CONSTRUCTION OF STATIONARY TIME SERIES VIA THE GIBBS SAMPLER WITH APPLICATION TO VOLATILITY MODELS

Michael K Pitt

and

Stephen G Walker

No 595

WARWICK ECONOMIC RESEARCH PAPERS

THE UNIVERSITY OF WARWICK

DEPARTMENT OF ECONOMICS

Construction of Stationary Time Series via the Gibbs Sampler with Application to Volatility Models

MICHAEL K PITT AND STEPHEN G WALKER

University of Warwick and University of Bath

M.K.Pitt@warwick.ac.uk s.g.walker@maths.bath.ac.uk

Abstract

In this paper, we provide a method for modelling stationary time series. We allow the family of marginal densities for the observations to be specified. Our approach is to construct the model with a specified marginal family and build the dependence structure around it. We show that the resulting time series is linear with a simple autocorrelation structure. In particular, we present an original application of the Gibbs sampler. We illustrate our approach by fitting a model to time series count data with a marginal Poisson-gamma density.

Some key words: ARCH, Exponential family, GARCH, Gibbs sampler, Filtering, Markov chains, Markov chain Monte Carlo , Stochastic Volatility

1 INTRODUCTION

In this paper we focus on a general method for constructing stationary time series models with marginals of choice. The construction of such time series is based on the Gibbs sampler. Although principally used for Bayesian inference (see, for example, Smith & Roberts (1993)), we intend to use the Gibbs sampler to design the model. Dynamic models are frequently derived in terms of the conditional updates. Whilst this construction is appealing, the marginal density (in cases where the time series is strictly stationary) of the observations belongs to a family of densities which is typically unknown. Our construction allows this family to be specified.

A key paper on the construction of stationary time series with specified marginal densities is Lawrence & Lewis (1985) and the debate about the starting point for modelling time series appears in the discussion of Lawrence & Lewis (1985). There is also an interesting discussion of stationary time series in Joe (1996) and in McDonald & Zucchini (1997, Chapter 1). The current state of the art in the construction of such models, from a probabilistic perspective, may be found in Jorgensen & Song (1998).

In a manner similar to Lawrence & Lewis (1985), we consider strictly stationary time series with a known marginal density and with linear expectations. Statistically this has several advantages. We can assign the correct stationary density to the initial observation, leading to efficient likelihood estimation. The linear expectations also enable the autocorrelations of the series to be easily obtained. In addition, point forecasting, several steps ahead, is also possible due to the linear expectations. The marginal interpretation of the model, particularly if exogenous covariates are included, is greatly facilitated by the fact that the marginal density of our observations is known. As for standard Gaussian dynamic models, the introduction of time dependence provides a direct generalisation of the assumptions of independence whilst retaining a specified density. These advantages will be made more explicit later in the paper.

In applications, there may be reasons for assuming a particular marginal density. Lawrence & Lewis (1985) consider a marginal Gamma density for wind speeds, as this appears to fit the data (once detrended) well marginally. For stationary processes, this forms a coherent modelling strategy as the marginal density may be suggested by a histogram and the dynamics, suggested by the correlogram, may then be incorporated into the model. In financial applications, the marginal density of stock and exchange rate returns is of interest. It is often argued, see for instance Campbell et al. (1997, Chapter 1), that whilst the first few moments (including the second moment) of returns are finite, many of the higher moments may not be. In particular, Praetz (1972) and Blattberg & Gonedes (1974) both model returns as independently identically distributed arguing that the scaled Student-t distribution is appropriate as a marginal model for both stock price indices and individual stock prices. This marginal density will be exploited in Sections 2·5 and 3.

The structure of this paper is as follows. In Section 2 we derive first-order autoregressive processes with specified marginal families. These processes may be used to directly model our observations in an analogous way to Gaussian autoregressive models. Alternatively, we may use the process to specify a subordinate or latent structure. The models therefore belong to the class of non-Gaussian state space models, see West & Harrison (1997). Within this section, we consider general autoregressive results for the exponential dispersion family, together with two examples from this family: Poisson and gamma marginals. We also consider inverse-gamma marginals. The stochastic volatility example, a non-Gaussian state space model, is considered in Section 2·5.

In Section 3, we examine the construction of models of ARCH(1)[1] type. For the changing variance model, we obtain a closed form update $f(y_{t+1}|y_t)$. The model is linear in y_t^2 and the marginal density of y is a scaled Student-t distribution. Other models of ARCH(1) structure are examined and a general result for the exponential family is established in Section 3·3. Section 4 deals with models of GARCH(1,1) type, allowing direct feedback of the observations from the preceding time step. Finally, in Section 5, a non-Gaussian state space model is proposed for modelling a series of dependent counts. This model, introduced in Section 2, is estimated and assessed using efficient Markov chain Monte Carlo sampling and recent particle filtering methods respectively. Finally we conclude in Section 6.

Throughout this paper, we provide an innovative approach to introducing strictly stationary processes. Some of these processes, such as the Poisson process of Section 2·2 are known but most, to our knowledge, are not. In particular, we attempt to unify the derivation of such models. Before considering the first-order autoregressive processes of Section 2, we outline our approach in the following section, Section 1·1, by examining some general properties of the Gibbs sampler. In particular, we show the standard Gaussian AR(1) process may be derived via the Gibbs sampler.

1·1 *Gibbs sampler*

In this paper we use the Gibbs sampler as a tool to construct models. The Gibbs sampler has been widely used in recent years for Bayesian inference, the literature beginning with Gelfand & Smith (1990). Reviews of the Gibbs sampler may be found in Tierney (1994), Chib & Greenberg (1995) and Gilks et al. (1996).

Suppose we wish to generate observations $y_1, y_2, \ldots$ from a Markov chain which starts in

[1] Autoregressive conditional heteroskedastic models, see Engle (1982).

equilibrium and has stationary density $f_Y(y)$. In order to do this we introduce an auxiliary variable x and consider the joint density

$$f_{X,Y}(x,y) = f_Y(y) f_{X|Y}(x|y).$$

Clearly the marginal density for y is $f_Y(y)$. Also,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

If we generate $y_1 \sim f_Y(y)$, then $x_1 \sim f_{X|Y}(x|y_1)$, $y_2 \sim f_{Y|X}(y|x_1)$ and so on, then $\{(x_t, y_t)\}$ will yield a stationary sequence from $f_{X,Y}(x,y)$. In particular, $\{y_t\}$ will form a stationary Markov sequence from $f_Y(y)$. The sequence of sampling recursively from each conditional density is known as the Gibbs sampler. It may be seen that the transition density of our Markov chain for y is

$$f(y_{t+1}|y_t) = \int f_{Y|X}(y_{t+1}|x) f_{X|Y}(x|y_t)\, dx.$$

In this paper, we construct a Markov chain for y_t which leads to plausible time series models, which may then be estimated and tested. We are regarding the y_t as observations. The "auxiliary" variable x plays no role in the model, except as a tool for obtaining a transition density $f(y_{t+1}|y_t)$, which ensures a marginal density of a specified form. The usual properties of the Gibbs sampler remain; in particular, the Markov chain is reversible.

As a simple example, let us consider the Gaussian density for our marginal. Thus we wish to form a Markov chain y_t with stationary density $f_Y(y) = N(y; \mu, \sigma^2)$. We introduce an auxiliary variable x and let $f_{X|Y}(x|y) = N(x; \phi y, 1)$. Thus

$$\begin{aligned} f_{X,Y}(x,y) &= f_Y(y)\, f_{X|Y}(x|y) \\ &= N(y; \mu, \sigma^2) \times N(x; \phi y, 1). \end{aligned} \quad (1{\cdot}1)$$

Therefore,

$$f_{Y|X}(y|x) = N\left(y; x\rho/\phi + (1-\rho)\mu, \rho/\phi^2\right),$$

where $\rho = \sigma^2\phi^2/(1+\sigma^2\phi^2)$. If we consider a Gibbs sampler generating $y_1 \sim f_Y(y)$, $x_1 \sim f_{X|Y}(x|y_1)$, $y_2 \sim f_{Y|X}(y|x_1)$ and so on, then the transition density for y is

$$\begin{aligned} f(y_{t+1}|y_t) &= \int f_{Y|X}(y_{t+1}|x) f_{X|Y}(x|y_t)\, dx \\ &= N\left(y_{t+1}; (1-\rho)\mu + \rho y_t, (1-\rho^2)\sigma^2\right). \end{aligned}$$

That is,

$$y_{t+1} = \mu(1-\rho) + \rho y_t + \varepsilon_t, \quad (1{\cdot}2)$$

where $\varepsilon_t \sim \mathrm{N}(0,(1-\rho^2)\sigma^2)$ and $y_1 \sim \mathrm{N}(\mu,\sigma^2)$. This is clearly the standard Gaussian AR(1) process. In this case, our approach of constructing a stationary Markov chain for a Gaussian target density is not necessary. However, the construction of stationary autoregressive processes without the use of the Gibbs sampler would be difficult for many of the examples we consider later on.

2 AUTOREGRESSIVE PROCESSES

In this section, we present some examples of the construction of stationary Markov chains via the Gibbs sampler. We consider models of the form

$$y_{t+1} = \mu(1-\rho) + \rho y_t + \varepsilon_t, \tag{2·1}$$

where ε_t is a martingale difference sequence. We show that the model in (2·1) may be used to generate any distribution which is a member of the exponential family. The general approach for the exponential dispersion family is discussed in Section 2·1. Two examples which lie in this family, the Poisson and Gamma, are then considered. We also focus on the inverse Gamma marginal in Section 2·4.

