On Bayesian analysis of non-linear continuous-time autoregression models

O. Stramer[*] and G.O. Roberts[†]

January 19, 2004

Abstract

This paper introduces a method for performing fully Bayesian inference for non-linear conditional auto-regressive continuous-time models, based on a finite skeleton of observations. Our approach uses MCMC and involves imputing data from times at which observations are not made. It uses a reparameterisation technique for the missing data, and due to the non-Markovian nature of the models, it is necessary to adopt an *overlapping blocks* scheme for sequentially updating segments of missing data. We illustrate the methodology using both simulated data and a data set from the S & P 500 index.

1   Introduction

We consider modeling a class of continuous-time non-linear autoregressive models, with additive noise of constant instantaneous variance, called the NLCAR(p) models. A NLCAR(p) process, is defined in Tsai & Chan (2000) to be a solution of the p-th order linear differential equation

$$dX^{(p-1)}(t) = g(\mathbf{X}(t); \theta)dt + \sigma dW(t), \tag{1}$$

where $\mathbf{X}(t) = (X(t), X^{(1)}(t), \ldots X^{(p-1)}(t))'$, $g(\mathbf{X}(t), \theta)$ is the drift, θ and $\sigma > 0$ are parameters, and $W(t)$ is a standard Brownian motion. We assume that $\mathbf{X}(0) = (x_0, x_1, \ldots, x_{p-1})$ and define $X^{(0)}(t)$ to be $X(t)$.

In particular this class includes continuous-time linear autoregressive (CAR) models and continuous-time threshold autoregressive (CTAR) models, defined in Brockwell(1994).

We consider modeling time series data obtained from sampling an underlying NLCAR process $X(t)$. We assume that the data are obtained from sampling $X(t)$ at discrete times $0 = t_0 < t_1 < \ldots < t_N$. (i.e. $X(t_i) = y_i$ for $i = 0, \ldots, N$).

[*]Postal Address: Department of Statistics and Actuarial Science, University of Iowa, Iowa City IA 52242, USA

[†]Postal Address: Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, England

The likelihood function of discrete-time data sampled from a NLCAR continuous-time process is usually unavailable. Some classical approaches for the inference of CTAR processes had been proposed in the literature (Tong and Yeung, 1991 and Brockwell, 1994). All methods involve assuming that the likelihood of the observations is Gaussian, and it is therefore of value to find methodology which can work with the exact likelihood.

If sufficiently fine data is available, then at least for known σ, its likelihood can be approximated by the Girsanov's formula, as we shall see in (5). However the data may not be sufficiently fine for the likelihood to be well-approximated.

This problem can be thought of as a missing data problem, and a natural and generally well-understood Bayesian solution to this problem is to use data augmentation (Tanner & Wong, 1987). Data-augmentation successively imputes missing data and updates parameters (θ, σ) from their conditional distribution given the augmented dataset. Recent papers which have applied this strategy in practice include Elerian et al. (2001), Eraker (2001), and Eraker et al. (2002).

However in the present context, due to very high posterior correlation between missing path and the volatility σ of the diffusion, data-augmentation can lead to arbitrary poor convergence properties. See Elerian et al. (2001), and Roberts & Stramer (2001) for $p = 1$.

In Roberts & Stramer (2001), a data reparameterisation scheme which breaks down dependency between the transformed missing paths and the volatility for $p = 1$ has been proposed. The algorithm iterates between updating the missing parameters, and the transformed missing data. One useful feature of the case $p = 1$ is that the missing data can be naturally partitioned into segments in between observed data, such that all segments are conditionally independent. This helps considerably in the construction of algorithms, and the approach adopted in Roberts & Stramer (2001) is a straightforward independence sampler procedure.

In this paper we generalise the reparameterisation scheme in Roberts & Stramer (2001) to the NLCAR(p), $p > 1$ models. In this case, the imputed diffusion sample paths between observed data are no-longer independent, but in fact are highly correlated, due to the fact that the smoothness of sample paths requires that left and right derivatives need to agree at

observed data. To avoid this problem we propose to use alternative approaches that involve block updating of collections of overlapping segments.

2 NLCAR models: Definition and Notations

A non-linear continuous-time autoregressive process $X(t)$ of order p, with $p > 0$, is defined in (1) or equivalently has the state space representation

$$dX(t) = X^{(1)}(t)dt$$

$$dX^{(1)}(t) = X^{(2)}(t)dt$$

$$\vdots \qquad (2)$$

$$dX^{(p-2)}(t) = X^{(p-1)}(t)dt$$

$$dX^{(p-1)}(t) = g(X(t), \ldots, X^{(p-1)}(t), \theta)dt + \sigma dW(t),$$

with $\mathbf{X}(0) = \mathbf{x} = (x_0, x_1, \ldots, x_{p-1})$.

When the instantaneous mean g is linear, i.e.

$$g(\mathbf{X}(t), \theta) = a_0 + a_1 X(t) + \ldots a_p X^{(p-1)}(t), \qquad (3)$$

the NLCAR(p) model become a CAR(p) model.

When

$$g(\mathbf{X}(t); \theta) = a_0(X(t)) + a_1(X(t))X(t) + \ldots a_{p-1}(X(t))X^{(p-1)}(t), \qquad (4)$$

where $a_0(x), a_1(x), \ldots, a_{p-1}(x)$ are piecewise constant functions with constant values on the intervals $[r_{i-1}, r_i)$, and $-\infty = r_0 < r_1 < \ldots < r_k = \infty$, the NLCAR(p) model become a continuous-time threshold autoregressive (CTAR) model, see for example Brockwell(1994).

We firstly note that if $\mathbf{X}(0) = \mathbf{x}$ is given, then we can write $\mathbf{X}(t) = (X(t), X^{(1)}(t), \ldots X^{(p-1)}(t))'$, in terms of $\{X^{(p-1)}(s), \ 0 \leq s \leq t\}$ using the relations, $X^{(p-2)}(t) = x_{p-2} + \int_0^t X^{(p-1)}(s)ds, \ldots X(t) = x_0 + \int_0^t X^{(1)}(s)ds.$

Using the same steps as in Brockwell (1994) for CTAR models, we can show that when g satisfies growth conditions, then (2) has a unique (in law) weak solution. Throughout this

paper we assume that $(X^{(p-1)}(t), W(t))$ on $(C[0,\infty), \mathcal{B}[0,\infty), P_\sigma, \{\mathcal{F}_t\})$ is a weak solution of (1), $0 \le t \le t_N$, with initial condition $\mathbf{X}(0) = \mathbf{x} = (y_0, \mathbf{0})$.

Let $\mathbb{W}_\sigma$ be the law of a standard Brownian motion with volatility σ defined on $(C[0,\infty), \mathcal{B}[0,\infty))$. Given $X(t)$ in the time interval $[0, t_N]$, the Radon-Nikodym derivative of the law of X with respect to $\mathbb{W}_\sigma$ is given by Girsanov's formula

$$\frac{dP_\sigma(X)}{d\mathbb{W}_\sigma} = G(\mathbf{X}, g, \theta, \sigma) \tag{5}$$

where

$$G(\mathbf{X}, g, \theta, \sigma) = \exp\left\{ \int_0^T \frac{g(\mathbf{X}(s), \theta)}{\sigma^2} dX^{(p-1)}(s) - \frac{1}{2} \int_0^T \frac{g^2(\mathbf{X}(s), \theta)}{\sigma^2} ds \right\}. \tag{6}$$

Throughout this paper we will adopt the following notational conventions, some of which have already been used. Processes transformed to have unit quadratic variation for their $(p-1)$th derivative, are all written with a tilde, e.g, $\tilde{X}$. Processes which are further transformed to be 0 at the times of observations, $\{t_i\}$, are written with a double tilde, e.g $\tilde{\tilde{X}}$. The same connections are used for the respective probability laws. We continue with a summary of some important notation which follows.

X	process of interest, defined as a solution to (1)
Z	"dominating process", defined as a solution to (1) with $g \equiv 0$
$\tilde{X}$	mean scaled process defined as $\tilde{X} = X/\sigma$
$\tilde{\tilde{X}}$	mean centered process defined as in (8)
H	proposed process needed for updating $\tilde{\tilde{X}}$
N	piecewise linear approximation to X used to construct H

3 A missing data problem

3.1 Preamble

If the complete data on a finite time interval, $\{X(t), \ 0 \le t \le t_N\}$, is given, then σ is completely identified through the usual quadratic variation identity:

$$\sigma^2 t_N = \int_0^{t_N} [dX^{(p-1)}(t)]^2, \quad a.s. \tag{7}$$

and therefore inference for θ can be carried out under the assumption that σ is known. Its likelihood $G(\mathbf{X}, g, \theta, \sigma)$ with respect to the distribution of a Brownian motion with volatility

σ is given in (6).

