

Bayesian Predictions of Low Count Time Series

B.P.M. MCCABE*and G.M. MARTIN†

September 15, 2004

*The School of Management, The University of Liverpool, Chatham Street, Liverpool, L69 7ZH, UK. Tel: 44 151 795 3705; Fax: 44 151 795 3001. (*Email: Brendan.McCabe@liverpool.ac.uk*).

†Correspondng Author, Department of Econometrics and Business Statistics, P.O. Box 11E, Monash University, Victoria, 3800, Australia. Tel.: 61 3 9905 1189; Fax: 61 3 9905 5474. (*Email: Gael.Martin@Buseco.monash.edu.au*).

Bayesian Predictions of Low Count Time Series

Abstract

The application of traditional forecasting methods to discrete count data yields forecasts that are non-coherent. That is, such methods produce non-integer point and interval predictions which violate the restrictions on the sample space of the integer variable. This paper presents a Bayesian methodology for producing coherent forecasts of low count time series. The forecasts are based on estimates of the p -step ahead predictive mass functions for a family of distributions nested in the integer-valued first-order autoregressive (INAR(1)) class. The predictive mass functions are constructed from convolutions of the unobserved components of the model, with uncertainty associated with both parameter values and model specification fully incorporated. The methodology is used to analyse Canadian wage loss claims data.

KEYWORDS: Forecasting; Discrete Time Series; INAR(1); Binomial Thinning; Bayesian Prediction; Bayesian Model Averaging.

JEL CODES: C11, C22, C53.

B.P.M. McCabe is Professor of Econometrics, Department of Economics and Accounting, School of Management, University of Liverpool, UK. He has previously taught at the Universities of British Columbia (Canada), Sydney (Australia) and Leeds (UK). His research interests are primarily in the area of time series econometrics.

G.M. Martin is Senior Lecturer, Department of Econometrics and Business Statistics, Monash University, Australia. Her research interests include Bayesian methods, time series econometrics and financial econometrics.

1 INTRODUCTION

The focus of this paper is on the analysis and prediction of data that enumerate infrequent events over time. An obvious feature of such data is that the values, or counts, are often low and, hence, not amenable to analysis via time series models designed for continuous random variables. Whilst the literature dealing with continuous variables has a long history, it is only in the last decade or so that much consideration has been given to time series variables that are discrete, including those that assume low count values. Good progress has been made in the specification of models for such data (see Fahrmeir and Tutz, 1994; McDonald and Zucchini, 1997; Cameron and Trevedi, 1998; and Winkelmann, 2000, for reviews). Nevertheless, there is still little consensus on which models are best used in practice, with no one particular model, or model class, proving itself applicable to a wide range of problems. This is in stark contrast to the pivotal role played by the Box-Jenkins Gaussian autoregressive moving average (ARMA) methodology for continuous variables.

Part of the reason for the lack of dominance of any one discrete model class is the relative paucity of inferential techniques in the discrete time series setting. That is, to be of practical use, a model needs to be allied with reliable techniques for estimation, testing and prediction, and the development of such techniques is still limited in the case of count time series. In particular, the development of forecasting methods that are coherent, in the sense of producing only integer forecasts of the count variable, has been lacking. Rather, the standard approach adopted in count data models has been to apply conventional forecasting methods, based on conditional expectations, and to accommodate parameter uncertainty by means of prediction intervals constructed around the point forecasts. Such an approach is obviously not forecast coherent in that both point and interval forecasts can assume non-integer values.

Recent work by Freeland and McCabe (2004a) attempts to address the issue of coherent prediction of count variables by proposing point forecasts in the context of the Poisson autoregressive (PAR) model, based on the integer-valued median of the forecast distribution. Estimation uncertainty is accommodated by producing a maximum likelihood-based estimate of the forecast distribution and constructing confidence intervals around the estimated probabilities. Whilst a feasible method for producing coherent forecasts, this solution is somewhat problem-specific. Moreover, the confidence intervals for the predictive probabilities have asymptotic justification only. In contrast, the present paper tackles the problem of forecasting counts using the Bayesian paradigm, whereby the predictive probability mass function (pmf), defined only over the support of the discrete count variable, is a natural outcome of Bayes Theorem. Not only parameter uncertainty, but also uncertainty due to the specification of the model itself, is directly incorporated into the predictive pmf, with there being no need either for the production of additional interval estimates or for qualifying statements concerning the dependence of the results on a particular model. The results are also valid for any sample size.

The predictive methodology is developed within the context of the INAR(1) class of model,

whereby the counts evolve as a birth and death (or survivorship) process. That is, the count value at time t may be considered to be the sum of new arrivals at time t and survivors from time $t - 1$. In the typical application of this model, the arrivals process is fully specified, often as Poisson, in which case the model is equivalent to the PAR model alluded to above. In contrast, we allow the arrivals to follow any distribution within a specified finite set of distributions in the integer class. This distributional flexibility is an acknowledgement of the fact that for low count time series models there is no compelling distributional assumption for the innovations that compares with the normal distribution in the continuous case. The forecasts are to be based on an estimate of the p -step ahead predictive pmf, which assigns probabilities to the different possible values of the random variable in its discrete support. Unwanted elements in the conditioning set of the predictive pmf are eliminated by using Bayesian methods to average over uncertainty about the parameter values as well as uncertainty about the specification of the arrivals process.

The INAR(1) model serves to induce dependence between the observations directly, via the survivorship component of the model. In this sense it can be viewed as falling within the class of *observation-driven* models. The alternative *parameter-driven* models introduce dynamics in the counts indirectly by specifying time-varying parameters as a function of an underlying latent process. Applications of Bayesian methodology to count times series data have focussed primarily on this latter class of model. For example, West, Harrison and Migon (1985), West and Harrison (1997), Chib, Greenberg and Winkelmann (1998), Durbin and Koopman (2000) and Chib and Winkelmann (2001) all apply Bayesian inferential procedures to dynamic latent factor models of count time series. Whilst such models are very flexible, the link between dependence in the latent variable and dependence in the observation sequence can sometimes be weak (see McCabe, Martin and Freeland, 2004). Moreover, the addition to the set of unknowns of an unobservable latent factor for each observation, although manageable via the techniques of Kalman filtering and/or Markov chain Monte Carlo (MCMC) sampling, does markedly increase the computational burden. In contrast, analysis of the parsimonious specifications considered in this paper requires only low-dimensional numerical integration techniques.

An outline of the remainder of paper is as follows. In Section 2, we begin by outlining the form of the p -step ahead predictive pmf appropriate for any discrete random variable. These prediction results are specialized in Section 3 to the INAR(1) class of models. We focus on three alternative arrivals distributions, the Poisson, the binomial and the negative binomial, the latter two being under- and over-dispersed respectively relative to the Poisson. The method is used, in Section 4, to analyse a set of data on the number of workers receiving wage loss benefits due to injuries received in the logging industry in British Columbia, Canada. Some conclusions are given in Section 5.

2 THE P -STEP AHEAD PREDICTIVE PMF

We begin by specifying a random variable Y_t that can assume only count values $\{0, 1, 2, \dots\}$ at each point in time $t = 1, 2, \dots, T$. The model generating Y_t is assumed to be any one within a set of K

models that are entertained, with those models denoted by M_k , $k = 1, 2, \dots, K$. Given the vector of observed data, $\mathbf{y} = (y_1, y_2, \dots, y_T)'$, the p -step ahead predictive pmf is defined as

$$P(Y_{T+p} = y_{T+p} | \mathbf{y}) = \sum_{k=1}^K P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k) P(M_k | \mathbf{y}), \quad (1)$$

where $P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k)$ is the k th model-specific p -step ahead predictive pmf and $P(M_k | \mathbf{y})$ is the posterior probability of model M_k . The posterior model probabilities, $P(M_k | \mathbf{y})$, $k = 1, 2, \dots, K$, are constructed via the posterior odds ratios for models M_2, \dots, M_K , relative to reference model M_1 . The posterior odds ratio for model M_k relative to M_1 , $PO_{k,1}$, is given by

$$PO_{k,1} = \frac{P(M_k | \mathbf{y})}{P(M_1 | \mathbf{y})} = \frac{P(M_k)}{P(M_1)} \times BF_{k,1}, \quad k = 2, 3, \dots, K, \quad (2)$$

where $P(M_k)/P(M_1)$ defines the prior odds ratio and

$$BF_{k,1} = \frac{P(\mathbf{y} | M_k)}{P(\mathbf{y} | M_1)} \quad (3)$$

defines the Bayes factor for M_k versus M_1 . The Bayes factor, constructed as the ratio of the marginal likelihood of model M_k ,

$$P(\mathbf{y} | M_k) = \int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k | M_k) p(\boldsymbol{\theta}_k | M_k) d\boldsymbol{\theta}_k, \quad (4)$$

to that of M_1 , measures the degree for support in the data for M_k relative to M_1 . The marginal likelihood for M_k is, in turn, defined as the expectation of the likelihood under M_k , $\ell(\boldsymbol{\theta}_k | M_k)$, with respect to the prior under M_k . The likelihood function reflects the form of the joint pmf for Y_t , $t = 1, 2, \dots, T$, under M_k . That is

$$\ell(\boldsymbol{\theta}_k | M_k) = P(\mathbf{y} | M_k, \boldsymbol{\theta}_k). \quad (5)$$

The prior probability density function (pdf) in (4), $p(\boldsymbol{\theta}_k | M_k)$, expresses the prior beliefs about the unknown parameters in M_k .

