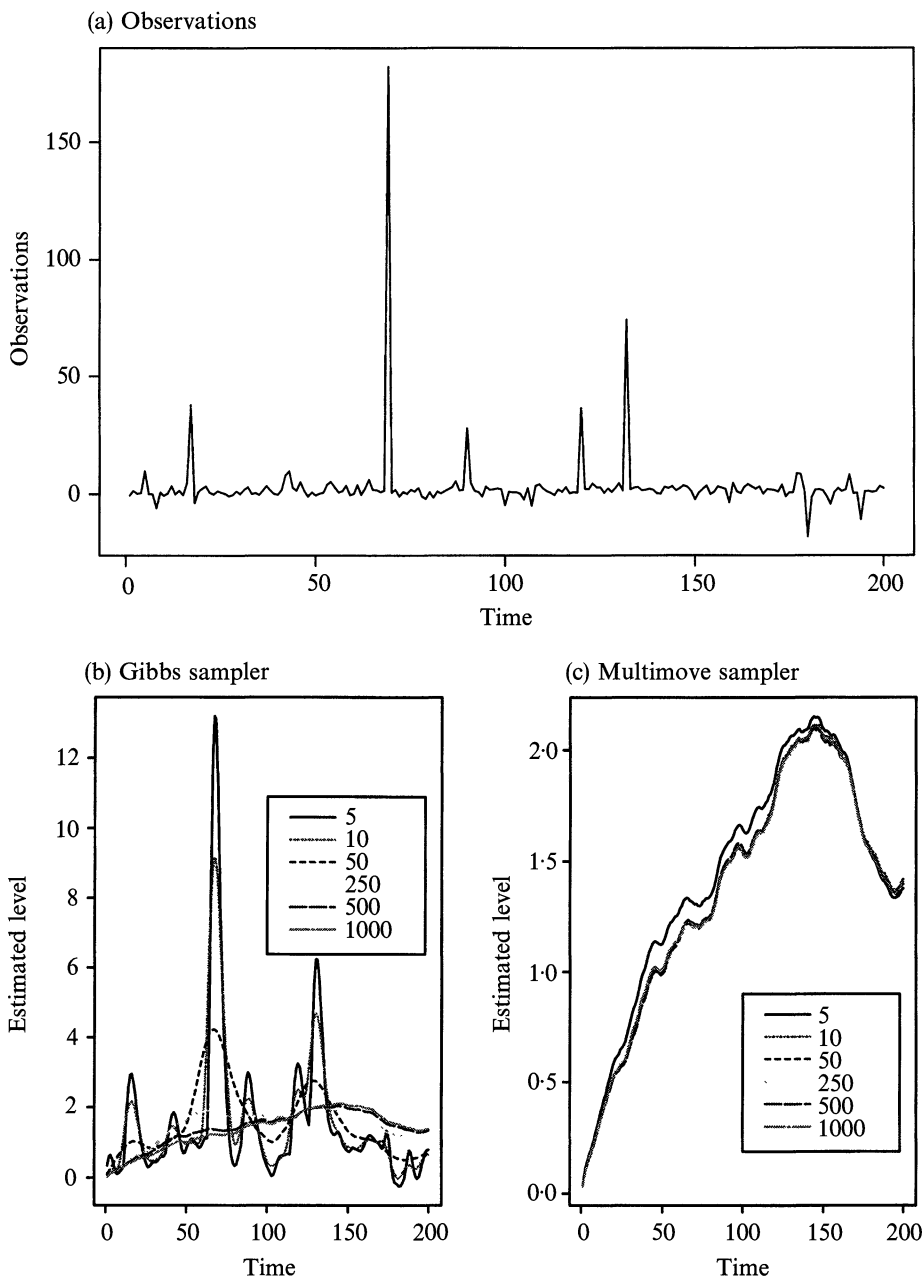(b) Gibbs sampler

(c) Multimove sampler

Fig. 2. (a) Simulated series with Cauchy measurement error. (b) Gibbs estimate, after displayed iterations. (c) Multimove estimate, after displayed iterations.

of success θ_1 and θ_2, respectively. The variances in the conditional distributions are given by

$$\sigma_{\varepsilon t}^2 = \begin{cases} \sigma_{\varepsilon_1}^2 & \text{if } \omega_{1t} = 0, \\ \sigma_{\varepsilon_2}^2 & \text{if } \omega_{1t} = 1, \end{cases} \qquad \sigma_{\eta t}^2 = \begin{cases} \sigma_{\eta_1}^2 & \text{if } \omega_{2t} = 0, \\ \sigma_{\eta_2}^2 & \text{if } \omega_{2t} = 1. \end{cases}$$

The statistical analysis of this model follows precisely the same lines as for the Student's

t model, for we can use (3) to simulate from $\alpha | \omega, Y_n$, while

$$f(\omega | \alpha^*, Y_n) = \prod_{t=1}^{n} f(\omega_{1t} | \alpha_t^*, y_t) f(\omega_{2t} | \alpha_t^*, \alpha_{t-1}^*). \tag{10}$$

As

$$f(\omega_{1t} | \alpha_t^*, y_t) \propto f(y_t | \omega_{1t}, \alpha_t^*) f(\omega_{1t}),$$

and the constant of proportionality can be evaluated as ω_{1t} has discrete support, we can simulate from (10) by drawing $2n$ uniform random variables.

3. MULTIPLICATIVE MODELS

Suppose that $\omega_t | \omega_{t-1}$ is Gaussian and not influenced by α and that the observations and states are linear in ω. A simple case of this is

$$y_t = \omega_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

$$\alpha_t = \alpha_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \tag{11}$$

$$\omega_t = \rho \omega_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2),$$

which might be called a multiplicative state space model. The partially linear structure to this model means the above argument for the Gaussian states can be used again. First, if we condition on the ω then the model is a Gaussian state space and so we can simulate α^*. Then, conditioning on the α^*, the observations and the transition equation for ω make up a Gaussian state space and so we can simulate using

$$f(\omega | \alpha^*, Y_n) = f(\omega_n | \alpha^*, Y_n) \prod_{t=1}^{n-1} f(\omega_t | \omega_{t+1}, \alpha^*, Y_t). \tag{12}$$

Going to and from one group of states to the next yields a valid sampler so long as $\rho < 1$. If ρ is positive then $\alpha | Y_n$ and $\omega | Y_n$ will be quite correlated and so this sampler may take some time to converge. However, most of the correlation in $\delta | Y_n$ has been conditioned out and so it will be much faster than the equivalent single-step Gibbs sampler.