Thus we would like to find a way of completing the missing path. We write $\mathbf{X}_{\text{mis}}$ for the sample path of $\mathbf{X}$ on $[0, t_N]$ excluding the observed observations $y = \{X(t_i),\ 1 \le i \le N\}$. Data-augmentation successively imputes missing datasets $\mathbf{X}_{\text{mis}}$ and updates parameters (θ, σ) from their conditional distribution given the augmented dataset.

In practice, we have to impute a finite number of points from the diffusion sample path $\mathbf{X}(t)$. For simplicity, we shall impute $\mathbf{X}\left(t_i + \frac{\Delta t_i}{m} k\right)$, $\Delta t_i = t_{i+1} - t_i$, for $i = 0, \ldots, N-1$, and $k = 1, \ldots, m-1$ given $X(t_i)$, $i = 0, \ldots, N$. The choice of m will be important, since it will be important to choose m sufficiently large that the diffusion and its likelihood are well approximated.

Recent papers which have applied this strategy in practice include Elerian et al. (2001), Eraker (2001), and Eraker et al. (2002).

However in the present context, the data-augmentation technique has a problem which can dramatically affect algorithm mixing. For large m, the number of iterations of the algorithm needed to achieve convergence is known to be $O(m)$. See Elerian et al. (2001), and Roberts & Stramer (2001) for $p = 1$. This is essentially caused by very high posterior correlation between σ and the missing path. See (7).

In many contexts, this approach will be sufficient, but in problems where g is highly non-linear between observed data points, the approximations involved can still be unacceptably poor.

3.2 Reparameterisation Scheme

In Roberts & Stramer (2001), a data reparameterisation scheme for $p = 1$ has been proposed which has the property that the algorithm does not degenerate as $m \to \infty$.

The reparameterisation scheme in Roberts & Stramer (2001) proceeds as follows. We set $\tilde{X}(s) = X(s)/\sigma$, and then apply a piecewise linear transformation to $\tilde{X}$, to produce $\tilde{\tilde{X}}$ ensuring that $\tilde{\tilde{X}}(t_i) = 0$ at all observed data points. This is essentially the unique transformation of the missing data that allows us to construct an irreducible Markov chain Monte Carlo algorithm for simulating from the appropriate posterior distribution. The

algorithm iterates between updating the missing parameters, and the transformed missing data.

We now generalise the reparameterisation scheme of Roberts & Stramer (2001) to the NLCAR(p), $p > 1$ models. We set $\tilde{X}(s) = X(s)/\sigma$, and then apply a piecewise linear transformation to $\tilde{X}$, to produce $\tilde{\tilde{X}}$, ensuring that $\tilde{\tilde{X}}(t_i) = 0$ at all observed data points, as follows.

$$\tilde{\tilde{X}}(t) = \tilde{X}(t) - E(\tilde{Z}(t)|\tilde{Z}(t_i) = \tilde{y}_i, \ i = 0, \ldots, N), \tag{8}$$

where $\tilde{y}_i = y_i/\sigma$ and $\mathbf{Z}(t)$ is a solution of (2) with $g \equiv 0$, and $\mathbf{Z}(0) = (y_0, \mathbf{0})$. Clearly $\tilde{\tilde{X}}(t_i) = 0$ for all $i = 0, \ldots, N$.

For $p = 1$, $\tilde{Z}(t)$ is a standard Brownian motion with a starting point $\tilde{y}_0$ and

$$\tilde{\tilde{X}}(s) = \tilde{X}(s) - \frac{(t_i - s)\tilde{y}_{i-1} + (s - t_{i-1})\tilde{y}_{t_i}}{t_i - t_{i-1}}, \quad t_{i-1} < s < t_i.$$

Returning to the case of more general p, we recall that we are assuming that $\mathbf{X}(0) = (y_0, \mathbf{0})$. Therefore, using integration by parts, it is easy to check that if $y_0 = 0$ then

$$X(t) = \frac{\int_0^t X^{(p-1)}(s)(t - s)^{p-2}}{(p-2)!} ds. \tag{9}$$

Thus, since $\tilde{Z}^{(p-1)}$ is a standard Brownian motion, $\tilde{Z}$ is a Gaussian process and $E(\tilde{Z}(t)|\tilde{Z}(t_i) = \tilde{y}_i, \ i = 0, \ldots, N)$ is a linear combination of $\tilde{y} = y/\sigma$. Here (and analogously elsewhere) we adopt the shorthand notation, $y = \{y_i, 1 \le i \le N\}$.

We now provide a theoretical justification to the construction $\tilde{\tilde{X}}(t)$. We firstly introduce the following definitions and notation.

Let $\mathbf{Z}$ be a strong solution to (2) with $g \equiv 0$ and $\mathbf{Z}(0) = (y_0, \mathbf{0})$. Let $\mathbb{Z}^\sigma$ denotes the law of $Z(t)$, $0 \le t \le t_N$ and let $\mathbb{Z}_y^\sigma$ denotes the law of $Z(t)$, $0 \le t \le t_N$ conditioned on $Z(t_i) = y_i$ for $i = 0, \ldots, N$. In particular, when $\sigma = 1$ and $y = \mathbf{0}$, we denote $\mathbb{Z}_{\mathbf{0}}^1$ by $\mathbb{Z}_0$.

Denote the N-dimensional Lebesgue density of Y under P_σ by $k(t_N, y, g, \theta, \sigma)$ and under the measure induced by Z by $f(t_N, y, \sigma)$.

Now we can factorise the measure $\mathbb{Z}^\sigma$ as follows.

$$\mathbb{Z}^\sigma = \mathbb{Z}_y^\sigma \times \mathrm{leb}_N(y) \ \times \ f(t_N, y, \sigma). \tag{10}$$

Thus, the conditional density of X_{mis} with respect to $\mathbb{Z}_y^\sigma$ is

$$\frac{dP_\sigma}{d\mathbb{Z}_y^\sigma}(X_{\text{mis}}|y) = G(\mathbf{X}, g, \theta, \sigma)\frac{f(t_N, y, \sigma)}{k(t_N, y, g, \theta, \sigma)} \tag{11}$$

$$\propto G(\mathbf{X}, g, \theta, \sigma) .$$

The dominating measure $\mathbb{Z}_y^\sigma$ depends on σ. In order to circumvent this problem, we factorise the distribution of P_σ in such a way that the dominating measure is independent of σ.

We use the transformation $\tilde{\tilde{\mathbf{X}}}$, defined as in (8) to write,

$$dP_\sigma(\tilde{\tilde{X}}_{\text{mis}}, y) = G(\nu(\tilde{\tilde{\mathbf{X}}}), \mu, \theta, 1)f(t_N, y, \sigma)d\mathbb{Z}_0 \otimes \text{leb}_N, \tag{12}$$

where

$$\mu(\mathbf{X}(t), \theta, \sigma) = \frac{g(\sigma\mathbf{X}(t), \theta)}{\sigma} \tag{13}$$

and $\nu(\tilde{\tilde{\mathbf{X}}})$ just inverts (8) so that

$$\nu(\tilde{\tilde{\mathbf{X}}})(t) := \tilde{\tilde{\mathbf{X}}}(t) + E(\tilde{\mathbf{Z}}(t)|\tilde{Z}(t_i) = \tilde{y}_i, \ i = 1, \ldots, N) \tag{14}$$

The dominating measure $\mathbb{Z}_0$ is independent of σ. By (12) we can easily write down the conditional distribution of the transformed missing data:

$$\frac{dP_\sigma(\tilde{\tilde{X}}_{\text{mis}}|y)}{d\mathbb{Z}_0} \propto G\{\nu(\tilde{\tilde{\mathbf{X}}}), t_N, \mu, 1, \theta\} \tag{15}$$

4 The MCMC algorithm

4.1 General

The Metropolis-Hastings algorithm iterates between updating the missing parameters, and the transformed missing data according to their conditional posterior distributions.

