ACTUARIAL RESEARCH CLEARING HOUSE
1992 VOL. 1

Analyzing Nonlinear and Non-Gaussian Actuarial Time Series

BRADLEY P. CARLIN[1]

July 23, 1991

ABSTRACT

Insurance professionals have long recognized the value of time series techniques in analyzing sequential business and economic data. However, the usual class of Box-Jenkins (1976) models are sometimes not flexible enough to adequately describe many practically-arising time series. By way of contrast, the state space model offers a solution to multivariate modeling, forecasting, and smoothing of time series, but the computations associated with this model have been too heavy in the past to justify their usage. Recently, however, Carlin, Polson and Stoffer (1992) have showed how state space models of this type may be more easily fit using Monte Carlo integration techniques. In this paper we give a brief review of these techniques, and subsequently illustrate their usefulness in settings of interest to practicing actuaries. The models employed allow for the possibilities of nonnormal errors and nonlinear functionals in the state equation, the observational equation, or both. Missing data problems (including the k-step ahead prediction problem) are also easily incorporated into this framework. We illustrate the broad applicability of our approach with several examples and data compiled by the National Center for Health Statistics.

[1]Bradley P. Carlin is Assistant Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455. The author thanks Dr. Myron Katzoff of the National Center for Health Statistics for access to the dataset analyzed herein, and Dr. Nicholas Polson of the University of Chicago and Dr. David Stoffer of the University of Pittsburgh for substantial assistance during their longtime collaborations with the author on problems of this kind.

1. INTRODUCTION

Over the past several years, time series analysis has become one of the practicing actuary's standard tools for studying and understanding sequential data. These techniques have proved especially valuable for analyzing economic indicator series, such as stock prices and interest rates over time, and also mortality and morbidity series, such as monthly death or disability rates from various diseases by age and sex. In both of the above examples, the ability to react quickly to changes in conditions, as well as to reliably forecast future trends, is critical to the well-being of a financial institution.

Unfortunately, because the appearance of time series methods on the actuarial syllabus has been fairly recent, many actuaries are unaware of the power afforded by these techniques. Further, most of those who have studied time series analysis have only seen the most basic linear, Gaussian (normal) models developed and Box and Jenkins (1976), as covered on the exam syllabus in the textbook by Miller and Wichern (1977). This is unfortunate, as many of the time series encountered in practice cannot be described by such simple models. For example, a standard model might take the form

$$y_t = h(y_{t-1}) + \epsilon_t,\ t = 1, \ldots, n$$

where $h(y_{t-1}) = \phi y_{t-1}$, a linear function, and the ϵ_t's are assumed to be a "white noise" process, i.e. independent Gaussian errors with mean 0. However, a more realistic assumption for many datasets might be to let $h(y_{t-1})$ be a *nonlinear* function of y_{t-1}, or to adopt a *non-Gaussian* error distribution for the ϵ_t's. The above assumptions enable simple updating of estimates via the usual Kalman filter, but are frequently found in practice to be too restrictive for realistic data analysis.

As an example, consider the ischemic heart disease mortality data in Figure 1(a). These are the final National Center for Health Statistics (NCHS) estimates of the monthly death rates for men

aged 25 to 34 over the period January 1979 – December 1986. As the trend is clearly nonstationary, a standard Box-Jenkins autoregressive integrated moving average (ARIMA) model analysis would proceed by differencing the series to achieve stationarity; a histogram of these differenced death rates is given in Figure 1(b). Notice that this distribution does not appear to be normally distributed; a more sharply peaked distribution with much heavier tails, such as the double exponential distribution, seems to be called for. Section 4 below offers further examples from the realm of mortality rate time series that illustrate such advanced modeling conditions.

In the remainder of this paper, we first review the applicable statistical methodology for the analysis of morbidity, mortality and other health related data that are known to exhibit nonlinear and non-Gaussian behavior. In particular, an adaptive Monte Carlo computational technique known as the Gibbs sampler is proposed as a mechanism for implementing a conceptually and computationally simple solution in such situations. This method will help implement our second goal: the solution of several difficult modeling problems that have heretofore been intractable using traditional computational methods. Examples of such problems include explicit incorporation of covariates (such as age and sex), heteroscedasticity of errors over time, multivariate analysis (including simultaneous modeling of both preliminary and final mortality estimates), the modeling of asymmetric densities on the positive real line (as might be appropriate for death rates), and model choice criteria for selecting the best model from many. Finally, we offer specific numerical examples which illustrate the relevance of the methodology, giving details concerning our FORTRAN implementation. We suggest that the availability and easy programmability of advanced time series modeling techniques will lead to higher quality estimation and prediction of time-dependent quantities relevant to financial institutions managing risk.

2. Model specification and the Gibbs sampler

2.1 Model Specification

The state space model has become a powerful tool for modeling and forecasting dynamic systems. These models, in conjunction with the Kalman filter, have been used in a wide range of applications from many disciplines including biology, economics, and engineering, and consequently have become of increasing interest to statisticians. These models are particularly amenable to a Bayesian approach, since the time-ordered arrival of the data means that the notion of updating prior knowledge in the presence of new data arises quite naturally. Good summaries of the work in this area appear in West, Harrison and Migon (1985) and West and Harrison (1989).

We will consider the state space model:

$$\begin{aligned} x_t &= F_t x_{t-1} + u_t, \qquad \text{and} \\ y_t &= H_t x_t + v_t, \quad t = 1, \ldots, n \end{aligned} \tag{1}$$

where x_t is the $p \times 1$ state vector, y_t is the $q \times 1$ observation vector, F_t is a $p \times p$ matrix of constants, and H_t is a $q \times p$ matrix of constants. Let $\mathbf{y} = (y_1, \ldots, y_n)$ denote the observed data, $\mathbf{x} = (x_1, \ldots, x_n)$ the (unknown) elements of the state, and x_0 the initial state. Typically, u_t and v_t are taken as independent and identically distributed, with $u_t \sim N_p(0, \Sigma)$ and $v_t \sim N_q(0, \Upsilon)$, where N_p denotes the p-dimensional normal distribution. Also, the matrices F_t, H_t, Σ, and Υ are generally assumed to be known.

In a recent paper, Carlin, Polson and Stoffer (1992) developed methodology for modeling the nonnormality of the u_t, the v_t, or both. A further departure from the model specification (1) was to allow for asymmetric densities on $\Re^+$ (e.g. gamma or Weibull), unknown and possibly unequal variances in the state or observational equation, and unknown parameters in the transition matrices

F_t and H_t. In general, we may allow for nonlinear functional forms by writing

$$\begin{aligned} x_t &= f_t(x_{t-1}) + u_t, \qquad \text{and} \\ y_t &= h_t(x_t) + v_t, \quad t = 1, \ldots, n, \end{aligned} \tag{2}$$

where $f_t(\cdot)$ and $h_t(\cdot)$ are given, but perhaps depend on some unknown parameters. Finally, the experimenter may wish to entertain a variety of possible nonlinear functional forms or choices of error distributions, resulting in a model choice problem.

In general, the likelihood specification for our model, suppressing the conditioning on $(\mu_0, \Sigma_0, F_t, H_t)$, is given by

$$p(y_1, \ldots, y_n, x_0, x_1, \ldots, x_n | \Sigma, \Upsilon) = g_1(x_0|\mu_0, \Sigma_0) \prod_{t=1}^{n} g_1(x_t|x_{t-1}, \Sigma) \prod_{t=1}^{n} g_2(y_t|x_t, \Upsilon) \tag{3}$$

for some densities $g_1(\cdot)$ and $g_2(\cdot)$. Specifically, we model g_1 and g_2 by letting

$$\begin{aligned} g_1(x_t|x_{t-1}, \Sigma) &= \textstyle\int_\Lambda p(x_t|x_{t-1}, \lambda_t, \Sigma) p_1(\lambda_t) d\lambda_t, \qquad \text{and} \\ g_2(y_t|x_t, \Upsilon) &= \textstyle\int_\Omega p(y_t|x_t, \omega_t, \Upsilon) p_2(\omega_t) d\omega_t, \qquad t = 1, \ldots, n\ , \end{aligned} \tag{4}$$

where we depart from the usual Gaussian assumption by assuming that, conditional on the nuisance parameters λ and ω,

$$\begin{aligned} x_t|x_{t-1}, \lambda_t, \Sigma &\sim N(f_t(x_{t-1}), \lambda_t \Sigma), \qquad \text{and} \\ y_t|x_t, \omega_t, \Upsilon &\sim N(h_t(x_t), \omega_t \Upsilon), \qquad t = 1, \ldots, n\ . \end{aligned} \tag{5}$$

Of course if $h_t(x_t) = H_t x_t$ and $f_t(x_{t-1}) = F_t x_{t-1}$, we have the linear model (1). Note that, by varying $p_1(\lambda_t)$ and $p_2(\omega_t)$, the distributions g_1 and g_2 are scale mixtures of multivariate normals for each t, thus enabling a wide variety of nonnormal error densities to emerge in (3). For example,

in the univariate case (where we denote Σ and Υ by σ and τ, respectively) the distributions $x_t|x_{t-1}, \sigma$ and $y_t|x_t, \tau$ can be double exponential, logistic, exponential power, or t densities (see Andrews and Mallows, 1974, West, 1987, and Carlin and Polson, 1991). In the multivariate case a rich class of densities emerges including the r-dimensional hyperbolic distribution (see Barndorff-Neilsen and Halgreen, 1977). Note that we are assuming $p(\lambda, \omega) = \prod_{t=1}^{n} p_1(\lambda_t)p_2(\omega_t)$, so that the densities $x_t|x_{t-1}, \Sigma$ and $y_t|x_t, \Upsilon$ are possibly different scale mixtures of normals. A further easily incorporated extension is to allow for different densities as t varies, $t = 1, \ldots, n$.

While we plan to investigate methodology for error densities on the positive real line, for the purpose of illustration the rest of this proposal will focus on modeling in the symmetric errors case using a nonlinear, multivariate scale mixture state space model. The key to the approach is the introduction of the (generally high dimensional) nuisance parameters λ and ω and the structure (5) which, as we shall now see, lends itself naturally to the Gibbs sampler, our computational tool.

