Model selection for time series of count data

Naif Alzahrani (Lancaster University), Peter Neal (Lancaster University)*,
Simon E.F. Spencer (University of Warwick), Trevelyan J. McKinley (University of Exeter)
and Panayiota Touloupou (University of Warwick)

January 8, 2018

Abstract

Selecting between competing statistical models is a challenging problem especially when the competing models are non-nested. An effective algorithm is developed in a Bayesian framework for selecting between a parameter-driven autoregressive Poisson regression model and an observation-driven integer valued autoregressive model when modeling time series count data. In order to achieve this a particle MCMC algorithm for the autoregressive Poisson regression model is introduced. The particle filter underpinning the particle MCMC algorithm plays a key role in estimating the marginal likelihood of the autoregressive Poisson regression model via importance sampling and is also utilised to estimate the DIC. The performance of the model selection algorithms are assessed via a simulation study. Two real-life data sets, monthly US polio cases (1970-1983) and monthly benefit claims from the logging industry to the British Columbia Workers Compensation Board (1985-1994) are successfully analysed.

Key words:autoregressive Poisson regression model; INAR model; INGARCH model; marginal likelihood; MCMC; particle filter.

MSC classification: 62M10; 62F15.

1 Introduction

There are a plethora of integer valued time series models for modelling low count time series data. There are two broad class of approaches for constructing integer valued time series models, observation-driven (*e.g.* McKenzie (2003); Neal and Subba Rao (2007); Enciso-Mora *et al.* (2009a)) and parameter-driven (*e.g.* Davis *et al.* (2003)) models, see Davis *et al.* (2015) for an overview. The INAR(p), the p^{th}

*Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, United Kingdom. email: p.neal@lancaster.ac.uk. tel: (+44) 1524 593848.

order integer autoregressive model is a prime example of an observation driven model. These models are motivated by real-valued time series models, primarily ARMA(p, q) (autoregressive-moving average) models and the desire to adapt such models to an integer-valued scenario. A time-series $\{X_t; t \in \mathbb{Z}\}$ is said to follow an INAR(p) model if, for $t \in \mathbb{Z}$,

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + Z_t, \tag{1.1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ are the autoregressive parameters, $\circ$ denotes a *thinning operator* and $\{Z_t\}$ are independent (typically identically distributed), integer-valued random variables. Note that if in (1.1) $\circ$ represented multiplication and the $\{Z_t\}$ were Gaussian, then we would recover a standard real-valued AR(p) process. The thinning operator, a generalised Steutel and van Harn operator (Steutel and van Harn (1979)) ensures $\alpha_i \circ X_{t-i}$ is an integer valued random variable. A binomial thinning operator is the most common choice such that $\alpha_i \circ X_{t-i} \equiv \text{Bin}(X_{t-i}, \alpha_i)$. The most common choice of Z_t is a Poisson distribution with mean λ, which combined with the binomial thinning operator and the condition $\sum_{i=1}^{p} \alpha_i < 1$ leads to the stationary distribution of X_t being Poisson with mean $\lambda/(1 - \sum_{i=1}^{p} \alpha_i)$. Parameter-driven models are based on the observed counts $\{X_t\}$ being driven by some underlying, unobserved latent process, $\{Y_t\}$, Durbin and Koopman (2000); Davis *et al.* (2003), for example a real-valued ARMA(p, q) process, see Dunsmuir (2015). With Poisson distributed counts, a log-link function is used to link the latent process, Y_t, and the observed count process, X_t. This results in a generalised linear model with

$$X_t | Y_t \sim \text{Po}\left(\mu \exp(Y_t)\right). \tag{1.2}$$

The observation and parameter driven models described above can be extended in many ways, for example the inclusion of time dependent covariates $\mathbf{z}_t$ into the INAR(p) parameters (Enciso-Mora *et al.* (2009b)) or into (1.2) to give $\mu = \exp(\mathbf{z}_t^T \boldsymbol{\beta})$, Davis *et al.* (2003). Other examples are the development of INARMA(p, q) extensions of (1.1), see Neal and Subba Rao (2007) and INGARCH models Fokianos (2011), where for $t \geq 1$,

$$X_t | \mathcal{F}_{t-1}^{X,\lambda} \sim \text{Po}\left(\lambda_t\right), \tag{1.3}$$

with $\lambda_t = \mu + a\lambda_{t-1} + bX_{t-1}$ and $\mathcal{F}_{t-1}^{X,\lambda}$ is the σ-field generated by $\{X_0, X_1, \ldots, X_t, \lambda_0\}$. The INGARCH model seeks to mimic the behaviour of GARCH models with alternative forms of λ_t considered in Fokianos (2011). It should be noted that for $a = 0$, the INGARCH model reduces to an INAR(1) model with $b \circ X_{t-1} \sim \text{Po}(bX_{t-1})$ and $Z_t \sim \text{Po}(\mu)$. For parameter driven models there are alternative latent process formulations such as replacing $\exp(Y_t)$ by θ_t where θ_t is a Markovian process satisfying

$$\theta_t = \frac{\theta_{t-1}}{\gamma}\text{Beta}(\gamma\alpha_{t-1}, (1-\gamma)\alpha_{t-1}), \tag{1.4}$$

see Aktekin *et al.* (2013) and Aktekin *et al.* (2017). Negative binomially distributed counts as opposed to Poisson distributed counts can also be included in (1.2), see for example Windle *et al.* (2013).

Given the wide range of models for integer valued time series a key question is, what is the most appropriate model for a given data set? This leads onto a secondary question of the appropriate order p for an INAR(p) model or an AR(p) autoregressive latent process. For INAR(p) models, efficient reversible jump MCMC algorithms (Green (1995)) have been developed in Enciso-Mora *et al.* (2009a) and Enciso-Mora *et al.* (2009b) for determining the order p of the model and the inclusion/exlusion of covariates. Reversible jump MCMC could also be employed for determining the most appropriate order of an AR(p) autoregressive latent process. However it is far more difficult to employ reversible jump MCMC for comparing between different classes of models due to the need to develop an efficient trans-dimensional moves between different models, see Brooks *et al.* (2003). Therefore in this work we focus primarily on choosing between different classes of integer valued time series models although we illustrate our approach for determining the model order p.

In this paper we consider model selection in a Bayesian framework, via direct computation of the marginal likelihood, also known as the model evidence, and alternatively using the DIC, *deviance information criterion*, Spiegelhalter *et al.* (2002). We focus for illustration purposes on three models; the INAR(p) model (1.1), the AR(p) Poisson regression model given by (1.2) with $\{Y_t\}$ being an AR(p) process, and the INGARCH model (1.3). To estimate the marginal likelihood, we extend the two stage algorithm given in Touloupou *et al.* (2017), which first estimates the posterior distribution using MCMC and then uses a parametric approximation of the posterior distribution to estimate the marginal likelihood via importance sampling. This leads to the two key novel contributions of this paper. Firstly, we introduce a particle MCMC algorithm (Andrieu *et al.* (2010)) for estimating the parameters of the AR(p) Poisson regression model. This involves using a particle filter (Gordon *et al.* (1993)) to estimate the likelihood, $\pi(\mathbf{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the parameters of the model. The use of the particle filter to estimate $\pi(\mathbf{x}|\boldsymbol{\theta})$ is then exploited both in the effective estimation of the marginal likelihood using the algorithm of Touloupou *et al.* (2017) and also in giving a mechanism for estimating the DIC without the need to resort to data augmentation and the problems that this potentially entails, Celeux *et al.* (2006).