For a fully Bayesian framework, we shall fix priors on all unknown parameters. We shall assume that σ and θ have continuous prior densities $p_\sigma(\cdot)$ and $p_\theta(\theta)$ respectively on $\mathbb{R}$. As a result, and using (12), we can write the posterior conditional density of θ and the posterior conditional density of σ as follows.

$$p(\theta|\tilde{\tilde{\mathbf{X}}}_{\text{mis}}, y, \sigma) \propto G\{\nu(\tilde{\tilde{X}}^{(p-1)}), t_N, \mu, 1, \theta\}p_\theta(\theta), \tag{16}$$

where ν and μ are defined as in (14) and (13) respectively.

$$p(\sigma|\tilde{\tilde{\mathbf{X}}}_{\mathrm{mis}}, y, \theta) \propto G\{\nu(\tilde{\tilde{\mathbf{X}}}), t_N, \mu, 1, \theta\} f(t_N, y, \sigma) \, p_\sigma(\sigma) \, . \tag{17}$$

The conditional distributions (15), (16) and (17) provide the basis for the algorithm.

Updating θ and σ can be carried out using standard Metropolis-Hastings steps. For updating σ, it is important to note that $\nu(\tilde{\tilde{\mathbf{X}}})$ is a function of σ. The calculation of $\nu(\tilde{\tilde{\mathbf{X}}})$ will be given in the Appendix. Updating $\tilde{\tilde{\mathbf{X}}}$ is less straightforward and will be discussed in Section 4.2

4.2 Updating $\tilde{\tilde{\mathbf{X}}}$

We propose a new process $\tilde{\tilde{\mathbf{H}}}(s)$, $0 \le s \le t_N$, from a proposal conditional diffusion distribution $\mathbb{P}_{\tilde{H}}$ which is the law obtained by conditioning the proposed diffusion process $\{\tilde{N}(s),$ $0 \le s \le t_N\}$, with $\tilde{\tilde{\mathbf{N}}}(0) = \mathbf{0}$, on $\tilde{N}(t_i) = 0$, for $i = 1, \ldots, N$. We require that $\mathbb{Z}_0$ be absolutely continuous with respect to $\mathbb{P}_{\tilde{H}}$.

One natural choice for $\tilde{N}$ is the Gaussian process Z, where $\mathbf{Z}$ is defined as a strong solution to (2), with $g \equiv 0$, $\sigma = 1$ and $\mathbf{Z}(0) = \mathbf{0}$. For this choice of N, $\mathbb{P}_{\tilde{H}}$ is simply $\mathbb{Z}_0$. Simulation of $\tilde{\tilde{\mathbf{H}}}$, which is a Gaussian process, can be carried out via a full multivariate analysis on a suitable discretisation.

An improvement on the choice of $\tilde{\tilde{N}}$ would be to use a linear process on $[t_{i-1}, \ t_i]$. In this paper we describe one natural choice for $\tilde{\tilde{N}}$ when $X(t)$ is a CTAR model. Recall that a CTAR model is defined as a solution to (2) where g is defined as in (4). Let $\mathbf{N}$ be a solution to (2) with drift

$$g^L(\mathbf{N}(t); \theta) = a_0(N(t_i)) + a_1(N(t_i))N(t) + \ldots a_{p-1}(N(t_i))N^{(p-1)}(t), \quad t_i \le t < t_{i+1} \tag{18}$$

with $\mathbf{N}(0) = (y_0, \mathbf{0})$. Note that $N(t)$ is a CAR process on $[t_i, t_{i+1})$. $\tilde{H}$ is the process obtained by conditioning the proposed diffusion process $\{\tilde{N}(s), 0 \le s \le t_N\}$, with $\tilde{\mathbf{N}}(0) = (\tilde{y}_0, \mathbf{0})$, on $\tilde{N}(t_i) = \tilde{y}_i$ for $i = 1, \ldots, N$. Now let $\tilde{\tilde{H}}$ be defined as

$$\tilde{\tilde{H}}(t) = \tilde{H}(t) - E(\tilde{Z}(t)|\tilde{Z}(t_i) = \tilde{y}_i, \ i = 0, \ldots, N)$$

Then the Metropolis-Hastings step proceeds as follows. Suppose that the current state of the missing data on (t_0, t_N) is $\tilde{\tilde{X}}_{\mathrm{mis}}(t_0, t_N)$. Then the algorithm proposes a new segment of

8

missing data $\tilde{\tilde{H}}$ on the interval $(0, t_N)$. This proposal is accepted with probability

$$\alpha\{\tilde{\tilde{X}}_{\mathrm{mis}}(t_0, t_N), \tilde{\tilde{H}}\} = \min\left(1, \frac{G[\nu(\tilde{\tilde{H}}); \mu, 1]}{G[\nu(\tilde{\tilde{H}}); \mu^L, 1]} \frac{G[\nu\{\tilde{\tilde{X}}_{\mathrm{mis}}(t_0, t_N)\}; \mu^L, 1]}{G[\nu\{\tilde{\tilde{X}}_{\mathrm{mis}}(t_0, t_N)\}; \mu, 1]}\right) \ . \tag{19}$$

where μ is defined as in (13) and μ^L is defined as in (13) with $g = g^L$.

4.3 Generating the proposal

We now describe how to simulate $\tilde{\tilde{\mathbf{H}}}$ on $[0, t_N]$. In practice, we have to impute a finite number of points from $\tilde{\tilde{\mathbf{H}}}$. We shall impute $m - 1$ imputed times between (t_j, t_{j+1}). Let $\mathbf{N}^h = \{\mathbf{N}(kh) : \quad k = 0, \ldots, Nm\}$, where $h = \frac{\Delta}{m}$, $\Delta t_i = t_{i+1} - t_i$, and let $y^t = \{y_j : \quad t_j \le t\}$.

From Carter and Kohn (1994) we have,

$$p(\mathbf{N}^h | y^{t_N}) = p\{\mathbf{N}(t_N) | y^{t_N}\} \prod_{k=1}^{Nm-1} p\{\mathbf{N}(kh) | y^{kh}, \mathbf{N}((k+1)h)\}. \tag{20}$$

We now show that $p\{\mathbf{N}(t_N)|y^{t_N}\}$ and $p\{\mathbf{N}(kh)|y^{kh}, \mathbf{N}((k+1)h)\}$ are Gaussian densities and thus in order to generate $\mathbf{N}$, it is sufficient to compute $E(\mathbf{N}(t_N)|y^{t_N})$, $\mathrm{var}(\mathbf{N}(t_N)|y^{t_N})$ and

$$E(\mathbf{N}(kh)|y^{kh}, \mathbf{N}((k+1)h)), \quad \mathrm{var}(\mathbf{N}(kh)|y^{kh}, \mathbf{N}((k+1)h)), \quad (k = 1, \ldots, Nm - 1).$$

Let

$$\mathbf{N}(k|j) = E(\mathbf{N}(kh)|y^{jh}), \quad S(k|j) = \mathrm{var}(\mathbf{N}(kh)|y^{jh}).$$

$\mathbf{N}(t)$ is a CAR process on each sub-interval $[t_j, t_{j+1})$. Thus, $\mathbf{N}$ is a solution to (2) with drift g^L defined as in (18) or equivalently as a solution to

$$d\mathbf{N}(t) = (A(N(t_i))\mathbf{N}(t) + \mathbf{e}a_0(N(t_i))) \, dt + \mathbf{e}dW(t), \qquad t_i \le t < t_{i+1} \tag{21}$$

where

$$A(x) := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_1(x) & a_2(x) & & \cdots & a_p(x) \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

where $a_0(x), a_1(x), \ldots, a_{p-1}(x)$ are defined as in (4).

From (21) we have that

$$E(\mathbf{N}((k+1)h)|\mathbf{N}(kh)) = e^{A(N(t_i)h)}\mathbf{N}(kh) + \frac{a_0(N(t_i))}{a_1(N(t_i))}(e^{A(N(t_i))h} - I)\mathbf{b} \tag{22}$$

$$\mathrm{var}(\mathbf{N}((k+1)h)|\mathbf{N}(kh)) = \int_0^h e^{A(N(t_i))u}\mathbf{e}\mathbf{e}'e^{A(N(t_i))'u}du, \tag{23}$$

for $t_i \leq kh, (k+1)h \leq t_{i+1}$, where $\mathbf{b}' = (1, 0, \ldots, 0)$ and $\mathbf{e}' = (0, \ldots, 0, 1)$. The calculation of the integrals can be done using the spectral representation of the matrix $A(N(t_i))$ (Jones, 1985) or as in Tsai and Chan (2000a).