2.2 Implementation of the Gibbs sampler

The Gibbs sampler is a Monte Carlo integration method which proceeds by a Markovian updating scheme. It is essentially a modification of the Metropolis algorithm (Metropolis et. al., 1953), developed formally by Geman and Geman (1984) in the context of image restoration. In the statistical framework, Tanner and Wong (1987) used essentially this algorithm in their substitution sampling approach. Most recently, Gelfand and Smith (1990) developed the Gibbs sampler for general settings; the reader is referred to that paper for a discussion of the method and its properties. To summarize the method briefly, suppose we have a collection of k (possibly vector-valued) random variables $U_1, \ldots, U_k$ whose complete conditional distributions, denoted generically by $f(U_s|U_r, r \neq s), s = 1, \ldots, k$, are available for sampling. Here, "available" means that samples may be generated by some method, given values of the appropriate condi-

tioning random variables. Under mild conditions (see Besag, 1974), these complete conditional distributions uniquely determine the full joint distribution, $f(U_1,\ldots,U_k)$, and hence all marginal distributions $f(U_s), s = 1,\ldots,k$. The Gibbs sampler generates samples from these marginal distributions as follows: Given an arbitrary starting set of values $U_{1(0)},\ldots,U_{k(0)}$, we draw $U_{1(1)}$ from $f(U_1|U_{2(0)},\ldots,U_{k(0)})$, then $U_{2(1)}$ from $f(U_2|U_{1(1)},U_{3(0)},\ldots,U_{k(0)})$, and so on up to $U_{k(1)}$ from $f(U_k|U_{1(1)},\ldots,U_{k-1(1)})$ to complete one iteration of the scheme. After l such iterations we obtain $(U_{1(l)},\ldots,U_{k(l)})$. Geman and Geman (1984) show under mild conditions that this k-tuple converges in distribution to a random observation from $f(U_1,\ldots,U_k)$ as $l \to \infty$. For this reason, in the sequel we suppress the (l) subscript, assuming that l is sufficiently large for the generated sample to be thought of as a realization from the joint distribution. Now, replicating the entire process in parallel G times provides i.i.d. k-tuples $(U_1^{(g)},\ldots,U_k^{(g)})$, $g = 1,\ldots,G$ from the joint distribution. These observations can then be used for estimation of any of the marginal densities. In particular, if $f(U_s|U_r, r \neq s)$ is available in closed form, then

$$\hat{f}(U_s) = \frac{1}{G}\sum_{g=1}^{G} f(U_s|U_r^{(g)}, r \neq s)\,. \tag{6}$$

In the context of our state space models, in order to implement the Gibbs sampler we require samples from the following complete conditional distributions:

- $x_t|x_{j\neq t},\boldsymbol{\lambda},\omega,\Sigma,\Upsilon,\mathbf{y},\ t = 0,\ldots,n$
- $\omega_t|\omega_{j\neq t},\boldsymbol{\lambda},\Sigma,\Upsilon,\mathbf{y},\mathbf{x},x_0 \sim \omega_t|\Upsilon,y_t,x_t,\ t = 1,\ldots,n$
- $\lambda_t|\lambda_{j\neq t},\omega,\Sigma,\Upsilon,\mathbf{y},\mathbf{x},x_0 \sim \lambda_t|\Sigma,x_t,x_{t-1},\ t = 1,\ldots,n$
- $\Sigma|\boldsymbol{\lambda},\omega,\Upsilon,\mathbf{y},\mathbf{x},x_0 \sim \Sigma|\boldsymbol{\lambda},\mathbf{y},\mathbf{x},x_0$
- $\Upsilon|\boldsymbol{\lambda},\omega,\Sigma,\mathbf{y},\mathbf{x},x_0 \sim \Upsilon|\omega,\mathbf{y},\mathbf{x}$

We now consider the first two distributions above. The third follows in a similar manner to the second. The last two, under conjugate priors, follow from standard normal and Wishart distribution theory, due to the conditioning on λ and ω.

First, under model (1), we prove a lemma which determines the set of conditionals, $x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y}, t = 1, \ldots, n$. The nonlinear case (2) will be illustrated in Example 2.2 below.

LEMMA. The complete conditional distribution $x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y}$ is $N_p(B_t b_t, B_t)$, where

$$B_t^{-1} = \frac{\Sigma^{-1}}{\lambda_t} + \frac{H_t^T \Upsilon^{-1} H_t}{\omega_t} + \frac{F_{t+1}^T \Sigma^{-1} F_{t+1}}{\lambda_{t+1}}, \text{ and } b_t^T = \frac{x_{t-1}^T F_t^T \Sigma^{-1}}{\lambda_t} + \frac{y_t^T \Upsilon^{-1} H_t}{\omega_t} + \frac{x_{t+1}^T \Sigma^{-1} F_{t+1}}{\lambda_{t+1}} . \quad (7)$$

PROOF. By Bayes theorem, the required exponent is a sum of three terms, that is, modulo a normalizing constant, $-2 \log f(x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y})$ is

$$\frac{1}{\lambda_t}(x_t - F_t x_{t-1})^T \Sigma^{-1}(x_t - F_t x_{t-1}) + \frac{1}{\omega_t}(y_t - H_t x_t)^T \Upsilon^{-1}(y_t - H_t x_t) + \frac{1}{\lambda_{t+1}}(x_{t+1} - F_{t+1} x_t)^T \Sigma^{-1}(x_{t+1} - F_{t+1} x_t)$$

which on manipulation gives the desired result. □

Note that adjustments will need to be made to formula (7) for the cases $t = 0$ and $t = n$ due to slight modifications and deletions in the likelihood for these "endpoint" cases. We illustrate these modifications in Example 2.1.

Now consider the determination of $\omega_t|\omega_{j\neq t}, \lambda, \Sigma, \Upsilon, \mathbf{y}, \mathbf{x} \sim \omega_t|\Upsilon, y_t, x_t, \ t = 1, \ldots, n$. By Bayes theorem, $\omega_t|\Upsilon, y_t, x_t \propto f(y_t|x_t, \omega_t, \Upsilon) p_2(\omega_t)$. But by (4) the normalization constant is known, and is given by $g_2(y_t|x_t, \Upsilon)$. Hence the complete conditional for ω_t is of known functional form. Generation of the required samples may be done directly if this form is a standard density; otherwise, a carefully selected rejection method may be employed.

We now turn to two examples.

Example 2.1: Univariate linear model. For the purpose of illustration, consider model (1)

with $p = q = 1$ and $\Sigma_0 = \sigma_0^2, \Sigma = \sigma^2$, $\Upsilon = \tau^2$, $H_t = H$ and $F_t = F$. Using the above lemma and taking care with the endpoint cases, we have $x_t|x_{j\neq t}, \lambda, \omega, \sigma, \tau, \mathbf{y} \sim N(B_t b_t, B_t)$ where

$$B_t^{-1} = \begin{cases} \frac{1}{\sigma_0^2} + \frac{F^2}{\sigma^2 \lambda_1} & t = 0 \\ \frac{1}{\sigma^2}\left(\frac{1}{\lambda_t} + \frac{F^2}{\lambda_{t+1}}\right) + \frac{H^2}{\tau^2 \omega_t}, & t = 1, \ldots, n-1 \\ \frac{1}{\sigma^2 \lambda_n} + \frac{H^2}{\tau^2 \omega_n}, & t = n \end{cases}$$

and

$$b_t = \begin{cases} \frac{\mu_0}{\sigma_0^2} + \frac{F x_1}{\sigma^2 \lambda_1}, & t = 0 \\ \frac{F}{\sigma^2}\left(\frac{x_{t-1}}{\lambda_t} + \frac{x_{t+1}}{\lambda_{t+1}}\right) + \frac{H y_t}{\tau^2 \omega_t}, & t = 1, \ldots, n-1 \\ \frac{F x_{n-1}}{\sigma^2 \lambda_n} + \frac{H y_n}{\tau^2 \omega_n}, & t = n \end{cases}$$

The complete conditionals for σ^2 and τ^2 are obtained as follows. Assuming the independent *a priori* specifications $\sigma^2 \sim IG(a_0, b_0)$ and $\tau^2 \sim IG(c_0, d_0)$, where IG denotes the inverse (reciprocal) gamma distribution, then

$$\begin{aligned} \sigma^2|\boldsymbol{\lambda}, \mathbf{y}, \mathbf{x}, x_0 &\sim IG\left(a_0 + \tfrac{n}{2}, \left\{\tfrac{1}{b_0} + \tfrac{1}{2}\textstyle\sum_{t=1}^{n}(x_t - F x_{t-1})^2/\lambda_t\right\}^{-1}\right), \text{ and} \\ \tau^2|\boldsymbol{\omega}, \mathbf{y}, \mathbf{x} &\sim IG\left(c_0 + \tfrac{n}{2}, \left\{\tfrac{1}{d_0} + \tfrac{1}{2}\textstyle\sum_{t=1}^{n}(y_t - H x_t)^2/\omega_t\right\}^{-1}\right). \end{aligned} \tag{8}$$

For the ω complete conditionals, suppose we wish to model $\mathbf{y}|\mathbf{x}, \tau$ as a product of double exponentials. The necessary *a priori* specification for ω_t is then $\omega_t \sim Expo(2)$, the exponential distribution having mean 2. Since $y_t|x_t, \omega_t, \tau \sim N(H x_t, \omega_t \tau^2)$, the complete conditional for ω_t is then

$$\omega_t|\tau, \mathbf{y}, \mathbf{x} \propto \omega_t^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\omega_t + \frac{(y_t - H_t x_t)^2}{\omega_t \tau^2}\right)\right), \tag{9}$$

that is, $\omega_t|\tau, \mathbf{y}, \mathbf{x} \sim GIG\left(\frac{1}{2}, 1, \left(\frac{y_t - H_t x_t}{\tau}\right)^2\right)$, where GIG denotes the generalized inverse Gaussian distribution (see Devroye, 1986, p. 478). In order to sample from this density, we note that it is

the reciprocal of an Inverse Gaussian $\left(\left|\frac{\tau}{y_t - H_t x_t}\right|, 1\right)$, a density from which we may easily sample.

A similar approach to the one just described could be used to model nonnormality in the state equation via the λ complete conditionals. Finally, if F or H are thought of as unknown parameters (as is often the case in practice), then their complete conditional distributions will also be required in order to implement the Gibbs sampler (see Example 4.1 below). □

Example 2.2: Nonlinear model. We now determine the distributions $x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y}$ for model (2), the nonlinearity presenting no further complications in the remaining complete conditional distributions. We consider separately the three cases where nonlinearity occurs in the state equation, the observation equation, or both.