The k th model-specific predictive pmf is defined as

$$P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k) = \int_{\boldsymbol{\theta}_k} P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | \mathbf{y}, M_k) d\boldsymbol{\theta}_k, \quad (6)$$

where $P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k, \boldsymbol{\theta}_k)$ is the p -step ahead predictive pmf, *conditional* on $\boldsymbol{\theta}_k$, and $p(\boldsymbol{\theta}_k | \mathbf{y}, M_k)$ is the joint posterior pdf for the unknown parameter vector $\boldsymbol{\theta}_k$, with the joint posterior given, via Bayes Theorem, as

$$p(\boldsymbol{\theta}_k | \mathbf{y}, M_k) \propto \ell(\boldsymbol{\theta}_k | M_k) p(\boldsymbol{\theta}_k | M_k). \quad (7)$$

As is clear from the expressions in (6) and (1), the p -step ahead predictive pmf averages over both parameter uncertainty, conditional on an assumed model, as quantified by $p(\boldsymbol{\theta}_k | \mathbf{y}, M_k)$, and model uncertainty, as quantified by the distribution of posterior model probabilities $P(M_k | \mathbf{y})$,

$k = 1, 2, \dots, K$. That is, the Bayesian method enables $\boldsymbol{\theta}_k$ and M_k to be eliminated from the conditioning set of the conditional mass function, $P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k, \boldsymbol{\theta}_k)$, in a way that reflects both prior and sample information on both of these unknown components of the data generating process.¹

Evaluation of the predictive pmf in (1) requires a numerical approach, with the precise details of that approach depending on the nature of the models in the model set. If the parameter set for each model is small, $p(\boldsymbol{\theta}_k | \mathbf{y}, M_k)$ can be evaluated using deterministic numerical integration. The k th model-specific p -step ahead predictive pmf in (6) can then be estimated as a weighted average of the conditional mass functions, with the weights being the probability “mass” assigned to each grid point in the numerically evaluated posterior. Using the rectangular numerical integration method for the sake of illustration, the k th model-specific predictive pmf is estimated as

$$\hat{P}(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k) = h \sum_{i=1}^{N_k} P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k, \boldsymbol{\theta}_k^{(i)}) p(\boldsymbol{\theta}_k^{(i)} | \mathbf{y}, M_k), \quad (8)$$

where N_k is the number of gridpoints used in evaluating the density $p(\boldsymbol{\theta}_k | \mathbf{y}, M_k)$, h is the grid width and $p(\boldsymbol{\theta}_k^{(i)} | \mathbf{y}, M_k)$ denotes the ordinate of $p(\boldsymbol{\theta}_k | \mathbf{y}, M_k)$ at grid value $\boldsymbol{\theta}_k^{(i)}$. When $\boldsymbol{\theta}_k$ is of low dimension, the marginal likelihood for model M_k in (4) can also be computed directly using deterministic integration, with the posterior odds ratios and model probabilities subsequently produced using simple calculations. Denoting the estimated model probabilities by $\hat{P}(M_k | \mathbf{y})$, the model-averaged predictive pmf in (1) is then estimated as

$$\hat{P}(Y_{T+p} = y_{T+p} | \mathbf{y}) = \sum_{k=1}^K \hat{P}(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k) \hat{P}(M_k | \mathbf{y}). \quad (9)$$

In the case where the parameter set of one or more of the models in the model set is large, simulation methods such as Importance sampling or MCMC sampling could be used to produce a sample from $p(\boldsymbol{\theta}_k | \mathbf{y}, M_k)$, with $P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k)$ estimated as an average of the conditional mass functions over the parameter draws. Estimates of the marginal likelihoods and, hence, the model probabilities, could also be constructed from the simulation output; see, for example, Chib (1995).

In the case of the models adopted in this paper, the first method of calculation described above, based on low-dimensional numerical integration techniques, is used. More details of those calculations are provided at the appropriate points below.

3 PREDICTION IN THE INAR(1) FAMILY

We demonstrate the methodology outlined in the previous section using a generalization of the INAR(1) model for count time series introduced by Al-Osh and Alzaid (1987) and McKenzie (1988) and as further investigated by Ronning and Jung (1992), Brännäs (1994), Freeland and

McCabe (2004a,b) and Jung and Tremayne (2003), amongst others. Let Y_1, Y_2, \dots, Y_T be a series of dependent counts generated according to the following model,

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \quad (10)$$

where the arrivals process $\{\varepsilon_t\}_{t=1}^{\infty}$ is a series of independently and identically distributed (iid) random variables defined on the support $\{0, 1, 2, \dots\}$. Inference is to be conditioned on the initial value, $Y_1 = y_1$. The thinning operator “ \circ ” is defined as follows. Given Y_{t-1} ,

$$\alpha \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} B_{it}, \quad (11)$$

where $B_{1t}, B_{2t}, \dots, B_{Y_{t-1}t}$ are iid Bernoulli random variables with

$$P(B_{it} = 1) = 1 - P(B_{it} = 0) = \alpha. \quad (12)$$

It is further assumed that B_{jt} and ε_t are independent for all j . Since $\alpha \circ Y_{t-1}$ given $Y_{t-1} = y_{t-1}$ is a sum of iid Bernoulli random variables it follows that it has a binomial distribution with parameters α and y_{t-1} i.e. is $Bin(\alpha, y_{t-1})$. The Binomial distribution thus determines the death/survivorship process. This model may also be interpreted as an infinite server queue. The model is stationary for $0 \leq \alpha < 1$; see Grunwald, Hyndman, Tedesco and Tweedie (2000). It is possible to include higher order lags in the INAR model, as well to expand the model to cater for a moving average structure. However, such augmented models are difficult to interpret as a birth and death process; see Freeland (1998) for details. For the applications considered in this paper the physical interpretation of the first order INAR model is clear cut.

Whilst applications of the INAR(1) model typically assume that the arrivals process, ε_t , is Poisson distributed, we allow ε_t to be any arbitrary (discrete) distribution, within a specified finite set of distributions.²The model set, M_k , $k = 1, 2, \dots, K$, is then to be defined according to the alternative distributional assumptions adopted for $\{\varepsilon_t\}_{t=1}^{\infty}$. We focus on three such models for the arrival process, Poisson, binomial and negative binomial. These three models are appropriate, respectively, for arrivals that are equi-dispersed (mean and variance equal), under-dispersed (variance less than mean) and over-dispersed (variance greater than mean). As such, they constitute a reasonably broad coverage of possible arrivals processes. The methodology that we present in Section 3.1 is, however, appropriate for any arrivals process embedded within the INAR(1) class. In particular, it is readily adaptable to the case where a Poisson process with covariates is used to model over-dispersion in the data. However, since the focus of this paper is on forecasting (and not on assessing the relative impacts of possible covariates) we choose to model over-dispersion via the more parsimonious negative binomial distribution and thereby avoid the additional unknown parameters that would be associated with the use of covariates. In Section 3.2 we describe how the methodology can be adapted to cater for more general specifications than the INAR(1) model.