From (22) and (23) we have that for $t_i \leq kh, (k+1)h \leq t_{i+1}$,

$$\mathbf{N}((k+1)h) + \frac{a_0(N(t_i))}{a_1(N(t_i))}\mathbf{b} = e^{A(N(t_i))h}\left(\mathbf{N}(kh) - \frac{a_0(N(t_i))}{a_1(N(t_i))}\right) + \mathbf{Z}_k, \tag{24}$$

where $\{\mathbf{Z}_k, \ k = 0, \ldots Nm - 1\}$ is an independent sequence of Gaussian random vectors with mean $E(\mathbf{Z}_k) = \mathbf{0}$ and covariance matrix

$$E[\mathbf{Z}_k\mathbf{Z}_k'] = \int_0^h e^{A(N(t_i))u}\mathbf{e}\mathbf{e}'e^{A(N(t_i))'u}du, \quad t_i \leq kh < t_{i+1}.$$

(24) is in precisely the form needed for application of the Kalman recursions (see e.g. Brockwell and Davis (1991), Chapter 12). From these recursions we can easily compute $\mathbf{N}(k|j)$ and $\mathbf{S}(k|j)$, for $j \leq k$ and in particular $\mathbf{N}(N|N)$ and $\mathbf{S}(N|N)$. To find, $E(\mathbf{N}(kh)|y^{t_j}, \mathbf{N}((k+1)h))$ and $\mathrm{var}(kh)|y^{t_j}, \mathbf{N}((k+1)h))$ for $t_i \leq kh < t_{i+1}$, we note that conditioned on y^{t_i},

$$\begin{bmatrix} \mathbf{N}(kh) \\ \mathbf{N}((k+1)h) \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{N}(kh|y^{t_j}) \\ \mathbf{N}((k+1)h|y^{t_j}) \end{bmatrix}, \begin{bmatrix} \mathbf{S}(k|k) & \mathbf{S}(k|k)e^{A'(N(t_j))h} \\ e^{A(N(t_j))h}\mathbf{S}(k|k) & \mathbf{S}(k+1|k) \end{bmatrix}\right).$$

See Tsai and Chan (2000b) for CAR models.

4.4   Block Sampling

Updating $\tilde{\tilde{\mathbf{X}}}$ could lead to a very law acceptance rate, due to the discrepancy between the proposed latent process component and the true latent component. To circumvent this problem, we often split up the latent process into blocks, and cycle through each block in turn for updating. This increases the acceptance rate of each move, because there is less latent variables in each block, limiting the discrepancy between the proposed latent process component and the true latent component. See an unpublished D.Phil. thesis from Nuffield College, Oxford by O. Elerian.

For $p = 1$, the imputed blocks between observed data are conditionally independent. However, for $p > 1$ the imputed diffusion sample paths between observed data are no-longer independent, but in-fact are highly correlated. It turns out that updating segments between observed data leads to an algorithm which degenerates to reducibility as $m \to \infty$, since

$$X^{(i+1)}(t_j) = \lim_{m \to \infty} \frac{X^{(i)}(t_j + \frac{1}{m}) - X^{(i)}(t_j)}{1/m} = \lim_{m \to \infty} \frac{X^{(i)}(t_j) - X^{(i)}(t_j - \frac{1}{m})}{1/m}, \quad i = 0, \ldots, p - 2.$$

Alternative approaches involving block updating of collections of overlapping segments avoid this problem. We propose to update $\tilde{X}(s)$, for $t_i \leq s \leq t_{i+2}$ for $i = 0, \ldots, N - 2$.

4.5 The Wilkinson scheme

We recall that basic schemes which impute $\{\mathbf{X}(t) : 0 \leq t \leq t_N\}$ and then update θ and σ are reducible. The reparameterisation scheme circumvents this problem.

An alternative methodology for this problem has recently been developed in Wilkinson (2001). It is proposed in Wilkinson (2001) to break down the correlation structure between $\mathbf{X}$ and its parameters by grouping $\mathbf{X}$ and σ in the Metropolis-Hastings algorithm. Although it seems unlikely that the Wilkinson approach will be able to deal effectively with large data sets (since it loses the advantage of having independence of imputed blocks between observed data), it will be more widely applicable for extremely smooth time-series.

5 Examples

5.1 Example 1: CTAR(2) process

$$dX^{(1)}(t) = g(X(t), X^{(1)}(t); \theta)dt + \sigma dW(t),$$

with

$$g(X(t), X^{(1)}(t); \theta) =$$
$$\begin{cases} a_0^1 + a_1^1 X(t) + a_2^1 X^{(1)}(t) & X_0(t) < r \\ a_0^2 + a_1^2 X(t) + a_2^2 X^{(1)}(t) & X_0(t) \geq r \end{cases}$$

Simulated Data: Data were simulated from the above model with $r = 0$, $a_0^1 = -0.5$, $a_1^1 = -0.5$, $a_2^1 = -1.0$, $a_0^2 = 0.9$, $a_1^2 = -3.0$, $a_2^2 = -2.5$, $\sigma^2 = 0.36$ at times kh, $k = 0, 1, \ldots, Nm$ with $h = 0.001$, $N = 1000$ and $m = 1000$ by assuming linearity on each time

interval $[kh, \ (k+1)h)$. After generating $1,000,000$ data points we recorded every 1000th point and thus obtained a sample path of $1,000$ observations at times $t = 0, 1, \ldots, 999$.

Prior: We assumed normal density priors for a_i^1, a_i^2

for $i = 0, 1, 2$, Uniform prior for r and $\mathrm{InvGa}(\alpha, \beta)$ prior, for σ^2.

<u>Algorithms</u>

We shall term the reparameterisation method introduced in Roberts & Stramer (2001) as the R method. We refer to

- the algorithm that updates all missing data at once as method Ra,

- the algorithm that updates overlapping blocks as Rb,

- The Wilkinson algorithm as W.

We applied the algorithms Ra, Rb, and W to the simulated data for $12,000$ iterations. Figure 1 show trace plots of the first 5000 steps taken by the Ra and Rb algorithm for the parameters (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 40$. The graphs show that for this example the Rb algorithm perform better. The Rb method with $m = 40$ algorithm converged quite rapidly from a distant starting point. The acceptance rates for proposed sample paths with $m = 40$ were 0.001 for the Ra scheme, 0.006 for the W scheme and 0.972 for the Rb scheme.

Figure 2 shows the autocorrelation plots in the Markov chain Monte Carlo output with $m = 40$ for (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ after a burn-in period of 2000 iterations. The autocorrelation plots demonstrate the advantage of using the Rb algorithm, dying down pretty quickly when using the Rb algorithm.

Figure 3 shows the histograms plots for (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ after a burn-in period of 2000 iterations. The histograms show that the discretisation with $m < 40$ is not sufficiently fine. m needs to be at least 40.