First, suppose that $h_t(x_t) = H_t x_t$, but the state equation is nonlinear. Then $x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y}$ $\propto w_1(x_t) N_p(B_{1t} b_{1t}, B_{1t})$ where

$$B_{1t}^{-1} = \frac{\Sigma^{-1}}{\lambda_t} + \frac{H_t^T \Upsilon^{-1} H_t}{\omega_t} \quad , \qquad b_{1t}^T = \frac{f_t(x_{t-1})^T \Sigma^{-1}}{\lambda_t} + \frac{y_t^T \Upsilon^{-1} H_t}{\omega_t} \tag{10}$$

and $w_1(x_t) = \exp\left(-\frac{1}{2\lambda_{t+1}}(x_{t+1} - f_t(x_t))^T \Sigma^{-1} (x_{t+1} - f_t(x_t))\right)$. But clearly $0 \leq w_1(x_t) \leq 1$ for all x_t, and so the distribution from which we want to sample is dominated by the $N(B_{1t} b_{1t}, B_{1t})$ density. Hence, we may use rejection sampling (see for example Devroye, 1986, section II.3) in order to obtain a random observation from the required complete conditional. That is, we sample an observation x_t from a $N(B_{1t} b_{1t}, B_{1t})$ density, and subsequently accept it with probability $w_1(x_t)$.

Of course, this algorithm may be rather inefficient if the $w_1(x_t)$ are close to 0; in such cases, more sophisticated envelope functions may be needed (see for example Gilks and Wild, 1991, or Carlin and Gelfand, 1992). Such envelope functions are often normal or t densities chosen to be as similar to the desired complete conditional as possible, thus enabling more efficient rejection sampling. However, the experimenter needs to take care that such an envelope function does in

fact "blanket" the complete conditional distribution for all x_t. Uncertainty about this condition, along with the need to recompute the envelope at each iteration and each replication of the Gibbs sampler (since the values we condition on change as the algorithm progesses) make such approaches unattractive unless the naive method described above is prohibitively slow.

Secondly, suppose that $f_t(x_{t-1}) = F_t x_{t-1}$, but that now the observational equation is nonlinear. Then $x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y} \propto w_2(x_t) N_p(B_{2t} b_{2t}, B_{2t})$ where

$$B_{2t}^{-1} = \frac{\Sigma^{-1}}{\lambda_t} + \frac{F_{t+1}^T \Sigma^{-1} F_{t+1}}{\lambda_{t+1}} \quad , \qquad b_{2t}^T = \frac{x_{t-1}^T F_t^T \Sigma^{-1}}{\lambda_t} + \frac{x_{t+1}^T \Sigma^{-1} F_{t+1}}{\lambda_{t+1}}$$

and $w_2(x_t) = \exp\left(-\frac{1}{2\omega_t}(y_t - h_t(x_t))^T \Upsilon^{-1}(y_t - h_t(x_t))\right)$, and again rejection may be employed. Finally, when both components are nonlinear, $x_t|x_{j\neq t}, \lambda, \omega, \Sigma, \Upsilon, \mathbf{y} \propto w_1(x_t) w_2(x_t) N_p(f_t(x_{t-1}), \lambda_t \Sigma)$. Thus we sample a $N_p(f_t(x_{t-1}), \lambda_t \Sigma)$ random variable and accept it with probability $w_1(x_t) w_2(x_t)$. □

3. Estimated marginal posterior densities

With all the complete conditionals available for sampling, it now remains to show how to estimate the marginal posterior densities of the quantities of interest using the generated Gibbs samples. If we denote this collection by $\{(x_t^{(g)}, t = 0, \ldots, n), (\lambda_t^{(g)}, t = 1, \ldots, n), (\omega_t^{(g)}, t = 1, \ldots, n), \Sigma^{(g)}, \Upsilon^{(g)}, g = 1, \ldots, G\}$, then we may use (6) to obtain

$$\hat{p}(x_t|\mathbf{y}) = \frac{1}{G} \sum_{g=1}^{G} p(x_t|x_{t-1}^{(g)}, x_{t+1}^{(g)}, \lambda_t^{(g)}, \lambda_{t+1}^{(g)}, \omega_t^{(g)}, \Sigma^{(g)}, \Upsilon^{(g)}, y_t). \tag{11}$$

Note that this of course assumes that the x_t complete conditional distribution is available in closed form. If this is not the case (as in Example 2.2 above), an alternative would be to simply compute a kernel density estimate using the $\{x_t^{(g)}\}$ samples themselves. Another approach would be to obtain the G standardizing constants necessary in equation (11) by univariate numerical integration,

perhaps a simple trapezoidal rule. While a bit more work, this latter approach would almost certainly produce a better density estimate as it does not discard the functional form used to obtain the $\{x_i^{(g)}\}$ iterates.

We note that equation (11) could be used to obtain a marginal posterior density estimate for x_{n+1} provided y_{n+1} was available, offering a solution to the so-called filtering problem. If y_{n+1} is not yet available, the problem becomes one of one-step ahead prediction, and can be solved by a slight modification of the Gibbs algorithm. In fact, the k-step ahead prediction problem can be easily handled as follows: Suppose we desire an estimate of $p(x_{n+k}|\mathbf{y})$ where again $\mathbf{y} = (y_1, \ldots, y_n)$ and $y_{n+1}, \ldots, y_{n+k}$ have not yet been observed. We simply add $\{x_{n+1}, \ldots, x_{n+k}, y_{n+1}, \ldots, y_{n+k}, \lambda_{n+1}, \ldots, \lambda_{n+k}, \omega_{n+1}, \ldots, \omega_{n+k}\}$ to the Gibbs sampler as $4k$ additional unknown parameters. The complete conditional distributions for the new x's are again obtained using the lemma in section 2 above, where now of course the upper "endpoint" condition pertains to x_{n+k} instead of x_n. Similarly, the complete conditionals for the new λ's and ω's arise in a manner exactly analogous to that described in section 2. Finally, the complete conditional distributions for the new y variables come directly from the model specification, namely

$$y_{n+t}|\{x_i, \lambda_i, \omega_i\}_{i=1}^{n+k}, x_0, \Sigma, \Upsilon, \mathbf{y} \sim y_{n+t}|x_{n+t}, \omega_{n+t}, \Upsilon \sim N(h_{n+t}(x_{n+t}), \omega_{n+t}\Upsilon) \ , \ t = 1, \ldots, k \ .$$

We now simply run the Gibbs sampler as usual, obtaining for any $i \in \{1, \ldots, k\}$ the slightly modified version of (11),

$$\hat{p}(x_{n+i}|\mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G} p(x_{n+i}|x_{n+i-1}^{(g)}, x_{n+i+1}^{(g)}, \lambda_{n+i}^{(g)}, \omega_{n+i}^{(g)}, \Sigma^{(g)}, \Upsilon^{(g)}, y_{n+i}^{(g)}), \tag{12}$$

the primary difference being the dependence on the generated "data" values $\{y_{n+i}^{(g)}, g = 1, \ldots, G\}$ rather than an observed value y_{n+i}. Of course, as these future y_{n+i} values become available, we

simply use these values in lieu of sampled values $y_{n+i}^{(g)}$ and rerun the algorithm – a computationally simple solution to the filtering problem. Examples 4.2 and 4.3 offer an illustration of this process.

4. Numerical Examples

Example 4.1: Univariate linear model. Consider again the model presented in Example 2.1. We apply this model to the aforementioned data displayed in Figure 1(a). Recall that these data are final NCHS estimates of the numbers of deaths due to ischemic heart disease per 100.000 men aged 25-34. The counts are given monthly over the period from January 1979 to December 1986. As this series appears nonstationary, a standard time series analysis would likely fit an ARIMA model using normal errors to the differenced series, which is plotted as asterisks versus time in Figure 2. Fitting a standard AR(1) model using the statistical package MINITAB (1989), we obtained an autoregressive parameter estimate of -0.58, with a standard error of 0.09. MINITAB also produces predicted values for the time series, which are plotted as a solid line in Figure 2. While this one parameter model is quite simple, the predicted values it produces can be rather poor (see those in the vicinity of month 60, for example). We shall compare the fit of this model with that of our state space model.

We assume that the estimates y_t in the data are unbiased for the true monthly death rates x_t, and thus set $H = 1$. The plot of the differenced data suggest that the simple exponential model given by $F_t = F$ is not unreasonable; however, we wish to treat F as an unknown parameter. This can be easily incorporated into the framework developed in section 2 by assuming that $F \sim N(\mu_F, \sigma_F^2)$ and noting that the complete conditional is given by $F|\boldsymbol{\lambda}, \omega, \sigma, \tau, \mathbf{x}, \mathbf{y} \sim N(B_F b_F, B_F)$, where

$$B_F^{-1} = \frac{1}{\sigma^2}\sum_{t=1}^{n}\frac{x_{t-1}^2}{\lambda_t} + \frac{1}{\sigma_F^2}, \quad \text{and} \quad b_F = \frac{1}{\sigma^2}\sum_{t=1}^{n}\frac{x_t x_{t-1}}{\lambda_t} + \frac{\mu_F}{\sigma_F^2}.$$

Looking again at Figure 2, note that a change in the variance (heteroscedasticity of errors) seems

to take place near month 25, and again around month 65. While this might suggest including a separate σ_t^2 or τ_t^2 parameter for each time point t, such overparametrization seems excessive for these data. Thus, as in section 2, we assume $\sigma_t^2 = \sigma^2$ and $\tau_t^2 = \tau^2$ for all t. Placing independent inverse gamma priors with parameters (a_0, b_0) and (c_0, d_0) on σ^2 and τ^2, respectively, their complete conditional distributions are again given by (8). Similarly, the complete conditionals for $x_t, t = 0, \ldots, n$ are the same as those in Example 2.1 above (recall we employ a $N(\mu_0, \sigma_0^2)$ prior on x_0). The densities $p(x_t|\mathbf{y})$ may be estimated using equation (11) with the argument $F^{(g)}$ added to the list of conditioning arguments, since F is no longer known but instead a component of the sampler.