3.1 The p -Step Ahead Conditional Predictive Pmf

Construction of the estimated p -step ahead model-averaged predictive pmf in (9) requires evaluation of the conditional pmf, $P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k, \boldsymbol{\theta}_k)$, in the integrand in (6). In the case of the INAR(1) model, the form of this function is determined by the structure of the model in (10). Note that in this model, Y_t is composed of two random components, the complement of the death (i.e. the survivorship) component $\alpha \circ Y_{t-1} | Y_{t-1}$, and the arrivals (birth) component ε_t , and that these two components are not (individually) observed. The distribution of Y_t given $Y_{t-1} = y_{t-1}$, conditional on arrivals model M_k and associated parameter vector $\boldsymbol{\theta}_k$, is thus given by the convolution of the two random components, as follows,

$$P(Y_t = y_t | y_{t-1}, M_k, \boldsymbol{\theta}_k) = \sum_{s=0}^{\min(y_t, y_{t-1})} P[B_{y_{t-1}}^\alpha = s] P_{M_k}(\varepsilon_t = y_t - s), \quad (13)$$

where $P_{M_k}(\varepsilon_t = y_t - s)$ denotes the pmf for the arrival ε_t under M_k , and we use the notation B_b^a to denote a variable that is distributed as $Bin(a, b)$. Equally, under thinning, the distribution of Y_t given $Y_{t-p} = y_{t-p}$ is the convolution of a $Bin(\alpha^p, y_{t-p})$ distribution and the distribution of $\sum_{j=0}^{p-1} \alpha^j \circ \varepsilon_{t-j}$, where the latter represents a sum of thinned arrivals processes. Thus

$$P(Y_t = y_t | y_{t-p}, M_k, \boldsymbol{\theta}_k) = \sum_{s=0}^{\min(y_t, y_{t-p})} P[B_{y_{t-p}}^{\alpha^p} = s] P\left[\sum_{j=0}^{p-1} \alpha^j \circ \varepsilon_{t-j} = y_t - s\right]. \quad (14)$$

Also, since the arrivals are iid,

$$P\left[\sum_{j=0}^{p-1} \alpha^j \circ \varepsilon_{t-j} = y_t - s\right] = \{P[\alpha^{p-1} \circ \varepsilon_{t-j} = y_t - s]\} * \dots * \{P[\alpha^0 \circ \varepsilon_{t-j} = y_t - s]\}, \quad (15)$$

where the convolution operator is defined as $\{c_n\} = \{a_n\} * \{b_n\} = \sum_{l=0}^n a_l b_{n-l}$ for sequences $\{a_n\}$, $\{b_n\}$, with $n = y_t - s$ in this case. For example, when $p = 2$, $\{a_n\} = \{P[\alpha^1 \circ \varepsilon_{t-j} = n]\}$ and $\{b_n\} = \{P[\alpha^0 \circ \varepsilon_{t-j} = \varepsilon_{t-j} = n]\}$ and

$$\begin{aligned} P\left[\sum_{j=0}^1 \alpha^j \circ \varepsilon_{t-j} = y_t - s\right] &= P[\alpha \circ \varepsilon_{t-1} + \varepsilon_t = y_t - s] \\ &= \sum_{l=0}^{y_t-s} P[\alpha \circ \varepsilon_{t-1} = l] P_{M_k}(\varepsilon_t = y_t - s - l) \\ &= \sum_{l=0}^{y_t-s} \left[\sum_{m=l}^{\infty} P[B_m^\alpha = l] P_{M_k}(\varepsilon_{t-1} = m) \right] \\ &\quad \times P_{M_k}(\varepsilon_t = y_t - s - l), \end{aligned} \quad (16)$$

where the expression in square brackets in the last line in (16) follows from mixing the conditional binomial variate B_m^α over the marginal distribution of the arrivals process, $P_{M_k}(\varepsilon_{t-1} = m)$. The

convolution operator is commutative and thus the iteration required in (15) may be carried out in any order. For $j = 0, \dots, p-1$,

$$P[\alpha^j \circ \varepsilon_{t-j} = l] = \left[\sum_{m=l}^{\infty} P[B_m^{\alpha^j} = l] P_{M_k}[\varepsilon_{t-j} = m] \right]. \quad (17)$$

Given the Markovian nature of the INAR(1) model, the p -step ahead conditional predictive pmf, $P(Y_{T+p} = y_{T+p} | \mathbf{y}, M_k, \boldsymbol{\theta}_k)$, is equivalent to $P(Y_{T+p} = y_T | y_T, M_k, \boldsymbol{\theta}_k)$, which is, in turn, given by

$$P(Y_{T+p} = y_{T+p} | y_T, M_k, \boldsymbol{\theta}_k) = \sum_{s=0}^{\min(y_{T+p}, y_T)} P[B_{y_T}^{\alpha^p} = s] P \left[\sum_{j=0}^{p-1} \alpha^j \circ \varepsilon_{T+p-j} = y_{T+p} - s \right]. \quad (18)$$

Each component in (18) is evaluated using (15) and (17), with the appropriate adjustments made for the use of lead rather than lag notation, and for the fact that the predictive pmf is conditioned on the last value in the sample, y_T .

The precise form of the conditional distribution in (18) depends of course on the distributional assumption, M_k , adopted for the arrivals process. We demonstrate the form of (18) in the case of the three distributions of interest, Poisson, binomial and negative binomial.

3.1.1 Poisson Arrivals (M_1)

Model M_1 involves the assumption that $\{\varepsilon_t\}_{t=1}^{\infty}$ is a series of independently distributed Poisson random variables with mean λ . That is

$$P_{M_1}(\varepsilon_t = w) = \frac{e^{-\lambda} \lambda^w}{w!}, \quad w = 0, 1, 2, \dots \quad (19)$$

for all $t = 1, 2, \dots, T$. The binomial thinning process in (10) combined with a Poisson arrivals process is commonly referred to as the Poisson autoregressive (PAR) model. Given a mean parameter λ for the Poisson arrivals, model M_1 has a two-dimensional parameter vector $\boldsymbol{\theta}_1 = (\alpha, \lambda)'$. In the Poisson case, each thinned arrival, $\alpha^j \circ \varepsilon_{T+p-j}$ in (18), is also Poisson, as is the sum of the p such independent variates that appears as the second term on the right hand side of the expression.³ Thus, in this case, (18) collapses to

$$\begin{aligned} P(Y_{T+p} = y_{T+p} | y_T, M_1, \boldsymbol{\theta}_1) &= \sum_{s=0}^{\min(y_{T+p}, y_T)} P[B_{y_T}^{\alpha^p} = s] \times \\ &\times \frac{1}{(y_{T+p} - s)!} \exp \left\{ -\lambda \frac{1 - \alpha^p}{1 - \alpha} \right\} \left(\lambda \frac{1 - \alpha^p}{1 - \alpha} \right)^{y_{T+p} - s}. \end{aligned} \quad (20)$$

3.1.2 Binomial Arrivals (M_2)

Model M_2 assumes the $\{\varepsilon_t\}_{t=1}^{\infty}$ to be a series of independently distributed binomial variates, with probability of ‘success’, q , in each of n trials. That is

$$P_{M_2}(\varepsilon_t = w) = \binom{n}{w} q^w (1 - q)^{(n-w)}; \quad w = 0, 1, 2, \dots, n, \quad (21)$$

for all $t = 1, 2, \dots, T$. We refer to the binomial thinning process in (10), combined with binomial arrivals, as the binomial autoregressive (BAR) model, with this model being characterized by the three-dimensional parameter vector $\boldsymbol{\theta}_2 = (\alpha, q, n)'$. In this case, there is no simplification of the p -step ahead conditional distribution in (18), with evaluation of that distribution requiring repeated iteration of the relevant convolution formulae, with $P_{M_k}(\cdot)$ being as given in (21). For example, when $p = 2$, (16) becomes

$$P[\varepsilon_{T+2} + \alpha \circ \varepsilon_{T+1} = y_{T+2} - s] = \sum_{l=0}^{y_{T+2}-s} P[B_{\varepsilon_{T+1}}^{\alpha} = l] P[B_n^q = y_{T+2} - s - l], \quad (22)$$

where

$$P[B_{\varepsilon_{T+1}}^{\alpha} = l] = \sum_{m=l}^{\infty} P[B_m^{\alpha} = l] P[B_n^q = m]. \quad (23)$$

Substituting (22) and (23) into (18) leads to a 2-step ahead conditional predictive pmf of the form

$$\begin{aligned} P(Y_{T+2} = y_{T+2} | y_T, M_2, \boldsymbol{\theta}_2) &= \sum_{s=0}^{\min(y_{T+2}, y_T)} \{P[B_{y_T}^{\alpha^2} = s]\} \times \\ &\sum_{l=0}^{y_{T+2}-s} \sum_{m=l}^{\infty} P[B_m^{\alpha} = l] P[B_n^q = m] \\ &\times P[B_n^q = y_{T+2} - s - l]. \end{aligned} \quad (24)$$