The remainder of this paper is structured as follows. In Section 2 we introduce the particle MCMC algorithm for the AR(p) Poisson regression model. Given that Neal and Subba Rao (2007) provides an effective data augmentation MCMC algorithm for INAR(p) models we utilise the algorithm provided there in our analysis, whilst in Section 3 we give brief details of an MCMC algorithm for INGARCH model which is particularly straightforward to implement as no data augmentation is required. In Section 4 we present the generic approach to model selection which is employed for all three integer valued time series models under consideration. In Section 5, we conduct a simulation study which demonstrates the ability of the approaches described in Section 4 for determining the *true* model. The simulation study also provides insights into the AR(p) Poisson regression model and issues associated with identifying the autoregressive parameters in the latent process. In Section 6 we apply the AR(p) Poisson regression, INAR(p) and INGARCH models to two real-life data sets, monthly US polio cases (1970-1983) and monthly benefit claims from the logging industry to the British Columbia Workers Compensation Board (1985-1994). We show that an AR(1) Poisson regression model is preferred for the Polio data, and the inclusion of covariates proposed by Zeger (1988) for the data lead to only a small increase in the marginal likelihood. By contrast the INGARCH model is preferred for benefit claims data with significant evidence for the inclusion of a summer effect. All the data sets analysed in Sections 5 and 6 along with the R code used for the analysis are provided as supplementary material. Finally in Section 7 we make some concluding observations.

2 AR(p) Poisson regression model

In this Section we introduce an adaptive, particle MCMC algorithm for obtaining samples from the posterior distribution of AR(p) Poisson regression models (Zeger (1988), Davis *et al.* (2003)). The AR(p) Poisson regression model assumes that we observe a (Poisson) count process $X_1, X_2, \ldots, X_n$ which depends upon a (typically unobserved) latent AR(p) process $Y_1, Y_2, \ldots, Y_n$. Specifically, we assume that

$$X_t | Y_t \sim \text{Po}(\mu_t \exp(Y_t)) \tag{2.1}$$

$$Y_t = \sum_{i=1}^{p} a_i Y_{t-i} + e_t, \tag{2.2}$$

where $\mu_t = \exp(\mathbf{z}_t^T \boldsymbol{\beta})$ depends upon k explanatory variables $\mathbf{z}_t = (1, z_t^1, \ldots, z_t^k)$ and unknown regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ and the $\{e_t\}$s are independent and identically distributed according to $N(0, \tau^2)$. The parameters of interest are $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{a}, \tau)$, where $\mathbf{a} = (a_1, a_2, \ldots, a_p)$. In the absence of explanatory variables ($k = 0$), we will set $\mu_t = \exp(\beta_0) = \phi$ and replace $\boldsymbol{\beta}$ by ϕ.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote a realisation of n observations from $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, then we are interested in $\pi(\boldsymbol{\theta}|\mathbf{x}) = \pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})/\pi(\mathbf{x})$. Direct computation of $\pi(\mathbf{x}|\boldsymbol{\theta})$ for an MCMC algorithm is not possible as it involves integrating over the unobserved latent process $\mathbf{y}$. A data augmentation MCMC algorithm could be constructed to obtain samples from $\pi(\boldsymbol{\theta}, \mathbf{y}|\mathbf{x})$ by alternately updating $\boldsymbol{\theta}$ given $\mathbf{y}$ and $\mathbf{x}$, and then updating $\mathbf{y}$ given $\boldsymbol{\theta}$ and $\mathbf{x}$. This raises the question of how to efficiently update $\mathbf{y}$? For state-space models block updating of $\mathbf{y}$, via the forward filtering-backward sampling algorithm, was found to be effective, Carter and Kohn (1994) and Frühwirth-Schnatter (1994). For Gaussian state-space models this leads to a Gibbs sampling algorithm. A similar approach could be applied for the AR(p) Poisson regression model but the strong dependence between $\boldsymbol{\theta}$ and $\mathbf{y}$ can lead to poor mixing of the sampler. Therefore instead we use a particle MCMC algorithm (Andrieu *et al.* (2010)) to obtain samples from $\pi(\boldsymbol{\theta}|\mathbf{x})$. The particle MCMC algorithm uses forward filtering of $\mathbf{y}$ to estimate $\pi(\mathbf{x}|\boldsymbol{\theta})$ without the backward sampling step to choose a representative latent process $\mathbf{y}$. This speeds up the process by removing the backward sampling step and removes any problems caused by dependence between $\boldsymbol{\theta}$ and $\mathbf{y}$.

We proceed by outlining the particle filter for estimating $\pi(\mathbf{x}|\boldsymbol{\theta})$ before outlining how this is utilised within the particle MCMC algorithm.

For $M \geq 1$, a particle filter for estimating $\pi(\mathbf{x}|\boldsymbol{\theta})$ with M particles works as follows. Let $\mathbf{y}_{IN} = (y_{-(p-1)}, y_{-(p-2)}, \ldots, y_0)$ and generate M copies of $\mathbf{y}_{IN}$, denoted $\mathbf{s}_0^1, \mathbf{s}_0^2, \ldots, \mathbf{s}_0^M$ from $\pi(\mathbf{y}_{IN}|\boldsymbol{\theta})$. This is particularly straightforward if $p = 1$ as $y_0|\boldsymbol{\theta} \sim N(0, \tau^2(1-a^2)^{-1})$ but otherwise we can utilise the innovation algorithm (Brockwell and Davis (1996), pages 172, 175) to sample $\mathbf{y}_{IN}$ from the stationary distribution of the AR(p) process. For $j = 1, 2, \ldots, M$, set $w_0^j = 1/M$. Then for $t = 1, 2, \ldots, n$, we perform the following particle filter steps, for $j = 1, 2, \ldots, M$:

1. Sample K from $\{1, 2, \ldots, M\}$ with $P(K = k) = w_{t-1}^k / \sum_{l=1}^M w_{t-1}^l$.
2. Sample $y_t^j \sim N(\sum_{i=1}^p a_j s_{t-1,p+1-i}^k, 1/\tau)$.
3. Set $\mathbf{s}_t^j = (s_{t-1,2}^k, \ldots, s_{t-1,p}^k, y_t^j)$ and

$$w_t^j = \pi(x_t|y_t^j, \boldsymbol{\beta}) = \frac{\exp(\mathbf{z}_t^T\boldsymbol{\beta} + y_t^j)^{x_t}}{x_t!} \exp(-\exp(\mathbf{z}_t^T\boldsymbol{\beta} + y_t^j)). \tag{2.3}$$

For each $t = 1, 2, \ldots, n$, $\{\mathbf{s}_{t-1}^k\}$ form a weighted sample of size M from $\pi(\mathbf{y}_{(t-p):(t-1)}|\mathbf{x}_{1:(t-1)}, \boldsymbol{\theta})$ with weight w_{t-1}^k attached to particle $\mathbf{s}_{t-1}^k$. Thus the first step of the algorithm samples a realisation for $\mathbf{y}_{(t-p):(t-1)}^j$ from $\pi(\mathbf{y}_{(t-p):(t-1)}|\mathbf{x}_{1:(t-1)}, \boldsymbol{\theta})$. Since y_t is independent of $\mathbf{x}_{1:t-1}$ given $\mathbf{y}_{(t-p):(t-1)}$, the second step samples y_t^j from $\pi(y_t|\mathbf{y}_{(t-p):(t-1)}^j, \mathbf{x}_{1:(t-1)}, \boldsymbol{\theta})$. For the final step of the algorithm we note that

$$\begin{aligned}
\pi(\mathbf{y}_{(t-p):t}|\mathbf{x}_{1:t}, \boldsymbol{\theta}) &= \frac{\pi(x_t|\mathbf{y}_{(t-p):t}, \mathbf{x}_{1:t-1}, \boldsymbol{\theta})\pi(y_t|\mathbf{y}_{(t-p):(t-1)}, \mathbf{x}_{1:t-1}, \boldsymbol{\theta})\pi(\mathbf{y}_{(t-p):(t-1)}|\mathbf{x}_{1:t-1}, \boldsymbol{\theta})}{\pi(x_t|\mathbf{x}_{1:t-1}, \boldsymbol{\theta})} \\
&= \pi(x_t|y_t, \boldsymbol{\beta})\frac{\pi(y_t|\mathbf{y}_{(t-p):(t-1)}, \boldsymbol{\theta})\pi(\mathbf{y}_{(t-p):(t-1)}|\mathbf{x}_{1:t-1}, \boldsymbol{\theta})}{\pi(x_t|\mathbf{x}_{1:t-1}, \boldsymbol{\theta})}.
\end{aligned} \tag{2.4}$$