2·1 *Exponential dispersion family*

Recently, Joe (1996) developed and constructed an AR(1) model with marginal distribution a member of the infinitely divisible convolution-closed exponential family (see also Jorgensen & Song (1998)). Here we demonstrate the construction has a Gibbs sampler representation. Consider the exponential dispersion model with density function

$$f_Y(y;\theta,\lambda) = c(y;\lambda)\exp\{y\theta - \lambda\kappa(\theta)\},$$

where $y \in \mathbb{R}$ and $\lambda \in \Lambda \subset \mathbb{R}_+$. We introduce an auxiliary variable x, as in (1·1), and define the joint density

$$f_{X,Y}(x,y) = f_Y(y;\theta,\lambda) f_{X|Y}(x;\lambda_1,\lambda_2,y)\,,$$

where

$$f_{X|Y}(x;\lambda_1,\lambda_2,y) = \frac{c(x;\lambda_1)c(y-x;\lambda_2)}{c(y;\lambda)}$$

and $\lambda_1 = \alpha\lambda$, $\lambda_2 = (1-\alpha)\lambda$ with $0<\alpha<1$. The full conditional for $f_{X|Y}(\cdot|\cdot)$ is given above, and we write a random variable from this conditional density as $A(y,\alpha)$, which is a 'thinned' variable (see Joe (1996); Jorgensen & Song (1998), for further details). The full conditional for $f_{Y|X}(\cdot|\cdot)$ can be obtained from the constructive form

$$y = x + \epsilon$$

where ϵ, independent of x, has density function $f_Y(\epsilon; \theta, (1-\alpha)\lambda)$. Consequently, the Gibbs sampler can be constructed via

$$y_t = A_t(y_{t-1}, \alpha) + \epsilon_t,$$

which is precisely the construction of Joe (1996). According to Joe (1996), $\mathrm{E}\{A_t(y, \alpha)\} = \alpha y$ so

$$\mathrm{E}(y_{t+1}|y_t) = \alpha y_t + (1-\alpha)\lambda\kappa'(\theta),$$

and the autocorrelation function is given by α^r. In this respect, it is seen that our approach for constructing first-order stationary processes based on the Gibbs sampler generalises the approach of Joe (1996).

2·2 *Poisson marginals*

Here we consider the joint density, through the introduction of an auxiliary variable x, given by

$$\begin{aligned} f_{Y,X}(y, x) &= f_Y(y) f_{X|Y}(x|y) \\ &= Po(y; \lambda)\ bi(x; y, p), \end{aligned}$$

where bi represents the binomial distribution and Po the Poisson distribution. Hence, constructively, $y_{t+1} = x + \Pi(\overline{p}\lambda)$, where $\overline{p} = 1 - p$ and $\Pi(\alpha)$ denotes a Poisson random variable with mean α. Hence, if $B(n, p)$ denotes a binomial random variable with parameters n and p, then

$$y_{t+1} = B(y_t, p) + \Pi(\overline{p}\lambda).$$

This model has an easy to interpret form as a "simple-death with immigration" process. If the population at time t is y_t, the population at time $t+1$ is the sum of the remaining alive members of the population from y_t, that is $B(y_t, p)$, and the new members who survive a unit of time, which is the $\Pi(\bar{p}\lambda)$ random variable.

2·3 *Gamma marginals*

Here we look for a stationary sequence with a gamma marginal distribution, say $ga(y; \nu, \beta)$. We consider the joint density

$$\begin{aligned} f_{Y,X}(y, x) &= f_Y(y) f_{X|Y}(x|y) \\ &= ga(y; \nu, \beta)\, Po(x; \alpha y). \end{aligned}$$

Then

$$y_{t+1} \sim ga(\nu + \Pi(\alpha y_t), \beta + \alpha),$$

where, again, $\Pi(\lambda)$ denotes a Poisson random variable with mean λ. Therefore,

$$\mathrm{E}(y_{t+1}|y_t) = \mu(1-\rho) + \rho y_t,$$

where $\rho = \alpha/(\beta+\alpha)$ and $\mu = \nu/\beta$. It is immediate that the autocorrelation function of y_t at lag τ is ρ^τ.

2·4 *Inverse Gamma marginals*

Here we consider a stationary sequence with an inverse-gamma marginal distribution, $Ig(y;\nu,\beta)$. Note, this is not a member of the exponential family. Nevertheless, we can introduce a conjugate style auxiliary variable, x. Consider the joint density,

$$\begin{aligned} f_{Y,X}(y,x) &= f_Y(y) f_{X|Y}(x|y) \\ &= Ig(y;\nu,\beta)\ ga(x;\alpha,y^{-1}). \end{aligned}$$

We obtain $f_{Y|X}(y_{t+1}|x) = Ig(y_{t+1};\nu+\alpha,\beta+x)$ and hence

$$\mathrm{E}(y_{t+1}|y_t) = \frac{\beta+\alpha y_t}{\nu+\alpha-1} = \mu(1-\rho)+\rho y_t,$$

where $\rho = \alpha/(\nu+\alpha-1)$ and $\mu = \beta/(\nu-1)$. and μ is the mean of the marginal distribution. The autocorrelation function of y_t is provided by $\rho(\tau) = \rho^\tau$.

Consequently, we have a first order autoregressive model which is linear in expectation and has inverse-gamma marginals. This provides a useful evolution equation for the variance in a stochastic volatility model.

2·5 *State space form*

It is possible to use our autoregressive models as latent or subordinated processes for the evolution of a parameter. These models are therefore of a *state space* type, see for example West & Harrison (1997) and Harvey (1993). Examples of such models include the stochastic volatility model, see Shephard (1996), and the conditional Poisson models considered by Durbin & Koopman (1997). Specifically, we are considering models of the type

$$\begin{aligned} y_t &\sim f(y_t|s_t) \\ s_{t+1} &\sim f(s_{t+1}|s_t), \end{aligned}$$

so that the observation y_t is independent conditional upon the corresponding state s_t. Here our notation has changed. We keep y as the observation variable. Now s is the state parameter, is unobserved, and for the models we consider, is a Markov process of the kind considered in Section 2.1, where it was labelled y. We shall illustrate by examining the construction of a stochastic volatility model. In Section 5, we consider the estimation of a state space model for count data.

The discrete time stochastic volatility model, see Jacquier et al. (1994) and Shephard (1996), has been used to model the clustering of variance over time for returns. The model assumes that the log variance follows a Gaussian autoregressive process. A notable exception to this assumption in continuous time is given by Barndorff-Nielsen & Shephard (2001). The implicit assumption of the standard model is that marginally the variance is log-normally distributed. The competing GARCH models have linear dynamics in the variance but the stationary density is unknown.

To illustrate the latent models obtainable from constructions described in Section 2, let us consider the volatility model, writing $s = \sigma^2$,

$$\begin{aligned} y_t &\sim N(0, \sigma_t^2) \\ \sigma_{t+1}^2 &= \mu(1-\rho) + \rho\sigma_t^2 + \xi_t, \end{aligned}$$

where the evolution for σ_t^2 is as defined in Section 2·4. In fact, we use a special case of inverse-gamma model, $Ig(\nu/2, \beta^2/2)$. The point here is that marginally $y_t \sim \beta t_\nu$, a scaled Student-t distribution. Since $y_t \sim N(0, \sigma_t^2)$, we have $y_t^2 = \sigma_t^2 \varepsilon_t^2$, where $\varepsilon_t^2 \sim \chi_1^2$. Thus $y_t^2 = \sigma_t^2 + v_t$, where $v_t = y_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1)^2$. Consequently, we may obtain the autocorrelation function of y_t^2, see Harvey (1993, page 31), as

$$\rho(\tau) = \frac{\rho^\tau}{(1 + \sigma_V^2/\sigma_1^2)},$$

where $\rho = 2\alpha/(\nu + 2\alpha - 2)$, $\sigma_1^2 = 2\beta^4/\{(\nu-2)^2(\nu-4)\}$ and $\sigma_V^2 = 2\beta^4/\{(\nu-4)(\nu-2)\}$, leading to $\rho(\tau) = \rho^\tau/(\nu-1)$, provided $\nu > 4$, the requirement for the variance of σ_t^2 to exist. The autocorrelation function clearly has an interpretable form. As the degrees of freedom $\nu \to \infty$, thinner tails (until y_t is marginally Gaussian), ρ becomes smaller, as does the baseline multiplier $(\nu-1)^{-1}$. As ν becomes smaller, the baseline autocorrelation becomes larger. The autocorrelation function in y^2 is equivalent to an ARMA(1,1) model. This is the same as for the highly successful GARCH(1,1) model. However, our model is more akin to the stochastic volatility models, in the sense that we have a separate noise term in the evolution of the variance. In the stochastic volatility (1) model, the marginal distribution however is a mixture of a normal with the log-variance being normal.

2·6 *Estimation*

The estimation of the autoregressive processes is not particularly difficult from a Bayesian perspective, since we can sample the auxiliary variables given the parameters and the data. Then we can sample the parameters given both the auxiliary variables and the data. Hence, we can us

Markov chain Monte Carlo methods for sampling from posterior distributions. The estimation methods for non-Gaussian state space models, however, are non-trivial. The posterior correlation between the states and the parameters can cause Markov chain Monte Carlo methods to mix poorly. Blocking methods, such as those used in Shephard & Pitt (1997), which sample large parts of the state space simultaneously can, to some extent, circumvent such problems. For example, in the stochastic volatility case is Section 2.3, we would wish to sample from the conditional posterior of a block of states, say $f(\sigma_t, \ldots, \sigma_{t+k}|\sigma_{t-1}, \sigma_{t+k+1}, \psi, y_t, \ldots, y_{t+k})$, where $\psi = (\nu, \beta, \alpha)'$. In Shephard & Pitt (1997), a Gaussian proposal formed by taking a second order Taylor expansion around the mode of this conditional posterior is used a an independence Metropolis-Hastings proposal. However, due to the fact that the state evolution is no longer Gaussian in the models we are now considering, it is unlikely that large blocks would be accepted frequently. The classical estimation approach via importance sampling (Durbin & Koopman (1997)) would encounter similar problems. Therefore, we employ a different Markov chain Monte Carlo strategy in this paper. This is illustrated in Section 5 for a count data which arises from a Poisson distribution whose intensity parameter is determined by a gamma autoregression.

3 ARCH(1)-TYPE MODELS

Here we consider the ARCH(1) class of models, first introduced for changing variance by Engle (1982). It is generally assumed that the conditional distribution of the measurement, y, is known, conditional on the parameter θ, for all time points t, say $f(y_t|\theta_t)$. We consider a Markov chain of the type shown by the top graph of Figure 1, where σ^2 replaces θ. Recent ideas concerning Markov chains of ARCH type, particularly for modelling volatility appear in Barndorff-Nielsen (1997) and Vidoni (1998). In neither of these cases is there a known stationary form for the evolution.