3.1.3 Negative Binomial Arrivals (M_3)

Model M_3 assumes the $\{\varepsilon_t\}_{t=1}^{\infty}$ to be a series of independently distributed negative binomial variates, with probability of ‘success’, π , in any one trial. The negative binomial random variable is defined as the number of failures, w , prior to the r th success, with pmf given by

$$P_{M_3}(\varepsilon_t = w) = \binom{r+w-1}{w} \pi^r (1-\pi)^w; \quad w = 0, 1, 2, \dots, \quad (25)$$

for all $t = 1, 2, \dots, T$. We refer to the binomial thinning process in (10), combined with negative binomial arrivals, as the negative binomial autoregressive (NBAR) model, with this model being characterized by the three-dimensional parameter vector $\boldsymbol{\theta}_3 = (\alpha, \pi, r)'$. As with the binomial arrivals, there is no simplification of the p -step ahead conditional distribution in (18), with evaluation of that distribution requiring repeated iteration of the relevant convolution formulae, with $P_{M_k}(\cdot)$ now being as given in (25). For $p = 2$, the conditional predictive pmf is

$$\begin{aligned} P(Y_{T+2} = y_{T+2} | y_T, M_3, \boldsymbol{\theta}_3) &= \sum_{s=0}^{\min(y_{T+2}, y_T)} \{P[B_{y_T}^{\alpha^2} = s]\} \times \\ &\sum_{l=0}^{y_{T+2}-s} \sum_{m=l}^{\infty} P[B_m^{\alpha} = l] P[NB_r^{\pi} = m] \\ &\times P[NB_r^{\pi} = y_{T+2} - s - l], \end{aligned} \quad (26)$$

where NB_r^{π} denotes a negative binomial variable with parameters π and r .

3.2 Numerical Evaluation of the p -Step Ahead Conditional Predictive Pmf

Although the focus of the paper is on the application of the proposed prediction methodology to the INAR(1) family of models, the methodology is able to be generalized to any model class. The key feature of the method is the use of the p -step conditional predictive pmf in (6) as the basic building block for the production of predictive probabilities. In the present paper, we have derived the analytical form of this mass function for any model in the INAR(1) class that retains the binomial thinning specification for the survivorship component. Such analytical results may not be readily derivable in the case of more general models. In this case, the method would be based on a numerical evaluation of the conditional predictive pmf. For example, evaluation of the one-step ahead conditional predictive pmf would require repeated simulation of Y_{T+1} from the specified model, at a given point in the parameter space, given the observed value of Y_T . The draws of Y_{T+1} would then be used to produce an estimate of $P(Y_{T+1} = y_{T+1} | y_T, M_k, \theta_k)$, with this estimated mass function then being used in the remaining steps of the algorithm instead of the analytical mass function. Numerical evaluation of the p -step ahead conditional function would proceed in a similar fashion, except for being conditioned on $p - 1$ additional out-of-sample values, $Y_{T+p-1}, Y_{T+p-2}, \dots, Y_{T+1}$, which would, in turn, be produced from the estimated mass functions for the corresponding periods.

4 EMPIRICAL APPLICATION

4.1 Data Description

We apply the proposed methodology to a data set that has been obtained from the Workers Compensation Board (WCB) of the Province of British Columbia, Canada. The same data set is also analyzed in Freeland and McCabe (2004a) via the application of classical inferential procedures to the PAR model. Further comparisons with the results in that paper are made in Section 4.4 below. The data set comprises 120 monthly counts of workers collecting Wage Loss Benefits for burn injuries received whilst working in a particular segment of the logging industry in British Columbia. Clearly these data may be considered as a birth and death (or survivorship) process. That is, at any period of time t , the observed number of claimants, X_t , can be viewed as the sum of the number of claimants from the previous period, X_{t-1} , who continue to collect benefits (or to survive in the claims queue), and the number of newly injured workers, ε_t . Since burns occur quite infrequently, the data set comprises very low count values, with values of 0, 1 and 2 only occurring. The sample autocorrelation function indicates significant first-order autocorrelation, indicating that there is indeed dependence to be modelled. All descriptive details of the dataset can be found in Freeland and McCabe (2004a). However, for convenience, the main features of the data are reported in Table 1,

Table 1 here

A comparison of the sample mean and variance in Table 1 indicates that the data are marginally underdispersed, suggesting that the binomial may be the appropriate choice of distribution for the arrivals process. However, the arrivals process is latent, with the observed data shedding light on a combination of both it and the latent thinning process. Conditional on the use of binomial thinning to cater for the autocorrelation in the data, the methodology proposed in the paper allows for freedom of choice regarding the distribution adopted for the independent arrivals, with no one model assumed to be more appropriate than another, a priori. Via the model averaging process, the final forecasts reflect the posterior probability assigned to each model in the choice set. With reference to the notation introduced in Section 3 we include the PAR, BAR and NBAR models in the model set.

4.2 Priors

The first step in the construction of the p -step ahead predictive pmf in (9) is the estimation of the three model-specific posterior pdf's according to (7), which requires, in turn, the specification of model-specific priors, $p(\boldsymbol{\theta}_k|M_k)$, $k = 1, 2, 3$. We have chosen to base the analysis on uniform priors for the parameters of all models. Whilst a well-established literature exists concerning the impact of different forms of noninformative priors for parameters in time series models for continuous variables (see, for example, the Special Issue of Journal of Applied Econometrics, Volume 6, No. 4, 1991), no comparable results exist for the discrete case. Moreover, the complexity of the likelihood function in the case of the models entertained here would appear to preclude, in general, the derivation of closed-form expressions for any likelihood-based noninformative priors, such as Jeffreys or reference priors (see Kass and Wasserman, 1996, for a summary), as well as the production of conjugate informative priors. The exception to this is the PAR model, for which an explicit expression for the information matrix and, hence, for independent Jeffreys priors for the two parameters α and λ , is available; see Freeland and McCabe (2004b).

The issue of prior specification is made even more difficult in the current context by the nature of some of the unknown parameters, in particular the parameters n and r that characterize the BAR and NBAR models respectively. In typical applications of the binomial and negative binomial distributions to observed processes, such quantities would not be viewed as unknowns. Hence, there is no well-established, intuitively plausible approach to the specification of prior information (whether noninformative or informative) for such quantities. Uniform priors thus represent a natural starting point in the Bayesian analysis of these types of models, with the subtleties of prior specification left for future work.

In the case of three of the unknown parameters, there is a natural truncation of the parameter space, in which case the uniform priors are in fact proper. Specifically, the binomial thinning

parameter α , as well as the parameters describing the probability of success in the binomial and negative binomial arrivals distributions (q and π respectively), are all constrained to lie between zero and one.⁴The parameters n and r cannot be less than one, and are truncated from below accordingly, whilst the Poisson parameter λ is obviously truncated from below at zero. There are no natural upper bounds on the three parameters λ , n and r . As such, appropriate upper bounds are determined empirically, in such a way that virtually all likelihood mass is encompassed. The use of finite bounds for λ , n and r means that the priors for these three parameters are, like those for α , q and π , proper. However, the somewhat arbitrary nature of the upper bound specification must be acknowledged. As discussed in the next section, the impact on the posterior model probabilities of this arbitrariness is largely obviated via a particular modification to the Bayes Factor. Robustness of the results to the prior specifications is assessed in Section 4.4.

4.3 Computational Details

Given the adoption of truncated uniform priors for all parameters, the joint posterior pdf in (7) becomes

$$p(\boldsymbol{\theta}_k|\mathbf{y}, M_k) \propto \ell(\boldsymbol{\theta}_k|M_k)I_{M_k}, \quad (27)$$

for $k = 1, 2, 3$, where I_{M_k} denotes the indicator function that imposes the necessary a-priori restrictions on the parameter space of M_k . These restrictions, in addition to the boundedness of the likelihood function, ensure that the joint posterior in (27) is proper. The likelihood function in (27) is defined by

$$\begin{aligned} \ell(\boldsymbol{\theta}_k|M_k) &= P(\mathbf{y}|M_k, \boldsymbol{\theta}_k) \\ &= \prod_{t=2}^T P(Y_t = y_t|y_{t-1}, M_k, \boldsymbol{\theta}_k), \end{aligned} \quad (28)$$

conditional on the initial value, $Y_1 = y_1$. Each component in the product in (28) is given by (13), with the relevant model-specific mass function for ε_t specified, in turn, by (19), (21) or (25). Given the low dimension of the parameter space for all three models, each model-specific posterior pdf in (27) is normalized using simple rectangular integration. That is, for $k = 1, 2, 3$, we have $p(\boldsymbol{\theta}_k|\mathbf{y}, M_k) = c\ell(\boldsymbol{\theta}_k|M_k)I_{M_k}$, where c denotes the numerical integrating constant. For example, for the BAR model ($k = 2$), we have

$$c = \left[w_1 w_2 w_3 \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} \sum_{j_3=1}^{m_3} \ell(\alpha^{(j_1)}, q^{(j_2)}, n^{(j_3)}|M_2)I_{M_2} \right]^{-1},$$

with m_1 , m_2 and m_3 denoting the number of grid points for α , q and n respectively, $\alpha^{(j_1)}$, $q^{(j_2)}$ and $n^{(j_3)}$ denoting the grid values, and $w_1 = 0.05$, $w_2 = 0.01$ and $w_3 = 1$ the associated grid widths. The marginal posteriors associated with (27) are computed using successive applications of rectangular integration to the normalized joint posterior. The k th model-specific predictive

pmf is estimated as in (8), with the p -step ahead conditional predictive mass functions specified as described in Section 3.1. With reference to the notation used in (8), for the BAR model for example, $h = w_1 w_2 w_3$, $N_2 = m_1 m_2 m_3$ and $\boldsymbol{\theta}_k^{(i)}$ denotes the i th grid point in the joint space for α , q and n .