In order to demonstrate the approach to nonnormal error distributions, consider the two models $\mathcal{M}_1$ and $\mathcal{M}_2$ given by

$$\mathcal{M}_1 : u_t \sim N(0, \sigma^2), v_t \sim N(0, \tau^2), \text{ and } \mathcal{M}_2 : u_t \sim DE(0, \sigma), v_t \sim DE(0, \tau).$$

For $\mathcal{M}_1$, we take $\lambda_t = \omega_t = 1$ with probability one for all $t = 1, \ldots, n$, leading to complete conditional distributions for λ_t and ω_t which are also degenerate at the value 1. For $\mathcal{M}_2$, we take both the λ_t and ω_t to be independently distributed *a priori* as $Expo(2)$ random variables, leading to the complete conditionals

$$\lambda_t \sim GIG\left(\frac{1}{2}, 1, \left(\frac{x_t - Fx_{t-1}}{\sigma}\right)^2\right), \text{ and } \omega_t \sim GIG\left(\frac{1}{2}, 1, \left(\frac{y_t - x_t}{\tau}\right)^2\right),$$

in a manner similar to that surrounding equation (9). We complete the specification of the prior on x_0 by setting $\mu_0 = 0.0$ and $\sigma_0 = 0.2$ (the differenced series should be centered close to 0), and $\mu_F = -0.5$ and $\sigma_F = 0.25$ (past experience with data of this type suggests that F should be negative with high probability). We take a rather vague prior on σ^2, having both prior mean and prior standard deviation equal to $(10)^2$ (i.e., $a_0 = 3$, $b_0 = 0.005$). Finally, we take an informative

prior on τ^2, having mean and standard deviation equal to $(0.1)^2$ (i.e., $c_0 = 3$, $d_0 = 50$). The resulting low variability in the observational equation will make our results more comparable to those obtained from a standard AR(1) model.

For our analysis, we ran the Gibbs sampler for $l = 50$ iterations on each model separately, obtaining the two model-specific density estimates $\hat{p}(F|\mathbf{y}, \mathcal{M}_i)$ shown in Figure 1(c). In each case, our algorithm used $G = 500$ parallel replications per iteration, and convergence was judged both by monitoring sample moments of the Gibbs values themselves and by plotting successive density estimates for the inflation constant F. We see that the normal errors assumption produces a posterior distribution for F which is slightly more variable, and generally suggests slightly smaller values for F. The mode of the DE errors model posterior distribution, -0.425, provides a point estimate of F; the point estimate for the normal errors model is slightly more negative, in line with the AR(1) estimate of -0.58. Notice that there is a reasonable amount of uncertainty associated with our F estimate: Figure 1(c) shows that values between -1 and 0 remain plausible in light of the data. This of course is consistent with the large amount of noise evident in the data (Figure 2). A fully Bayesian approach would involve obtaining estimates of the posterior probabilities $p(\mathcal{M}_i|\mathbf{y})$, $i = 1, 2$, leading to a Bayes factor between the normal and DE models; a Gibbs sampling approach useful in choosing amongst competing error distributions is discussed by Carlin and Polson (1991). Overall, the preliminary results obtained here indicate that the assumption of normal errors is not a grossly misleading one.

Calculations similar to those undertaken for F could also be performed for all of the x_t states. In particular, since in this case the x_t complete conditionals are available in closed form, equation (11) could be used to obtain point estimates and credible sets for each x_t. However, rough point and interval estimates for any parameter θ may be obtained by simply taking appropriate functions or quantiles of the $\{\theta^{(g)}, g = 1, \ldots, G\}$ iterates themselves. For example, a point estimate of x_t is

given by $\sum_{g=1}^{G} x_t^{(g)}/G$, and the .025 and .975 empirical percentiles of the $x_t^{(g)}$ distribution provide a 95% interval estimate for x_t. These estimates are plotted as dashed lines in Figure 2. The posterior means are extremely close to the observed y_t values; the sum of the squared discrepancies (residuals) between the two is only 0.002, as compared with 0.330 for the AR(1) model fit above. Of course, this is not really a "fair" comparison since the state space model has many more parameters – our main point here is to illustrate that predicted values and confidence limits are produced automatically as a by-product of the algorithm. In fact, since all the random generation is "one-for-one" (no rejection algorithms are needed), the algorithm is surprisingly fast, a typical run taking no more than 10 minutes using FORTRAN on a DECStation 3100.

Example 4.2: Bivariate linear model. In monitoring numbers of deaths by sex and cause over time, the NCHS actually first assembles preliminary rate estimates $y_t^{(p)}$, and later revises them into final estimates $y_t^{(f)}$. The final estimates are based on a much larger sample of individuals, and as such are much more precise than the preliminary ones as well as being less variable across t. But of course the preliminary estimates are not devoid of information about the true underlying rate states x_t, and so we might seek methodology for combining these two sources of information. The general linear framework outlined in Section 2 above offers such a methodology: we simply take $p = 1$ and $q = 2$ in model (1), defining the vector $y_t = (y_t^{(p)}, y_t^{(f)})^T$ and taking $H_t = (1, 1)^T$ (i.e., we assume that both the preliminary and final estimates are unbiased for x_t). The true state x_t is still univariate, so we again have $\Sigma_0 = \sigma_0^2$ and $\Sigma = \sigma^2$. Mathematically we would like to assume the y_t covariance matrix, Υ, to be diagonal, but this is perhaps unjustified in that the preliminary data is included when compiling the final estimates, leading to dependence between $y_t^{(p)}$ and $y_t^{(f)}$.

As an concrete illustration, consider the septicemia mortality rate data for mean aged 75-84 which is plotted versus time in Figure 3. A steadily increasing pattern in apparent, again meaning that a standard ARIMA analysis would require differencing to produce stationarity before modeling

could begin. However, our state space approach enables direct modeling of the mean structure on the original scale, rather than removing all such structure by differencing. In this way we can preserve the interpretability of our results. Again the simple linear growth model $F_t = F$ seems justified, and for simplicity we adopt the normal errors model ($\lambda = \omega = \mathbf{0}$).

We now list the complete conditional distributions necessary for implementation of the Gibbs sampler in this case. First for x_t, using the results of the Lemma in Section 2 and taking care with the endpoint cases we have that $x_t | x_{j \neq t}, y_t^*, \sigma^2, \tau^2 \sim N(B_t \mathbf{b}_t, B_t)$, where

$$\begin{array}{lll} B_t^{-1} = \sigma_0^{-2} + F^2\sigma^{-2}, & \mathbf{b}_t = \mu_0\sigma_0^{-2} + F\sigma^{-2}x_1, & t = 0 \\ B_t^{-1} = \sigma^{-2}(1+F^2) + \mathbf{1}^T\Upsilon^{-1}\mathbf{1}, & \mathbf{b}_t = F\sigma^{-2}(x_{t-1}+x_{t+1}) + y_t^T\Upsilon^{-1}\mathbf{1}, & t = 1,\ldots,n-1 \\ B_t^{-1} = \sigma^{-2} + \mathbf{1}^T\Upsilon^{-1}\mathbf{1}, & \mathbf{b}_t = F\sigma^{-2}x_{n-1} + y_n^T\Upsilon^{-1}\mathbf{1}, & t = n \end{array} \quad , \quad (13)$$

where $\mathbf{1}$ denotes a 2×1 vector of 1's. Next, for σ^2 and Υ we assume the independent *a priori* specifications $\sigma^2 \sim IG(a_0, b_0)$ and $\Upsilon^{-1} \sim W((\rho R)^{-1}, \rho)$, where W denotes the Wishart distribution, the usual conjugate prior distribution for covariance matrices (see Box and Tiao, 1973, p. 427). Standard calculations again lead to the complete conditionals

$$\sigma^2 | \mathbf{x}, x_0 \sim IG\left(a_0 + \tfrac{n}{2}, \left\{\tfrac{1}{b_0} + \tfrac{1}{2}\textstyle\sum_{t=1}^n (x_t - Fx_{t-1})^2\right\}^{-1}\right) \quad , \text{ and}$$

$$\Upsilon^{-1} | \mathbf{y}, \mathbf{x} \sim W\left\{\left[\textstyle\sum_{t=1}^n (y_t - x_t\mathbf{1})(y_t - x_t\mathbf{1})^T + \rho R\right]^{-1} \quad , n + \rho\right\} \quad .$$

Generation from the Wishart distribution may be accomplished via an algorithm due to Odell and Fieveson (1966); this algorithm is outlined in the 2×2 case by Gelfand et al. (1990). Finally, we again wish to treat F as an unknown parameter, assigning it the $N(\mu_F, \sigma_F^2)$ prior distribution. The complete conditional for F is thus of the same form as given in Example 4.1 above.

For the values of the model hyperparameters, we simply chose $\mu_0 = 4$ and $\sigma_0^2 = 1$, corresponding

very roughly to the observed level of the y_t series at $t = 0$. We also took $a_0 = 3$ and $b_0 = 2$, implying a prior mean and standard deviation of $(0.5)^2$ for σ^2. Since the matrix R^{-1} is the prior expected value of Υ^{-1}, and the preliminary estimates $y_t^{(p)}$ are clearly more variable than the final ones $y_t^{(f)}$, we chose $R = Diag(2^2, (0.5)^2)$ (roughly the prior mean of Υ). While this implies the oversimplification that the two elements of y_t are uncorrelated, we shall adopt the small prior precision value of $\rho = 2$, resulting in a vague prior which will allow the data to suggest the correct value of $Corr(Y_t^{(p)}, Y_t^{(f)})$. Finally, we chose $\mu_F = 1.01$ and $\sigma_F^2 = 0.05$, indicating a very vague belief in a one percent monthly upward drift for our series.

Running $G = 500$ parallel replications of the Gibbs sampler for $l = 50$ iterations each, we obtained Gibbs iterates $\{x_{tj}, j = 1, \ldots, G\}$. As in the previous example, we obtained simple point estimates of the x_t posterior means as averages of these iterates, and a 95% posterior credible set for x_t via the .025 and .975 empirical percentiles of the x_{tj} distributions. These point and interval estimates are plotted in Figure 3. Notice that our model has had the desired smoothing effect throughout the series. The estimated posterior mean of Υ is given by

$$\hat{E}[\Upsilon|\mathbf{y}] = \begin{pmatrix} 13.25 & 2.73 \\ 2.73 & 2.86 \end{pmatrix},$$

implying a modest posterior correlation between the preliminary and final estimates of 0.44.