4. MARKOV SWITCHING MODEL

Switching regime models have generated considerable attention in the statistics and econometrics literature. Recently Hamilton (1989) has written about using auto-regressions and a Markov switching regime to model the US business cycle. The use of autoregression allows the computation of the likelihood in a relatively straightforward manner. When the autoregression is replaced by ARMA or unobserved components, this approach no longer works.

To focus ideas we will again exploit a simple model with

$$y_t = \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \quad \alpha_t = \alpha_{t-1} + \gamma + \beta \omega_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2). \tag{13}$$

Here ε_t and η_s are independent from the ω_u process for all t, s and u. ω_t is assumed to be a binary Markov chain, with transition probabilities

$$\begin{array}{cc} & \begin{array}{cc} \omega_t = 0 & \omega_t = 1 \end{array} \\ \begin{array}{c} \omega_{t-1} = 0 \\ \omega_{t-1} = 1 \end{array} & \begin{pmatrix} 1 - \theta_1 & \theta_1 \\ \theta_2 & 1 - \theta_2 \end{pmatrix}. \end{array} \tag{14}$$

The motivation for this model is that the slope of the trend is either γ or $\gamma + \beta$, which in the US GDP case corresponds to a recession or boom. The slope state, ω_t, shifts with probabilities controlled by θ_1 and θ_2.

The simulation from $\alpha \mid \omega$, Y_n can be carried out using (3). Using (12), the drawing of random variables from $\omega \mid \alpha^*$, Y_n is also straightforward for the filtering equations take on the form, if we write $\alpha^t = (\alpha'_1, \ldots, \alpha'_t)'$,

$$\text{pr}\,(w_t \mid \alpha^{t-1}, Y_{t-1}) = \text{pr}\,(w_t \mid \alpha^{t-1}) = \sum_{w_{t-1}} \text{pr}\,(w_t \mid w_{t-1})\,\text{pr}\,(w_{t-1} \mid \alpha^{t-1}),$$

$$\text{pr}\,(w_t \mid \alpha^t, Y_t) = \text{pr}\,(w_t \mid \alpha^t) \propto \text{pr}\,(\alpha_t, \alpha_{t-1} \mid w_t)\,\text{pr}\,(w_t \mid \alpha^{t-1}). \tag{15}$$

This is soluble because of the discrete support of w_t, which means the summation and the constant of proportionality can be evaluated. These equations allow $\text{pr}\,(w_n \mid \alpha^*, Y_n)$ to be computed. All that is additionally required is the calculation

$$\text{pr}\,(w_t \mid w_{t+1}, \alpha^*, Y_n) \propto \text{pr}\,(w_t \mid \alpha^{*t})\,\text{pr}\,(w_{t+1} \mid w_t).$$

5. Stochastic variance models

The stochastic variance model is given by

$$y_t = \varepsilon_t \exp\,(\alpha_t/2), \quad \varepsilon_t \sim NID(0, 1),$$

$$\alpha_t = \phi \alpha_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2). \tag{16}$$

This model is a discrete time approximation to a model which is highly motivated in the finance literature. It is used to generalise the Black–Scholes option pricing equation to allow the variance to change over time (Hull & White, 1987; Chesney & Scott, 1989; Melino & Turnbull, 1990).

The transformation

$$\log y_t^2 = \alpha_t + \log \varepsilon_t^2,$$

gives a linear but non-Gaussian state space. Harvey, Ruiz & Shephard (1994) employed a Kalman filter to estimate the unobservable states and a quasi-likelihood to perform parameter estimation. This is suboptimal as $\log \varepsilon_t^2$ is far away from its normal approximation $N(-1{\cdot}2704, 4{\cdot}9348)$, which in turn means the implied quasi-likelihood estimator has poor small sample properties.

Here we approximate the density of $z_t = \log \varepsilon_t^2$, the log χ_1^2 distribution, by a mixture density

$$f_Z(z) \simeq \sum_{i=1}^{K} \pi_i f_{z \mid \omega^i}(z \mid \omega^i), \tag{17}$$

where $z \mid \omega^i \sim N(\omega^i - 1{\cdot}2704, \sigma^2)$. We select $\omega^1, \ldots, \omega^K$, $\pi_1, \ldots, \pi_K$ and σ^2 to make the approximation 'sufficiently good'. In practice, we might make the first 4 or 6 moments agree exactly and require that the approximating density lie within a certain distance of the true density. See Titterington, Smith & Makov (1985, p. 133) for the appropriate techniques.

Clearly,

$$\log y_t^2 \mid \omega_t, \alpha_t \sim N(\alpha_t - 1{\cdot}2704 + \omega_t, \sigma^2), \tag{18}$$

where ω_t denotes the single element of $\omega^1, \ldots, \omega^K$ corresponding to time t. We can sample

from $\alpha|\omega, Y_n$, which has the same distribution as $\alpha|\omega, X_n$, where $X_n = (\log y_1^2, \ldots, \log y_n^2)'$. The second part of the multimove sampler requires us to simulate from

$$f(\omega|\alpha^*, X_n) = \prod_{t=1}^{n} f(\omega_t|\alpha_t^*, \log y_t^2) \propto \prod_{t=1}^{n} f(\log y_t^2|\omega_t, \alpha_t^*) f_\omega(\omega_t). \tag{19}$$

We can simulate from $\omega_t|\log y_t^2, \alpha_t^*$ using a rejection algorithm, drawing with replacement ω_t from $\omega^1, \ldots, \omega^K$ with probability $\pi_1, \ldots, \pi_K$ respectively.