To illustrate the arbitrariness in the Bayes factors that results from the arbitrary prior upper bounds for certain of the parameters, denote by c_1 and c_2 respectively the integrating constants associated with particular upper bounds for the priors on the parameters λ and n that characterize M_1 and M_2 respectively. The Bayes factor for M_2 versus M_1 is given by

$$\begin{aligned} BF_{2,1} &= \frac{P(\mathbf{y}|M_2)}{P(\mathbf{y}|M_1)} \\ &= \frac{\int_{\boldsymbol{\theta}_2} \ell(\boldsymbol{\theta}_2|M_2)p(\boldsymbol{\theta}_2|M_2)d\boldsymbol{\theta}_2}{\int_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1|M_1)p(\boldsymbol{\theta}_1|M_1)d\boldsymbol{\theta}_1} \\ &= \frac{c_2 \int_{\boldsymbol{\theta}_2} \ell(\boldsymbol{\theta}_2|M_2)f(\boldsymbol{\theta}_2|M_2)d\boldsymbol{\theta}_2}{c_1 \int_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1|M_1)f(\boldsymbol{\theta}_1|M_1)d\boldsymbol{\theta}_1}, \end{aligned} \quad (29)$$

where $f(\boldsymbol{\theta}_2|M_2)$ denotes the kernel of $p(\boldsymbol{\theta}_2|M_2)$ up to the arbitrary scale factor c_2 and $f(\boldsymbol{\theta}_1|M_1)$ denotes the kernel of $p(\boldsymbol{\theta}_1|M_1)$ up to the arbitrary scale factor c_1 . The Bayes factor in (29) is directly affected by the ratio c_2/c_1 . The corresponding Bayes factors for M_3 versus M_1 will be similarly affected by an arbitrary ratio c_3/c_1 , where c_3 is the integrating constant associated with a particular upper bound on the parameter r in M_3 . The posterior model probabilities, constructed via (2), with the added condition that $P(M_1|\mathbf{y}) + P(M_2|\mathbf{y}) + P(M_3|\mathbf{y}) = 1$, have arbitrary values as a consequence.

In order to offset this problem, the approach of O'Hagan (1995) is adopted. Specifically, given a scalar b , a scaled version of the marginal likelihood is constructed for each M_k ,

$$\begin{aligned} P(\mathbf{y}, b|M_k) &= \frac{\int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)p(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}{\int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)^b p(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k} \\ &= \frac{c_k \int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)f(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}{c_k \int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)^b f(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k} \\ &= \frac{\int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)f(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}{\int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)^b f(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}, \end{aligned} \quad (30)$$

where $f(\boldsymbol{\theta}_k|M_k)$ denotes the kernel of $p(\boldsymbol{\theta}_k|M_k)$ up to the arbitrary integrating constant c_k , $k = 1, 2, 3$. These scaled marginal likelihoods are used to produce so-called *fractional* Bayes factors as

$$BF(b)_{k,1} = \frac{P(\mathbf{y}, b|M_k)}{P(\mathbf{y}, b|M_1)}; \quad k = 2, 3, \quad (31)$$

on which the posterior model probabilities are then based. The cancellation of each of the model-specific integrating constants in (30) means that these arbitrary constants do not impact directly on the Bayes factors and posterior model probabilities. However, there is still a potential element of arbitrariness associated with the choice of b in (30). Following O'Hagan, b is chosen to be small,

but still large enough to ensure robustness to the precise specification of the prior which, in the current context, means robustness to the upper bounds specified for λ , n and r .⁵To provide further confirmation of the robustness of the results to the choice of value for b , in the empirical section below we report results for alternative values of b .

We follow the convention in the Bayesian literature of specifying equal prior probabilities for the alternative models, that is, $P(M_k) = 1/3$ for $k = 1, 2, 3$, in which case the posterior odds ratio in (2) is equated to the fractional Bayes Factor in (31). The estimated posterior model probabilities to be used in (9) are then produced as

$$\widehat{P}(M_k|\mathbf{y}) = \frac{\widehat{P}(\mathbf{y}, b|M_k)}{\sum_{j=1}^3 \widehat{P}(\mathbf{y}, b|M_j)}; \quad k = 1, 2, 3.$$

Each element, $\widehat{P}(\mathbf{y}, b|M_k)$, $k = 1, 2, 3$, is calculated by applying rectangular numerical integration to the numerator and denominator in (30). Note that the numerator in (30) is equivalent to the integrating constant already calculated in the context of normalizing the joint posterior in (27) and, hence, requires no additional computation.

4.4 Empirical Results

Tables 2, 3 and 4 summarize the results associated with estimating the three alternative specifications, PAR, BAR and NBAR. Estimation is based on the first 118 of the 120 observations in the dataset, with the final two observations reserved for an assessment of predictive accuracy. Marginal posterior modes and means are reported for each parameter, as well as 95% Highest Probability Density (HPD) intervals.⁶The last line in each table records the posterior probability for the model. Model probabilities are reported for two values of b in (31), namely $b = \ln(T)/T \approx 5/120$ and $b = \sqrt{T}/T \approx 11/120$, where these two values of b tally, in turn, with the alternative rules given in O'Hagan (1995, Section 6) for specifying this parameter. Probabilities were computed for further values of b , but not reported due to their similarity to the results recorded in the tables. The marginal posterior densities for the parameters of each model are graphed in Figures 1, 2 and 3 respectively.

Table 2 here

Table 3 here

Table 4 here

The first thing to note regarding the results in Tables 2, 3 and 4 is the closeness of the estimates of the correlation parameter α across models. All point estimates of α range between 0.2 and 0.25, with all three interval estimates placing α clearly within the stationary region. For the PAR model,

the modal estimate of the Poisson mean (and variance), λ , is equal to 0.141, with an HPD interval of (0.081,0.221). The Maximum Likelihood estimates of α and λ in the PAR analysis of Freeland and McCabe (2004a) are equal to 0.240 and 0.134 respectively, with associated 95% confidence intervals of (0.007, 0.472) and (0.064, 0.204). All of these classical estimates are broadly consistent with the Bayesian results reported in Table 2, which is not surprising given the use of uniform priors.

The estimated BAR parameters imply a point estimate for the mean of the arrivals process of 0.093 and a variance that is slightly smaller, at 0.083. The NBAR parameter estimates are of course associated with overdispersion in the arrival process, with mean and variance estimates of 0.141 and 0.151 respectively.⁷As would be anticipated given the large number of zeros in the dataset, the posterior estimates of the binomial success probability, q , are very low, with the HPD interval, for example, ranging from 0.001 to only 0.221. The point estimates of the number of trials associated with the binomial arrivals process are also low, with the modal point estimate being equal to the lower bound of 1. That said, the HPD interval for n indicates that non-negligible probability mass is distributed across a wide range of values greater than 1. A similar feature characterizes the marginal posterior for the negative binomial parameter r , with the long upper tail in the density producing both a high mean value relative to the mode and a wide HPD interval. The marked positive skewness of the n and r densities is also evident in the respective graphs in 2 and 3. The appearance of very low counts in the dataset is associated with a very high value for π , since a high value for this parameter ensures that very low numbers of failures occur before the r th success, for any value of $r > 0$.

Figure 1 here

Figure 2 here

Figure 3 here

The posterior model probabilities favour the BAR model, a result that is consistent with the fact that the raw data exhibits slight under-dispersion. Reasonable probability is however assigned to the PAR model, with slightly less to the NBAR model. It is of interest to note that the joint information matrix test conducted in Freeland and McCabe (2004a) yields a p -value of 0.81 for the PAR specification, indicating that the PAR model is acceptable for this data set according to classical criteria.