5.2 Example 2: CTAR(3) process

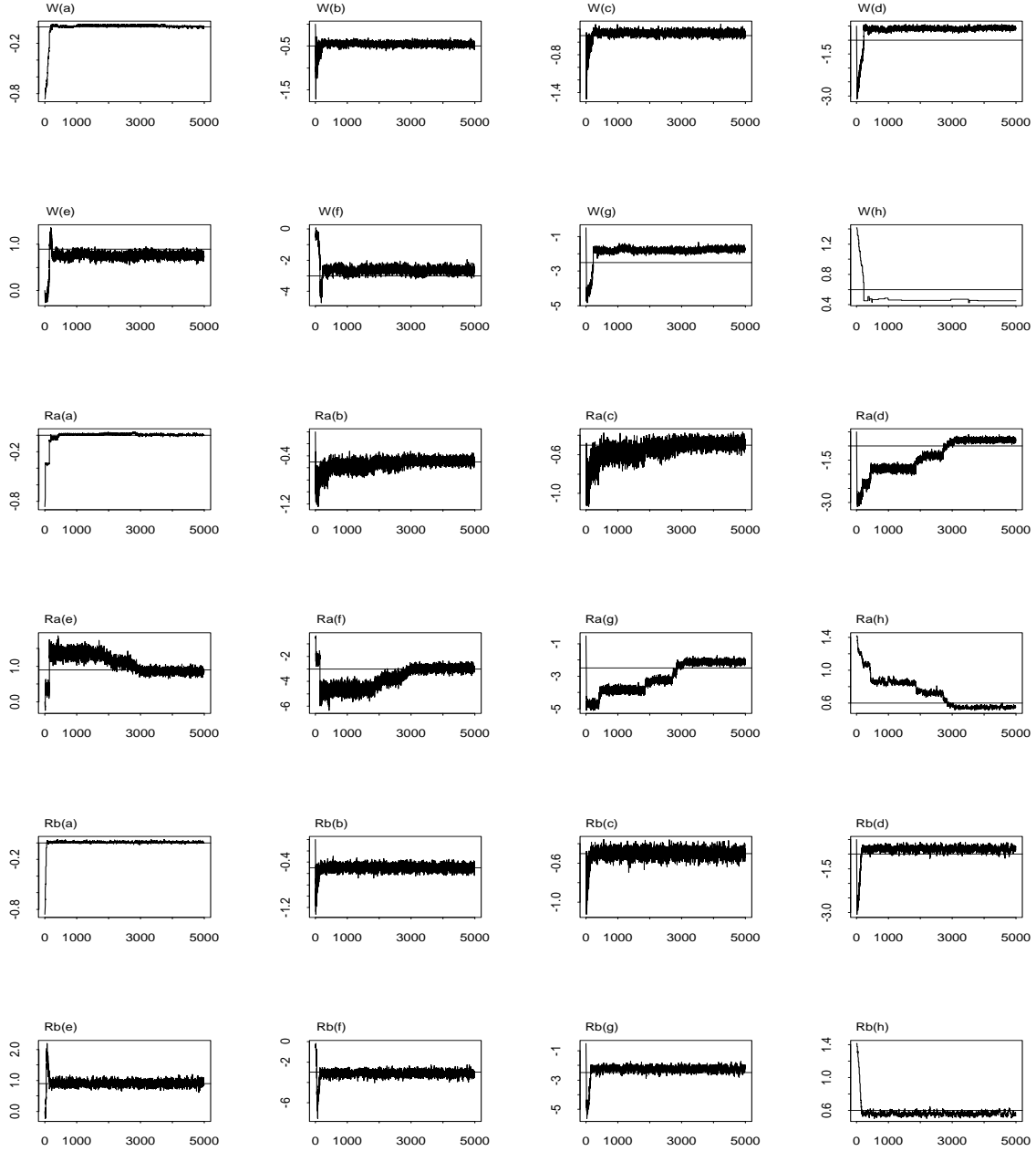$$dX^{(2)}(t) = g(X(t), X^{(1)}(t), X^{(2)}(t); \theta)dt + \sigma dW(t),$$

Figure 1: Trace plots of the steps taken by the algorithms used on the simulated example of Subsection 3.1: (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 40$ using algorithms W, Ra, and Rb. The horizontal line corresponds to the true value.
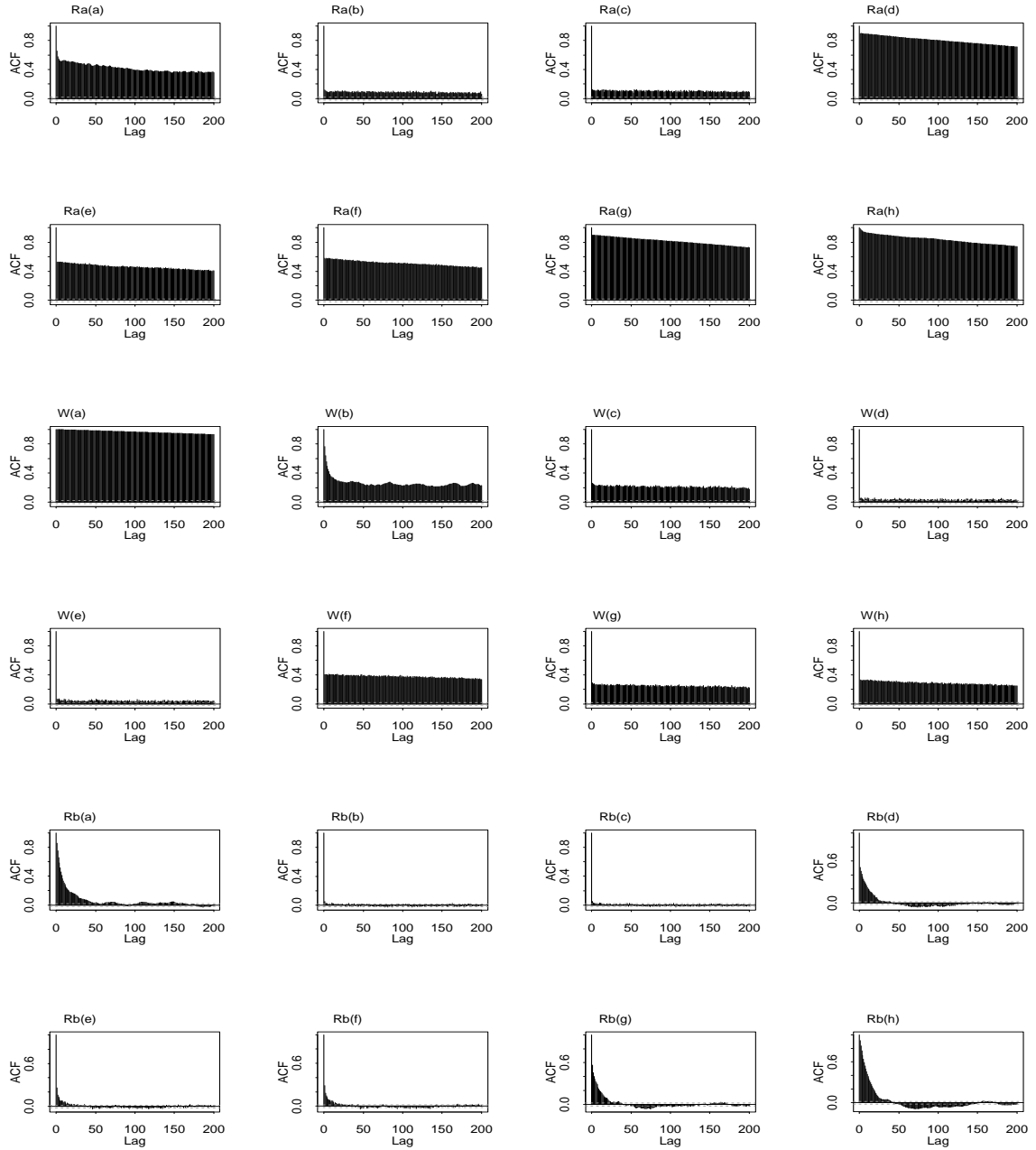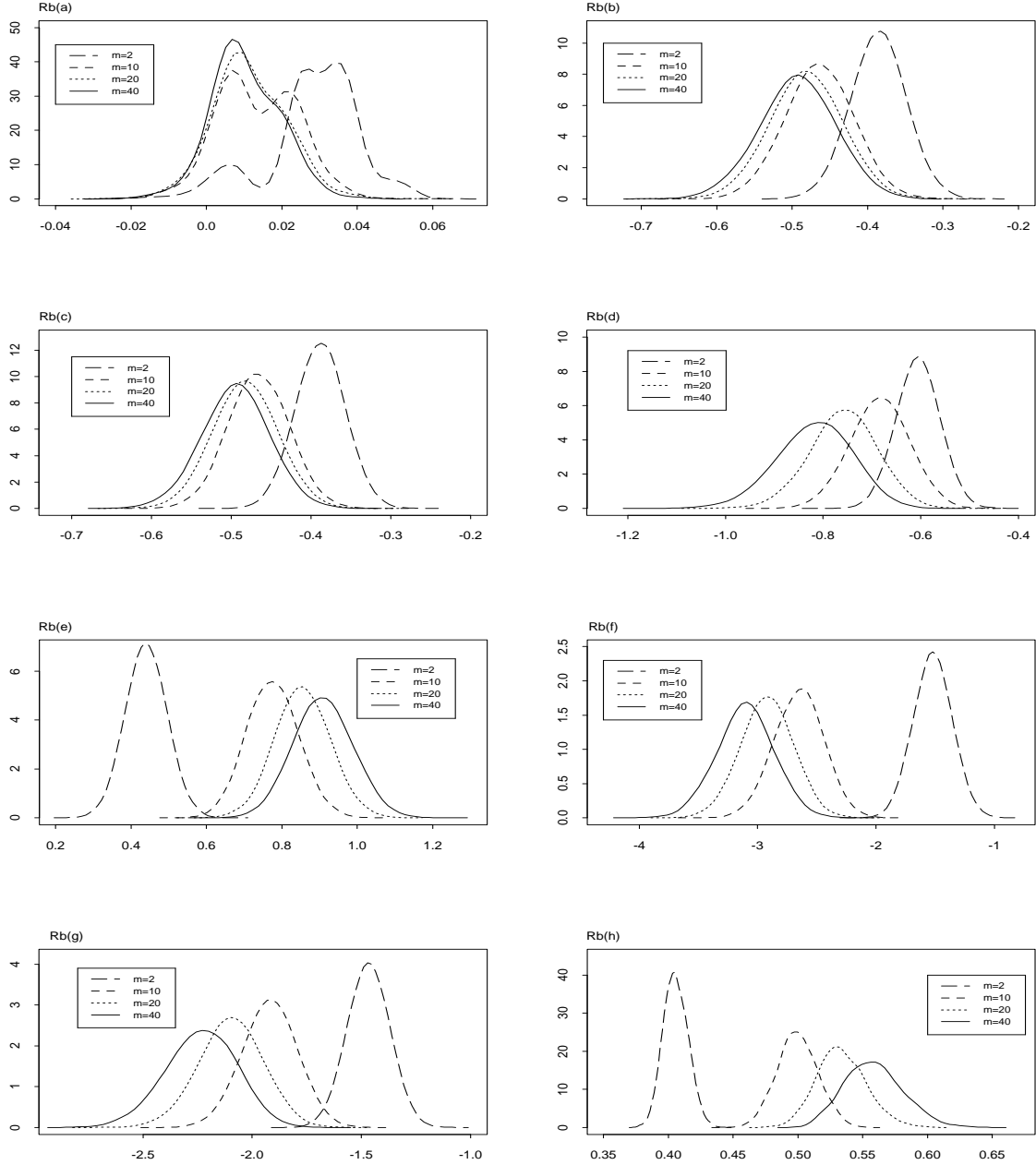
Figure 2: Autocorrelation plots for the algorithms used on the simulated example of Subsection 3.1: (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 40$ using algorithms W, Ra, and Rb.