6. PARTIAL NON-GAUSSIAN STATE SPACE

All the models discussed above, and many more, fit within the following framework which we call the partial non-Gaussian state space:

$$y_t = Z_t(\omega_t)\alpha_t + d_t(\omega_t) + \varepsilon_t(\omega_t), \quad \varepsilon_t(\omega_t)|\omega_t \sim N\{0, H_t(\omega_t)\},$$
$$\alpha_t = T_t(\omega_t)\alpha_{t-1} + c_t(\omega_t) + \eta_t(\omega_t), \quad \eta_t(\omega_t)|\omega_t \sim N\{0, Q_t(\omega_t)\}, \tag{20}$$

where $\omega = (\omega_1', \ldots, \omega_n')'$ is a Markov process, that is

$$f(\omega|\omega_0) = \prod_{t=1}^{n} f(\omega_t|\omega_{t-1}). \tag{21}$$

Throughout, $\varepsilon_t(\omega_t)|\omega_t$ and $\eta_s(\omega_s)|\omega_s$ are assumed independent of one another for all values of t and s.

This framework allows the introduction of non-Gaussianity and nonlinearities into the model while still allowing a certain ease of analysis. It is a special case of the set-up suggested by Kitagawa (1987). However, unlike Kitagawa's model, this approach does not exhibit a prohibitive computational explosion as the size of the state, α_t, expands.

The smoothing density $\delta|Y_n$, where $\delta = (\alpha', \omega')'$, can be simulated using a multimove sampler. The first step is to draw from $\alpha|\omega, Y_n$. As before

$$f(\alpha|\omega, Y_n) = f(\alpha_n|\omega, Y_n) \prod_{t=1}^{n-1} f(\alpha_t|\alpha_{t+1}, \omega, Y_t), \tag{22}$$

which are all Gaussian densities. The first of them, $f(\alpha_n|\omega, Y_n)$, is given by the Kalman filter, (6), while

$$\alpha_t|\alpha_{t+1}^*, \omega, Y_t \sim N\{a_t + P_{t,t+1|t}P_{t+1|t}^{-1}(\alpha_{t+1}^* - a_{t+1|t}), P_{t|t} - P_{t,t+1|t}P_{t+1|t}^{-1}P_{t+1,t|t}\} \tag{23}$$

can be found from a single run of the Kalman filter. Computationally the only problem is that, at each $t = n-1, n-2, \ldots, 1$, we have to compute the inverse of $P_{t+1|t}$, a matrix the dimension of the state vector.

This result implies that $\alpha|\omega, Y_n$ can be simulated routinely. The task of simulating from $\omega|\alpha^*, Y_n$ is not so easy in general, but there are some special cases where a relatively complete solution can be offered.

One case is where $\omega_t|\omega_{t-1}$ is non-Gaussian but where, given α^*, the model has a conjugate filter for ω. A class of models which possess this property is based on the gamma-beta transition models suggested by Smith & Miller (1986). An example of this is Shephard's (1994) local scale model. This will be illustrated by

$$y_t = \alpha_t + \varepsilon_t(\omega_t), \quad \varepsilon_t|\omega_t \sim N(0, \omega_t^{-1}), \quad \alpha_t = \alpha_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$
$$\omega_t = e^{r_t}\omega_{t-1}\eta_t, \quad \eta_t \sim \text{Beta}\{\lambda a_{t-1}, (1-\lambda)a_{t-1}\}, \quad r_t = -E\log\eta_t, \tag{24}$$

where a_{t-1} is defined in (25). Here the logarithm of the measurement equation precision follows a random walk. Then

$$\omega_t | \alpha^*, Y_t \sim G(a_t, b_t), \quad a_t = (\lambda a_{t-1|t-1}) + \tfrac{1}{2}, \quad b_t = (e^{-r_t} b_{t-1|t-1}) + \tfrac{1}{2}(y_t - \alpha_t^*)^2. \quad (25)$$

We can use a rejection algorithm to simulate from

$$f(\omega_t | \omega_{t+1}^*, \alpha^*, Y_t) \propto f(\omega_t | \alpha_t^*, y_t) f(\omega_{t+1}^* | \omega_t), \quad (26)$$

which is straightforward using a gamma random number generator.

Thus results on simple conjugate time series models can be exploited inside quite complicated models, widening their use.

For more general $f(\omega_t | \omega_{t-1})$, the following cases may be useful.

Case 6·1: ω_t are serially independent. We can simulate from

$$f(\omega | \alpha^*, Y_n) = \prod_{t=1}^{n} f(\omega_t | \alpha_t^*, \alpha_{t-1}^*, y_t) \propto \prod_{t=1}^{n} f(y_t | \alpha_t^*, \omega_t) f(\alpha_t^* | \alpha_{t-1}^*, \omega_t) f(\omega_t) \quad (27)$$

using a rejection algorithm if we can bound $f(y_t | \alpha_t^*, \omega_t) f(\alpha_t^* | \alpha_{t-1}^*, \omega_t)$. If this is not possible then we can fall back on the use of the Metropolis algorithm.

Case 6·2: Univariate general case. In general for a nonconjugate $\omega_t | \omega_{t-1}$ we can use a single-step Gibbs sampler based on the conditionals

$$f(\omega_t | \omega_{t-1}^*, \omega_{t+1}, \alpha^*, Y_n) = f(\omega_t | \omega_{t-1}^*, \omega_{t+1}, \alpha_{t-1}^*, \alpha_t^*, y_t)$$

$$\propto f(y_t, \alpha_{t-1}^*, \alpha_t^* | \omega_t) f(\omega_t | \omega_{t-1}^*, \omega_{t+1})$$

$$\propto f(y_t | \alpha_t^*, \omega_t) f(\alpha_t^* | \alpha_{t-1}^*, \omega_t) f(\omega_{t+1} | \omega_t) f(\omega_t | \omega_{t-1}^*), \quad (28)$$

working our way forwards through the data, simulating $\omega_1, \ldots, \omega_n$ in turn. Ways of carrying out rejection sampling from densities such as these are discussed by Carlin et al. (1992). When this is not directly possible then the Metropolis algorithm can be used (Shephard, 1993).