In order to illustrate the solution to the one-step-ahead prediction problem (i.e., find the marginal posterior density of x_{n+1} given $\mathbf{y}$), suppose that up until now we have observed only the first 83 y_t values in Figure 3, so that $\mathbf{y} = (y_1, \ldots, y_{83})^T$. Now notice that the nonappearance of y_{84} and x_{85} in the likelihood (3) means that a Monte Carlo mixture density estimate is available as

$$\hat{p}(x_{84}|\mathbf{y}) = \frac{1}{G}\sum_{j=1}^{G} N\left(F_j x_{(83,j)}\ ,\ \sigma_j^2\right).$$

This estimate is plotted as the solid line in Figure 4(b). Now suppose that y_{84} has become available (in our dataset, $y_{84} = (8.994, 8.443)^T$), and so we may wish to solve the filtering problem (i.e., find the marginal posterior density of x_{84} given $\mathbf{y}$ *and* y_{84}). This is easily done in our context by including x_{84} as an additional parameter in the sampling order and rerunning the algorithm. With Gibbs samples for this parameter now available, we can obtain a mixture density estimate simply by mixing the appropriate complete conditional distributions from equation (13), namely

$$\hat{p}(x_{84}|\mathbf{y}, y_{84}) = \frac{1}{G}\sum_{j=1}^{G} N\left(B_{(84,j)}\mathbf{b}_{(84,j)},\ B_{(84,j)}\right),$$

where $B^{-1}_{(84,j)} = \sigma_j^{-2} + \mathbf{1}^T\Upsilon_j^{-1}\mathbf{1}$ and $\mathbf{b}_{(84,j)} = F_j\sigma_j^{-2}x_{(83,j)} + y^T_{(84,j)}\Upsilon_j^{-1}\mathbf{1}$. This estimate is plotted as the dashed line in Figure 4(b). Notice that the filtering posterior is less variable (due to the addition of the new information) and centered at a slightly higher value (closer to the observed values of $y_t^{(p)}$ and $y_t^{(f)}$).

In a simliar vein, Figure 4(a) plots two estimated marginal posteriors for F, one conditional on $\mathbf{y}$ ("prediction") and the other conditional on $\mathbf{y}$ and y_{84} ("filtering"). Unlike the plots for x_{84}, since all the y_t's have a direct impact on the posterior for F, we would expect it to be only slightly less diffuse for filtering than for prediction. This is indeed what we see in Figure 4(a).

Example 4.3: Univariate nonstationary growth model. The y and x values displayed as solid lines in Figure 5 were generated according to the model

$$\begin{aligned} x_t &= \alpha x_{t-1} + \beta x_{t-1}/(1 + x_{t-1}^2) + \gamma\cos(1.2(t-1)) + u_t, \qquad \text{and} \\ y_t &= x_t^2/20 + v_t, \qquad\qquad t = 1, \ldots, 100 \end{aligned} \tag{14}$$

where $x_0 \equiv 0$, the u_t are independent random variables having a t-distribution with $\nu = 10$ degrees of freedom, mean zero, and variance 10, and the v_t are distributed as $N(0,1)$ random variables

independent of the u_t, $t = 1, \ldots, 100$. In the rejoinder to his paper, Kitagawa (1987) fit a non-Gaussian filter and smoother to data generated from this model where the u_t and v_t were both Gaussian white noise sequences with these same means and variances, and the values $\alpha = .5, \beta = 25$, and $\gamma = 8$ assumed known. We shall use these values for α, β and γ in our study but assume they are unknown to the experimenter, and obtain marginal posterior densities for all three. In addition, we shall obtain an estimate of $p(x_{101}|\mathbf{y})$, the density of the one-step ahead predicted state.

To implement the Gibbs sampler we follow the model outlined in Example 2.2, where $p = q = 1$. We again assume $\sigma^2 \sim IG(a_0, b_0)$ and $\tau^2 \sim IG(c_0, d_0)$), which again leads to inverse gamma complete conditionals of a form similar to that given in equation (8). Next, by letting $\nu/\lambda_t \sim \chi^2_\nu$, we get that marginally, $u_t|\sigma \sim t(0, \sigma, \nu)$ as required, leading to the complete conditional $\lambda_t|\sigma, \alpha, \beta, \gamma, \mathbf{y}, \mathbf{x}, x_0$ being distributed as

$$IG\left(\frac{\nu+1}{2}\,,\, 2\left\{\frac{[x_t - \alpha x_{t-1} - \beta x_{t-1}/(1+x_{t-1}^2) - \gamma\cos(1.2(t-1))]^2}{\sigma^2} + \nu\right\}^{-1}\right),\ t = 1, \ldots, 101.$$

Since we are assuming the observation noise to be Gaussian, we may take $\omega_t \equiv 1$, $t = 1, \ldots, 101$. Turning to the x_t complete conditionals and again making the prior assumption $x_0 \sim N(\mu_0, \sigma_0^2)$, we note that the nonlinear structure in both the state and observational equations precludes closed form complete conditionals, but we may use the rejection algorithm discussed in Example 2.2 to generate the necessary samples. That is, we generate x_t from a $N(\alpha x_{t-1} + \beta x_{t-1}/(1+x_{t-1}^2) + \gamma\cos(1.2(t-1))\,,\, \lambda_t\sigma^2)$ distribution and accept it with probability $w_1(x_t)w_2(x_t)$, where $w_1(x_t) = \exp\left\{\frac{1}{2\lambda_{t+1}\sigma^2}(x_{t+1} - \alpha x_t + \beta x_t/(1+x_t^2) + \gamma\cos(1.2t))^2\right\}$ and $w_2(x_t) = \exp\left\{\frac{1}{2\omega_t\tau^2}(y_t - x_t^2/20)^2\right\}$ for $t = 1, \ldots, 100$. For $t = 0$, we generate $x_t \sim N(\mu_0, \sigma_0^2)$ and accept with probability $w_1(x_t)$; for $t = 101$ we generate x_t as usual but accept with probability $w_2(x_t)$. Note that this last complete conditional depends on y_{101}, a "data" value which is not observed but instead generated according

to *its* complete conditional distribution, which of course is $N(x_{101}^2/20\ ,\ \omega_t\tau^2)$.

Finally, for the prior on the state equation model parameters we suppose that $(\alpha,\beta,\gamma)^T \sim N_3((\mu_\alpha,\mu_\beta,\mu_\gamma)^T, V)$ where $V = Diag(\sigma_\alpha^2,\sigma_\beta^2,\sigma_\gamma^2)$. This enables complete conditionals of the form $N(Bb, B)$, where for α,

$$B^{-1} = \frac{1}{\sigma_\alpha^2} + \frac{1}{\sigma^2}\sum_{t=1}^{101}\frac{x_{t-1}^2}{\lambda_t}\ ,\ \text{and}\ b = \frac{\mu_\alpha}{\sigma_\alpha^2} + \frac{1}{\sigma^2}\sum_{t=1}^{101}\frac{x_{t-1}}{\lambda_t}\left(x_t - \beta\frac{x_{t-1}}{1+x_{t-1}^2} - \gamma\cos(1.2(t-1))\right)\ ,$$

whilst for β,

$$B^{-1} = \frac{1}{\sigma_\beta^2}+\frac{1}{\sigma^2}\sum_{t=1}^{101}\frac{x_{t-1}^2}{\lambda_t(1+x_{t-1}^2)^2}\ ,\ \text{and}\ b = \frac{\mu_\beta}{\sigma_\beta^2}+\frac{1}{\sigma^2}\sum_{t=1}^{101}\frac{x_{t-1}}{\lambda_t(1+x_{t-1}^2)}\left[x_t - \alpha x_{t-1} - \gamma\cos(1.2(t-1))\right]\ ,$$

and finally for γ,

$$B^{-1} = \frac{1}{\sigma_\gamma^2}+\frac{1}{\sigma^2}\sum_{t=1}^{101}\frac{\cos^2(1.2(t-1))}{\lambda_t}\ ,\ \text{and}\ b = \frac{\mu_\gamma}{\sigma_\gamma^2}+\frac{1}{\sigma^2}\sum_{t=1}^{101}\frac{\cos(1.2(t-1))}{\lambda_t}\left(x_t - \alpha x_{t-1} - \beta\frac{x_{t-1}}{1+x_{t-1}^2}\right)\ .$$

For this example we took $\mu_0 = 0$ and $\sigma_0^2 = 10$, $a_0 = 3$ and $b_0 = .05$ (so that the prior on σ^2 has mean and standard deviation equal to 10), and $c_0 = 3$ and $d_0 = .5$ (so that the prior on τ^2 has mean and standard deviation equal to 1). We also chose $\mu_\alpha = .5, \mu_\beta = 25, \mu_\gamma = 8, \sigma_\alpha = .25, \sigma_\beta = 10$, and $\sigma_\gamma = 4$. We then ran the Gibbs sampler for $l = 50$ iterations, using $G = 500$ parallel replications per iteration. The generation cycle in this case involves updating $3(101) + 7 = 310$ parameters per iteration, 102 of which (the x's) must be sampled via rejection, thus substantially adding to the computational burden; however, programming effort is still quite minimal. Figures 6 (a) - (c) show the resulting marginal posterior density estimates of the form given in (6) for α, β, and γ. Note that this estimation is quite unambiguous, the posteriors being centered nearly at the true parameter values and fairly tightly concentrated. To compute the marginal posterior density of

x_{101}, we could use equation (12) with $n = 100$ and $i = 1$, but the nonappearance of y_{101} and x_{102} in the likelihood implies that we may take advantage of the simplified conditional density given in (5), obtaining the estimated density as

$$\hat{p}(x_{101}|\mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G} N\left(\alpha^{(g)}x_{100}^{(g)} + \beta^{(g)}x_{100}^{(g)}/(1+(x_{100}^{(g)})^2) + \gamma^{(g)}\cos(120)\,,\ \lambda_{101}(\sigma^{(g)})^2\right)(x_{101})\,. \quad (15)$$

This estimate is given in Figure 7(a). Nervous about the validity of the bimodal shape of this posterior, we constructed a histogram of the actual Gibbs values $\{x_{101}^{(g)}, g = 1,\ldots,G\}$, shown in Figure 7(b), which also supports a bimodal shape. Looking again at the pattern of the true x values in Figure 5(b), the reason for the bimodality becomes apparent: the system is currently near the zero point, and is likely to drop back down into the negative realm, as it has done most recently. However, there is a substantial probability that the system will now return to the positive realm, explaining the second "bump."