Figure 3: Histograms for the algorithms used on the simulated example of Subsection 3.1: (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 2, 10, 20, 40$ using algorithm Rb. True values are $r = 0$, $a_0^1 = -0.5$, $a_1^1 = -0.5$, $a_2^1 = -1.0$, $a_0^2 = 0.9$, $a_1^2 = -3.0$, $a_2^2 = -2.5$, $\sigma^2 = 0.36$.

where

$$g(\mathbf{X}(t), \theta) =$$
$$\begin{cases} a_0^1 + a_1^1 X_0(t) + a_2^1 X_1(t) + a_3^1 X_2(t) & X_0(t) < r \\ \\ a_0^2 + a_1^2 X_0(t) + a_2^2 X_1(t) + a_3^2 X_2(t) & X_0(t) \geq r. \end{cases}$$

Data were simulated from the above model at times kh, $k = 0, 1, \ldots, 999$ with $r = 0$, $a_0^1 = 0$, $a_1^1 = -0.2$, $a_2^1 = -1.2$, $a_3^1 = -0.5$ $a_0^2 = 0$, $a_1^2 = -1.0$, $a_2^2 = -1.9$, $a_3^2 = -1$, and $\sigma^2 = 1$.

We assumed normal density priors for a_i^1, a_i^2 for $i = 0, 1, 2, 3$, Uniform prior for r and InvGa(α, β) prior, for σ^2. All of our results proved to be robust to the change of means and variances in this prior.

We applied the algorithms Rb to the simulated data for $20,000$ iterations.

Figure 4 shows the autocorrelation plots in the Markov chain Monte Carlo output with $m = 50$ for (b1) a_0^1, (b2) a_1^1, (b3) a_2^1, (b4) a_3^1, (a1) a_0^2, (a2) a_1^2, (a3) a_2^2, (a4) a_3^2, (r) r, and (s) σ after a burn-in period of 5000 iterations.

Figure 5 shows the histograms plots with $m = 20$ and $m = 50$ for (b1) a_0^1, (b2) a_1^1, (b3) a_2^1, (b4) a_3^1, (a1) a_0^2, (a2) a_1^2, (a3) a_2^2, (a4) a_3^2, (r) r, and (s) σ after a burn-in period of 5000 iterations.

6   Some example

In this section we briefly outline two examples where we have applied the methodology introduced earlier in the paper.

6.1   An example from the S&P 500 index

We applied our algorithm, the Rb algorithm, to the S&P500 index data. The observations are daily from 3 January 1994 to 25 March 2002. The dataset is publicly available at http://www.economagic.com/em-cgi/data.exe/sp/day-sp500c.

We fitted CTAR(p), $p = 1, 2$ to the series of the first 700 relative daily percent change. Our MCMC output suggests that for $p = 1$, the threshold is below the minimum of the data which implies that the CAR(1) model performs better than the CTAR(1) model for this data-set.