When ω_t is multidimensional the problems of drawing random variables from the relevant densities becomes much harder and probably should be addressed on a case by case basis.

7. Estimation using simulation

7·1. *The states*

Having obtained M replications from $\delta | Y_n$, perhaps M consecutive realisations from the multistep sampler which will not be independent, we need to be able to estimate the characteristics of $\alpha_t | Y_n$ and $\omega_t | Y_n$. We write the ith replication of $\omega | Y_n$ as $\omega^{(i)}$. A density estimator of $\alpha_t | Y_n$ is

$$M^{-1} \sum_{i=1}^{M} f(\alpha_t | \omega^{(i)}, Y_n),$$

while the corresponding mean, $E(\alpha_t | Y_n)$, can be estimated by

$$M^{-1} \sum_{i=1}^{M} E(\alpha_t | \omega^{(i)}, Y_n).$$

Since $\alpha | \omega^{(i)}, Y_n$ is Gaussian there exist exact smoothing algorithms for the mean and

variance of this distribution. Following a single pass of the Kalman filter, the backwards smoothing recursions take on the form

$$e_{t|n} = y_t - Z_t a_{t|n}, \quad n_{t|n} = a_{t|n} - T_t a_{t-1|n}, \quad a_{t|n} = a_{t|t-1} + P_{t|t-1} r_{t-1},$$

$$P_{t|n} = P_{t|t-1} - P_{t|t-1} N_{t-1} P_{t|t-1}, \quad r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t$$

$$(29)$$

(De Jong, 1989). The Kalman filter will already have been run in order to carry out the simulations of $\alpha | \omega$, Y_n and so this adds no extra burden. The smoothing algorithms here are more computationally involved than the Kalman filter; however the cost of running them is still small compared to the expense of the drawing of the simulations.

In general, analytic smoothing algorithms for $\omega | \alpha^*$, Y_n will not exist and so simpler averages of the replications $\omega^{(i)}$ will sometimes have to be used in reporting summaries of $\omega | Y_n$.

7·2. *One-step ahead prediction densities*

The likelihood is given by

$$f(y_1, \ldots, y_n | Y_0; \theta) = \prod_{t=1}^{n} f(y_t | Y_{t-1}; \theta).$$

Also

$$f(y_t | Y_{t-1}; \theta) = \int f(y_t | \Omega_{t-1}, Y_{t-1}; \theta) f(\Omega_{t-1} | Y_{t-1}) \, d\Omega_{t-1}, \qquad (30)$$

where $\Omega_{t-1} = (\omega_1', \ldots, \omega_{t-1}')'$. As we can use multimove sampling to draw from $\Omega_{t-1} | Y_{t-1}$, we can estimate (30) by

$$M^{-1} \sum_{i=1}^{M} f(y_t | \Omega_{t-1}^{(i)}, Y_{t-1}; \theta),$$

where $\Omega_{t-1}^{(i)}$ are the simulated values. Recall that $f(y_t | \Omega_{t-1}^{(i)}, Y_{t-1}; \theta)$ has a simple form for $y_t | \Omega_{t-1}, Y_{t-1}$, is Gaussian and its mean and variance are given by the Kalman filter (6).

As t increases we have to lengthen Ω and each iteration of the Gibbs sampler, in effect, performs a Kalman filter. Thus direct likelihood evaluation may not be the most computationally efficient way of maximising the likelihood. However, it is still useful to compute the likelihood to gauge the fit of the model, as well as to allow simple hypothesis testing. Further, the one-step ahead prediction densities are useful for diagnostic checking. These points will be illustrated in § 9.

8. PARAMETRIC ESTIMATION
8·1. *General*

It is important to present methods which deal effectively with the unknown parameters θ. We discuss two ways; a Bayesian approach and a simulated EM algorithm.

The Bayesian approach has been well documented elsewhere (Smith & Roberts, 1993), and so we discuss it only briefly here. It requires the specification of a prior over θ, written $f_\theta(\theta)$. Then the multimove sampler is amended as follows:

Multimove Bayesian sampler
1. Generate a candidate point $\delta = (\alpha', \omega', \theta')'$.
2. Draw $\alpha*$ from the density $f(\alpha | \omega, \theta, Y_n)$.
2. Draw $\omega*$ from the density $f(\omega | \alpha*, \theta, Y_n)$.
4. Draw $\theta*$ from the density $f(\theta | \omega*, \alpha*, Y_n)$.
5. Write $\delta = (\alpha*', \omega*', \theta*')'$ and go to 2.

This sampler will converge to replications from $\delta | Y_n$ which will contain drawings from the posterior density $\theta | Y_n$. The posterior density can be estimated by

$$M^{-1} \sum_{i=1}^{M} f(\theta | \omega^{(i)}, \alpha^{(i)}, Y_n).$$

If $f_\theta(\theta)$ is flat, this estimate can be used to find the maximum likelihood estimator.

8·2. EM *algorithm*

An alternative approach to find the maximum likelihood estimator uses an EM algorithm. This follows naturally from using smoothing techniques.

Recall

$$\log f(Y_n | Y_0; \theta) = \log f(\delta, Y_n | Y_0; \theta) - \log f(\delta | Y_n; \theta), \tag{31}$$

so that integrating both sides with respect to the smoothing density $f(\delta | Y_n; \theta^i)$, where θ^i is some arbitrary point, gives

$$\log f(Y_n | Y_0; \theta) = Q(\theta, \theta^i | Y_n) - H(\theta, \theta^i | Y_n), \tag{32}$$

where

$$Q(\theta, \theta^i | Y_n) = \int \log f(\delta, Y_n | Y_0; \theta) f(\delta | Y_n; \theta^i) \, d\delta, \tag{33}$$

$$H(\theta, \theta^i | Y_n) = \int \log f(\delta | Y_n; \theta) f(\delta | Y_n; \theta^i) \, d\delta. \tag{34}$$

The usual EM algorithm argument (Dempster, Laird & Rubin, 1977) implies that, if we select θ^{i+1} so that $Q(\theta, \theta^i | Y_n)$ is maximised with respect to θ, then

$$\log f(Y_n | Y_0; \theta^{i+1}) \geqslant \log f(Y_n | Y_0; \theta^i).$$

Iteration leads to the maximisation of the likelihood.