Therefore given that we have sampled $\mathbf{y}_{(t-p):t}^j$ from $\pi(\mathbf{y}_{(t-p):t}|\mathbf{x}_{1:t-1}, \boldsymbol{\theta})$ in steps 1 and 2, $w_t = \pi(x_t|y_t^j, \boldsymbol{\beta})$ gives the relative weight for the observation to have come from $\pi(\mathbf{y}_{(t-p):t}|\mathbf{x}_{1:t}, \boldsymbol{\theta})$. Given that the weight w_t^j does not depend upon y_{t-p}^j we can trivially integrate it out to obtain $(\mathbf{y}_{t-(p-1):t}^j, w_t^j)$ to take forward to the time point $t+1$. For our purposes the key benefit of the particle filter algorithm is that we can write

$$\pi(\mathbf{x}|\boldsymbol{\theta}) = \prod_{t=1}^n \pi(x_t|\mathbf{x}_{1:(t-1)}, \boldsymbol{\theta}), \tag{2.5}$$

with a by-product of the algorithm being that $P_t = M^{-1}\sum_{k=1}^M w_t^k$ is an unbiased estimate of $\pi(x_t|\mathbf{x}_{1:(t-1)},\boldsymbol{\theta})$ $(t=1,2,\ldots,n)$. Hence we estimate $\pi(\mathbf{x}|\boldsymbol{\theta})$ by $\prod_{t=1}^n P_t$. The observation that P_t is an unbiased estimate of $\pi(x_t|\mathbf{x}_{1:(t-1)},\boldsymbol{\theta})$ follows from steps 1 and 2 of the particle filter algorithm and

$$\pi(x_t|\mathbf{x}_{1:(t-1)},\boldsymbol{\theta}) = \int \left\{ \int \pi(x_t|y_t,\boldsymbol{\beta})\pi(y_t|\mathbf{y}_{(t-p):(t-1)},\boldsymbol{\theta})\, dy_t \right\} \pi(\mathbf{y}_{(t-p):(t-1)}|\mathbf{x}_{1:t-1},\boldsymbol{\theta})\, d\mathbf{y}_{(t-p):(t-1)}. \tag{2.6}$$

The particle filter will over time experience particle degeneracy, see Carvalho *et al.* (2010), and this will be exasperated by outlier observations. The main effect of the particle degeneracy will be to increase the variance of the estimate of $\pi(\mathbf{x}|\boldsymbol{\theta})$. This problem can be alleviated by choosing larger M, with more particles making the particle filter more computationally intensive but reducing the variance of the estimator. We observe that the variance of the estimator is inversely proportional to M. Given that our focus is simply on using the particle filter to estimate $\pi(\mathbf{x}|\boldsymbol{\theta})$ rather than to recover $\mathbf{y}$, particle degeneracy was not observed to present a problem throughout the analyses presented below.

The particle MCMC algorithm is constructed by embedding the particle filter within a random walk Metropolis algorithm to update $\boldsymbol{\theta}$. We initialize with parameters $\boldsymbol{\theta}$ and compute L, an estimate of the likelihood, $\pi(\mathbf{x}|\boldsymbol{\theta})$, using the particle filter. Then at each iteration we propose a new set of parameters $\boldsymbol{\theta}' \sim N(\boldsymbol{\theta},\Sigma)$ with proposal variance Σ. Provided that $\boldsymbol{\theta}'$ satisfies $|\sum_{i=1}^p a_i'| < 1$ and $\tau' > 0$, we estimate $\pi(\mathbf{x}|\boldsymbol{\theta}')$ by L', computed using the particle filter. The proposed move is then accepted with probability

$$\min\left\{1, \frac{L'\pi(\boldsymbol{\theta}')}{L\pi(\boldsymbol{\theta})}\right\} \tag{2.7}$$

with $(\boldsymbol{\theta},L)$ set equal to $(\boldsymbol{\theta}',L')$, if the proposed move is accepted, and $(\boldsymbol{\theta},L)$ remains unchanged otherwise. The efficiency of the algorithm depends upon the choice of Σ and M. The larger M is, the smaller the Monte Carlo error in estimating $\pi(\mathbf{x}|\boldsymbol{\theta})$, but the longer each iteration of the MCMC takes. Thus for a fixed computational cost there is the need to balance the choice of M and I, the number of iterations of the MCMC algorithm. In this paper with n (the length of the data) between 100 and 200, $M=100$ performs well in balancing small Monte Carlo error in the estimation of $\pi(\mathbf{x}|\boldsymbol{\theta})$ with fast implementation of the MCMC algorithm. The optimal choice of Σ is $h_d\Sigma_0$, where Σ_0 is the covariance matrix of the posterior distribution and h_d is a scaling coefficient depending upon on the dimension of d. We initialise

$\Sigma = s^2 I_d$, where I_d is the d-dimensional identity matrix and s is a small scalar. Then at regular intervals (three equally spaced points in this paper) during the burn-in we re-estimate Σ_0 and set $\Sigma = h_d \Sigma_0$. For large d, the optimal choice is $h_d = 2.38^2/d$ if $\pi(\mathbf{x}|\boldsymbol{\theta})$ is known (Roberts *et al.* (1997)). This changes if $\pi(\mathbf{x}|\boldsymbol{\theta})$ is replaced by an unbiased estimator, see Sherlock *et al.* (2014). However, provided that the Monte Carlo error of the particle filter estimate of $\pi(\mathbf{x}|\boldsymbol{\theta})$ is not large, $h_d = 2.38^2/d$ performs well and is successfully implemented throughout this paper.

3 INGARCH model

In this Section we briefly discuss an MCMC algorithm for the INGARCH model. Given that for the INGARCH,

$$X_t|\mathcal{F}_{t-1}^{X,\lambda} \sim \text{Po}\left(\lambda_t\right), \tag{3.1}$$

with $\lambda_t = \mu + a\lambda_{t-1} + bX_{t-1}$, for observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ from $\{X_t\}$, the likelihood satisfies

$$\begin{aligned} \pi(\mathbf{x}|\mu, a, b, \lambda_0, x_0) &= \prod_{t=1}^{n} \pi(x_t|\mu, a, b, \lambda_{t-1}, x_{t-1}) \\ &= \frac{\lambda_t^{x_t}}{x_t!} \exp\left(-\lambda_t\right). \end{aligned} \tag{3.2}$$

Consequently, no data augmentation is required for analysing this model using MCMC, and given priors on the parameters $\boldsymbol{\theta} = (\mu, a, b, \lambda_0)$ we can construct a random walk Metropolis algorithm to explore $\pi(\boldsymbol{\theta}|\mathbf{x})$. We choose independent gamma priors for μ and λ_0 and a uniform prior on the simplex $a, b > 0$ and $a + b < 1$ for (a, b). The priors on (a, b) are chosen to ensure that the INGARCH model is stationary, see Fokianos (2011). For tuning the proposal variance Σ for the random walk Metropolis algorithm to obtain sample from the posterior distribution, we take the same approach as in Section 2, of tuning Σ automatically during the burn-in period.

4 Model selection

In this Section we consider model selection tools for choosing between competing integer valued time series models. We highlight a range of model selection tools in the Bayesian paradigm.

Reversible jump MCMC Green (1995) which extends MCMC to allow trans-dimensional moves enabling the comparison of different models within a single MCMC algorithm. Reversible jump MCMC is particularly well suited for moving between nested models where effective trans-dimensional moves can be identified and has successfully been applied to INARMA(p, q) models Enciso-Mora *et al.* (2009a) and INAR(p) models with covariates Enciso-Mora *et al.* (2009b) to determine both model order and whether or not to include covariates in the model. Reversible jump MCMC is more challenging to implement when comparing non-nested models as it is much harder to identify effective trans-dimensional moves between the competing models, see Brooks *et al.* (2003).