Curious about the effect knowledge of y_{101} would have on the posterior for x_{101}, we repeated the above analysis using the observed value $y_{101} = 4.55$. In computing the marginal posterior for x_{101}, we are now solving the filtering problem. The addition of y_{101} to the likelihood means obtaining this marginal posterior by simple mixing as in equation (15) is no longer available, and we must resort to mixing the full posteriors as in equation (12). The normalization constants needed for each term of this sum were computed using a trapezoidal approximation. Figures 8(a) and (b) show the resulting estimated posterior and actual Gibbs samples, respectively, from running $l = 50$ iterations of $G = 2500$ replications each (the larger G being required to obtain the same level of accuracy with the more complicated density estimation procedure). We see that the bimodal shape observed in Figure 7 has become more exaggerated, the additional information provided by y_{101} leading to a tighter distribution for both modes. The peaks have also shifted to the left by roughly

5 units; interestingly, the true value $x_{101} = -9.05$ is very close to the location of the first mode ($x = -8.86$). The ability to effectively handle bimodalities is one of the features of Monte Carlo integration methods like the Gibbs sampler; analytic approximations such as Laplace's method (see Tierney and Kadane, 1986) are generally not recommended for use in such situations.

While calculations similar to those undertaken for x_{101} could also be performed for all of the remaining x_t states, to save time we simply calculate the rough x_t point estimates $\sum_{g=1}^{G} x_t^{(g)}/G$ mentioned earlier, and plot them as dashed lines in Figure 5(b). They perform surprisingly well, and on the whole seem quite competitive with those obtained by Kitagawa (1987, p. 1062), especially given our assumption of nonnormal errors in the state space and that $\alpha, \beta, \gamma, \sigma$ and τ were all unknown.

References

[1] ANDREWS, D.F. and MALLOWS, C.L. (1974). Scale mixtures of normality. *J.R. Statist. Soc., Ser. B*, **36**, 99-102.

[2] BARNDORFF-NEILSEN, O.E. and HALGREEN, C. (1977). Infinte divisibility of the hyperbolic and generalized inverse gaussian distributions. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **38**, 309-311.

[3] BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc., Ser. B*, **36**, 192-236.

[4] BOX, G.E.P. and JENKINS, G.M. (1976). *Time Series Analysis: Forecasting and Control*, revised edition. Oakland, CA: Holden-Day.

[5] BOX, G.E.P. and TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

[6] BROCKWELL, P.J. and DAVIS, R.A. (1987). *Time Series: Theory and Methods.* New York: Springer Verlag.

[7] CARLIN, B.P. and GELFAND, A.E. (1992). An iterative Monte Carlo method for nonconjugate Bayesian analysis. To appear *Statistics and Computing.*

[8] CARLIN, B.P. and POLSON, N.G. (1991). Inference for nonconjugate Bayesian models using the Gibbs sampler. To appear *Canadian Journal of Statistics.*

[9] CARLIN, B.P., POLSON, N.G., and STOFFER, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state space modeling. To appear *J. Amer. Statist. Assoc.*

[10] DEVROYE, L. (1986). *Non-uniform random variate generation.* New York: Springer-Verlag.

[11] GELFAND, A.E., HILLS, S.E., RACINE-POON, A., AND SMITH, A.F.M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *J. Amer. Statist. Assoc.*, **85**, 972-985.

[12] GELFAND, A.E. and SMITH, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398-409.

[13] GEMAN, S., and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

[14] GILKS, W.R. and WILD, P. (1991), Adaptive rejection sampling for Gibbs sampling. To appear *J.R. Statist. Soc., Ser. C (Applied Statistics).*

[15] KITAGAWA, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion). *J. Amer. Statist. Assoc.*, **82**, 1032-1064.

[16] MARTIN, R.D. and RAFTERY, A.E. (1987). Comment on "Non-Gaussian state-space modeling of nonstationary time series," by G. Kitagawa. *J. Amer. Statist. Assoc.*, **82**, 1044-50.

[17] METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H., and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chemical Physics*, **21**, 1087-1091.

[18] MILLER, R.B. and WICHERN, D.W. (1977). *Intermediate Business Statistics.* New York: Holt, Rinehart and Winston.

[19] *Minitab Reference Manual* (1989). State College, PA: Minitab, Inc.

[20] ODELL, P.L., AND FEIVESON, A.H. (1966). A numerical procedure to generate a sample covariance matrix. *J. Amer. Statist. Assoc.*, **61**, 198-203.

[21] SHUMWAY, R.H. and STOFFER, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Analysis*, **3**, 253-264.

[22] SPIEGELHALTER, D.J. and SMITH, A.F.M. (1982). Bayes factors for linear and log-linear models with vague prior information. *J. R. Statist. Soc., Ser. B*, **44**, 377-387.

[23] TANNER, M. A., and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.*, **82**, 528-550.

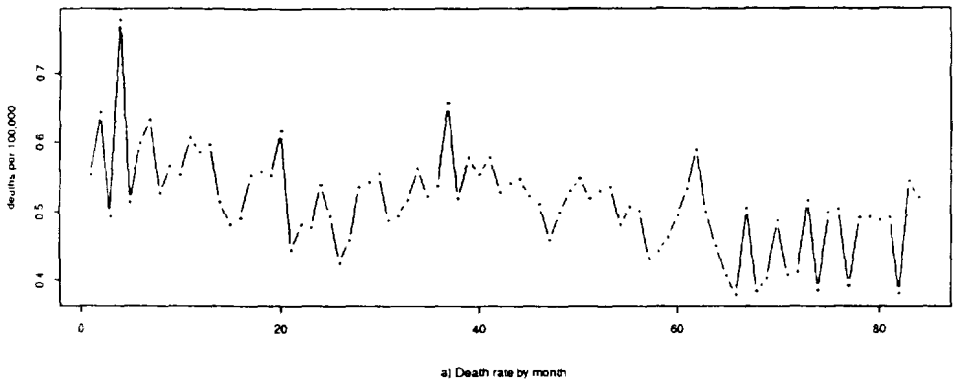
[24] TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, **81**, 82-86.

[25] WEST, M. (1987). On scale mixtures of normality. *Biometrika*, **74**, 694-697.

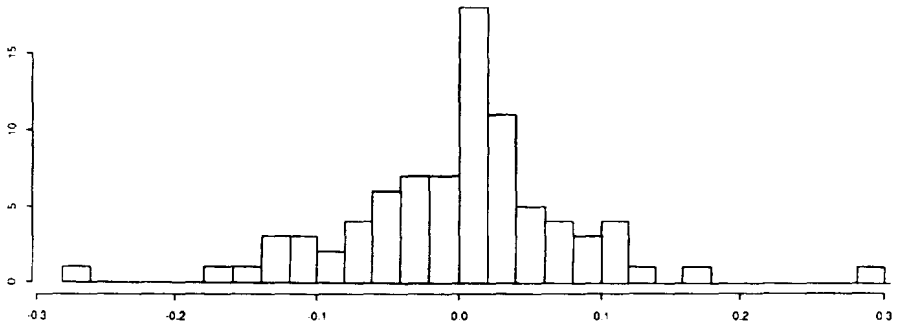
[26] WEST, M. and HARRISON, P.J. (1989). *Bayesian Forecasing and Dynamic Models.* New York: Springer-Verlag.

[27] WEST, M., HARRISON, P.J. and MIGON, H.S. (1985). Dynamic generalised linear models and Bayesian forecasting (with discussion). *J. Amer. Statist. Assoc.*, **80**, 73-97.

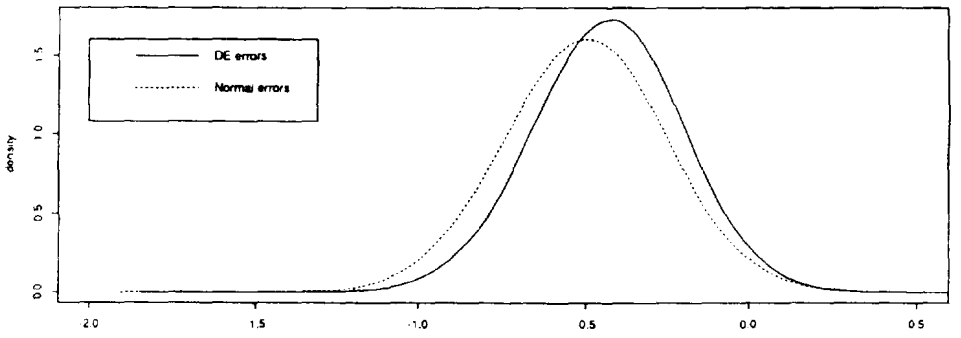
Figure 1. Ischemic heart disease data, men aged 25-34