Using the structure of the general partial non-Gaussian state space we can solve some of these integrals analytically. We write

$$Q(\theta, \theta^i | Y_n) = \int R(\theta, \theta^i | \omega, Y_n) f(\omega | Y_n; \theta^i) \, d\omega + \int \log f(\omega; \theta) f(\omega | Y_n; \theta^i) \, d\omega, \tag{35}$$

where

$$R(\theta, \theta^i | \omega, Y_n) = \int \log f(Y_n, \alpha | \omega, Y_0; \theta) f(\alpha | \omega, Y_n; \theta^i) \, d\alpha \tag{36}$$

(Harvey, 1989, p. 188). We focus on the special case where all the parameters in the Gaussian measurement and transition equations are in the matrices H_t and Q_t. Then,

suppressing dependence on ω, we can follow Koopman (1993) in writing

$$R(\theta, \theta^i) = -\frac{1}{2} \sum_{t=1}^{n} \log |H_t(\theta)| - \frac{1}{2} \sum_{t=1}^{n} \text{tr} \left[H_t(\theta)^{-1} \{ e_{t|n} e'_{t|n} + \text{cov}\,(\varepsilon_t - e_{t|n}) \} \right]$$

$$- \frac{1}{2} \sum_{t=1}^{n} \log |Q_t(\theta)| - \frac{1}{2} \sum_{t=1}^{n} \text{tr} \left[Q_t(\theta)^{-1} \{ n_{t|n} n'_{t|n} + \text{cov}\,(\eta_t - n_{t|n}) \} \right], \qquad (37)$$

where $E(\varepsilon_t | \omega, Y_n) = e_{t|n}$ and $E(\eta_t | \omega, Y_n) = n_{t|n}$. These terms, and the others in the above formula, are computed recursively through

$$e_{t|n} = H_t e_t, \quad \text{cov}\,(\varepsilon_t - e_{t|n}) = Q_t r_{t-1},$$
$$n_{t|n} = H_t - H_t D_t H_t, \quad \text{cov}\,(\eta_t - n_{t|n}) = Q_t - Q_t N_{t-1} Q_t \qquad (38)$$

where they are calculated by first running the Kalman filter, (6), and then starting with $r_n = 0$ and $N_n = 0$ using the backwards recursions

$$e_t = F_t^{-1} v_t - K'_t r_t, \quad r_{t-1} = Z'_t F_t^{-1} v_t + L'_t r_t,$$
$$D_t = F_t^{-1} + K'_t N_t K_t, \quad N_{t-1} = Z'_t F_t^{-1} Z_t + L'_t N_t L_t. \qquad (39)$$

Throughout the recursions and the associated Kalman filter, θ is taken to be θ^i.

8·3. *Simulated EM algorithm*

We can estimate $Q(\theta, \theta^i | Y_n)$ by using simulations from $\omega | Y_n; \theta^i$. Simulated EM algorithms have been used by, amongst others, Bresnahan (1981), Celeux & Diebolt (1985), Wei & Tanner (1990), Tanner (1991), Qian & Titterington (1991) and Shephard (1993). So

$$\hat{Q}(\theta, \theta^i | Y_n) = \frac{1}{M} \sum_{j=1}^{M} \{ R(\theta, \theta^i | \omega^{(j)}, Y_n) + \log f(\omega^{(j)}; \theta) \}, \qquad (40)$$

which is usually easy to manipulate. Typically the parameter space splits so that we can work with $\theta = (\chi', \tau')'$ so that

$$\hat{Q}(\theta, \theta^i | Y_n) = \frac{1}{M} \sum_{j=1}^{M} \{ R(\chi, \theta^i | \omega^{(j)}, Y_n) + \log f(\omega^{(j)}; \tau) \}, \qquad (41)$$

in which case the maximisation will be particularly straightforward. In some special cases τ may be a null set: this is true, for instance, in the stochastic variance model discussed in § 3. A similar type of argument can be used to deliver a direct estimator of the score vector (Koopman & Shephard, 1992).

9. NONLINEAR SEASONAL ADJUSTMENT

We illustrate the techniques suggested above by modelling the quarterly changes of M4 in the UK from 63Q2 to 82Q4 in Fig. 3(a); a series used in the Bank of England's 1992 report on seasonal adjustment procedures. The Bank's report compares three seasonal adjustment procedures: an in house moving average scheme, the US Bureau of the Census' X11 and a model based seasonal adjustment procedure which fits a Gaussian unobserved component model (Harvey, 1989, pp. 300–9). We focus on the last of these, to compare its fit with a new non-Gaussian model we propose.
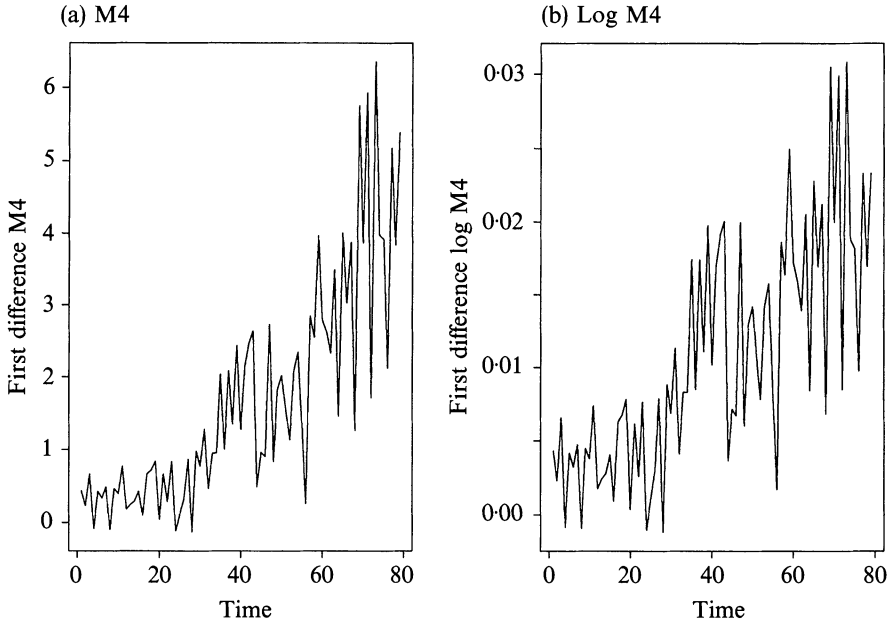
Fig. 3. Quarterly changes (a) in UK first difference M4 (£ billions): 63Q2 to 82Q4, and (b) in log of first difference M4.