There are a number of methods for estimating the marginal likelihood, $\pi(\mathbf{x})$, also known as the model evidence using MCMC output. These include the harmonic mean estimator Newton and Raftery (1994), Chib's method Chib (1995); Chib and Jeliazkov (2001), bridge sampling Meng and Wong (1996), power posteriors Friel and Pettitt (2008) and importance sampling Touloupou *et al.* (2017). Using a longitudinal epidemic example, Touloupou *et al.* (2017) showed that bridge sampling and the importance sampling based algorithm significantly outperformed the other methods in estimation of the marginal likelihood for a given cost. In this paper we show that the importance sampling based approach of Touloupou *et al.* (2017) is particularly well suited to the estimation of the marginal likelihood of time series models, utilising the particle filter for the AR(p) Poisson regression model introduced in Section 2 and similar particle filters for INAR(p) models for estimating the likelihood. For INAR(1) and INGARCH models the likelihood can easily be computed exactly.

The marginal likelihood is given by

$$\pi(\mathbf{x}) = \int \pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{4.1}$$

and is thus sensitive to the choice of prior distribution. In Spiegelhalter *et al.* (2002) the *deviance information criterion* DIC was introduced as a Bayesian information criteria for comparing models. The DIC is given, see Celeux *et al.* (2006) (2), by

$$\text{DIC} = -4E_{\boldsymbol{\theta}|\mathbf{x}}[\log \pi(\mathbf{x}|\boldsymbol{\theta})] + 2\log \pi(\mathbf{x}|\tilde{\boldsymbol{\theta}}), \tag{4.2}$$

where $\tilde{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ depending on $\mathbf{x}$. The posterior mean $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\mathbf{x}]$ is often a natural choice for

$\tilde{\boldsymbol{\theta}}$ especially when the posterior distribution is unimodal. Whilst the DIC has its critics, it has proved to be a popular tool for comparing competing models and, provided that $\pi(\mathbf{x}|\boldsymbol{\theta})$ can be calculated, the DIC can be computed as a straightforward appendage to an MCMC algorithm. Moreover, the DIC is far less sensitive to the choice of prior than the marginal likelihood. The direct estimation of $\pi(\mathbf{x}|\boldsymbol{\theta})$ given by the particle filter allows us to circumvent the DIC's sensitivity to missing data, Celeux *et al.* (2006).

In the following we focus on the importance sampling based estimation of the marginal likelihood Touloupou *et al.* (2017) and the DIC Spiegelhalter *et al.* (2002) for selecting between the competing models.

The algorithm of Touloupou *et al.* (2017) is a two-stage algorithm, where in stage one an estimate of the posterior distribution is obtained, typically using an MCMC sampler which will be the case in this paper. The posterior distribution is then approximated by a tractable probability density $q(\boldsymbol{\theta})$. Samples $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N$ from $q(\cdot)$ and for each i, we estimate $\pi(\mathbf{x}|\boldsymbol{\theta}_i)$ by $\widehat{\pi(\mathbf{x}|\boldsymbol{\theta}_i)}$ using a particle filter. Then

$$\widehat{\pi(\mathbf{x})} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\pi(\mathbf{x}|\boldsymbol{\theta}_i)} \frac{\pi(\boldsymbol{\theta}_i)}{q(\boldsymbol{\theta}_i)}, \tag{4.3}$$

is an unbiased estimate of $\pi(\mathbf{x})$. For the INAR(1) model it is possible to compute $\pi(\mathbf{x}|\boldsymbol{\theta}_i)$ exactly but for the AR(p) Poisson regression model, INAR(p) model ($p > 1$) and INGARCH model estimation of $\pi(\mathbf{x}|\boldsymbol{\theta}_i)$ is required. In Touloupou *et al.* (2017), Gaussian approximations of the posterior distributions were exploited through using multivariate Gaussian or t-distributions for $q(\cdot)$ along with the "defense mixture" proposal based upon Hesterberg (1995), where

$$q(\boldsymbol{\theta}) = (1-p)\phi(\boldsymbol{\theta}; \boldsymbol{\mu}, \Sigma) + p\pi(\boldsymbol{\theta}), \tag{4.4}$$

with $\boldsymbol{\mu}$ and Σ denoting the posterior mean vector and variance matrix, respectively, estimated from the MCMC output, $\phi(\cdot; \boldsymbol{\mu}, \Sigma)$ denoting the multivariate Gaussian density with parameters $\boldsymbol{\mu}$ and Σ and p is a mixture proportion between the multivariate Gaussian and the prior. The "defense mixture" proposal is simple to implement, robust with $\pi(\boldsymbol{\theta})/q(\boldsymbol{\theta})$ bounded above by p^{-1} and captures the key characteristics of the posterior distribution through the multivariate Gaussian approximation. Given the effectiveness of the "defense mixture" proposal with $p = 0.05$ demonstrated in Sections 5 and 6 for all the time series models we do not consider alternatives here.

In the case that $\pi(\mathbf{x}|\boldsymbol{\theta})$ can be computed, an unbiased estimate of the DIC can be obtained using MCMC samples $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{\tilde{N}}$ from $\pi(\boldsymbol{\theta}|\mathbf{x})$ by

$$\widehat{\text{DIC}} = -\frac{4}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \log \pi(\mathbf{x}|\boldsymbol{\theta}_i) + 2 \log \pi \left(\mathbf{x} \left| \bar{\boldsymbol{\theta}} \right.\right) \tag{4.5}$$

with $\bar{\boldsymbol{\theta}} = \tilde{N}^{-1} \sum_{i=1}^{\tilde{N}} \boldsymbol{\theta}_i$ and the last term on the right hand side of (4.5) is the only additional computation required to those computed as part of the MCMC algorithm. This becomes more complicated in situations where $\pi(\mathbf{x}|\boldsymbol{\theta})$ cannot be computed with Celeux *et al.* (2006) providing a number of candidates for DIC using data augmentation. By using a particle filter or alternative method to estimate $\pi(\mathbf{x}|\boldsymbol{\theta})$ directly, we can estimate the DIC by

$$\widehat{\text{DIC}} = -\frac{4}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \log \widehat{\pi(\mathbf{x}|\boldsymbol{\theta}_i)} + 2 \log \widehat{\pi \left(\mathbf{x} \left| \bar{\boldsymbol{\theta}} \right.\right)}. \tag{4.6}$$

This will lead to a biased estimator of the DIC as $E[\log \widehat{\pi(\mathbf{x}|\boldsymbol{\theta})}] \neq \log \pi(\mathbf{x}|\boldsymbol{\theta})$. However provided that $\widehat{\pi(\mathbf{x}|\boldsymbol{\theta})}$ has a small variance, the bias will be small. Let $\epsilon_\theta = \widehat{\pi(\mathbf{x}|\boldsymbol{\theta})} - \pi(\mathbf{x}|\boldsymbol{\theta})$, then $E[\epsilon_\theta] = 0$ and simple algebraic manipulation yields

$$E[\log \widehat{\pi(\mathbf{x}|\boldsymbol{\theta})}] \quad \approx \quad \log \pi(\mathbf{x}|\boldsymbol{\theta}) + \frac{1}{2\pi(\mathbf{x}|\boldsymbol{\theta})} E[\epsilon_\theta^2], \tag{4.7}$$

provided ϵ_θ is small relative to $\pi(\mathbf{x}|\boldsymbol{\theta})$. Moreover, we cannot simply use $\{\widehat{\pi(\mathbf{x}|\boldsymbol{\theta}_1)}, \widehat{\pi(\mathbf{x}|\boldsymbol{\theta}_2)}, \ldots, \widehat{\pi(\mathbf{x}|\boldsymbol{\theta}_{\tilde{N}})}\}$ generated by the MCMC algorithm to estimate the DIC using (4.6) as the sample generated will be biased, see Andrieu and Roberts (2009). Whilst this does not bias the estimation of the posterior distribution using the MCMC algorithm, it would effect the estimation of the DIC. Therefore as with computation of the marginal likelihood, we compute unbiased estimates of $\pi(\mathbf{x}|\boldsymbol{\theta}_k)$ $(k = 1, 2, \ldots, \tilde{N})$ using the particle filter.

We have that the computation of the DIC is also a two stage process and has similar computational cost to computing the marginal likelihood. Moreover, for the MCMC and the computation of the marginal likelihood and DIC the key cost is computing the likelihood using the particle filter and one iteration of the MCMC algorithms takes a similar length of time to perform as one of the computations in (4.3) and (4.5). Therefore given that in this paper 11,000-110,000 iterations are used per MCMC algorithm and throughout $N = \tilde{N} = 1000$, the additional computation cost of the marginal likelihood and DIC is small.