a) Death rate by month



b) Histogram of differenced death rates



c) Marginal posteriors for F, G = 500 , mode for DE model = -0.425

Figure 2. Differenced ischemic heart disease data, men aged 25-34

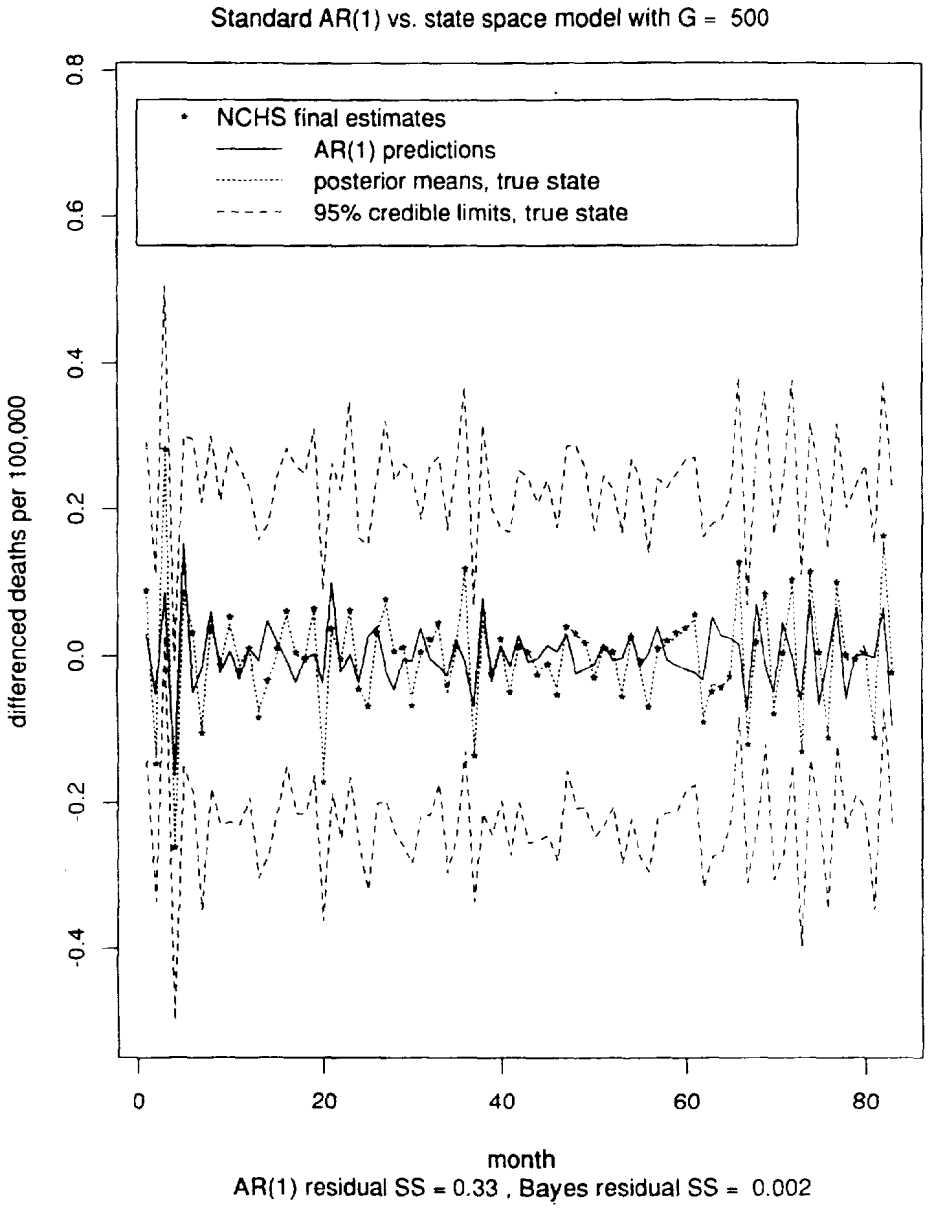


Figure 3. Septicemia data and estimates, men aged 75-84

G = 500 , y(84) assumed known

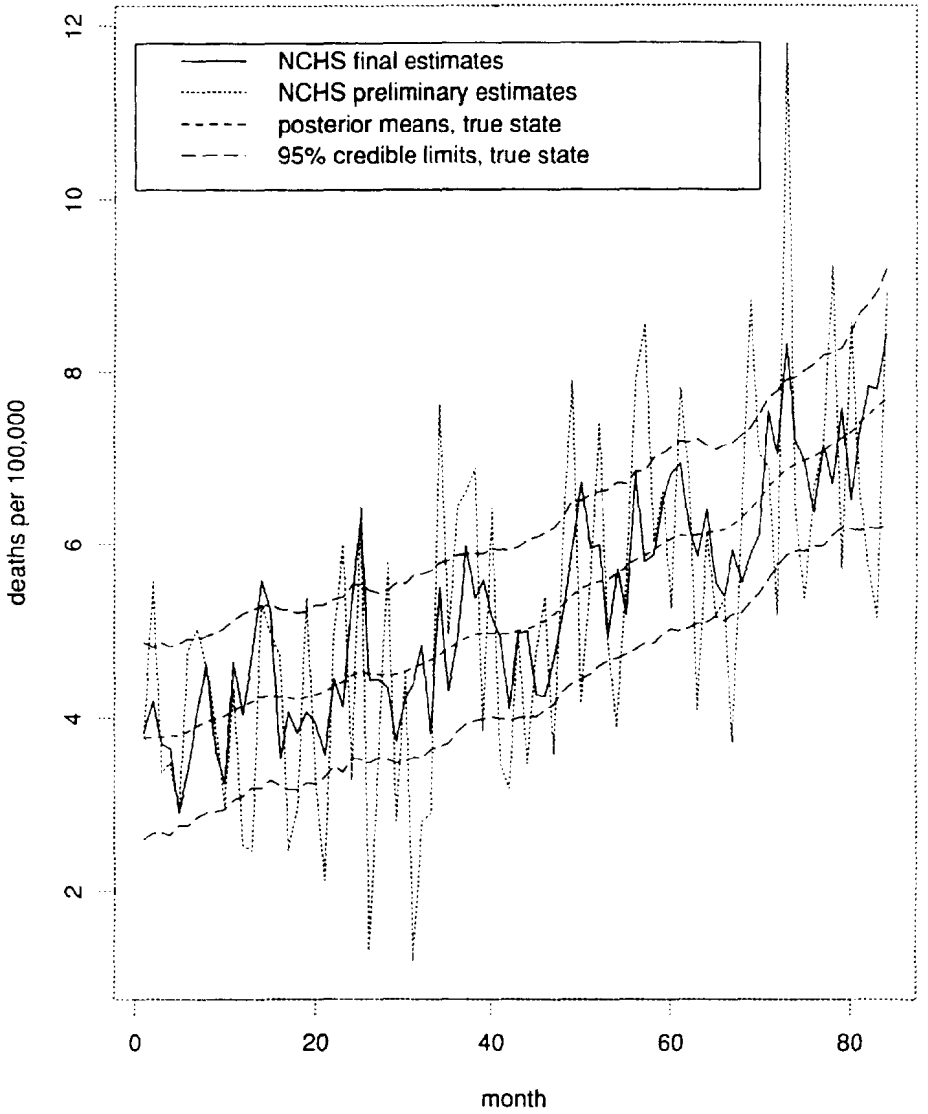
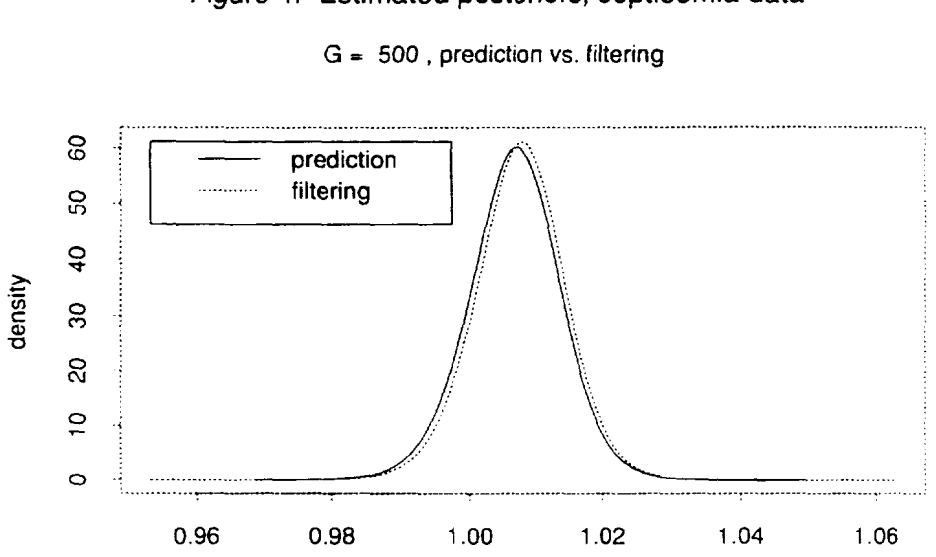
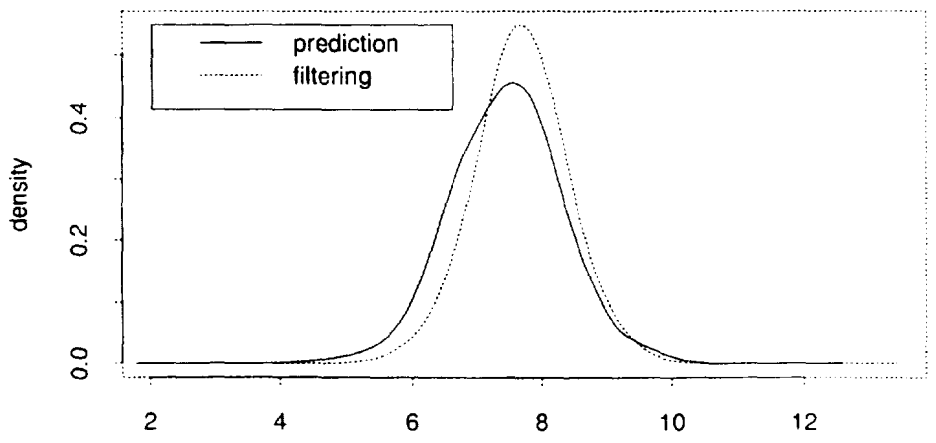