The Gaussian model fitted to M4 was

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2), \tag{42}$$

$$\gamma_t = -\gamma_{t-1} \ldots - \gamma_{t-3} + \omega_t, \quad \omega_t \sim NID(0, \sigma_\omega^2).$$

Here $\varepsilon_t, \eta_s, \zeta_u$ and ω_v are assumed independent for all t, s, u and v. The maximum likelihood estimates of the parameters are

$$\hat{\sigma}_\varepsilon^2 = 0 \cdot 21, \quad \hat{\sigma}_\eta^2 = 0 \cdot 041, \quad \hat{\sigma}_\zeta^2 = 0 \cdot 000007, \quad \hat{\sigma}_\omega^2 = 0 \cdot 020, \quad \log L(\hat{\theta}) = -79, \tag{43}$$

while the standard innovations-based diagnostic statistics (Harvey, 1989, pp. 258–60) take on the values

$$S = 0 \cdot 20, \quad K = 4 \cdot 12, \quad N(\chi_2^2) = 4 \cdot 39, \quad Q_{10}(\chi_9^2) = 9 \cdot 75, \quad H_{24}(F_{24,24}) = 17 \cdot 1. \tag{44}$$

These statistics refer to skewness, kurtosis, normality, serial correlation measured by the Box–Ljung statistic and heteroscedasticity, respectively. The diagnostics are satisfactory except the model allows too much variability in the initial part of the series while not allowing enough at the end. This is picked up in Fig. 4(a) and by the heteroscedasticity statistic H_{24}.

Figure 3(b) gives differences of the logged series. This is convenient as this transformation approximately equals the proportional change, which is probably the statistic of most interest. The innovations from fitting (42) to Fig. 3(b) are given in Fig. 4(b). Although they show some minor improvements over Fig. 4(a), they are still very poor. The diagnostic statistics and estimated components from this model are very similar to those of the linear
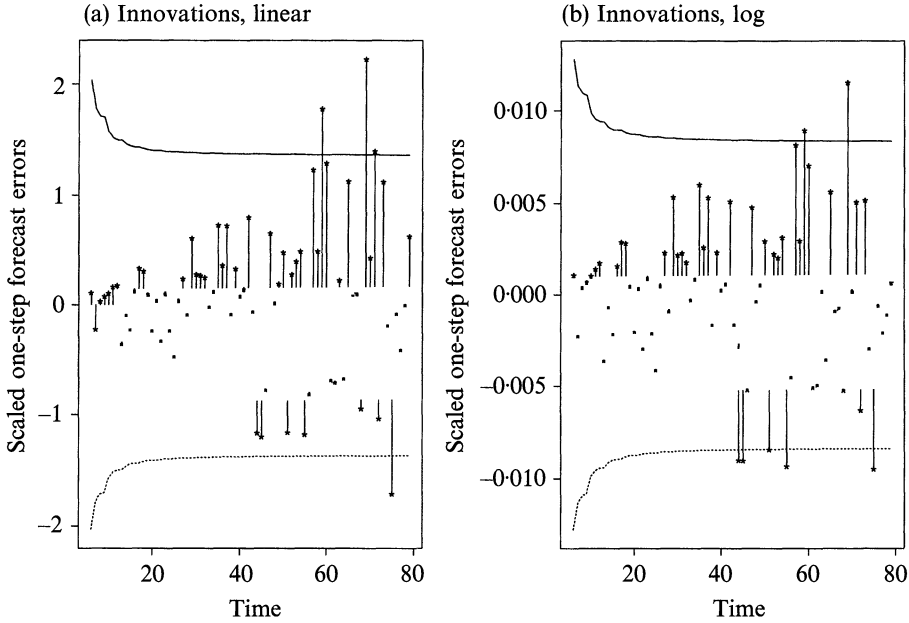
Fig. 4. Innovations (a) from linear model and corresponding 95% intervals, (b) from log-linear models and 95% intervals.

model. To try to improve the fit of the linear model we exploit a mixed multiplicative-additive unobserved component model. The model we use is

$$y_t = \mu_t(1 + \gamma_t^* + \varepsilon_t^*) + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2), \quad \varepsilon_t^* \sim NID(0, \sigma_{\varepsilon*}^2),$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2),$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2),$$

$$\gamma_t^* = -\gamma_{t-1}^* - \ldots - \gamma_{t-3}^* + \omega_t^*, \quad \omega_t^* \sim NID(0, \sigma_{\omega*}^2),$$

$$(45)$$

where ε_t, ε_s^*, η_u, ζ_v and ω_j^* are assumed independent for all t, s, u, v, i and j. Notice that unlike (42), (45) cannot be made stationary by transformation.

The γ_t^* is the multiplicative seasonal effect and enters the model as a proportion, moving the trend up and down by scaling. The total seasonal effect is $\mu_t\gamma_t^*$. Likewise the irregular term in this model is $\mu_t\varepsilon_t^* + \varepsilon_t$, making it heteroscedastic.

This is an example of the multiplicative models discussed briefly in § 3. We can write $\omega_t = (\gamma_t^*, \gamma_{t-1}^*, \gamma_{t-2}^*, \varepsilon_t^*)'$ to place the model into its partial non-Gaussian state space. So long as the variances are all strictly positive the multistep Gibbs sampler will converge to its required distribution. In practice the rate of convergence will depend crucially on σ_ε^2, a low value rendering the algorithm relatively slow.