The simplest ARCH model is the ARCH(1). This has an explicit conditional density update, given by

$$y_{t+1} \sim N(0, a + by_t^2), \quad a > 0, b \geq 0,$$

the parameter constraints ensuring that the variance remains positive. This can be considered in two stages, where $y_t \sim N(0, \sigma_t^2)$ and $\sigma_{t+1}^2 = a + by_t^2$. This updating allows the dependence representation shown by the top graph of Figure 1. However, it should be emphasised that the update for $\sigma_{t+1}^2|y_t$ is deterministic rather than stochastic. The attractive statistical features of the ARCH class of models is that the likelihood is available directly via the prediction decomposition. Estimation and testing are therefore quite straightforward. The stationary density for the ARCH(1) model, when it exists, is unknown. Nelson (1990) showed that the condition necessary for covariance stationarity for the ARCH(1) model is $b < 3.5622$. The ARCH class is generally thought not to allow for sufficient kurtosis to adequately model the marginal density of financial data; see, for example, Bollerslev (1987), and the discussion of Shephard (1996, Section 1.2.2). Bollerslev (1987) notes from the literature that "the general conclusion to emerge from most of these studies is that speculative price changes ... are well described by a unimodal symmetric distribution with fatter tails than the normal". In particular, Praetz (1972) and Blattberg & Gonedes (1974), argue that the scaled Student-t distribution is appropriate as a marginal model for both stock price indices and individual stock prices. Bollerslev (1987) introduced a Student-t for the *conditional* density of the ARCH model. However, this of course, does not lead to a marginal Student-t distribution.

The ARCH(1) volatility model provides the first example of our approach for a class of models for which the conditional density $f(y_t|\theta_t)$ is regarded as known. After consideration of a duration model of this type, we shall, in Section 3·3, provide a version for the exponential family.

General approach In the general form of the ARCH(1) model we consider here we have our observation y_t depending only upon a time varying parameter $s_t = \theta_t$. We now form our model in a similar manner to the methodology of Section 1·1. We consider the joint density

$$f_{\Theta,Y}(\theta, y) = f_{Y|\Theta}(y|\theta) f_\Theta(\theta).$$

In this case, the marginal density for y from this joint density is simply $f_Y(y)$, where

$$f_Y(y) = \int_\Theta f_{Y|\Theta}(y|\theta) f_\Theta(\theta) d\theta.$$

Let us suppose we also have the form of conditional density,

$$f_{\Theta|Y}(\theta|y) = \frac{f_{Y|\Theta}(y|\theta) f_\Theta(\theta)}{f_Y(y)}.$$

As in Section 1·1 we consider the Gibbs sampler approach as generating $y_0 \sim f_Y(y)$, $\theta_1 \sim f_{\Theta|Y}(\theta|y_0)$, $y_1 \sim f_{Y|\Theta}(y|\theta_1)$ and so on. Then $\{(\theta_t, y_t)\}$ will yield a stationary sequence from $f_{\Theta,Y}(\theta, y)$. In particular $\{y_t\}$ will yield a stationary sequence from $f_Y(y)$. This principle forms the basis of the following results.

3·1 *Volatility model*

We consider the joint density

$$\begin{aligned} f(y, \sigma^2) &= f(y|\sigma^2)\, f(\sigma^2) \\ &= N(y; 0, \sigma^2)\, Ig(\sigma^2; \nu/2, \nu\beta^2/2). \end{aligned}$$

Marginally, $f_Y(y) = t_\nu(y; 0, \beta^2)$, a scaled Student-$t$ distribution. We also have the conditional density $f(\sigma^2|y) = Ig(\sigma^2; \nu/2 + 1/2, \nu\beta^2/2 + y^2/2)$.

Following the Gibbs sampler construction, if $\sigma_0^2 \sim Ig(\nu/2, \nu\beta^2/2)$ and then we cycle through the two full conditionals, for $t = 1, \ldots, n$, we have constructed a stationary process for y_t, the unconditional density being a scaled Student-t; that is, marginally $y_t = \beta\, t_\nu$. This Gibbs sequence can itself be viewed as a model with a feedback interpretation[2]. Since $\{\ldots, y_t, \sigma_t^2, y_{t+1}, \ldots\}$ is a Markov chain, we can integrate σ_t^2 out to calculate the transition density

$$f(y_{t+1}|y_t) = \int f(y_{t+1}|\sigma_t^2) f(\sigma_t^2|y_t)\, d\sigma_t^2,$$

obtaining

$$y_{t+1} = \sqrt{\frac{y_t^2 + \nu\beta^2}{1+\nu}}\, s_{\nu+1}, \tag{3·1}$$

where $s_{\nu+1} \sim t_{\nu+1}$, a Student-t random variable with $\nu + 1$ degrees of freedom. In the original ARCH(1) model of Engle (1982) the $s_{\nu+1}$ random variable is replace with a standard normal random variable. Our model is therefore different from the ARCH(1) model of Engle (1982). However, our model arises as a restricted version of the extension provided by Bollerslev (1987). The t-ARCH(1) of Bollerslev (1987) is given by

$$y_{t+1} = \sqrt{a + by_t^2}\, s_m,$$

where $s_m \sim t_m$, a Student-t random variable with m degrees of freedom. Our t-ARCH model therefore arises by setting $a = \nu\beta^2/(1+\nu)$, $b = 1/(1+\nu)$ and constraining $bm = 1$. So a single restriction on the model of Bollerslev (1987) provides a known, scaled t marginal density. This new model is linear in y_t^2. We have

$$\mathrm{E}(y_{t+1}^2|y_t^2) = \frac{(y_t^2 + \beta^2\nu)}{(\nu - 1)},$$

which implies the autocorrelation function of y_t^2 is $\rho(\tau) = (\nu - 1)^{-\tau}$.

3·2 *Duration model*

Models for explaining times between successive trades have recently been of interest, due to the amount of data currently available on intra-daily market activity. The autoregressive conditional duration model of Engle & Russell (1998) has been introduced to explain the slowly changing mean structure of durations between trades. This work parallels the GARCH literature for changing volatility. Here we consider a Gibbs structure for constructing an observation driven model of this type.

[2]See the top graph in Figure 1. This gives the dependency structure for our ARCH(1) model. Note the arrows indicate the direction of dependence. These graphs are known as directed acyclic graphs (DAGs) in the statistical community.

The standard assumption made in this context is that y_i (the time between the ith and the $(i-1)$th trade) is exponential with parameter θ_i. The θ_i (as for the variance in the volatility model) is assumed to evolve over time. In the non-time varying context, it is often assumed the θ_i arise from a gamma distribution, $ga(\alpha, \beta)$ (see, for example, Lancaster (1990)). We will therefore take this as the marginal process for $\{\theta_i\}$. Now consider the joint density

$$\begin{aligned} f_{Y,\Theta}(y,\theta) &= f_{Y|\Theta}(y|\theta) f_\Theta(\theta) \\ &= Ep(y;\theta)\, ga(\theta;\alpha,\beta), \end{aligned}$$

where Ep denotes the exponential distribution. The conditional density if θ given y is $ga(\alpha + 1, \beta + y)$. The Gibbs sampler runs over the two conditional densities for θ_{i+1} and for y_{i+1}, iteratively. Further, it is clear that

$$\mathrm{E}(y_{i+1}|y_i) = \mathrm{E}(\theta_{i+1}^{-1}|y_i) = \frac{\beta + y_i}{\alpha},$$

similar to the specification of Engle & Russell (1998). We can integrate out θ, to obtain

$$f(y_{i+1}|y_i) = Gg(y_{i+1}; \alpha + 1, \beta + y_i, 1),$$

where Gg represents the gamma-gamma distribution (see, for example, Bernardo & Smith (1994)), allowing classical inference for the two parameters α and β via prediction decomposition.

3·3 *Exponential family*

More generally, we consider the situation when the conditional density of y_t given θ_t is a member of the exponential family;

$$f(y_t;\theta_t) = c(y_t) \exp\left\{y_t\theta_t - M(\theta_t)\right\},$$

so that $\mathrm{E}(y_t|\theta_t) = \mu(\theta_t) = M'(\theta_t)$. We choose $f(\theta_t)$ to be a member of the standard conjugate family, see Gutierrez-Pena & Smith (1997);

$$f(\theta_t) = \exp\left\{s\theta_t - \phi M(\theta_t) + \lambda(s,\phi)\right\},$$

where $\phi > 0$. Under mild regularity conditions (Diaconis & Ylvisaker (1979)), $\mathrm{E}\{\mu(\theta_t)\} = s/\phi$. Consequently,

$$\mathrm{E}(y_{t+1}|y_t) = \frac{s + y_t}{1 + \phi}$$

since

$$f(\theta_{t+1}|y_t) = \exp\{(s + y_t)\theta_{t+1} - (1+\phi)M(\theta_{t+1}) + \lambda(s + y_t, 1+\phi)\}$$

and $E(y_{t+1}|\theta_{t+1}) = \mu(\theta_{t+1})$. It is clear that we can calculate $f(y_{t+1}|y_t)$. As $\{\ldots, y_t, \theta_{t+1}, y_{t+1}, \ldots\}$ is a Markov chain we obtain,

$$
\begin{aligned}
f(y_{t+1}|y_t) &= \frac{f(y_{t+1}|\theta_{t+1}) f(\theta_{t+1}|y_t)}{f(\theta_{t+1}|y_t, y_{t+1})} \\
&= \frac{c(y_{t+1}) \exp\{\lambda(s + y_t, 1 + \phi)\}}{\exp\{\lambda(s + y_t + y_{t+1}, 2 + \phi)\}}
\end{aligned}
$$

and marginally $f_Y(y) = \exp\{\lambda(s, \phi)\}\, c(y_t) \exp\{-\lambda(s + y_t, 1 + \phi)\}$.

3·4 *Estimation*

Estimation for all models of the above type is reasonably straightforward via prediction decomposition, as $\{\ldots, y_t, y_{t+1}, y_{t+2}, \ldots\}$ is a Markov chain. Denoting the unknown fixed parameters by ψ, we have

$$
f(y;\psi) = f(y_1;\psi) \prod_{t=2}^{n} f(y_t|y_{t-1};\psi).
$$

The log-likelihood may be maximised with respect to ψ via usual numerical procedures. Note that likelihood inference is made more statistically efficient as we explicitly have the initial density $f(y_1|\psi)$ in the likelihood. This is a problem for ARCH and GARCH models as the unconditional density is unknown.