Table 5 reports the predictive results associated with each model, as well as the model-averaged predictive results. Model averaged results are produced for the two different sets of model probabilities associated with the two values of b cited earlier. The predictive output for each model is

the estimated probability associated with each value in the support of the count variable, conditional on the model. In principle, the support of the p -step ahead predictive pmf is defined over all integers. However, given that the frequency distribution of the data over the values 0, 1 and 2 is 0.839, 0.153 and 0.008, we have produced predictions only for the four integer values 0, 1, 2 and 3. The predictions reflect the proportions in the sample data and, via the non-zero value for α , the last observation in the sample, which is equal to 1. One-step and two-step ahead predictions are reported, in Panel A and Panel B respectively. The actual values of the count variable in periods 119 and 120 respectively are also reported.

Table 5 here

The predictions of all models are quite similar. One-step ahead, all three models assign a probability to the (observed) value of 1 that is larger than the relative frequency of 1 in the sample, due to the influence of the value of 1 observed at the end of the sample. The predictions still, however, reflect the high proportion of 0's in the sample and the relatively small proportions of 1's and 2's. Two-steps ahead, the influence of the last value in the sample is less marked, with the predictive pmf moving closer to the unconditional distribution of the data. The BAR model assigns more probability mass to a value of 1 in both periods than does either the PAR or NBAR models. In this sense, it could be said to be marginally more accurate in period 119 than the alternative models, given that a 1 was actually observed, and marginally less accurate in period 120, in which a 0 was observed. The model-averaging process means that the features of the predictions associated with each model are incorporated in the final predictive results, although the differences between the model-averaged and model-specific results are not marked as a consequence of the similarity in predictions across the different models. To three decimal places, the model-averaged predictives of using the sets different model probabilities associated with the different values of b in (31), are equivalent.

As a final exercise, the model-averaged one-step ahead predictives (associated with a single value of b , $b = 5/120$) are produced for the full data set, yielding the distribution of values 0.864, 0.128, 0.008 and 0.000 over the relevant support. These values can be compared with the estimated predictive probabilities produced in Freeland and McCabe (2004a) using the classical methodology: 0.875, 0.120, 0.008 and 0.000, with associated 95% confidence intervals of (0.812, 0.937), (0.063, 0.171), (0.000, 0.016) and (0.000, 0.001). The classical point estimate of the predictive pmf is seen to be very similar to the Bayesian predictive pmf. The classical point estimates of the predictive probabilities are of course conditioned on particular parameter estimates, with the uncertainty associated with the latter incorporated in the confidence intervals. The predictive estimates are also conditioned the PAR model specification. In contrast, the Bayesian predictive mass function itself directly reflects all such uncertainty.

5 CONCLUSIONS

The paper has presented a Bayesian methodology for producing coherent forecasts of low count time series. Although developed in the context of a specific family of models, the methodology is completely general. With the predictions coming directly from the estimated p -step ahead predictive pmf, a probability is assigned to each value in the support of the count variable. Interpretation of the predictions is thus clear-cut and in complete accordance with the discrete nature of the variable. The spread of the predictive probabilities over the support reflects the uncertainty associated with both the parameter values and the precise model specification.

The approach demonstrated in the paper has exploited analytical results pertaining to the p -step ahead predictive pmf's for the INAR(1) family. However, as noted in the text, the approach can be readily extended to the case where such results are not available. In this case, the numerical procedure is augmented with a component in which the predictive pmf, conditional on both a specific model and set of parameter values, is evaluated via simulation prior to being averaged across both the parameter and model space.

The method has been used to analyse and produce coherent forecasts for a set of low count benefits claim data. With the data generating process having a clear interpretation as a birth and death process, the INAR(1) is an appropriate choice of model. Uncertainty about the form of the arrivals process is catered for by allowing for distributions that are over- and under-dispersed relative to the Poisson. Point and interval estimates of the thinning parameter clearly indicate the presence of short-run dependence. The posterior probabilities associated with alternative arrivals processes slightly favour the binomial as the data is underdispersed. One-step and two-step-ahead forecasts are produced in which both parameter and model uncertainty have been integrated out. As such, the estimated predictive probabilities can be interpreted directly, without the need for the usual qualifications regarding conditioning on assumed parameter values and model specification.

Acknowledgements

This research has been supported by an Australian Research Council Discovery Grant. The authors would like to thank an Associate Editor, two referees and Rob Hyndman for some very helpful suggestions and comments on an earlier version of the paper, as well as participants at the Australasian Meetings of the Econometric Society in Sydney, July, 2003.

References

- [1] Al-Osh M.A., & Alzaid, A.A. (1987). First-order Integer Valued Autoregressive (INAR(1)) Process, *Journal of Time Series Analysis*, 8, 261-275.
- [2] Berger, J.O., & Pericchi, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the American Statistical Society*, 91, 109-122.
- [3] Berger, J.O., & Mortera, J. (1999). Default Bayes Factors for Nonnested Hypothesis Testing, *Journal of the American Statistical Society*, 94, 542-554.
- [4] Bockenholt, U. (1999). Mixed INAR(1) Poisson Regression Models: Analyzing Heterogeneity and Serial Dependencies in Longitudinal Count Data, *Journal of Econometrics*, 89, 317-338.
- [5] Brännäs, K. (1994). Estimation and Testing in Integer Valued AR(1) Models, *Umea Economic Studies No. 355*. University of Umea.
- [6] Brännäs, K., & Hellstrom, J. (2001). Generalized Integer-Valued Autoregression, *Econometric Reviews*, 20, 425-443.
- [7] Chib, S. (1995). Marginal Output from the Gibbs Output, *Journal of the American Statistical Association*, 90, 1313-1321.
- [8] Chib, S., Greenberg, E., & Winkelmann, R. (1998). Posterior Simulation and Bayes Factors in Panel Count Data Models, *Journal of Econometrics*, 86, 33-54.
- [9] Chib, S., & Winkelmann, R. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data, *Journal of Business and Economic Statistics*, 19, 428-435.
- [10] Durbin, J., & Koopman, S.J. (2000). Time Series Analysis of Non-Gaussian Observations Based on State Space Models From Both Classical and Bayesian Perspectives, *Journal of the Royal Statistical Society, B*, 60, 3-56.
- [11] Cameron, A.C., & Trevedi, P.K. (1998). *Regression Analysis of Count Data*, Cambridge University Press.
- [12] Fahrmeir, L., & Tutz, G. (1994). *Multivariate Statistical Modeling Based on Generalized Linear Models*, Springer-Verlag, New York.
- [13] Freeland, R.K. (1998). *Statistical Analysis of Discrete Time Series with Applications to the Analysis of Workers Compensation Claims Data*, Unpublished Ph.D. Thesis. The University of British Columbia, Canada.
- [14] Freeland, R.K., & McCabe, B.P.M. (2004a). Forecasting Discrete Valued Low Count Time Series, *International Journal of Forecasting*, 20, 427-434.

- [15] Freeland, R.K., & McCabe, B.P.M. (2004b). Analysis of Low Count Time Series Data by Poisson Autoregression, *Journal of Time Series Analysis*, 25, 701-722.
- [16] Grunwald, G.K., Hyndman, R.J., Tedesco, L., & Tweedie, R.L. (2000). Non-Gaussian Conditional Linear AR(1) Models, *Australian and New Zealand Journal of Statistics*, 42, 478-495.
- [17] Hoeting, J.A., Madigan, D., Raftery, A.E., & Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial, *Statistical Science*, 14, 382-417.
- [18] Jung, R.C., & Tremayne, A.R., (2003). Testing for Serial Dependence in Time Series Models of Counts, *Journal of Time Series Analysis*, 24, 65-84.
- [19] Kass, R.E., & Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules, *Journal of the American Statistical Association*, 91, 1343-1370.
- [20] Kleibergen, F. (2003). Invariant Bayesian Inference in Regression Models that is Robust Against the Jeffreys-Lindley's Paradox, mimeo, University of Amsterdam.
- [21] McCabe, B.P.M., Martin, G.M., & Freeland, R.K. (2004). Testing for Dependence in Non-Gaussian Time Series Data, *Working Paper 13/04, Department of Econometrics and Business Statistics, Monash University*.
- [22] McDonald, I.L. & Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*, Chapman and Hall, London.
- [23] McKenzie, E. (1986). Autoregressive Moving Average Processes with Negative Binomial and Geometric Marginal Distributions, *Advances in Applied Probability*, 18, 679-705.
- [24] McKenzie, E. (1988). Some ARMA Models for Dependent Sequences of Poisson Counts, *Advances in Applied Probability*, 20, 822-835.
- [25] O'Hagan, (1995). Fractional Bayes Factors for Model Comparison, *Journal of the Royal Statistical Society, B*, 57, 99-138.
- [26] Ronning, G., & Jung, R.C. (1992). Estimation of a First Order Autoregressive Process with Poisson Marginals for Count Data. In: L. Fahrmeir et al. (Eds.), *Advances in GLIM and Statistical Modelling*. Springer-Verlag, New York.
- [27] West, M., & Harrison, P.J. (1997). *Bayesian Forecasting and Dynamic Models*, Second Edition, Springer-Verlag, New York.
- [28] West, M., Harrison, P.J., & Migon, H.S. (1985). Dynamic Generalized Linear Models and Bayesian Forecasting, *Journal of the American Statistical Society*, 80, 73-83.
- [29] Winkelmann, R. (2000). *Econometric Analysis of Count Data*, Third Edition, Springer-Verlag, Berlin.