The simulated EM algorithm was used to perform parameter estimation, yielding

$$\hat{\sigma}_\varepsilon^2 = 0.004, \quad \hat{\sigma}_{\varepsilon*}^2 = 0.094, \quad \hat{\sigma}_\eta^2 = 0.01, \quad \hat{\sigma}_\zeta^2 = 0.00005, \quad \hat{\sigma}_{\omega*}^2 = 0.003, \quad \log L(\hat{\theta}) = -62.$$

$$(46)$$

Most of the irregular component has gone into ε_t^* which accounts for the large increase in the likelihood over the linear model. The variance of the multiplicative seasonal is lower than for the linear model, but this is mostly due to the scaling effect of the trend and so
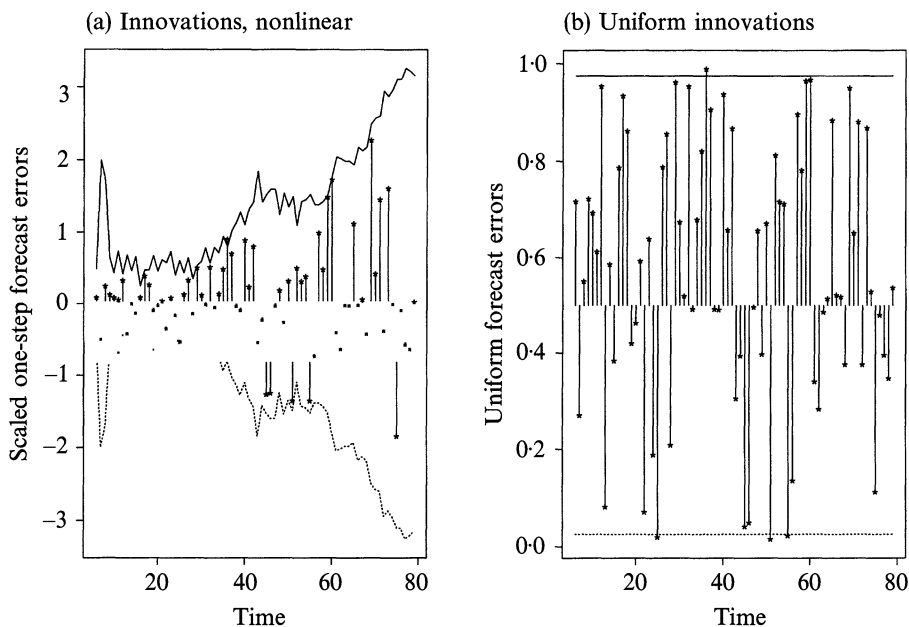
(a) Innovations, nonlinear

(b) Uniform innovations

Fig. 5. (a) Innovations from fitted nonlinear model and twice their mean squared errors.
(b) Uniform innovations, from Fig. 3.

this fitted model does not exhibit a dramatically more stable seasonal pattern than the linear one. The variance on the random walk part of the trend has fallen fourfold, while the slope variance has increased to compensate. Overall we will see below that this implies that the estimated trend is discernibly smoother than for the linear model.

The raw innovations from the fitted model, based on $y_t - y_t | Y_{t-1}$ ($t = 6, \ldots, 79$), and twice their mean squared errors are given in Fig. 5(a). The innovations are not normally distributed. The varying mean squared errors are the effect of the multiplicative irregular term.

Figure 5(b) is statistically more satisfactory. It is generated by using the model to randomly produce 1000 one-step ahead forecasts, for each t, and then recording how many of these were smaller than the actual observations. A naive diagnostic can then be based on the hypothesis that they should be independently uniformly distributed on the inverval 0 to 1 if the model is good: naive in that this distribution does not take into the account that 5 parameters have been estimated; 95% intervals are also drawn on this figure and indicate that the scatter of large or small forecasts is not obviously unusual.

The uniform variables displayed in Fig. 5(b) can also be mapped back into $N(0, 1)$ variables. These normal innovations were used to construct correlograms and probability plots to check for remaining structure in the data. They show very little activity. Taken collectively this battery of diagnostic graphs suggests that the model has fitted the data well, probably because of the addition of the multiplicative irregular term.

The estimated trends for the linear and nonlinear models are given in Fig. 6; the results for the log-linear model are similar to the linear. As suspected, the nonlinear model trend is considerably smoother. Of more interest is the estimated seasonal component given in Fig. 7. There the linear model component is rather different at the beginning of the series but is quite close to the nonlinear one at the end. Altogether the linear model does surprisingly well, given how poorly it actually fits the data.
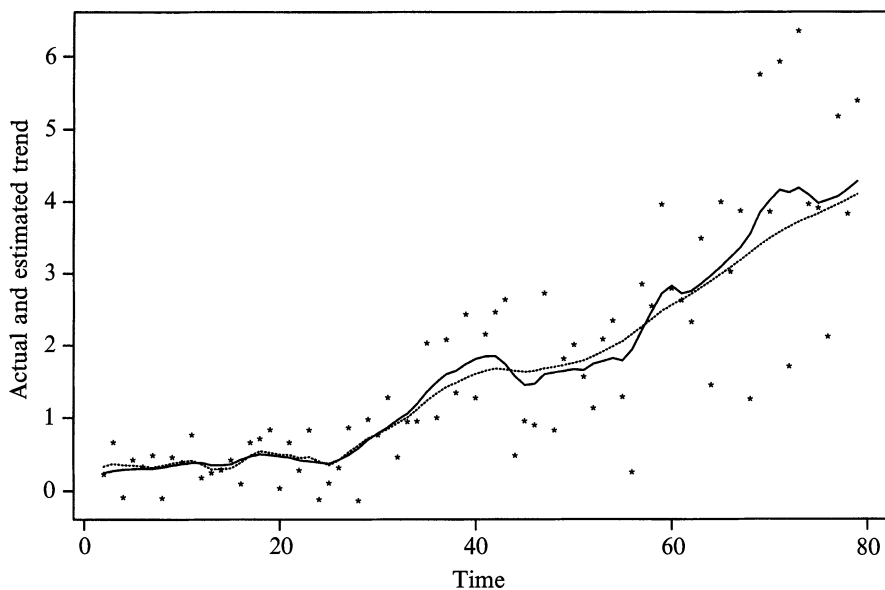
Fig. 6. Actual and estimated trend, linear (solid line) and nonlinear (dotted line).
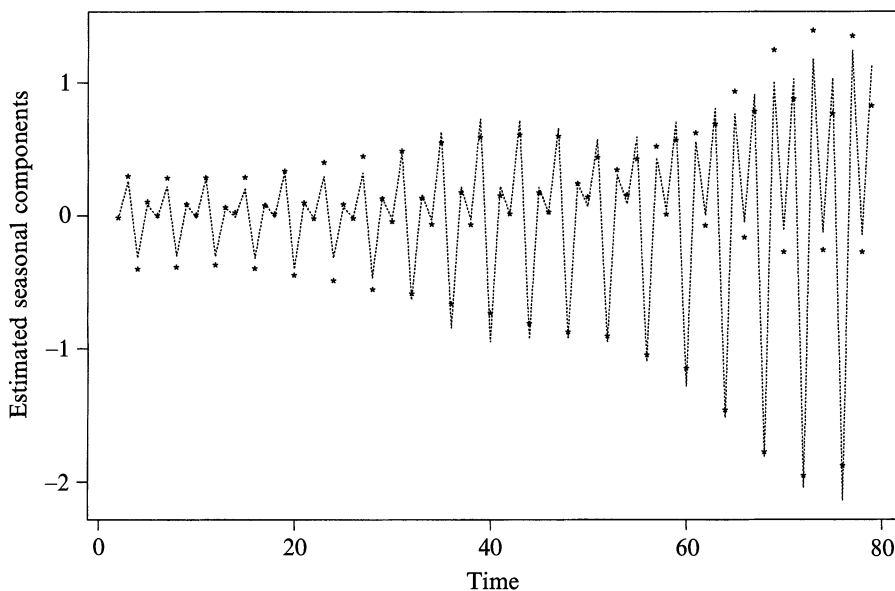