4 GARCH(1,1)-TYPE MODELS

The GARCH(1,1), generalised ARCH, model of Bollerslev (1986), is one of the most widely used for modelling volatility. The ARMA type structure allows sufficient memory for most volatility models applied to asset and exchange rate returns. There are many papers on the properties of the standard GARCH(1,1) model. A recent look at the existence of marginal moments and conditions for stationarity of the GARCH model is to be found in He & Tersvirta (1999). The conditions for strict stationarity are quite complicated and the marginal density, if one exists, is unknown. This is not the case for the models introduced in this section, which may usefully be thought of as GARCH processes with the addition of a (heteroskedastic) error term in the parameter evolution.

The GARCH(1,1) model differs from the state space formulation of Section 2·5 in that observations can feedback directly into the prediction of the next observation. The GARCH(1,1) model can be applied not only to volatility models, of course, but also to durations, see Engle & Russell (1998), count models and so on. Indeed, the GARCH model provides a general formation for the evolution of an unobserved parameter. The standard GARCH(1,1) model for volatility, Bollerslev (1986), is

$$y_t \sim N(0, \sigma_t^2)$$

$$\sigma_{t+1}^2 = a + by_t^2 + c\sigma_t^2.$$

Our models will differ from the above structure in that, in keeping with the spirit of this paper, the marginal density of y_t and σ_t^2 will be kept fixed and known. In the following subsections we consider the GARCH model for volatility, count data and the exponential family respectively.

4·1 *Volatility model*

Although the models introduced here are similar to GARCH(1,1), we are not able write down $f(y_{t+1}; \sigma_t^2, y_t)$ explicitly, as we can for the standard GARCH(1,1) model. Our aim is to find a model for which marginally y_t is Student-t and, if y_t conditional on σ_t^2 is $N(0, \sigma_t^2)$, then, for some a, b, c,

$$\mathrm{E}(\sigma_{t+1}^2 | y_t, \sigma_t^2) = a + by_t^2 + c\sigma_t^2.$$

Of course, we still need σ_t^2 to be a stationary $Ig(\nu/2, \nu/2)$ process. Following the development of the Gibbs sampler in Section 3, we introduce the joint density

$$f(y, \sigma, z) = N(y; 0, \sigma^2)\, Ig(\sigma^2; \nu/2, \nu\beta^2/2)\, ga(z; \alpha, \sigma^{-2}), \tag{4·1}$$

where z is an auxiliary variable; making no difference to our marginal process, but allowing longer range dependence. This GARCH model is based on a Gibbs sampler driven by the full conditionals $f(y|\sigma^2)$, $f(z|\sigma^2)$ and $f(\sigma^2|y, z)$. Adding time subscripts to explicitly model the process (illustrated in the middle graph of Figure 1) we update (from time t to $t+1$) as $f(\sigma_{t+1}^2 | y_t, z_t)$, $f(z_{t+1}|\sigma_{t+1}^2)$ and $f(y_{t+1}|\sigma_{t+1}^2)$. This satisfies our requirements, since we obtain

$$\mathrm{E}\left(\sigma_{t+1}^2 | y_t, \sigma_t^2\right) = \frac{y_t^2 + \nu\beta^2 + 2\alpha\sigma_t^2}{\nu + 2\alpha - 1}.$$

We have three parameters ν, β and α. The first two represent aspects of the marginal density, while again α represents persistence. When $\alpha = 0$, we obtain the same variance evolution as our t-ARCH model. As $\alpha \longrightarrow \infty$, it is clear we obtain increased dependence. It is immediate that we have $y_t = \beta\, t_\nu$ as the marginal density for observations. An approach to estimation of this model is outlined in the Appendix (Section 7·2). The update structure is given in the bottom graph of Figure 1.

4·2 *Poisson model*

In order to show the generality of this method, let us consider a GARCH(1,1) structure for a conditional Poisson model. We start with a joint density

$$f(y, \theta, z) = Po(y; \lambda)\, ga(\lambda; a, b)\, Po(z; \alpha\lambda),$$

so here the marginal density for y is Poisson-gamma, $Pg(a, b, 1)$, see Bernardo & Smith (1994, page 119). Adding time subscripts to explicitly model the process (illustrated in the middle graph of Figure 1 with λ replacing σ^2) we consider updating (from time t to $t+1$) via $f(\lambda_{t+1}|y_t, z_t)$, $f(z_{t+1}|\lambda_{t+1})$ and $f(y_{t+1}|\lambda_{t+1})$. We obtain

$$f(\lambda_{t+1}|y_t, z_t) = ga(\lambda_{t+1}; y_t + a + z_t, 1 + b + \alpha),$$

so $f(\lambda_{t+1}|y_t, \lambda_t)$ follows the evolution

$$\lambda_{t+1} \sim ga(\lambda_{t+1}; y_t + a + \Pi(\alpha\lambda_t), 1 + b + \alpha).$$

Therefore, we have

$$\mathrm{E}(\lambda_{t+1}|y_t, \lambda_t) = \frac{a + y_t + \alpha\lambda_t}{1 + b + \alpha}.$$

4·3 *Exponential family*

We now rely on the joint density

$$f(y|\theta)f(\theta)f(z|\theta).$$

More generally, we consider the situation when the conditional density of y_t given θ_t is from exponential family;

$$f(y_t; \theta_t) = c(y_t) \exp\left\{y_t\theta_t - M(\theta_t)\right\},$$

so $\mathrm{E}(y_t|\theta_t) = \mu(\theta_t) = M'(\theta_t)$. As previously, we choose the mixing density $f(\theta_t)$ to be a member of the standard conjugate family,

$$f(\theta_t) = \exp\left\{s\theta_t - \phi M(\theta_t) + \lambda(s, \phi)\right\}.$$

For the GARCH version, we introduce

$$f(z_t; \theta_t) = c_*(z_t) \exp\left\{z_t\theta_t - \rho M(\theta_t)\right\}.$$

Then

$$f(\theta_{t+1}|z_t, y_t) \propto \exp\{(s + y_t + z_t)\theta_{t+1} - (1 + \phi + \rho)M(\theta_{t+1})\}$$

and therefore

$$\mathrm{E}(y_{t+1}|y_t, \theta_t) = \mathrm{E}\,\mathrm{E}\{\mu(\theta_{t+1})|y_t, z_t\} = \frac{s + y_t + \rho\mu(\theta_t)}{1 + \phi + \rho}.$$

The stochastic volatility version follows by running the Gibbs sampler over $f(z, \theta)$ and drawing $f(y|\theta)$ at each iteration, see the bottom graph of Figure 1, replacing σ^2 by θ. In this case, we have

$$\mathrm{E}(\theta_{t+1}|\theta_t) = \frac{s + \rho\mu(\theta_t)}{\phi + \rho}.$$

We shall illustrate the application of our model and the associated methods by examining the data set considered by McDonald & Zucchini (1997, page 194-195). The data consists of weekly firearm homicides in Cape Town from 1 January 1986 to 31 December 1991 (313 observations in total), see Figure 4. These models are in the form of small counts. We shall assume that the observations are Poisson, conditional upon a parameter which evolves according to the gamma autoregressive process of Section 2·3. Formally we have,

$$\begin{aligned} y_t &\sim Po(x_t), & t = 1, \ldots, n \\ x_{t+1} &\sim ga(\nu + z_t; \beta + \alpha), & z_t \sim Po(\alpha x_t). \end{aligned} \qquad (5{\cdot}1)$$

Marginally, $x_t \sim ga(\nu, \beta)$ so the mean and variance of x_t are $\mu = \nu/\beta$ and $\sigma^2 = \nu/\beta^2$ respectively. The persistence parameter is $\rho = \alpha/(\beta + \alpha)$. Of course,

$$\mathrm{E}(x_{t+1}|x_t) = \frac{\nu + \alpha x_t}{\beta + \alpha} = \mu(1-\rho) + \rho x_t.$$

We may write $y_t = x_t + v_t$, where $v_t = y_t - x_t = \Pi(x_t) - x_t$, an uncorrelated zero mean process. Therefore the autocorrelation function of y_t is given by

$$\rho(\tau) = \frac{\rho^\tau}{1 + \beta}.$$

5·1 *Estimation Procedure*

Maximum likelihood for the model (5·1) is difficult due to the non-conjugate structure of the model. Quasi-maximum likelihood procedures are easier to apply. For instance, we can linearise the above system directly, obtaining,

$$\begin{aligned} y_t &= x_t + v_t, & \sigma_V^2 = \mu, \\ x_{t+1} &= \mu(1-\rho) + \rho x_t + \xi_t, & \sigma_\xi^2 = \sigma^2(1-\rho^2), \end{aligned} \qquad (5{\cdot}2)$$

where $\mu = \nu/\beta$ and $\sigma^2 = \nu/\beta^2$. We could then treat the martingale terms v_t and ξ_t as if they were independent Gaussian, $N(0, \sigma_V^2)$ and $N(0, \sigma_\xi^2)$. This would then allow maximum likelihood of the resulting state space form model to be carried out via the Kalman filter, see for example West & Harrison (1997). However, this approach would lead to biased estimation. For this reason, we choose a Bayesian Markov chain Monte Carlo approach for estimation of the parameters $\theta = (\nu, \beta, \alpha)'$ and the underlying state $x = (x_1, \ldots, x_n)'$. We obtain stationary samples from the posterior density $f(\theta, x|y)$. A thorough review of Markov chain Monte Carlo is given, for example, in Gilks et al. (1996). Here we outline the main steps of the Markov chain

Monte Carlo algorithm, relegating the details to the Appendix (Section 7.1). We retain our auxiliary variables $z = (z_1, \ldots, z_{n-1})'$ in the Markov chain Monte Carlo analysis. The sampling scheme works by iterating the following steps:

- $x \sim f(x|z, \theta; y)$:

$$f(x|z,\theta;y) = f(x_1|z_1,\theta;y_1) f(x_n|z_{n-1},\theta;y_n) \prod_{t=2}^{n-1} f(x_t|z_t, z_{t-1}, \theta; y_t).$$

  These univariate densities may be simulated from directly as they are of gamma form, see Appendix (Section 7.1).

- $z \sim f(z|x, \theta; y)$:

$$f(z|x,\theta;y) = \prod_{t=1}^{n-1} f(z_t|x_t, x_{t+1}, \theta).$$

  The univariate distributions may be bounded by a Poisson distribution allowing an efficient accept-reject algorithm, see Appendix (Section 7.1).