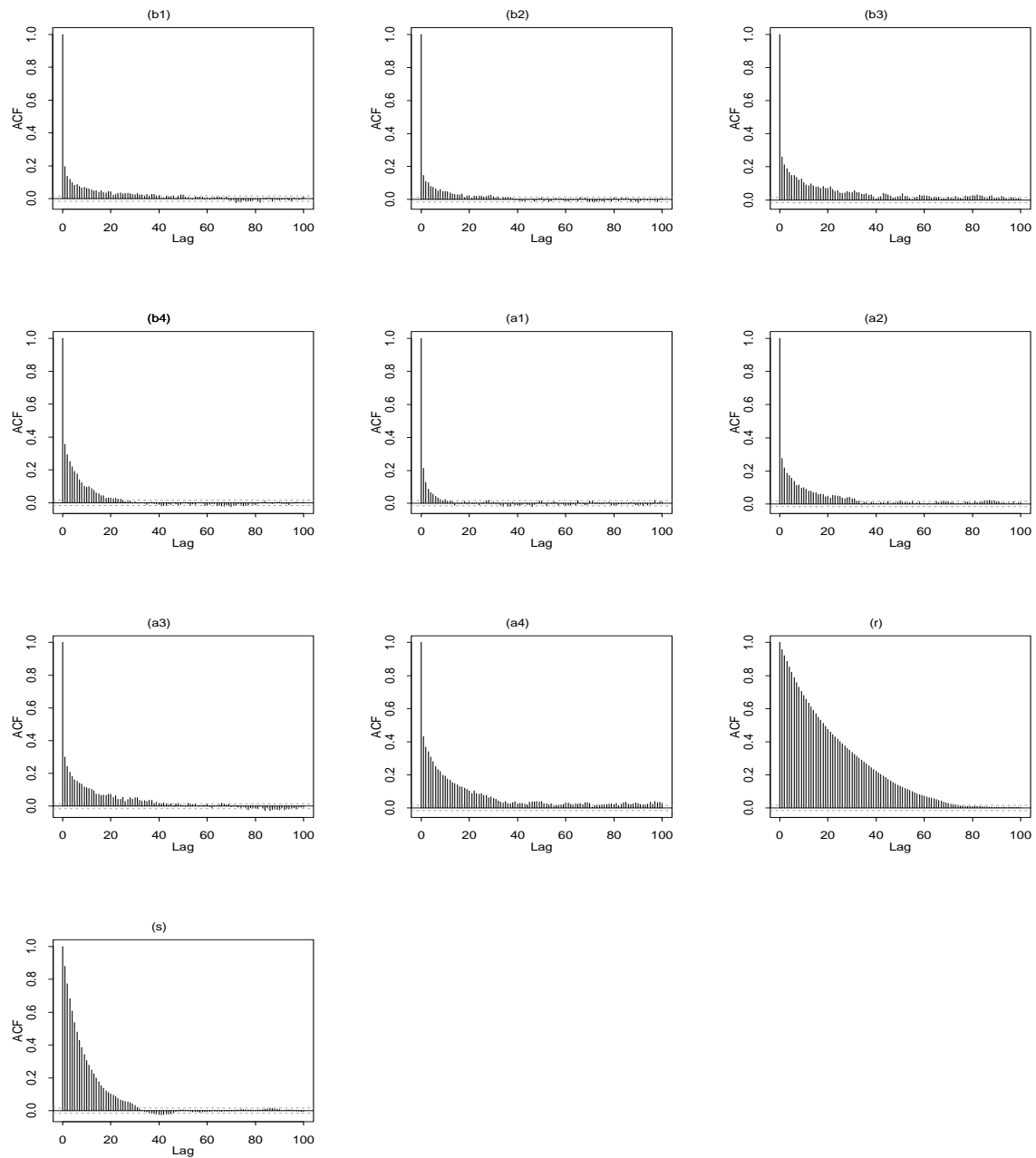Figure 4: Autocorrelation plots for the simulated example of Subsection 3.2: (b1) a_0^1, (b2) a_1^1, (b3) a_2^1, (b4) a_3^1, (a1) a_0^2, (a2) a_1^2, (a3) a_2^2, (a4) a_3^2, (r) r, and (s) σ with $m = 50$ using algorithm Rb.
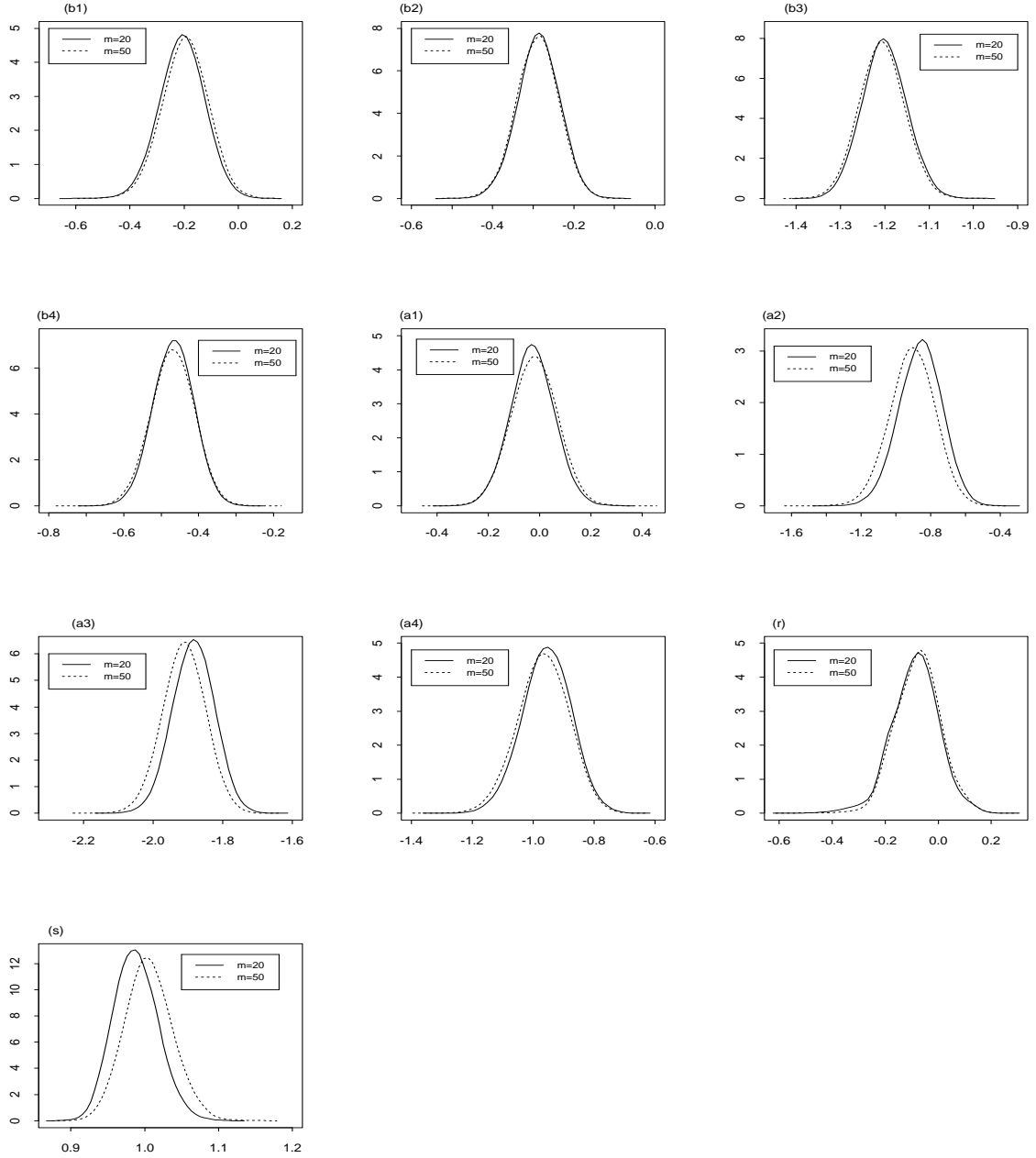
Figure 5: Histograms for the simulated example of Subsection 3.2: (b1) a_0^1, (b2) a_1^1, (b3) a_2^1, (b4) a_3^1, (a1) a_0^2, (a2) a_1^2, (a3) a_2^2, (a4) a_3^2, (r) r, and (s) σ with $m = 20$ and $m = 50$ using algorithm Rb

We then applied the Rb algorithm with $p = 2$ for $30,000$ iterations.

Figure 6 show trace plots of the first 10000 steps taken by the Rb algorithm for the parameters (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 30$.

Figure 7 shows the histograms plots for (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ after a burn-in period of 10000 iterations.

6.2 The Wolfe's sunspot data

We have used the 280 observations from the year $1700 - 1979$. We denote our observations by Y_t and the transformed data used by Tong (1983) by $X_t = 2(\sqrt{Y_t + 1} - 1)$. We apply the $CTAR(2)$ to the transformed data set. Our MCMC output suggests that the threshold is below the minimum of the data which implies that the $CAR(2)$ model performs better than the $CTAR(2)$ model for the transformed sunspot data.

We note that the Tsai & Chan (2000) test suggests that with $p = 1$ and $p = 2$ the transformed data is linear while with $p = 3, 4$ it is non-linear.

7 Discussion

We have shown that the overlapping blocks extension of the method introduced in [5] is an effective way of fitting NLCAR models. The approach combines a missing data reparameterisation technique with an overlapping blocks updating scheme as part of an MCMC algorithm.