Fig. 7. Estimated seasonal pattern, linear (points) and nonlinear (dotted line).

REFERENCES

ALSPACH, D. L. & SORENSON, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Auto. Control* **AC-17**, 439–48.

ANDERSON, B. D. O. & MOORE, J. B. (1979). *Optimal Filtering*. Englewood Cliffs: Prentice-Hall.

BRESNAHAN, T. F. (1981). Departures from marginal-cost pricing in the American automobile industry, estimates from 1977–1978. *J. Economet.* **17**, 201–27.

CARLIN, B. P., POLSON, N. G. & STOFFER, D. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *J. Am. Statist. Assoc.* **87**, 493–500.

CARTER, C. K. & KOHN, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**. To appear.

CELEUX, G. & DIEBOLT, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computat. Statist. Quart.* **2**, 73–82.

CHESNEY, M. & SCOTT, L. O. (1989). Pricing European options: a comparison of the modified Black-Scholes model and a random variance model. *J. Financ. Qualitat. Anal.* **24**, 267–84.

DE JONG, P. (1989). Smoothing the interpolation with the state space model. *J. Am. Statist. Assoc.* **84**, 1085–8.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc.* B **39**, 1–38.

FRUHWIRTH-SCHNATTER, S. (1994). Data augmentation and dynamic linear models. *J. Time Ser. Anal.* **15**, 183–202.

HAMILTON, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–84.

HARRISON, J. & STEVENS, C. F. (1976). Bayesian forecasting (with discussion). *J. R. Statist. Soc.* B **38**, 205–47.

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

HARVEY, A. C., RUIZ, E. & SHEPHARD, N. (1994). Multivariate stochastic variance models. *Rev. Econ. Stud.* **61**. To appear.

HULL, J. & WHITE, A. (1987). Hedging the risks for writing foreign currency options. *J. Int. Money Finance* **6**, 131–52.

KITAGAWA, G. (1987). Non-Gaussian state space modelling of non-stationary time series. *J. Am. Statist. Assoc.* **82**, 503–14.

KITAGAWA, G. (1989). Non-Gaussian seasonal adjustment. *Comp. Math. Applic.* **18**, 503–14.

KOOPMAN, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**, 117–26.

KOOPMAN, S. J. & SHEPHARD, N. (1992). Exact score for time series models in state space form. *Biometrika* **79**, 823–6.

MELINO, A. & TURNBULL, S. M. (1990). Pricing foreign currency options with stochastic volatility. *J. Economet.* **45**, 239–65.

PEÑA, D. & GUTTMAN, L. (1988). Bayesian approach to robustifying the Kalman filter. In *Bayesian Analysis of Time Series and Dynamic Models*, Ed. J. C. Spall, pp. 227–54. Cambridge, Mass.: Marcel Dekker.

QIAN, W. & TITTERINGTON, D. M. (1991). Estimation of parameters in hidden Markov Chain models. *Phil. Trans. R. Soc. Lond.* A **337**, 407–28.

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

SHEPHARD, N. (1993). Fitting non-linear time series models, with applications to stochastic variance models. *J. Appl. Economet.* **8**, 5135–58.

SHEPHARD, N. (1994). Local scale model: state space alternative to integrated GARCH processes. *J. Economet.* **60**, 181–202.

SMITH, A. F. M. & ROBERTS, G. (1993). Bayesian computations via the Gibbs sampler and related Markov Chain Monte Carlo methods. *J. R. Statist. Soc.* B **55**, 3–23.

SMITH, R. L. & MILLER, J. E. (1986). A non-Gaussian state space model and application to prediction records. *J. R. Statist. Soc.* B **48**, 79–88.

TANNER, M. A. (1981). *Tools for Statistical Inference, Observed Data and Data Augmentation Methods*, **67**. New York: Springer-Verlag.

TIERNEY, L. (1994). Markov Chains for exploring posterior distributions. *Ann. Statist.* **22**. To appear.

TITTERINGTON, D. M., SMITH, A. F. M. & MAKOV, U. E. (1985). *Statistical Analysis of Finite Markov Distributions*. Chichester: Wiley.

WEI, G. C. G. & TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Statist. Assoc.* **85**, 699–704.

WEST, M. & HARRISON, J. (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.