- $\theta \sim f(\theta|z, x; y)$:

$$\begin{aligned} f(\theta|z,x;y) &\propto f(z,x|\theta) f(\theta) \\ &= f(x_1|\theta) \prod_{t=1}^{n-1} f(x_{t+1}|z_t,\theta) f(z_t|x_t,\theta) f(\theta), \end{aligned}$$

  where $f(\theta)$ is the prior. This is the full conditional posterior for θ. A Student-t distribution is used as a proposal for each univariate element of $\theta = (\nu, \beta, \alpha)'$. This proposal is accepted or rejected as an independence Metropolis-Hastings candidate, see Appendix (Section 7.1).

- $\theta, x \sim f(\theta, x|z; y)$:

  We have

$$\begin{aligned} f(\theta,x|z;y) &= f(x|z,\theta;y) f(\theta|z;y) \\ &\propto f(y|x) f(z,x|\theta) f(\theta), \end{aligned}$$

  Therefore,

$$f(\theta|z;y) \propto \frac{f(y|x) f(z,x|\theta) f(\theta)}{f(x|z,\theta;y)}.$$

This last move allows reduced conditioning, effectively integrating out x, and is necessary to permit a reasonably efficient Markov chain Monte Carlo scheme. Again, a Student-t distribution is used as a proposal within an independence Metropolis-Hastings scheme.

5·2 *Results*

Let us denote the three parameter model given by (5·1) as $Pga(\nu,\beta;\alpha)$. We shall estimate this model for the data set of weekly firearm homicides described above. However, prior to this two simple alternative models were considered. In particular a simple one parameter Poisson model, $Po(\lambda)$, is considered together with a static Poisson model with gamma heterogeneity, $Pg(\nu,\beta)$. The maximum likelihood results for these models are given in Table 1. The diagnostics of the table will be examined in the following section. Note that there is a rise of 67.9 in the log-likelihood in going from the basic $Po(\lambda)$ model to the more general $Pg(\nu,\beta)$ distribution. As these are nested models, the usual χ^2 test is highly significant. Therefore, heterogeneity (or mixing) is present.

To estimate our $Pga(\nu,\beta;\alpha)$ model we transform to $(\mu,\sigma^2;\rho)$. For brevity, we display the resulting samples μ, the mean of x_t, resulting from our Markov chain Monte Carlo scheme in Figure 2. The Markov chain Monte Carlo sampler was run over 60 000 iterations. The scheme was efficient as can be seen from the, almost, independent draws of μ. The marginal posteriors for $(\nu,\beta;\alpha)'$ are given in Figure 2. The corresponding means and covariance matrix are given in Table 1. Figure 3 displays the marginal posteriors for $(\mu,\sigma^2;\rho)$, perhaps a more interpretable transformation. It can be seen that there is clear evidence of persistence in the time series. There is no appreciable marginal posterior mass for ρ below 0.65, however this parameter is also not very close to 1, indicating that shocks are not highly persistent.

The particle filter of Pitt & Shephard (1999) was applied to the time series. Representing $\mathcal{F}_t = (y_1,\ldots,y_t)'$, the observations up to and including that of time t, the particle filter draws samples from the filtering density $f(x_t|\mathcal{F}_t;\widehat{\theta})$. We fix $\widehat{\theta}$ at the posterior mean. In Figure 4, we display the filtered mean over time together with the posterior mean under the smoothing density $f(x_t|\mathcal{F}_n)$, obtained from the Markov chain Monte Carlo output. It can be seen from the top graph that both the filtered and smoothed estimates of the state follow the data closely. In addition, the bottom graph shows that the filtered and smoothed state means are in alignment. This is reassuring as a heuristic diagnostic since the filtered mean should be an unbiased estimate of the smoothed mean only if the model is correct.

The log-likelihood for our model may be efficiently estimated as a bi-product of the particle filter, see Pitt & Shephard (2000). We have an estimate of our one-step ahead prediction density,

$$\widehat{f}(y_{t+1}|\mathcal{F}_t;\widehat{\theta}) = \frac{1}{M}\sum_{k=1}^{M} f(y_{t+1}|x_{t+1}^k),$$

where $x_t^k \sim f(x_t|\mathcal{F}_t;\widehat{\theta})$ and $x_{t+1}^k \sim f(x_{t+1}|x_t^k;\widehat{\theta})$, $k = 1,\ldots,M$. We may now estimate the log-likelihood as,

$$\begin{aligned}\log L(\widehat{\theta}) &= \log f(y_1,\ldots,y_n|\widehat{\theta}) \\ &= \sum_{t=0}^{n-1} \log \widehat{f}(y_{t+1}|\mathcal{F}_t;\widehat{\theta}).\end{aligned}$$

This allows the calculation of the log-likelihood to be evaluated at the Bayesian mean $\widehat{\theta}$. In Table 1 we show the log-likelihood of our model. Interpreting classically, the likelihood ratio test indicates that the temporal parameter, α, is highly significant compared with the nested $Pg(\nu,\beta)$ model, for which $\alpha = 0$. This confirms the Bayesian result of the importance of the time series aspect.

5·3 *Diagnostics*

We now look for evidence of departures from our model. In particular, we look at tests under the null hypothesis that the model is correct, specifically that our observations at time t have associated density $f(y_t|\mathcal{F}_{t-1};\widehat{\theta})$. We have a discrete distribution function on the random variable Y_t, so $F(y_t) = P(Y_t \le y_t|\mathcal{F}_{t-1};\widehat{\theta})$. We are denoting our observations as y_t and the corresponding random variable as Y_t. We shall suppress the conditioning upon $\mathcal{F}_{t-1};\widehat{\theta}$ at this juncture since it does not affect our discussion. Suppose we sample u_t from

$$f(u_t|y_t) = \begin{cases} \frac{1}{\Pr(y_t)}, & F(y_t-1) < u_t \le F(y_t) \\ 0, & \text{otherwise}, \end{cases} \tag{5·3}$$

where $\Pr(y_t) = F(y_t) - F(y_t - 1)$. Note that the regions $[F(y_t-1),\, F(y_t)]$ provide a partition of $[0,1]$, for $y_t = 0,1,2,\ldots$. Under the hypothesis that $y_t \sim F(y_t)$, we have that

$$f(u_t) = \sum\nolimits_{y_t=0}^{\infty} f(u_t|y_t)\Pr(y_t) = 1,\ u_t \in [0,1],$$

as we have a partition of the space of u_t. Therefore, if our model, and parameters, are correct, $u_t \sim$U$(0,1)$, independently, over time. This forms the basis of our diagnostic tests. Any departure from our model, either marginally or dependence through time should be apparent by examining these residuals. We restrict ourselves to two tests. Firstly, we transform our u_t via $n_t = \Phi^{-1}(u_t)$. Under our null hypothesis these should be independently standard Gaussian. This allows us to perform a standard Portmanteau test for autocorrelation. Clearly, this allows us to test whether we have modelled temporal dependence adequately. Secondly, we perform a χ^2 test for goodness of fit on the u_t's, comparing with a standard uniform.

In order to sample u_t we need to calculate $F(y_t|\widehat{\theta})$. This is can be explicitly calculated for the $Po(\lambda)$ model and for the $Pg(\nu,\beta)$ distribution. Before commenting on the results, the details of estimating $F(y_{t+1}|\mathcal{F}_t;\widehat{\theta})$ for our $Pga(\nu,\beta;\alpha)$ time series model should be outlined. From the output of the particle filter which we recall yields samples from $f(x_t|\mathcal{F}_t;\hat{\theta})$, the filtering density, we have

$$\hat{F}(y_{t+1}|\mathcal{F}_t;\hat{\theta}) = \frac{1}{M}\sum_{k=1}^{M} F(y_{t+1}|x_{t+1}^k),$$

where $x_t^k \sim f(x_t|\mathcal{F}_t;\hat{\theta})$ and $x_{t+1}^k \sim f(x_{t+1}|x_t^k;\hat{\theta})$, $k = 1,\ldots,M$. Noting that $F(y_{t+1}|x_{t+1}^k)$ is just the distribution function for a Poisson, since $y_t \sim Po(x_t)$, we arrive at efficient estimates of the prediction distribution. This allows us to sample u_t via (5·3) for $t = 1,\ldots,n$.

The details out of the way, we can look at the results. The last two columns of Table 1 give the goodness of fit statistic for the three models and the Portmanteau statistic respectively. Corresponding to the table we have Figure 5, showing the Gaussian residuals, together with their correlogram and the standard quantile plot for the three models. It is clear from Table 1 that the simple independent Poisson model fails both tests at high significance. The independent Poisson-gamma model fails to capture dependence but does provide a good overall fit, indicated by the fact that at the 5% level the goodness of fit test is passed. This suggests that the marginal model of Poisson-gamma is a sensible choice. Our three parameter Poisson-gamma time series model passes both tests with large p-values. The residuals in Figure 5 indicate hardly any evidence of correlation over time and the quantile plot is close to linear. This indicates that marginally and conditionally our model fits the data well.

Poisson model $Po(\lambda)$

Par	ML	Var			$\log L$	GOF χ^2_{19} (p)	Q χ^2_{15} (p)
λ	2.6198	0.00837			-719.36	77.09 (0)	281.51 (0.0)

Poisson-gamma model $Pg(\nu,\beta)$

Par	ML	Var			$\log L$	GOF χ^2_{19} (p)	Q χ^2_{15} (p)
ν	2.1959	0.1108	0.0422		-651.5	28.40 (0.076)	216.93 (0.0)
β	0.8382	0.0422	0.0180				

Poisson-gamma stationary model $Pga(\nu,\beta;\alpha)$

Par	ML	Var			$\log L$	GOF χ^2_{19} (p)	Q χ^2_{15} (p)
ν	2.7769	0.28254	0.1006	0.5855			
β	1.0040	0.1006	0.0460	0.17188	-617.07	21.88 (0.290)	14.393 (0.496)
α	6.1773	0.5855	0.17188	6.3061			

Table 1: *Estimation results for the firearms homicide dataset. The goodness of fit statistics and the Portmanteau tests on the correlogram are reported for all three models.*

In this paper we have shown how it is possible to assign a fixed marginal density to a time series model and to build the dependence structure around this. We have considered three different classes of models; the ARCH(1), latent AR(1) and GARCH(1,1) classes. There is a degree of flexibility in this approach as we may choose which marginal family is most appropriate for the problem under consideration. The incorporation of dependence whilst retaining the marginal density clearly provides a natural generalisation over treating observations as independent. In addition, the specification of the unconditional density means that the assignment of a density for our initial observation is trivial. This can be more problematic for the observation driven models such as the GARCH(1,1) for which the unconditional density is unknown.