Figure 4. Estimated posteriors, septicemia data

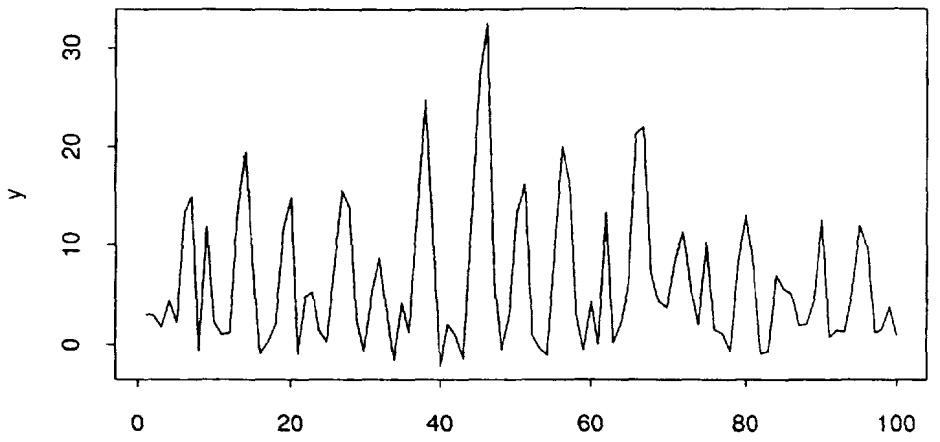


a) Marginal posterior for F; modes are 1.007 1.008

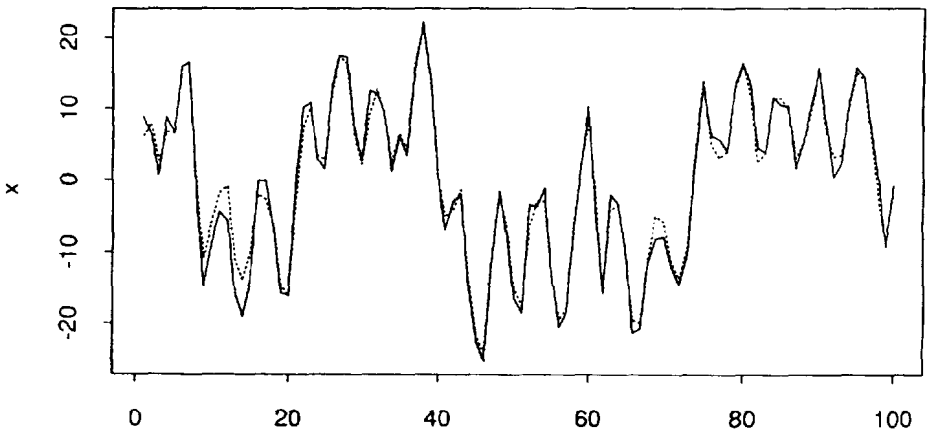


b) Marginal posterior for x(84); modes are 7.54 7.665

Figure 5. Data and estimates, Example 4.3

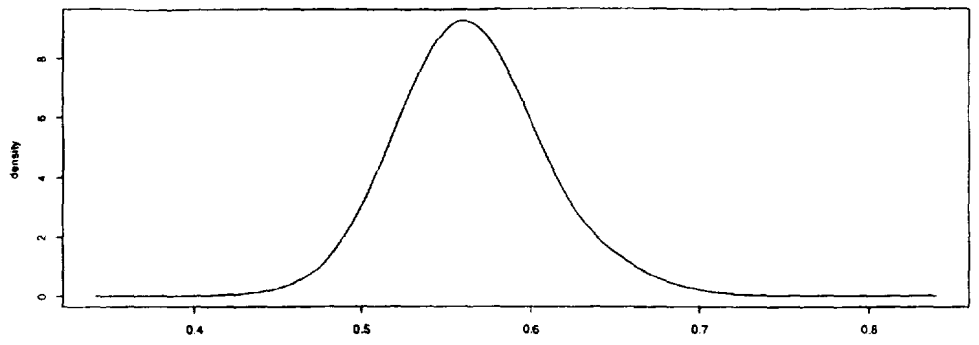


a) observed y values

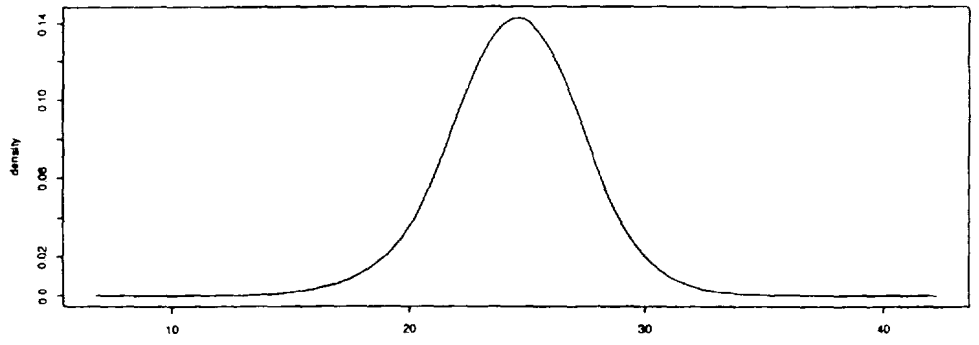


b) true x values (solid line) and point estimates (dashed line), y(101) unknown

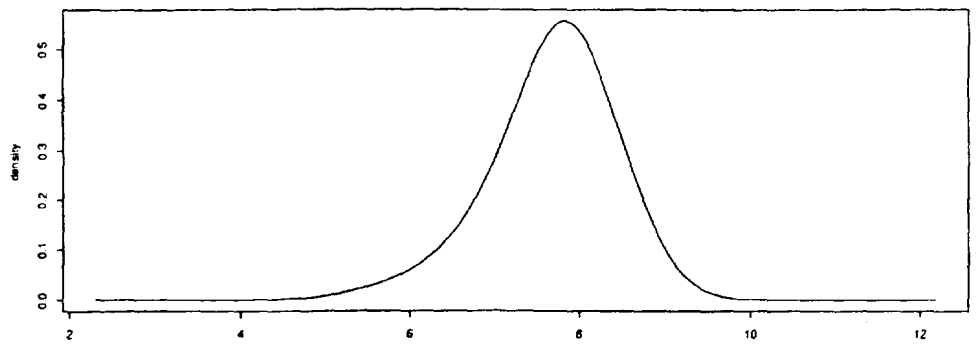
Figure 6. Estimated marginal posteriors, Example 4.3



a) Marginal posterior for alpha, G = 500 ; mode = 0.56

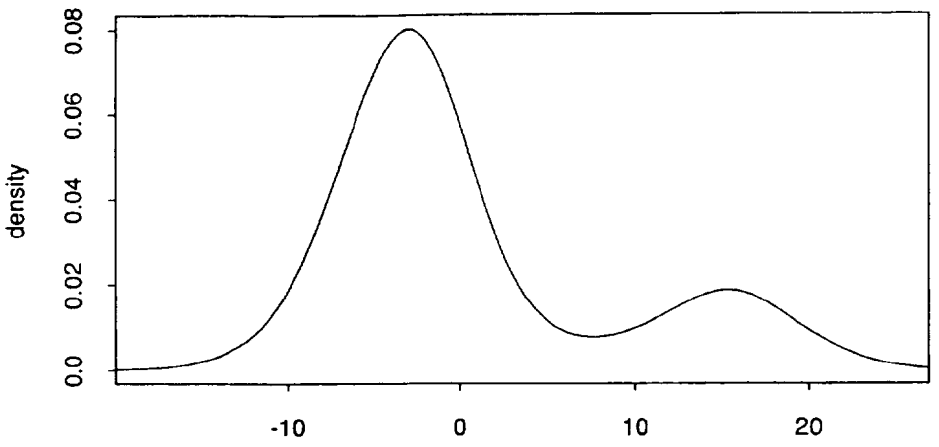


b) Marginal posterior for beta, G = 500 ; mode = 24.641

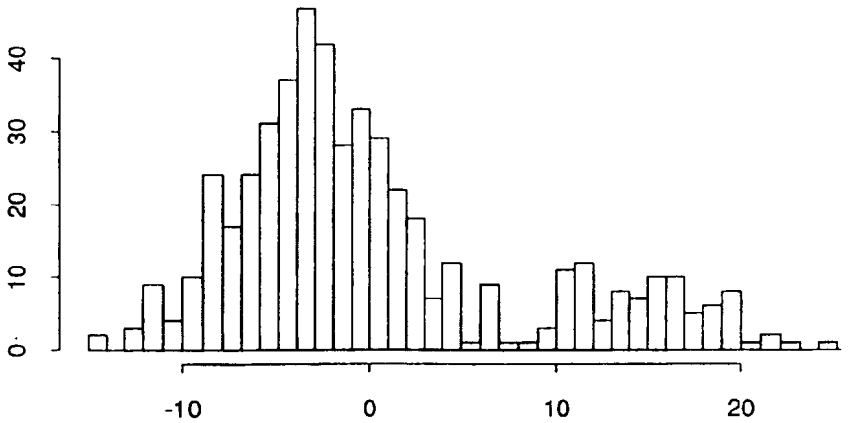


c) Marginal posterior for gamma, G = 500 ; mode = 7.826

Figure 7. One-step ahead prediction, Example 4.3

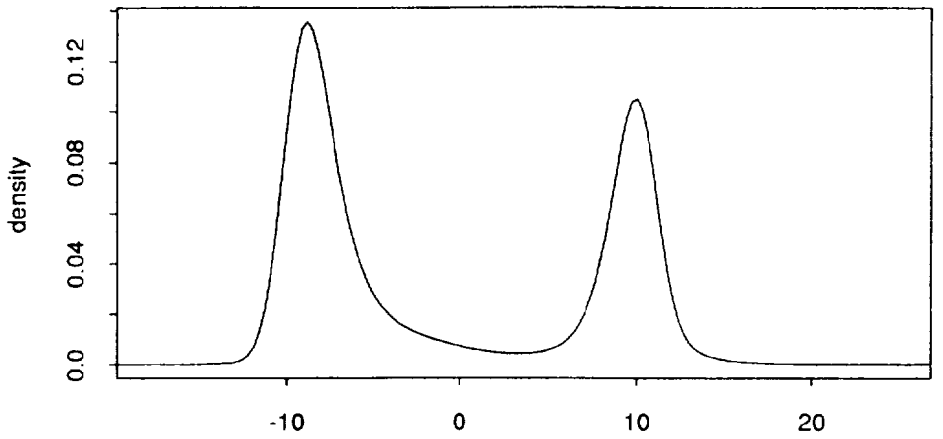


a) Marginal posterior for x(101), G = 500 ; mode = -2.829

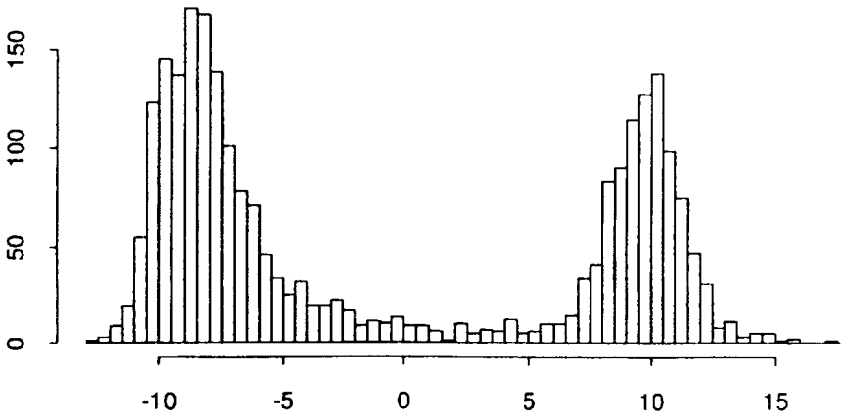


b) Histogram of Gibbs x(101) values, G = 500

Figure 8. Filtering, Example 4.3



a) Marginal posterior for x(101), G = 2500 ; mode = -8.859



b) Histogram of Gibbs x(101) values, G = 2500