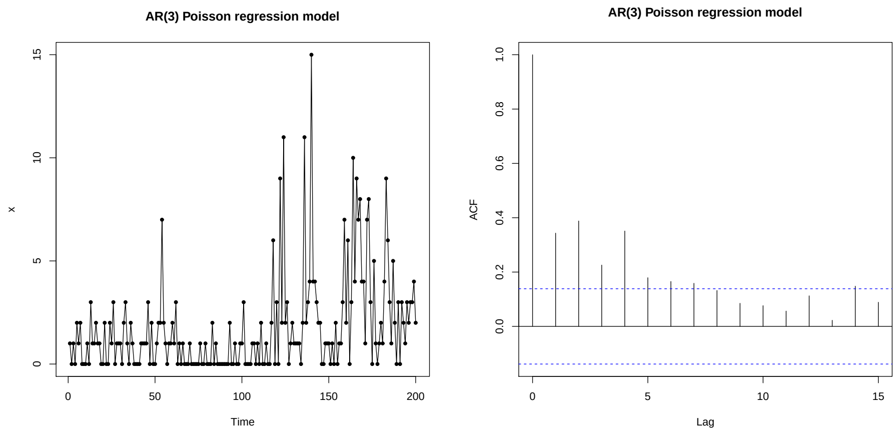

Figure 1: Left: Simulated data $\mathbf{x}$ from AR(3) Poisson regression model with $\mathbf{a} = (0.4, 0.25, 0.15)$, $\tau = 2$ and $\phi = 1$. Right: Acf plot of $\mathbf{x}$.

5 Simulation study

In this Section we present a simulation study which investigates the effectiveness of the model selection techniques on selecting the order p of an AR(p) Poisson regression model and of identifying the *true* model for three data sets with one data set each simulated from an AR(1) Poisson regression model, an INAR(1) model and an INGARCH model.

A time series of length 200 was generated from an AR(3) Poisson regression model with $\mathbf{a} = (0.4, 0.25, 0.15)$, $\tau^2 = 0.5$ and $\phi = 1$ (no covariate data). The data $\mathbf{x}$ along with an Acf (autocorrelation function) plot of $\mathbf{x}$ are shown in Figure 1. The data are over-dispersed with a mean and variance of 1.74 and 5.82, respectively. However, as noted in Davis *et al.* (2003) the correlation observed in the Poisson counts is less and often considerably less than the correlation observed in the underlying autoregressive process.

We ran the particle MCMC algorithm for the data with $p = 1, 2, 3$ and 4. We choose $N(0, 1)$ priors truncated to $(-1, 1)$ for the a_i's and Gamma$(1, 1)$ priors for τ and ϕ. In all cases the particle MCMC algorithm was run for 110,000 iterations. The final 50,000 iterations were retained to estimate the posterior distribution. The first 60,000 iterations were split into three blocks of 20,000 iterations with the posterior variance Σ estimated after each block and used within the random walk Metropolis algorithm as outlined in Section 2. The initial choice of Σ was $0.2^2 I_d$, where I_d is the $d \times d$ identity matrix. The marginal likelihood and DIC were then estimated using 1,000 samples from the defense mixture proposal distribution and posterior distribution (MCMC output), respectively, with $M = 100$ particles used. The exception being that $M = 1000$ particles were used to estimate $\log \pi(\mathbf{x}|\bar{\boldsymbol{\theta}})$, the latter term on the right hand side of (4.6) in the computation of the DIC to reduce the variance. For computing the DIC, every 50^{th} observation from the posterior sample is used. The parameter estimates (posterior means and standard deviations) and DIC values are recorded in Table 1.

From Table 1 we observe that the preferred order is $p = 2$ using the marginal likelihood and $p = 3$ using the DIC. However, the log marginal likelihood and DIC for $p = 2$ and $p = 3$ are close and hence we repeated the analysis 20 times to assess the robustness of the findings. In Figure 2, boxplots for the log marginal likelihood and DIC are given over the 20 replications. These show considerably less variability in the estimates of the log marginal likelihood than for the DIC with the standard error between 4 and 8 times smaller. Moreover, the marginal likelihood clearly identifies $p = 2$ as the preferred model, whereas the DIC demonstrates uncertainty between $p = 2$ and $p = 3$. We note that the estimation of ϕ and τ are fairly consistent across all orders of p with a similar value for $\sum_{i=1}^p a_i$, the sum of the autoregressive terms for all p. The posterior means of ϕ, τ and $\sum_{i=1}^p a_i$ are all close to the true values. We also observe that the estimates (posterior means and standard deviations) of all autoregressive terms included in the model are very similar with interestingly the AR(3) Poisson regression model having larger standard deviations for the autoregressive parameters than the other models. Similar behaviour was observed for other simulated data sets where the autoregressive parameters were decreasing with lag, in that, whilst ϕ, τ and $\sum_{i=1}^p a_i$ are estimated well, the individual a_i parameters are not estimated well with similar values for all autoregressive parameters. This highlights potential identifiability issues associated with the AR(p) Poisson regression model.

We now turn to the question of choosing between different classes of time series models. We simulated three data sets of length $n = 120$, presented in Figure 3 and provided as supplementary material:

- **Data Set 1** INAR(1) model with parameters $\alpha = 0.6$ and $\lambda = 1.2$.

p		ϕ	$\mathbf{a}$	τ	$\log \pi(\mathbf{x})$	DIC
1	Mean	1.145	0.801	0.597	-341.49	681.98
	SD	0.299	0.071	0.098	0.120	0.961
2	Mean	1.107	(0.422,0.415)	0.653	-338.57	675.61
	SD	0.345	(0.092,0.093)	0.096	0.134	1.026
3	Mean	1.066	(0.248,0.276,0.312)	0.572	-339.96	675.57
	SD	0.378	(0.414,0.388,0.386)	0.132	0.342	1.215
4	Mean	1.175	(0.211,0.223,0.227,0.216)	0.668	-343.24	682.77
	SD	0.505	(0.102,0.109,0.105,0.108))	0.100	0.186	1.327

Table 1: Parameter estimates, (log) marginal likelihood and DIC for simulated data set from AR(3) Poisson regression model fitting AR(p) Poisson regression model with $p = 1, 2, 3, 4$. Parameter estimates are based on a single MCMC run whilst (log) marginal likelihood and DIC estimates are based on means over 20 runs.

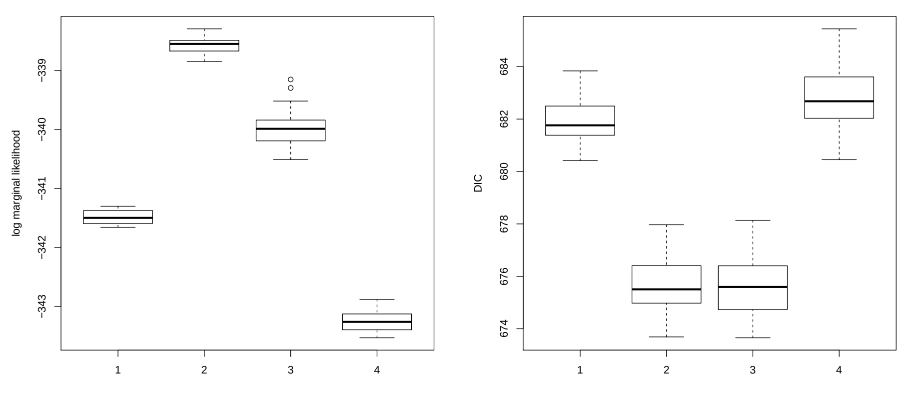

Figure 2: Log marginal likelihood (left) and DIC (right) for $p = 1, 2, 3, 4$ for AR(p) Poisson regression model fitted to $\mathbf{x}$.