Endnotes

1. See Hoeting, Madigan, Raftery and Volinsky (1999) for an overview of the use of Bayesian model averaging in prediction.
2. Analyses that adopt distributional assumptions for ε_t other than Poisson include McKenzie (1986), Bockenholt (1999) and Brannas and Hellstrom (2001). However, none of this work allows for a range of distinct distributional assumptions to be catered for simultaneously.
3. If the assumption of a fixed Y_1 is replaced by the assumption that Y_1 is Poisson with mean λ , then the stationary distribution of $Y_t|M_1, \theta_1$ is also Poisson, with mean $\lambda/(1 - \alpha)$.
4. Note that in the case of continuous autoregressive models, the use of a uniform prior for the autoregressive coefficient is associated with bias towards stationarity, when the true parameter is close to unity. Since the data set under study here has been found, via standard preliminary analysis, to exhibit clearly stationary behaviour, we hypothesize that any such bias, if it exists in this context, would not have a qualitative impact on the results.
5. For alternative approaches to Bayes factor calculation in the presence of noninformative priors see Berger and Pericchi (1996), Berger and Mortera (1999) and Kleibergen (2003).
6. We construct the 95% probability interval in such a way that the ordinates of the upper and lower bounds are as close as possible to being equal, subject to the restriction that the tail probabilities add to 5% and that each tail has non-zero probability. For the densities that are either truncated or multimodal this means that there may be ordinates within the interval that are smaller than ordinates beyond the interval. In this sense the interval is only an approximate HPD interval.
7. The estimates of the mean and variance of both the binomial and negative binomial arrivals processes are the modal values of the posterior densities for these two parameters. These densities are constructed by applying a kernel smoothing algorithm to the relevant functions of independent draws from the marginal posteriors of n and q , in the case of the binomial distribution, and r and π in the negative binomial case.
8. The k th order autocorrelation of the INAR(1) model is proportional to α^k , for $\alpha < 1$; see Grunwald, Hyndman, Tedesco and Tweedie (2000).
9. The mean estimates of the arrivals mean and variance indicate more overdispersion, with respective values of 3.779 and 7.814.

Table 1: Summary Statistics for the Burns Claims Data

Minimum Count	Maximum Count	Median	Mode	Mean	Variance
0	2	0	0	0.175	0.162

Table 2: Parameter Estimates for the PAR model

Parameter		
α	Mode	0.200
	Mean	0.234
	95% HPD	(0.001, 0.500)
λ	Mode	0.141
	Mean	0.146
	95% HPD	(0.081, 0.221)
$P(M_k \mathbf{y})$	$b = 5/120$	0.361
	$b = 11/120$	0.319

Table 3: Parameter Estimates for the BAR model

Parameter		
α	Mode	0.200
	Mean	0.238
	95% HPD	(0.001, 0.500)
n	Mode	1.000
	Mean	4.472 ^(a)
	95% HPD	(1.000, 18.000)
q	Mode	0.011
	Mean	0.078
	95% HPD	(0.001, 0.221)
Arrivals mean	Mode	0.093
Arrivals variance	Mode	0.083
$P(M_k \mathbf{y})$	$b = 5/120$	0.410
	$b = 11/120$	0.434

(a) The noninteger mean of the marginal density of n is reported for descriptive purposes only.

Table 4: Parameter Estimates for the NBAR model

Parameter		
α	Mode	0.200
	Mean	0.250
	95% HPD	(0.001, 0.500)
r	Mode	1.000
	Mean	23.711 ^(a)
	95% HPD	(1.000, 97.000)
π	Mode	0.998
	Mean	0.973
	95% HPD	(0.880, 0.999)
Arrivals mean	Mode	0.141
Arrivals variance	Mode	0.151
$P(M_k \mathbf{y})$	$b = 5/120$	0.229
	$b = 11/120$	0.248

(a) The noninteger mean of the marginal density of r is reported for descriptive purposes only.

Table 5: Predictive Probability Mass Functions

Panel A: One-step Ahead Predictive Probabilities					
Count Value	Model-Specific Predictive Probabilities			Model-Averaged Predictive Probabilities	
	PAR	BAR	NBAR	$b = 5/120$	$b = 11/120$
0	0.662	0.654	0.663	0.659	0.659
1	0.300	0.310	0.298	0.303	0.303
2	0.036	0.035	0.036	0.036	0.036
3	0.002	0.001	0.003	0.002	0.002
Actual Value = 1					
Panel B: Two-step Ahead Predictive Probabilities					
Count Value	Model-Specific Predictive Probabilities			Model-Averaged Predictive Probabilities	
	PAR	BAR	NBAR	$b = 5/120$	$b = 11/120$
0	0.780	0.775	0.777	0.777	0.777
1	0.195	0.206	0.195	0.199	0.199
2	0.023	0.018	0.025	0.022	0.022
3	0.002	0.001	0.002	0.002	0.002
Actual Value = 0					