The models obtained in this paper are all linear as the expectation of future observations is linear as a function of the previous parameters and observations. This lends interpretability to the models and allows direct evaluation of the autocorrelation. The inclusion of covariates has not been considered in this paper but is quite straightforward. In the example considered in Section 5 it was found that a simple three parameter state space model provided a very good fit to the data, both marginally and temporally. The Markov chain Monte Carlo estimation procedure and the use of the particle filter provides straightforward estimation and evaluation for models of this type.

6·1 *Acknowledgements*

We wish to thank a referee for constructive comments on a previous version of the paper.

References

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist* **24**, 1–14.

Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-gaussian ou based models and some of their uses in financial economics. *J. R. Statist. Soc. B* forthcoming.

Bernardo, J. M. & Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley, Chichester.

Blattberg, R. C. & Gonedes, N. J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *J. Business* **47**, 244–280.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *J. Econometrics* **51**, 307–327. Reprinted as pp. 42–60 in Engle, R.F.(1995), *ARCH: Selected Readings*, Oxford: Oxford University Press.

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Rev. Economics and Statistics* **69**, 542–47.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey.

Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican* **49**, 327–35.

Diaconis, P. & Ylvisaker, D. (1979). Conjugate prior for exponential families. *Ann. Statist.* **7**, 269–281.

Durbin, J. & Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika* **84**, 669–84.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* **50**, 987–1007. Reprinted as pp. 1–23 in Engle, R.F.(1995), *ARCH: Selected Readings*, Oxford: Oxford University Press.

Engle, R. F. & Russell, J. R. (1998). Forecasting transaction rates: the autoregressive conditional duration model. *Econometrica* **66**, 1127–1162.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.

Gilks, W. K., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gutierrez-Pena, E. & Smith, A. F. M. (1997). Exponential and Bayesian Conjugate Families: Reviews and Extensions. *Test* **6**, 1–90.

Harvey, A. C. (1993). *Time Series Models*. Harvester Wheatsheaf, Hemel Hempstead, 2nd edition.

He, C. & Tersvirta, T. (1999). Properties of moments of a family of garch processes. *J. Econometrics* **92**, 173–192.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *J. Business and Economic Statist.* **12**, 371–417.

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability* **33**, 664–677.

Jorgensen, B. & Song, P. (1998). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability* **35**, 78–92.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge.

Lawrence, A. J. & Lewis, P. A. W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *J. R. Statist. Soc. B* **47**, 165–202.

McDonald, I. & Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, London.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334. Reprinted as pp. 176–192 in Engle, R.F.(1995), *ARCH: Selected Readings*, Oxford: Oxford University Press.

Pitt, M. & Shephard, N. (2000). Auxiliary variable based particle filters. In Doucet, A., de Freitas, J. F. G., & Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*. Cambridge University Press.

Pitt, M. K. & Shephard, N. (1999). Filtering via simulation based on auxiliary particle filters. *J. American Statistical Association* **94**, 590–599.

Praetz, P. (1972). The distribution of share price changes. *J. Business* **45**, 49–55.

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

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V., & Barndorff-Nielson, O. E., editors, *Time Series Models in Econometrics, Finance and Other Fields*, pages 1–67. Chapman & Hall, London.

Shephard, N. & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**, 653–67.

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

Tierney, L. (1994). Markov Chains for exploring posterior distributions (with discussion). *Ann. Statist.* **21**, 1701–62.

Vidoni, P. (1998). Proper dispersion state space models for stochastic volatility. University of Udine, working paper.

West, M. & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 2 edition.

7 APPENDIX

7·1 *Markov chain Monte Carlo for Poisson-gamma model.*

Here we describe the details of the Markov chain Monte Carlo procedure of Section 5·1. Recall we have

$$\begin{aligned} y_t &\sim Po(x_t), & t = 1, \ldots, n, \\ x_{t+1} &\sim ga(\nu + z_t, \beta + \alpha), & z_t \sim Po(\alpha x_t). \end{aligned}$$

Our full data likelihood is therefore,

$$f(y,z,x|\theta) \propto f(y_1|x_1)f(x_1|\theta)\prod_{t=1}^{n-1} f(y_{t+1}|x_{t+1})f(x_{t+1}|z_t,\theta)f(z_t|x_t,\theta).$$

From this the following conditional steps can be derived.

Step 1: Sampling from $f(x|z,\theta;y)$. We obtain the closed forms,

$$\begin{aligned} f(x_t|z_t,z_{t-1},\theta) &= ga\left(x_t|y_t+z_t+z_{t-1}+\nu;1+\beta+2\alpha\right), \\ f(x_1|z_1,\theta) &= ga\left(x_1|y_t+z_1+\nu,1+\beta+\alpha\right), \\ f(x_n|z_{n-1},\theta) &= ga\left(x_n|y_n+z_{n-1}+\nu,1+\beta+\alpha\right). \end{aligned}$$

Step 2: Sampling from $f(z|x,\theta;y)$. An accept-reject scheme is used for our discrete valued auxiliary variables z_t. We obtain

$$f(z_t|x_t,x_{t+1},\theta) \propto f(x_{t+1}|z_t,\theta)f(z_t|x_t,\theta).$$

We get the simplification,

$$\log f(z_t|x_t,x_{t+1},\theta) = c + z_t\log(\lambda_t) - \log(z_t!) - \log\Gamma(\nu+z_t),$$

where

$$\log(\lambda_t) = \log(\alpha x_t) + \log(\beta+\alpha) + \log(x_{t+1}).$$

Let us denote $l(z_t) = -\log\Gamma(\nu+z_t)$. Then we can form an efficient accept-reject algorithm, see Ripley (1987), as follows, noting that $l(z_t)$ is concave (and can therefore be bounded by a straight line from above),

$$\begin{aligned} \log f(z_t|x_t,x_{t+1},\theta) &= c + z_t\log(\lambda_t) - \log(z_t!) + l(z_t) \\ &\leq c + z_t\log(\lambda_t) - \log(z_t!) + l(\widehat{z}_t) + l'(\widehat{z}_t)(z_t-\widehat{z}_t) \\ &= k - \mu_t + z_t\log(\mu_t) - \log(z_t!) \\ &= k + \log Po(z_t;\mu_t), \end{aligned}$$

where $\log\mu_t = \log(\lambda_t) + l'(\widehat{z}_t)$ and we take the expansion point as $\widehat{z}_t = \alpha(x_t+x_{t+1})/2$. We therefore sample from $\mathrm{Po}(\mu_t)$ until acceptance with probability

$$\log\Pr(\text{accepting } z_t) = l(z_t) - l(\widehat{z}_t) - l'(\widehat{z}_t)(z_t-\widehat{z}_t).$$

In practice we found that this proposal was accepted more than 99% of the time.

Step 3: Sampling from $f(\theta|z,x;y)$. In order to perform parameter sampling efficiently we transform from $\theta = (\nu,\beta,\alpha)'$ to $\widetilde{\theta} = (\mu,\sigma^2,\rho)'$ where $\mu = \frac{\nu}{\beta}$, and $\sigma^2 = \frac{\nu}{\beta^2}$ and the persistence parameter, $\rho = \frac{\alpha}{\beta+\alpha}$. This transformation is done to reduce the dependency between the parameters. We place our weak prior on $\widetilde{\theta}$, $f(\widetilde{\theta})$.

$$\begin{aligned} f(\widetilde{\theta}|z,x;y) &\propto f(z,x|\widetilde{\theta})f(\widetilde{\theta}) \\ &= f(\widetilde{\theta})f(x_1|\widetilde{\theta})\prod_{t=1}^{n-1} f(x_{t+1}|z_t,\widetilde{\theta})f(z_t|x_t,\widetilde{\theta}). \end{aligned} \tag{7·1}$$

We sample the elements of $\widetilde{\theta}$ in turn drawing from $f(\mu|\sigma^2,\rho,z,x;y)$, $f(\sigma^2|\mu,\rho,z,x;y)$ and $f(\rho|\mu,\sigma^2,z,x;y)$, noting that each of these distributions can be derived, up to proportionality, from (7·1). In each case we iterate to the mode of the log of (7·1), recording the mode as m, and the second derivative of the log of (7·1) $l''(m)$. We propose by drawing from a t-distribution with degrees of freedom 5, mean m and variance $-1/l''(m)$. We accept with probability given by the usual independence Metropolis Hastings expression. In practice, we find that moves are frequently accepted, yielding an efficient algorithm.

Step 4: Sampling from $f(\theta,x|z;y)$. Despite the fact that the proposed moves in the above step are frequently accepted, it is still the case that, due to high posterior correlation between the states, x and z, and the parameters, $\widetilde{\theta}$, the resulting Markov chain Monte Carlo method mixes quite slowly. To combat this we effectively integrate out the x states and obtain samples from $f(\widetilde{\theta},x|z;y)$. This is possible as we know the form of $f(x|z,\widetilde{\theta};y)$, given previously as the product of Gamma densities. We have,

$$\begin{aligned} f(\widetilde{\theta},x|z;y) &= f(x|z,\widetilde{\theta};y)f(\widetilde{\theta}|z;y) \\ &\propto f(y|x)f(z,x|\widetilde{\theta})f(\widetilde{\theta}). \end{aligned}$$

Therefore,

$$f(\widetilde{\theta}|z;y) \propto \frac{f(y|x)f(z,x|\widetilde{\theta})f(\widetilde{\theta})}{f(x|z,\widetilde{\theta};y)}.$$

We iterate through the elements of $\widetilde{\theta}$ in turn drawing from $f(\mu,x|\sigma^2,\rho,z;y)$, $f(\sigma^2,x|\mu,\rho,z;y)$ and $f(\rho,x|\mu,\sigma^2,z;y)$. The Metropolis scheme is the same in each case so we shall detail sampling from $f(\mu,x|\sigma^2,\rho,z;y)$.