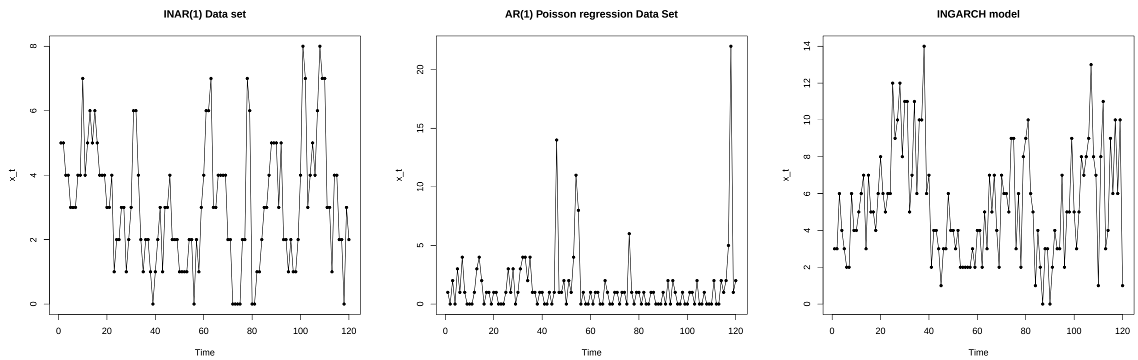

Figure 3: Simulated data sets: INAR(1) (left), AR(1) Poisson regression (centre) and INGARCH (right) models.

- **Data Set 2** AR(1) Poisson regression model with parameters $a = 0.6$, $\sigma = 1.0$ and $\mu = 1.0$.

- **Data Set 3** INGARCH model with parameters $\mu = 1$, $a = 0.4$, $b = 0.4$ and $\lambda_0 = 1$.

The simulation parameters were chosen to generate typical counts in the range 0 to 10 as observed in the real-life data sets with the length of the simulated data sets similar to the real-life data sets. We observe that the INAR(1) and INGARCH models generates data with few 0s whereas the AR(1) Poisson regression model tends to produce more 0s with a few spikes in the data where the count jumps up. A similar pattern is observed in the simulated data above from the AR(3) Poisson regression model. For the INAR(1) model we choose $U(0, 1)$ and Exp(1) priors for α and λ, respectively. For the AR(1) Poisson regression model we choose Exp(1) for the priors on ϕ and τ and a $N(0, 1)$ prior truncated to $(-1, 1)$ for a. Finally for the INGARCH model, we choose Exp(1) priors for μ and λ_0 and $U(0, 1)$ priors for a and b.

For all three data sets we ran the MCMC algorithms for 10,000 iterations after the burn-in periods of 15,000 iterations (3 blocks of 5,000 iterations to tune the random walk Metropolis proposal variance) for the particle MCMC algorithm for the AR(1) Poisson regression model and the INGARCH model MCMC and 1,000 iterations for the data augmented MCMC algorithm for the INAR(1) model. The marginal likelihood and DIC are again computed using 1,000 samples (for the DIC every 10^{th} observation from the posterior sample is used) with $M = 100$ for the particle filter for the AR(1) Poisson regression model. Note that for the INAR(1) model and INGARCH model, the likelihood $\pi(\mathbf{x}|\boldsymbol{\theta})$ can be computed exactly.

Data Set	INAR(1)	AR(1) Pois Reg	INGARCH
1	**-214.39** (0.005)	-239.77 (0.040)	-218.07 (0.044)
2	-257.02 (0.008)	**-193.88** (0.125)	-247.05 (0.035)
3	-293.01 (0.006)	-287.73 (0.056)	**-280.37** (0.031)

Data Set	INAR(1)	AR(1) Pois Reg	INGARCH
1	**423.82** (0.129)	472.24 (0.564)	428.26 (1.834)
2	508.43 (0.185)	**393.29** (1.604)	484.66 (3.565)
3	577.72 (0.181)	564.60 (0.638)	**551.90** (3.600)

Table 2: Mean (standard error) of log marginal likelihood (top) and DIC (bottom) for the INAR(1), AR(1) Poisson regression and INGARCH models applied to data sets 1, 2 and 3. The mean log marginal likelihood and DIC of the selected model are in bold type.

We repeated the estimation procedure 20 times for each data set and algorithm to test the robustness of the marginal likelihood and DIC estimates. The mean and standard errors of the log marginal likelihood and DIC are given in Table 2. In all cases the marginal likelihood and DIC identify the true model with the log marginal likelihood having considerably smaller standard errors than the DIC, up to 100 times smaller.

6 Analysis of data sets

6.1 Introduction

We illustrate the methodology with two examples; the monthly total number of polio cases in the USA from January 1970 to December 1983, Zeger (1988), Davis *et al.* (2000) and the monthly total number of injured logging workers claiming benefit from January 1985 to December 1994, Zhu and Joe (2006), Enciso-Mora *et al.* (2009b). The two data sets are chosen as they have previously been analysed by an AR(1) Poisson regression model (Polio data) and INAR(p) model (cut injury data) with the possible inclusion of covariates.

Throughout, unless otherwise stated, the following were used. For the AR(p) Poisson regression model Exp(1) prior is used for ϕ (no covariates in the model) and τ, a trauncated $N(0,1)$ prior is used for a and $N(0,1^2)$ priors for $\boldsymbol{\beta}$ (covariates included in the model). For the INAR(p) model, a uniform prior on $\boldsymbol{\alpha}$ on the simplex given by $\min \alpha_i > 0$ and $\sum_{i=1}^{p} \alpha_i < 1$ and Exp(1) prior for λ. Priors for the inclusion of covariate information into the INAR(p) model are discussed as required in Section 6.3. For the INGARCH model, Exp(1) priors for μ and λ_0 and uniform priors for a and b over the simplex $a, b > 0$ and $a + b < 1$. All MCMC algorithms were run to produce 50,000 iterations after burn-in with $30,000 = (3 \times 10,000)$ iterations burn-in for the AR(p) Poisson regression and INGARCH models and $10,000$ iterations burn-in for the INAR(p) algorithms. The estimates of the marginal likelihood and DIC were based on 1,000 samples from the defense mixture proposal distribution and every 50^{th} observation from the MCMC sample, respectively

6.2 Polio data

The polio data has disease case counts ranging from 0 to 14 with the majority being 0 or 1s and a mean of 1.3333. The data are given in Figure 4 along with an Acf plot for the data. The Polio data have been analysed with a linear trend and two sinusoidal functions corresponding to periods of 6 and 12 months, respectively. Specifically, Zeger (1988) and Davis *et al.* (2000) take the covariates to be

$$\mathbf{z}_t = \left(1, \frac{t'}{1000}, \cos\left(\frac{2\pi t'}{12}\right), \sin\left(\frac{2\pi t'}{12}\right), \cos\left(\frac{2\pi t'}{6}\right), \sin\left(\frac{2\pi t'}{6}\right)\right), \tag{6.1}$$

where $t' = t - 73$, a a linear trend with intercept January 1976. The Acf plot shows a significant lag-1 correlation in the data of 0.301 and also suggests the presence of 6 and 12 month dependence offering some support for the choice of sinusoidal covariates. Therefore we will use the covariates given in (6.1) when fitting models including covariates.

Our analysis of the data proceeds as follows. First we fit a selection of models without covariates, namely, an AR(p) Poisson regression model and INAR(p) model with $p = 1$ given that only the lag-1 correlation appears to be significant and INGARCH model. The results are presented in Table 3 and show overwhelming support in terms of both the log marginal likelihood and DIC for fitting an AR(1) Poisson regression model. In addition, we also considered an AR(2) Poisson regression model but the simpler AR(1) Poisson regression model was preferred. Observe that for the AR(2) Poisson regression model we have similar posterior means and standard deviations for both autoregressive components as observed in Section 5.