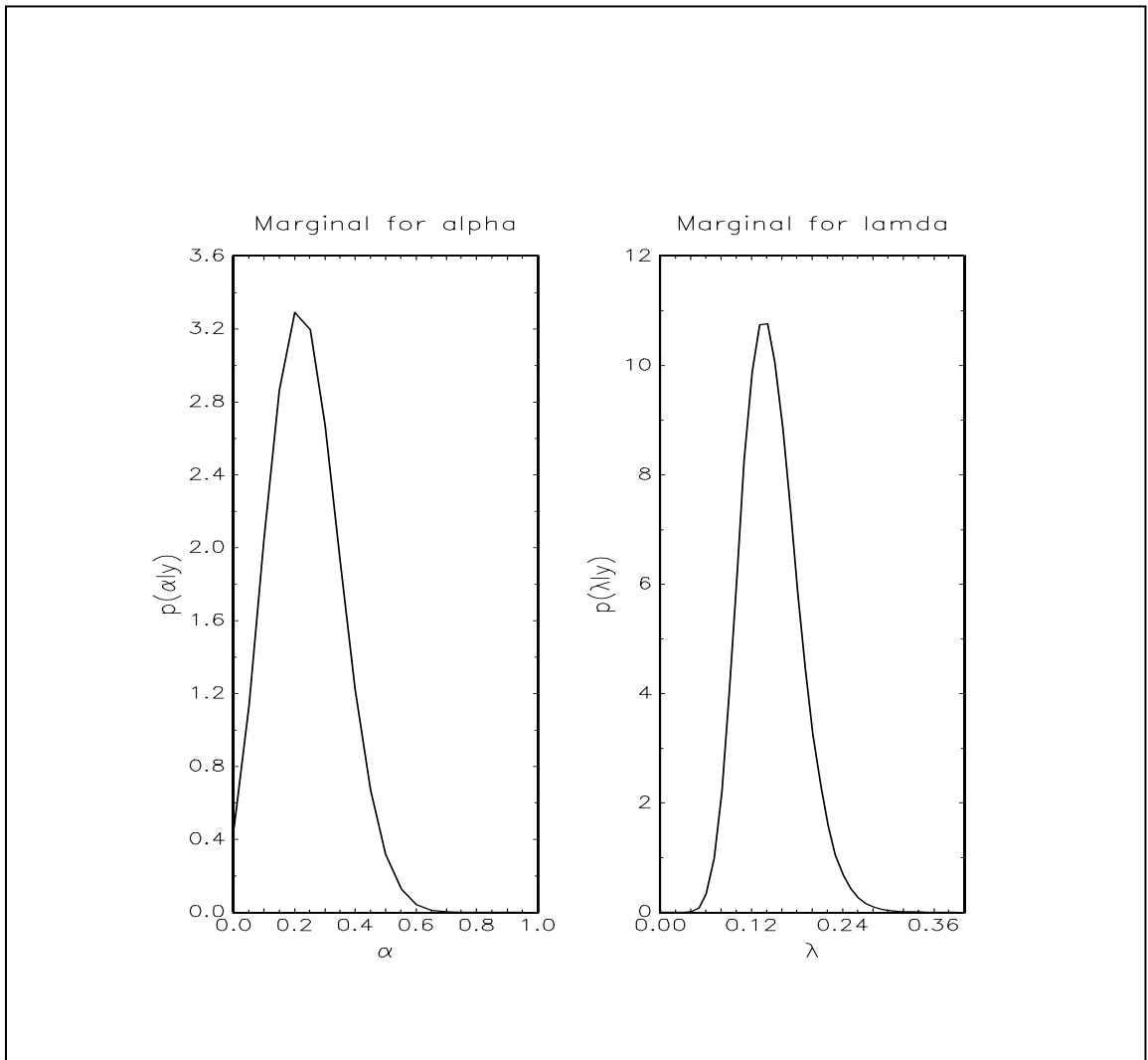


Figure 1: Poisson Arrivals

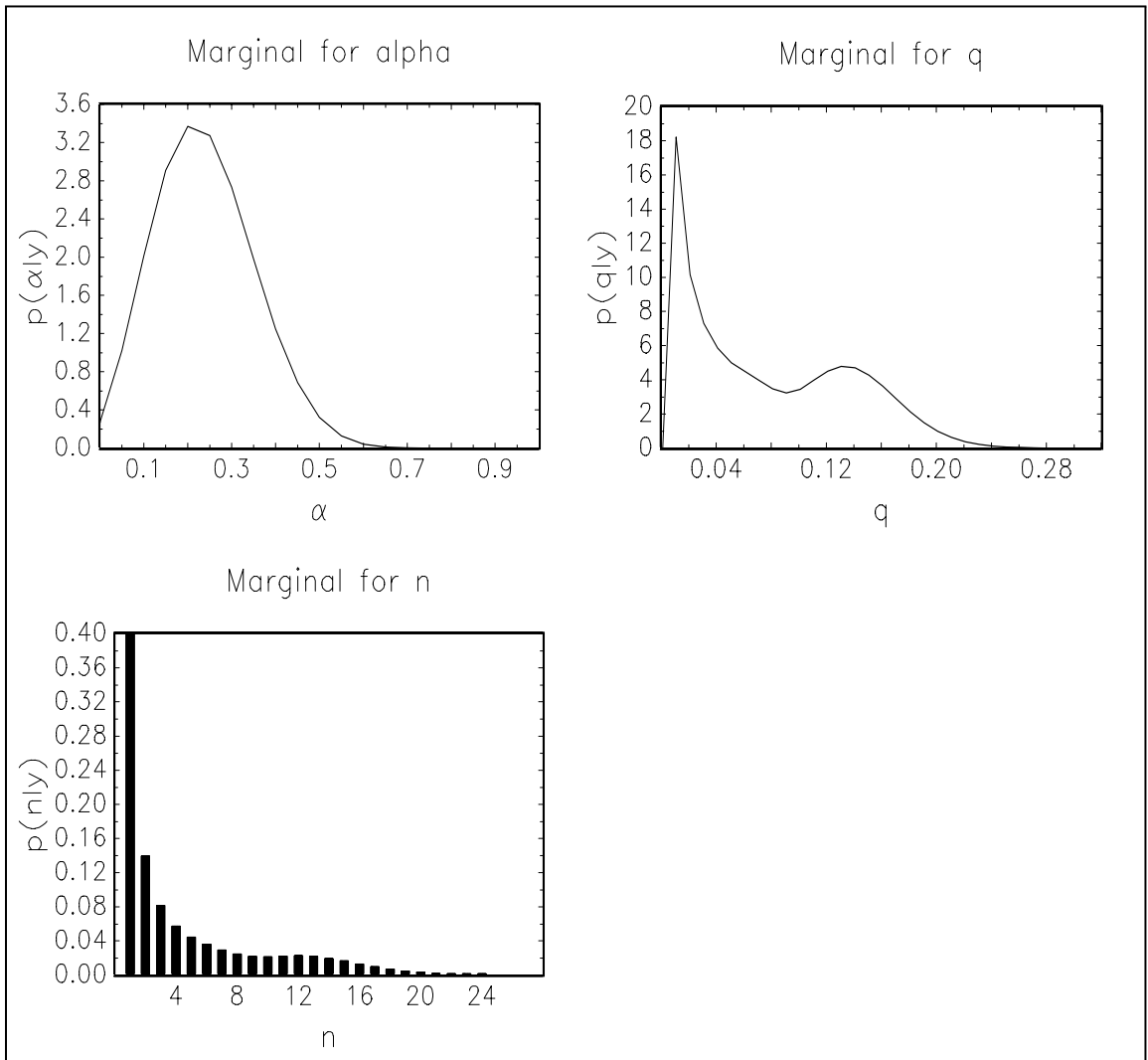


Figure 2: Binomial Arrivals

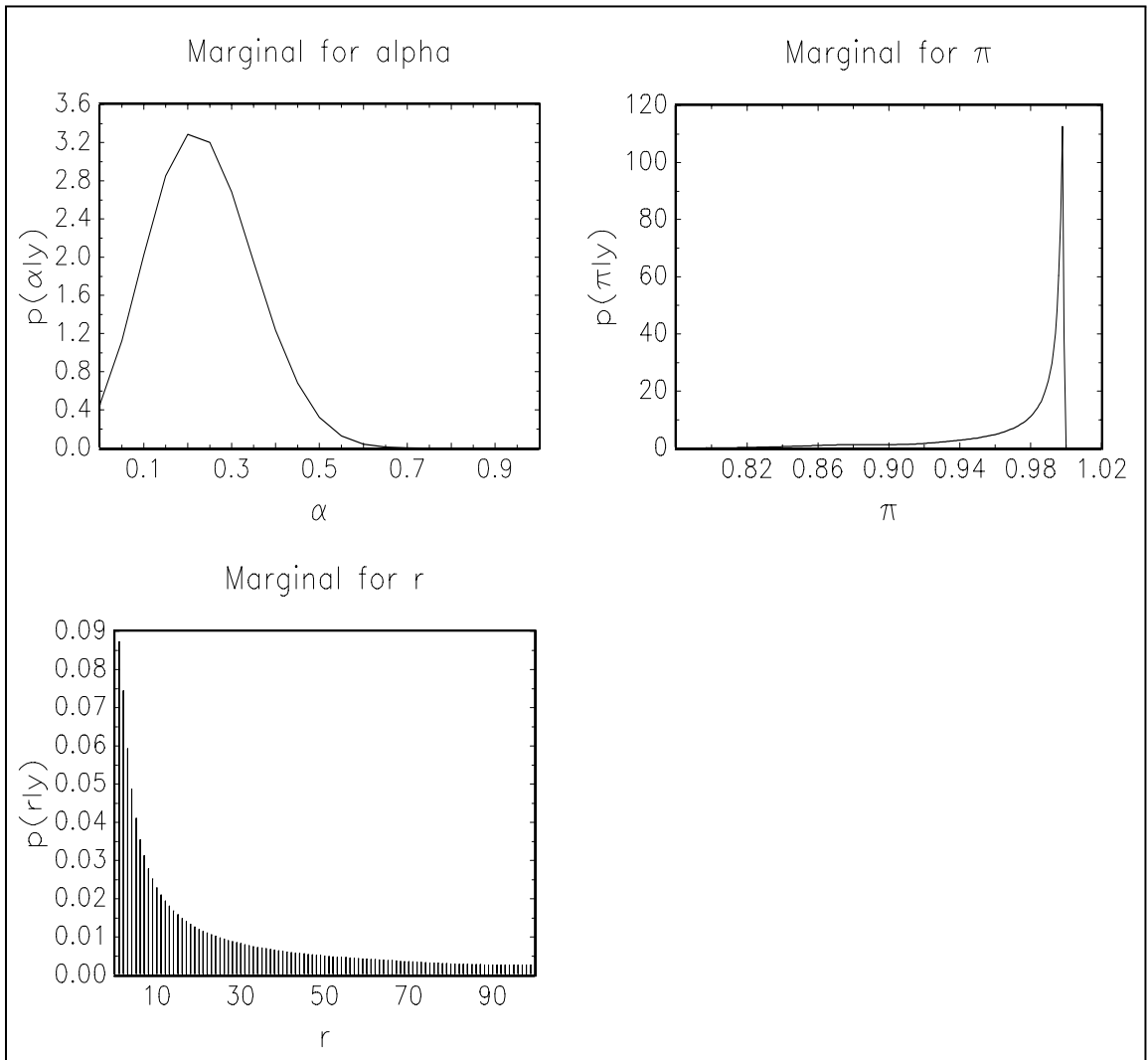


Figure 3: Negative Binomial Arrivals