We have $f(\mu,x|\sigma^2,\rho,z;y) = f(x|z,\mu,\sigma^2,\rho;y)f(\mu|\sigma^2,\rho,z;y)$. Note that $f(\mu|\sigma^2,\rho,z;y) \propto f(\widetilde{\theta}|z;y)$, given above. Let us denote $l(\mu) = \log f(\mu|\sigma^2,\rho,z;y)$. We iterate to the mode of $l(\mu)$, m, recording $l''(m)$. We sample our proposal $\mu^p \sim g(\mu|m,v)$, a t-density with mean m, variance $v = -1/l''(m)$ and degrees of freedom 5. We then sample $x^p \sim f(x|z,\mu,\sigma^2,\rho;y)$. Letting μ and

x represent out current values, we accept the proposed values, μ^p and x^p with log probability,

$$\log P(\mu, x \to \mu^p, x^p) = l(\mu^p) - \log g(\mu^p|m, v) - l(\mu) + \log g(\mu|m, v).$$

We find that this independence Metropolis proposal accepts moves with high probability.

7·2 *Estimation for GARCH Processes*

The estimation procedure for the GARCH models is aided by the conditional updating structure of the latent variables. Thus the Gibbs sampling scheme can be directly applied. We propose an Markov chain Monte Carlo procedure which allows blocking for our GARCH structure models of Section 4. Note that in this case we had

$$f(y, \sigma, z) = N(y; 0, \sigma^2)\, Ig(\sigma^2; \nu/2, \nu\beta^2/2)\, ga(z; \alpha, \sigma^{-2}),$$

from which all our conditional densities arise. In order carry out a general Markov chain Monte Carlo method we actually retain the auxiliary structure to perform estimation, rather than attempting to integrate variables out analytically. Let us denote the unknown parameters by θ. Then we can perform the following sweeps on a system, see middle graph in Figure 1;

1. for $t = 1, \dots, n$ sample z_t from $f(z_t|\sigma_t^2, \sigma_{t+1}^2, y_t) \propto f(z_t|\sigma_t^2) f(\sigma_{t+1}^2|z_t, y_t)$.
2. For $t = 1, \dots, n$ sample σ_t^2 from $f(\sigma_t^2|y_t, y_{t-1}, z_t, z_{t-1}) \propto f(y_t|\sigma_t^2) f(\sigma_t^2|z_{t-1}, y_{t-1}) f(z_t|\sigma_t^2)$.
3. Sample θ from

$$f(\theta|y, z, \sigma^2) \propto f(\theta) f(z_1|\theta) f(y_1|\theta) \prod_{t=2}^{n} f(y_t|\sigma_t^2; \theta) f(\sigma_t^2|z_{t-1}, y_{t-1}; \theta) f(z_t|\sigma_t^2; \theta).$$

The dependence upon the fixed parameters θ has been suppressed from the notation of steps (1) and (2). Note that in steps (1) and (2) we are, in fact, sampling directly from the reduced conditionals $f(z|\sigma^2)$ and $f(\sigma^2|y, z)$ where $z = (z_1, \dots, z_n)'$ and $\sigma^2 = (\sigma_1^2, \dots, \sigma_n^2)'$. Fortunately the conditional independence structure means that this task reduces to univariate draws which are usually straightforward. Indeed the densities in (2) will usually be of standard form and the densities of (1) can generally be sampled from efficiently using accept-reject sampling. The procedure in (3) may be speeded up by using the reduced conditional densities $f(y_t|z_t)$ and $f(z_{t+1}|z_t, y_t)$ given for the specific forms of Section 4. In this case we use the following density for step (3);

$$f(\theta|y, z) \propto f(\theta) f(z_1|\theta) f(y_1|\theta) \prod_{t=2}^{n} f(y_t|z_t; \theta) f(z_t|z_{t-1}; \theta).$$

Note that $f(\theta)$ is the prior for θ, whilst $f(z_1|\theta)$, $f(y_1|\theta)$ are the stationary densities for the first latent variable and initial observation respectively. For large data sets a Gaussian proposal on these parameters (inside a Metropolis step) should work well as the posterior will be asymptotically Normal.

7·3 *GARCH volatility model*

To illustrate how steps (1) and (2) of the above Gibbs procedure can be carried out efficiently in practice we illustrate by examining the GARCH structure for volatility. In this case step (2) may be carried out directly as

$$f(\sigma_t^2|y_t, y_{t-1}, z_t, z_{t-1}) = Ig\left(\sigma_t^2|\frac{1}{2}(\nu + 4\alpha + 2); \frac{1}{2}(\nu\beta^2 + y_t^2 + y_{t-1}^2 + 2z_t + 2z_{t-1})\right).$$

Step (1) proceeds in the following manner. Let us examine the log-density

$$\log f(z_t|\sigma_t^2, \sigma_{t+1}^2, y_t) = c + \log f(z_t|\sigma_t^2) + \log f(\sigma_{t+1}^2|z_t, y_t).$$

Letting $a = \nu\beta^2/2 + y_t^2/2$, $b = \alpha + 1/2 + \nu/2$ and $d = \frac{\sigma_t^2+\sigma_{t+1}^2}{\sigma_t^2\sigma_{t+1}^2}$, we have

$$\log f(z_t|\sigma_t^2, \sigma_{t+1}^2, y_t) = c - dz_t + (\alpha - 1)\log z_t + b\log(z_t + a),$$

where support is on $z_t \geq 0$. By changing variable to $x = z_t + a$, $x > a$, we have

$$\begin{aligned} \log f(x|\sigma_t^2, \sigma_{t+1}^2, y_t) &= c - dx + (\alpha - 1)\log(x - a) + b\log x \\ &\leq c - dx + (\alpha - 1)\log x + b\log x = k + \log g(x). \end{aligned}$$

We may sample easily from the density $g(x)$ which is $ga(x; \alpha + b; d)$, truncated to be greater a, then accepting the sample x with probability

$$\exp[(\alpha - 1)(\log(x - a) - \log x))].$$

This is the rejection algorithm, see Ripley (1987). We may then transform back to get $z_t = x - a$, yielding a direct sample from the required posterior.

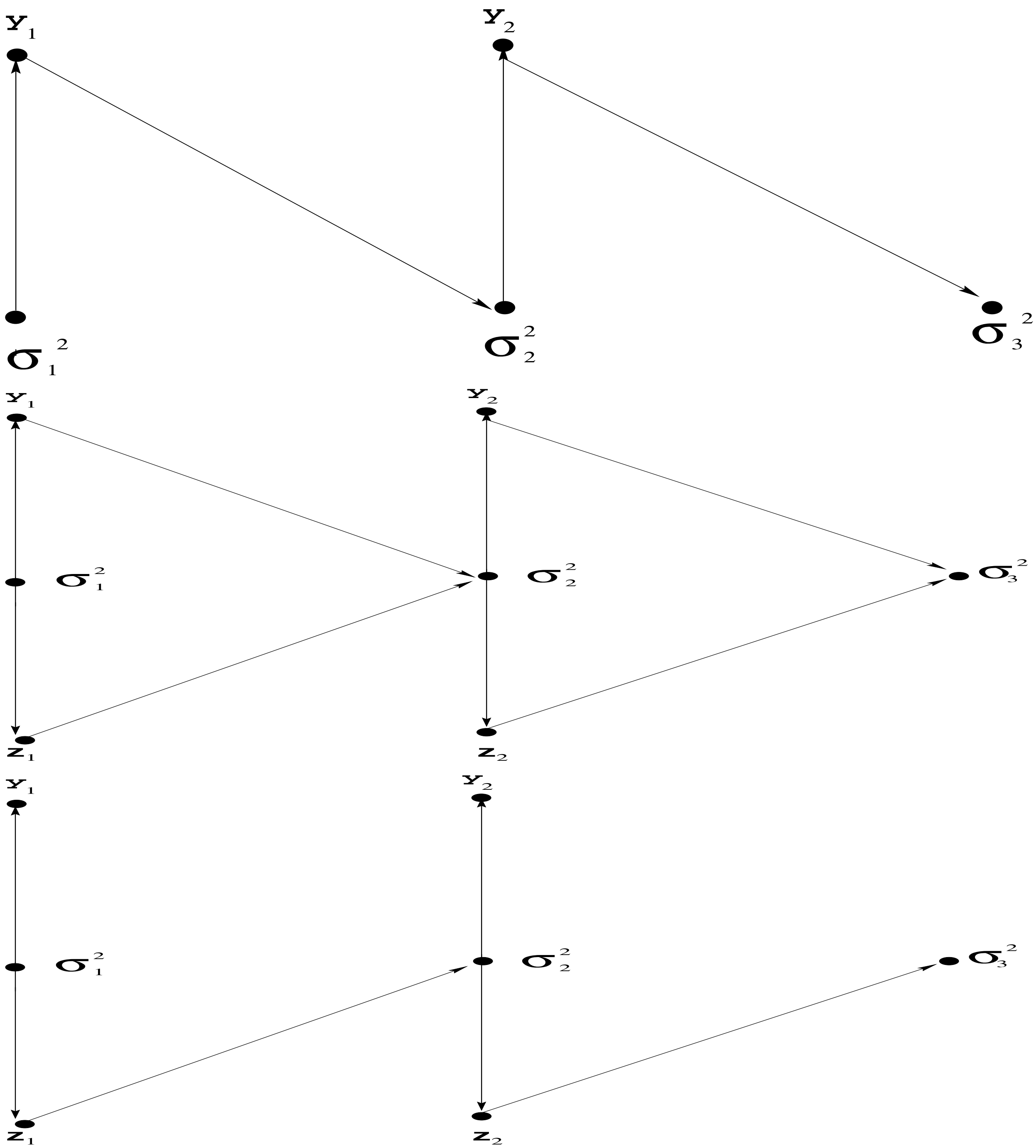

Figure 1: *Dependency structure for the ARCH(1), GARCH and Stochastic Volatility models.*