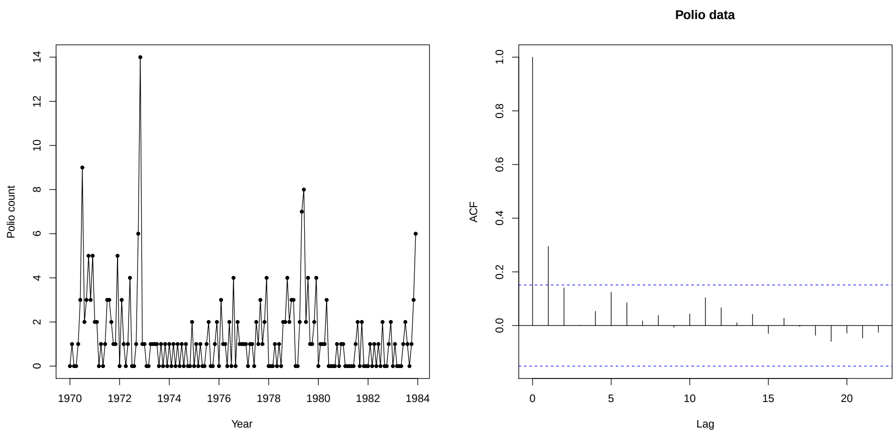

Figure 4: Left: Monthly cases of Polio in the USA from January 1970 to December 1983. Right: Autocorrelation function plot for the Polio data set.

Model	$\boldsymbol{\theta}$		Parameter estimates	$\log \pi(\mathbf{x})$	DIC
AR(1) Poisson	(ϕ, a, τ)	Mean	(0.947,0.601,0.683)	-263.50	524.99
		sd/se	(0.164,0.125,0.110)	0.069	0.867
AR(2) Poisson	(ϕ, a_1, a_2, τ)	Mean	(0.930,0.342,0.346,0.663)	-264.62	528.23
		sd/se	(0.210,0.194,0.197,0.140)	0.127	0.789
INAR(1)	(α, λ)	Mean	(0.187,1.100)	-293.86	582.07
		sd/se	(0.046,0.095)	0.007	0.127
INGARCH	(μ, a, b, λ_0)	Mean	(0.619,0.206,0.348,0.946)	-283.49	558.94
		sd/se	(0.152,0.119,0.068,0.920)	0.053	2.346

Table 3: Parameter estimates, (log) marginal likelihood and DIC for models without covariates fitted to the Polio data set.

Prior		$\boldsymbol{\beta}$	a	τ	$\log \pi(\mathbf{x})$	DIC
$N(0, 1^2)$	Mean	(-0.097, -0.384, 0.154, -0.464, 0.403, -0.005)	0.634	0.585	-263.07	515.18
	sd/se	(0.173, 0.965, 0.155,0.165,0.128, 0.133)	0.141	0.116	0.087	0.523
$N(0, 5^2)$	Mean	(-0.058,-3.090, 0.166,-0.472, 0.411, -0.0004)	0.618	0.581	-270.81	515.16
	sd/se	(0.186, 2.719, 0.149, 0.168, 0.128, 0.133)	0.152	0.117	0.091	0.921

Table 4: Parameter estimates, (log) marginal likelihood and DIC for AR(1) Poisson regression model with covariates.

Given that the AR(1) Poisson regression model is clearly the preferred model, we considered the inclusion of covariates (6.1) into this model only. We considered both $N(0, 1^2)$ and $N(0, 5^2)$ priors on the regression parameters $\boldsymbol{\beta}$ to test the sensitivity of the analysis to prior choice. The results are presented in Table 4. We observe that except for the trend β_2, the parameter estimates are very similar under both priors and as we would expect the larger the variance on the prior of the $\boldsymbol{\beta}$ coefficient, the greater the penalisation in the marginal likelihood calculations for the model with covariates. By comparison we note that the DIC does not penalise the models to the same extent with the inclusion of covariates leading to a much smaller DIC. This is as we would expect, since the DIC is largely unaffected by the prior distribution. The sensitivity of the slope coefficient β_2 to the choice of prior is due in large part to the rescaled time t' having a small range from $(-0.072, 0.095)$. We observe that despite the small difference between the marginal likelihoods for the AR(1) Poisson regression model with and without covariates (with $N(0, 1^2)$ prior on $\boldsymbol{\beta}$), the standard errors of the estimates are small. As such the smallest estimate of the marginal likelihood for the AR(1) Poisson regression model with covariates was larger than the largest estimate of the marginal likelihood for the AR(1) Poisson regression model without covariates. Thus consistent results were observed over all 20 replications of the analysis.

6.3 Cut injury data

The cut injury data has counts ranging from 1 to 21 with a mean of 6.1333. The data are given in Figure 5 along with an Acf plot for the data. The Acf plot shows that the first two or three lags are significant along with an annual (12 month lag) effect. The annual effect is capturing the fact that there are more claims in the summer (May-November) than in the winter (December-April) with all summer months

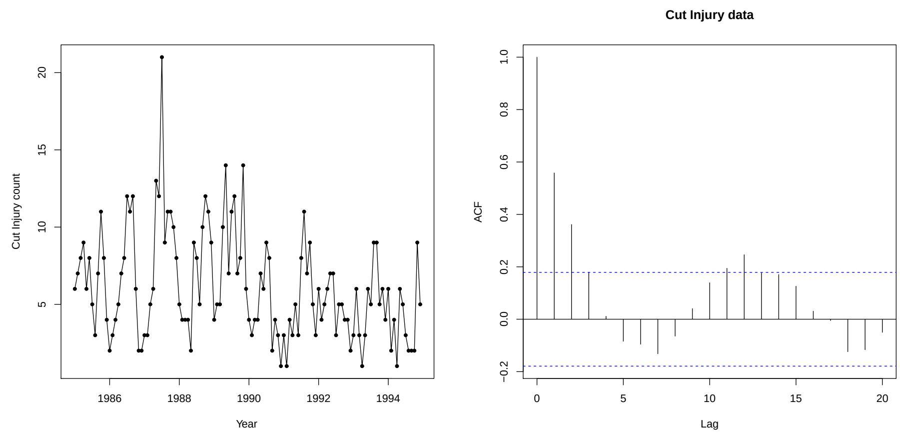

Figure 5: Left: Monthly cases of cut injury benefit claims from January 1985 to December 1994. Right: Autocorrelation function plot for the cut injury data set.

having a mean number of claims in excess of 7 and all winter months having having a mean number of claims below 5 (see Zhu and Joe (2006), Table II). Therefore when considering covariates we include a summer effect, s_t, which takes the value 1 in the summer and 0 in the winter.

Our analysis proceeds as in Section 6.2 by first considering fitting models without covariates. We fit AR(p) Poisson regression and INAR(p) models with $p = 1, 2, 3$ as well as the INGARCH model to the data. The results are presented in Table 5 and using the marginal likelihood show a preference for the INGARCH model over an INAR(2) model, the preferred INAR(p) model. The results are less conclusive than for the Polio data but show a clear preference for an INAR(p) model over an AR(p) regression model. We extend the analysis to incorporate the summer effect covariate restricting attention to the INGARCH model and INAR(p) models. For the INGARCH model we include the summer effect into the λ_t term via

$$\lambda_t = \mu + a\lambda_{t-1} + bX_{t-1} + cs_t. \tag{6.2}$$

Since we expect c to be a positive effect we place an Exp(1) prior on c. For the INAR(p) model with the summer effect covariate, we follow Enciso-Mora *et al.* (2009b) in taking

$$X_t = \sum_{i=1}^{p} \alpha_{i,t} X_{t-i} + Z_t, \tag{6.3}$$

Model	Summer	Mean $\log \pi(\mathbf{x})$	se $\log \pi(\mathbf{x})$	Mean DIC	se DIC
AR(1) Poisson	No	-300.59	0.049	589.95	0.769
AR(2) Poisson	No	-305.46	0.082	599.50	0.590
AR(3) Poisson	No	-308.45	0.181	601.62	0.757
INAR(1)	No	-298.35	0.005	588.27	0.106
INAR(2)	No	-295.85	0.074	585.81	0.779
INAR(3)	No	-296.64	0.117	584.15	1.422
INGARCH	No	-290.11	0.049	570.74	1.663
INAR(1)	Yes	-285.97	0.045	564.60	0.522
INAR(2)	Yes	-285.61	0.070	561.67	0.921
INAR(3)	Yes	-288.20	0.080	560.10	0.769
INGARCH	Yes	-283.51	0.042	556.47	2.704

Table 5: (Log) marginal likelihood and DIC for models fitted to the cut injury data set.

where $\alpha_{i,t} = \exp(\beta_0^i + \beta_1^i s_t)/\{1 + \exp(\beta_0^i + \beta_1^i s_t)\}$ and $Z_t \sim \text{Po}(\exp(\gamma_0 + \gamma_1 s_t))$. We place a $N(0, 1^2)$ prior on both the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ coefficients. We observe a clear improvement in the model fit by including a summer effect term with the INGARCH model still preferred to an INAR(2) model. It should be noted that the DIC penalises the order p in the INAR(p) model less than the prior in the marginal likelihood but otherwise similar conclusions are drawn with both approaches giving the same ordering. Once again we observe considerably smaller standard errors in the marginal likelihood calculations.