The methodology introduced in this paper can be extended in a number of ways. More general model types can easily considered within this framework. In this paper, we have assumed that initial states (including the relevant derivatives) are known. It is straightforward to generalise our methods to the case where initial conditions are uncertain. Furthermore, generalisations to situations where model parameters vary over time (for instance according to change-points which may or not be known) are in principle feasible, although increased computational overheads will be encountered for highly complex models. Models where volatility takes on different values in different regions, are more challenging for our methods since it is not easy to construct appropriate reparameterisation schemes.
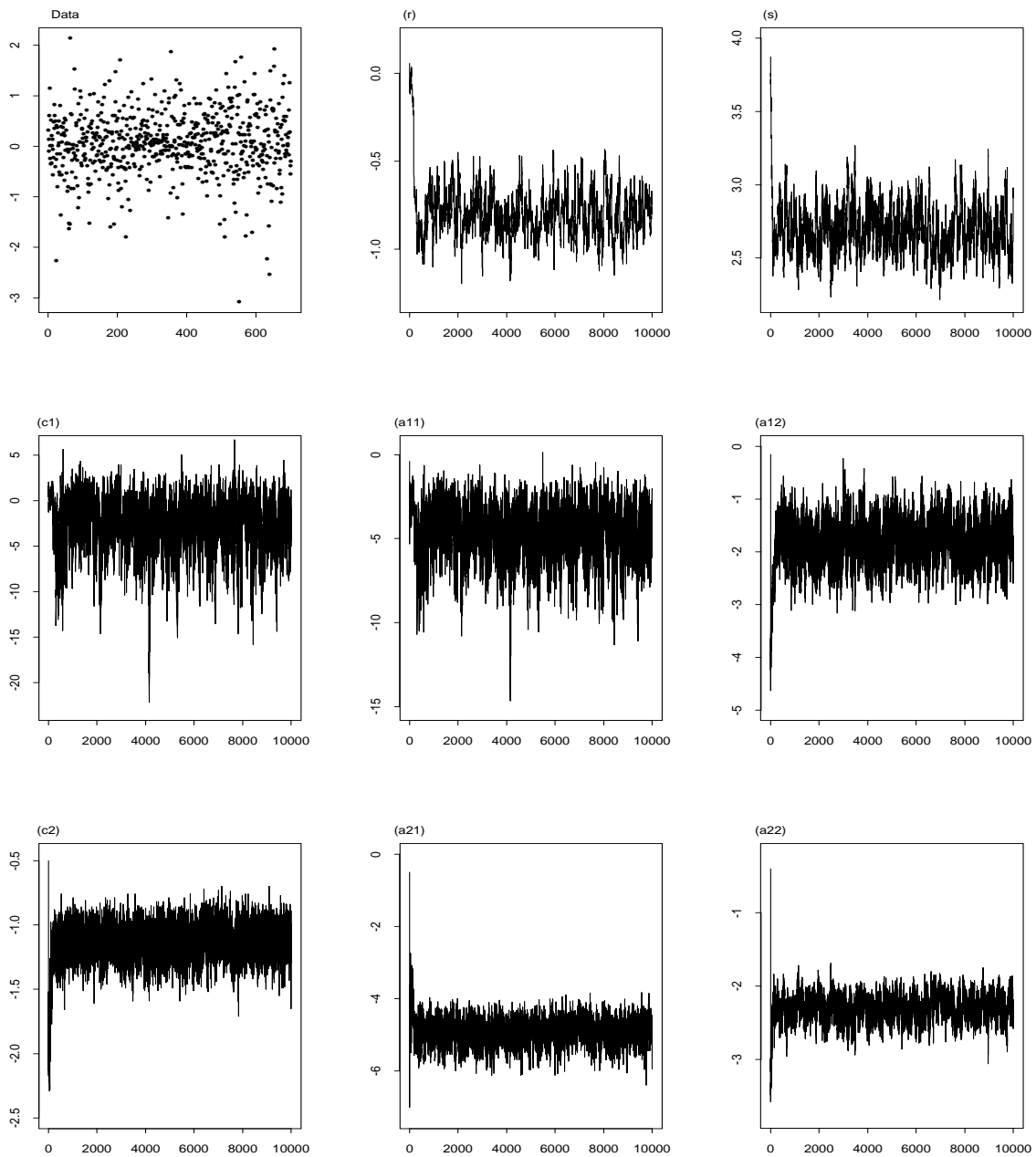
Figure 6: Trace plots of the steps taken by the algorithms used on the S&P 500 data of Section 4: (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 20$ using algorithm Rb.
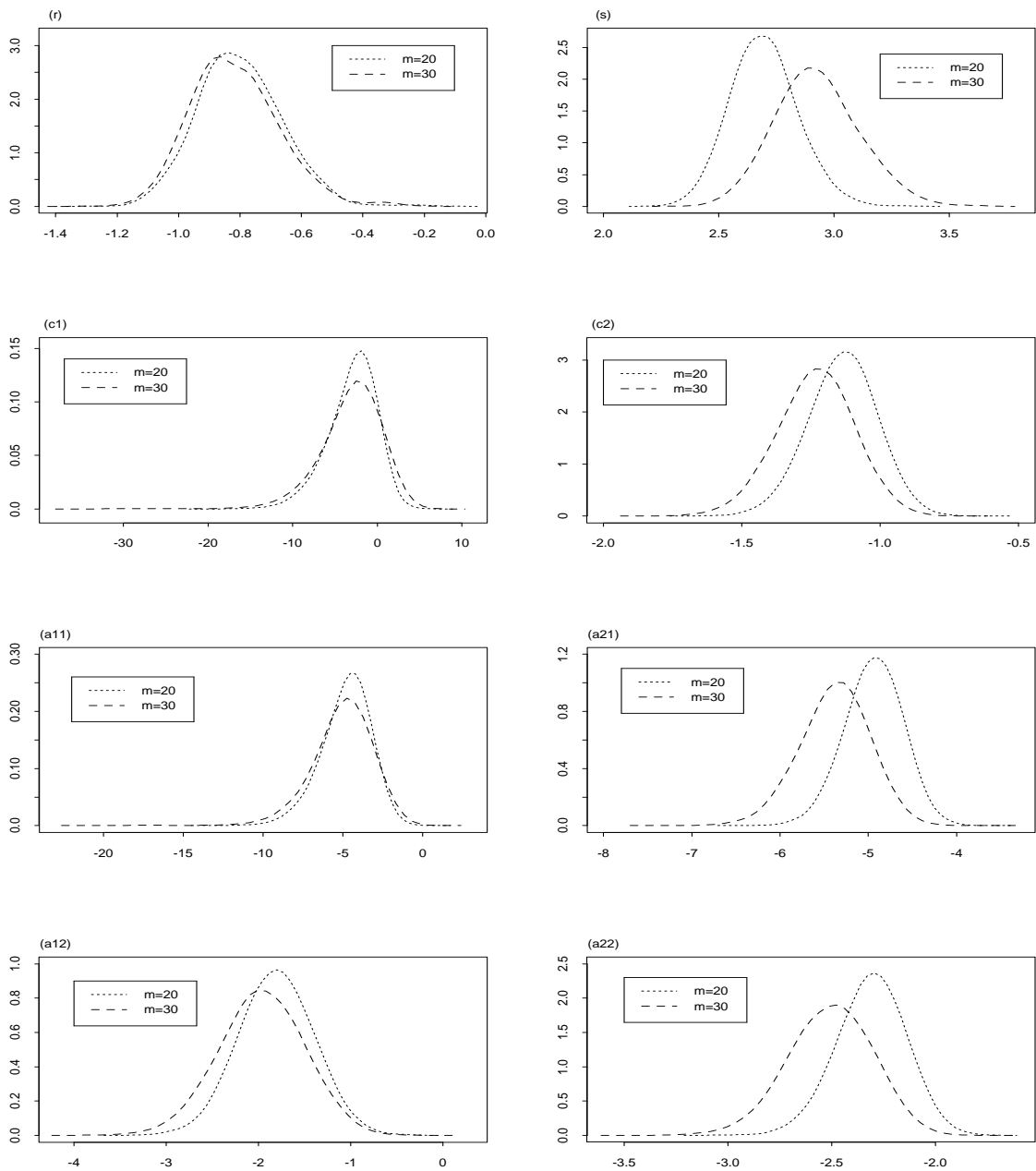
Figure 7: The S&P 500 example of Section 4: histograms (a) r, (b) a_0^1, (c) a_1^1, (d) a_2^1, (e) a_0^2, (f) a_1^2, (g) a_2^2, and (h) σ with $m = 10, 30$ using algorithm Rb.

Generalising to problems of model choice, inference on the order of a NLCAR(p) process can also be carried out using an extension of our framework. Our preferred methodology for this would require the use of reversible jump MCMC techniques.

From an MCMC implementational point of view, many important problems still need to be addressed. For instance there are many different choices of updating strategies for the various blocks of missing data. The method of overlapping blocks used here seems effective in our limited experience, but further investigation theoretically and in applications is clearly required.

References

[1] Brockwell, P. J. (1994). On continuous-time threshold ARMA processes. *Journal of Statistical Planning and Inference*, **39**, 291–303.

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

[3] Elerian, O.S., Chib, S. and Shephard, N. (2001) Likelihood inference for discretely observed non-linear diffusions. *Econometrica*, **69**, 959–993.

[4] Eraker, B. (2001) MCMC analysis of diffusion models with application to Finance. *Journal of Business and Economic Statistics*, **19**, 177–191.

[5] Eraker, B., Johannes, M. and Polson, N.G. (2002) The impact of jumps in returns and volatility. *Journal of Finance*, forthcoming.

[6] Roberts G.O. and Stramer, O. (2001). On inference for non-linear diffusion models using the Hastings-Metropolis algorithms. *Biometrika*, **88**, 603–621.

[7] Tsai, H. and Chan, K. S. (2000). Testing for non-linearity with partially observed time series. *Biometrika*, **87**, 805–821.

[8] Tsai, H. and Chan, K. S. (2002). A note on testing for non-linearity with partially observed time series. *Biometrika*, **1**, 245–250.

[9] Tong, H. (1990). Nonlinear time series: a dynamical system approach. Oxford: Oxford University Press.

[10] Tong, H. and Yeung, I. (1991). Threshold autoregressive modelling in continuous time. *Statistica Sinica*, **1**, 411–430.