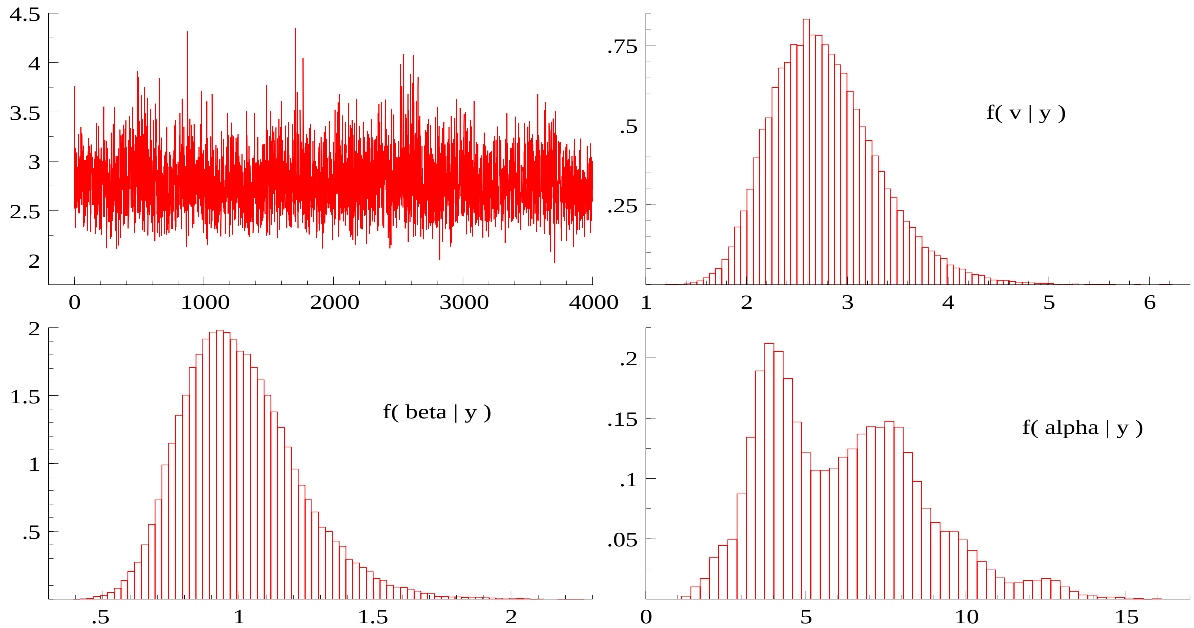

Figure 2: Top left: *60000 draws from the posterior of* μ. Top right: *The posterior of* ν. Bottom left: *The posterior of* β. Bottom right: *The posterior of* α.

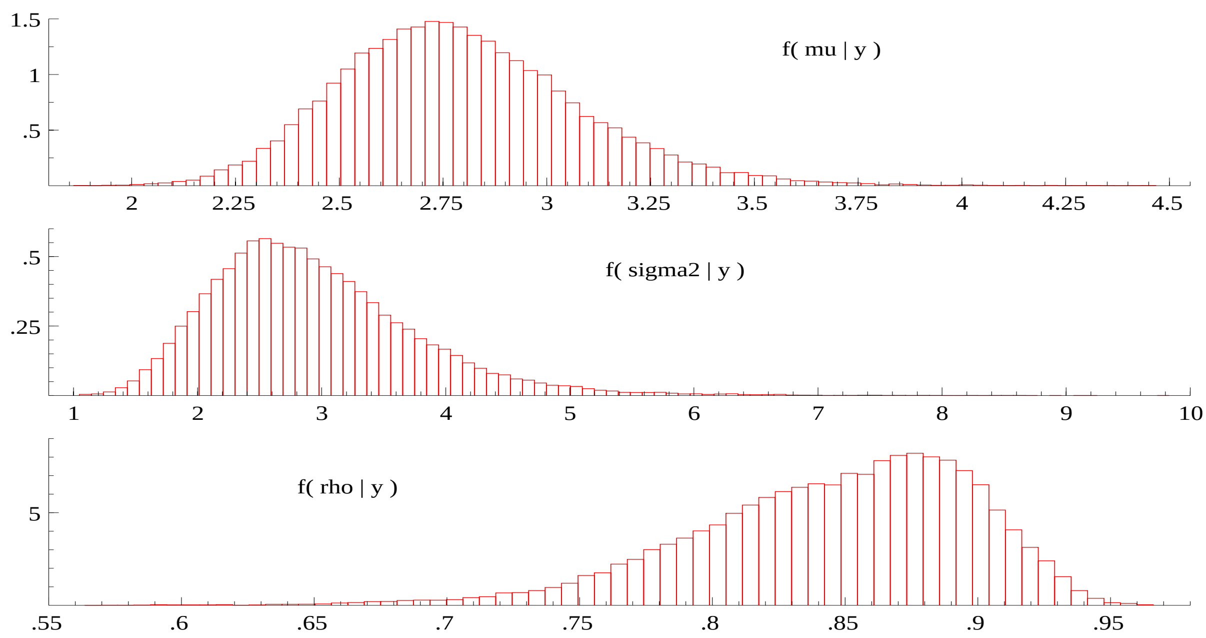

Figure 3: Top: *The posterior of* μ. Middle: *The posterior of* σ^2. Bottom: *The posterior of* ρ.

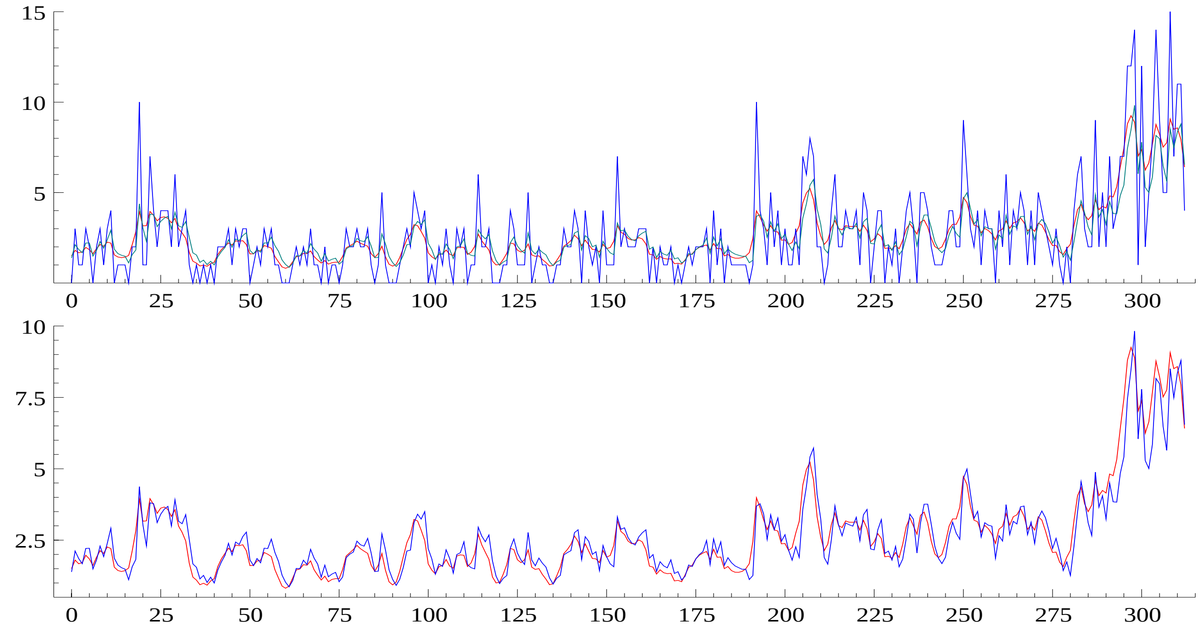

Figure 4: Top: *The number of firearm homicides per week in Cape Town from 1 January 1986 to 31 December 1991, together with (mean) smoothed and filtered estimates of the underlying Poisson parameter. Original Series: Mean=2.617, variance=6.568.* Bottom: *The mean of the filtered state (solid line) and the mean of the smoothed state (dashed line).*

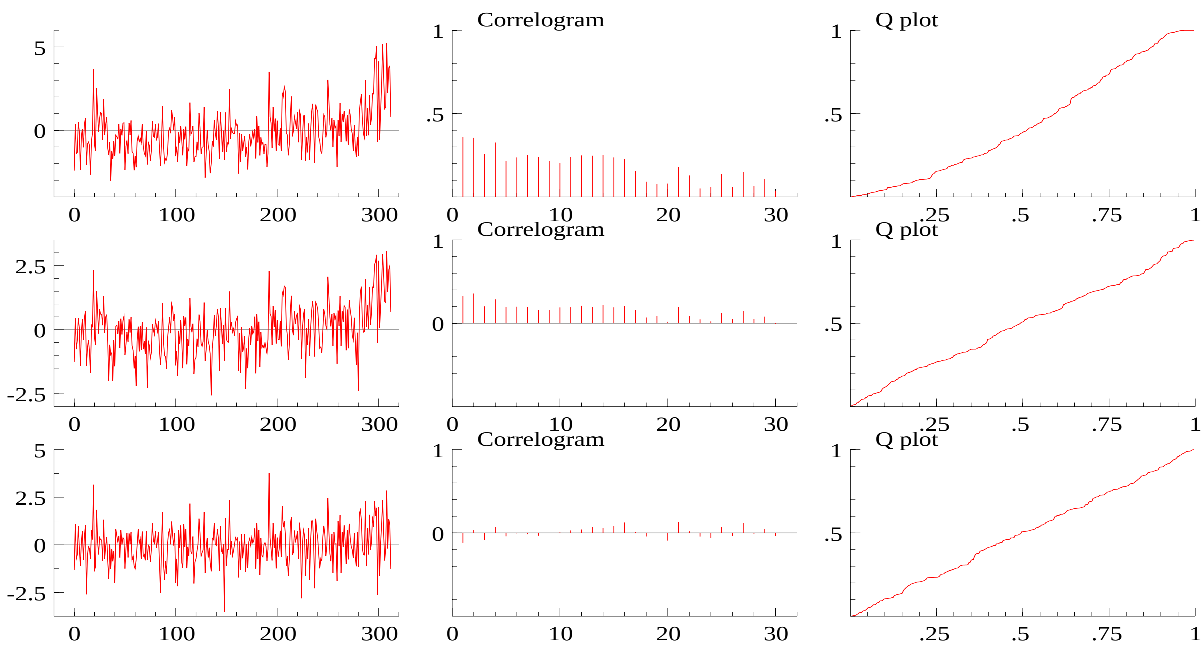

Figure 5: left to right, L: *The* n_t *residuals against time.* M: *The correlogram of the* n_t *residuals.* R: *The quantile plot of the* u_t residuals. TOP: *Po model.* MIDDLE: *Pg model.* BOTTOM: *Pga time series model.*