In Table 6, we summarise the parameter estimates for the three models with the highest marginal likelihoods, INAR(p) ($p = 1, 2$) and INGARCH models including the summer covariate. We again observed (not reported) that for the AR(p) Poisson regression model we had consistent estimates of ϕ and τ irrespective of the order p and that the sum of the autoregressive parameters had similar mean values for all orders.

7 Concluding remarks

In this paper we have shown how a particle filter algorithm can be successfully applied to estimate the likelihood, $\pi(\mathbf{x}|\boldsymbol{\theta})$, for an AR($p$) Poisson regression model. The particle filter is then utilised both within a particle MCMC algorithm and for computation of the marginal likelihood and the DIC for model selection. This has enabled us to conduct model selection both within AR(p) Poisson regression models

Model	Parameters		Estimates
INAR(1)	$(\beta_0^1, \beta_1^1, \gamma_0, \gamma_1)$	Mean	(-0.342,-0.111,0.818,0.711)
		sd	(0.368,0.454,0.191,0.214)
INAR(2)	$(\beta_0^1, \beta_0^2, \beta_1^1, \beta_1^2, \gamma_0, \gamma_1)$	Mean	(-0.653,-1.771,-0.073,-0.453,0.588,0.857)
		sd	(0.386, 0.520, 0.507, 0.681, 0.256, 0.296)
INGARCH	$(\mu, a, b, c, \lambda_0)$	Mean	(1.709,0.080, 0.490, 1.076,1.559)
		sd	(0.432, 0.063, 0.075, 1.060, 0.439)

Table 6: Parameter estimates for INAR(p) ($p = 1, 2$) and INGARCH models including the summer covariate applied to the cut injury data set.

to select the appropriate order p of the model and between AR(p) Poisson regression, INAR(p) and INGARCH models to choose the most appropriate model. In addition, the particle MCMC algorithm uncovered identifiability issues associated with the AR(p) Poisson regression model. A key benefit of the approaches for computing both the marginal likelihood and DIC is that it is computationally cheap relative to running the MCMC algorithm to obtain samples from the posterior distribution as observed at the end of Section 4 allowing model selection as a simple appendage to parameter estimation.

Acknowledgements

NA was supported by a PhD scholarship from the Saudi Arabian Government.

We thank an associate editor and two anonymous referees for insightful comments which helped improve the paper.

Supplementary Material

The R code used to produce the results in this paper along with the data sets analysed are included as supplementary material.

References

Aktekin, T., Polson, N. and Soyer, R. (2017). Sequential bayesian analysis of multivariate count data. *To appear in Bayesian Analysis.* 3

Aktekin, T., Soyer, R. and Xu, F. (2013) Assessment of mortgage default risk via Bayesian state space models. *Ann. Appl. Statist.* **7**, 1450–1473. 3

Andrieu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods (with discussion). *J. Roy. Soc. Series B.* **72**, 269–342. 4, 5

Andrieu, C. and Roberts, G.O. (2009) The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37**, 697–725. 11

Brockwell, P.J. and Davis, R.A. (1996) *Time Series: Theory and Methods.* Second Edition. Springer, London. 6

Brooks, S.P., Giudici, P. and Roberts, G.O. (2003) Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *J. Roy. Soc. Series B.* **65**, 3–39. 3, 9

Carter, C. K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika* **81**, 541–553. 5

Carvalho, C.M., Johannes, M.S., Lopes, H.F. and Polson, N.G. (2010) Particle Learning and Smoothing. *Statistical Science* **25**, 88–106. 7

Celeux, G., Forbes, F., Robert, C.P. and Titterington, D.M. (2006) Deviance Information Criteria for Missing Data Models. *Bayesian Analysis* **1**, 651–674. 4, 9, 10, 11

Chib, S. (1995) Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313–1321. 9

Chib, S. and Jeliazkov, I. (2001) Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* **96**, 270–281. 9

Davis, R.A., Dunsmuir, W.T. and Streett, S.B. (2003) Observation-driven models for Poisson counts. *Biometrika* **90**, 777–790. 1, 2, 5, 12

Davis, R.A., Dunsmuir, W.T. and Wang, Y. (2000) On autocorrelation in a Poisson regression model. *Biometrika* **87**, 491–505. 16, 17

Davis, R.A., Holan, S., Lund, R. and Ravishanker, N. (2015) *Handbook of Discrete-Valued Time Series.* Chapman and Hall/CRC. 1

Dunsmuir, W.T.M. (2015) Generalized Linear Autoregressive Moving Average Models. Chapter 3 in Handbook of Discrete-Valued Time Series, edited by Richard A. Davis, Scott H. Holan, Robert Lund, Nalini Ravishanker. pp.51–76. 2

Durbin, J. and Koopman, S.J. (2000) Time series analysis of non-Gaussian observations on state space models from both classical and Bayesian perspectives. *J. R. Statist. Soc. B* **62**, 3–56. 2

Enciso-Mora, V., Neal, P. and Subba Rao, T. (2009a) Efficient order selection algorithms for integer valued ARMA processes. *J. Time Ser. Anal.* **30**, 1–18. 1, 3, 9

Enciso-Mora, V., Neal, P.J. and Subba Rao, T. (2009b) Integer valued AR processes with explanatory variables. *Sankhya, Ser. B.* **71**, 248–263. 2, 3, 9, 16, 20

Fokianos, K. (2011) Some recent progress in count time series. *Statistics* **45**, 49–58. 2, 3, 8

Friel, N. and Pettitt, A. N. (2008) Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 589–607. 9

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *J. Time Series Anal.* **15**, 183–202. 5

Gordon, N.J., Salmond, D.J., and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F* **140**, 107-113. 4

Green P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732. 3, 9

Hesterberg, T. (1995) Weighted average importance sampling and defense mixture distributions. *Technometrics*, **37**, 185–194. 10

McKenzie, E. (2003) Discrete variate time series. *Stochastic Processes: Modelling and Simulation.* (eds. D. N. Shanbhag and C. R. Rao.) Elsevier. pp. 573–606. 1

Meng, X.-L. and Wong, W. H. (1996) Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, **6**, 831–860. 9

Neal, P. and Subba Rao, T. (2007) MCMC for integer-valued ARMA processes. *J. Time Ser. Anal.* **28**, 92–110. 1, 2, 4

Newton, M.A. and Raftery, A.E. (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Soc. Series B.* **56**, 3–48. 9

Roberts, G.O., Gelman, A. and Gilks, W.R. (1997) Weak convergence and optimal scaling of Random walk Metropolis algorithms. *Ann. Appl. Prob.* **7**, 110–120. 8

Sherlock, C., Thiery, A., Roberts, G. and Rosenthal, J. (2014) On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43**, 238–275. 8

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit. *J. Roy. Soc. Series B.* **64**, 583–639. 3, 9, 10

Steutel, F.W. and van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Annals of Probability* 7, 893–899. 2

Touloupou, P., Alzahrani, N., Neal, P., Spencer, S. and McKinley, T.J. (2017) Efficient model comparison techniques for models requiring large scale data augmentation. *To appear in Bayesian Analysis* 3, 4, 9, 10

Windle, J., Carvalho, C.M., Scott, J.G. and Sun, L. (2013) Efficient Data Augmentation in Dynamic Models for Binary and Count Data. arXiv:1308.0774 3

Zeger, S. (1988) A regression model for time series of counts. *Biometrika* **75**, 621–629. 4, 5, 16, 17

Zhu R. and Joe H. (2006) Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis* **27**, 725–738. 16, 20