ISSN 1440-771X



AUSTRALIA

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Coherent Predictions of Low Count Time Series

B.P.M. McCabe and G.M. Martin

Working Paper 8/2003

Coherent Predictions of Low Count Time Series

B.P.M. MCCABE and G.M. MARTIN*

April 9, 2003

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 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 two sets of Canadian wage loss claims data.

KEYWORDS: Forecasting; Discrete Time Series; INAR(1); 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, Chatham Street, Liverpool, L69 7ZH, Great Britain. (Email: Brendan.McCabe@liverpool.ac.uk). G.M. Martin is Senior Lecturer, Department of Econometrics and Business Statistics, P.O. Box, 11E, Monash University, Victoria, 3800, Australia. (Email: Gael.Martin@Buseco.monash.edu.au).

.

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 in only 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, and Cameron and Trevedi 1998, 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 which are coherent, in the sense of producing only integer forecasts of the count variable, is still in its infancy (see, for example, Freeland and McCabe 2002a).

In this paper we develop a general methodology for producing coherent predictions of low count data. The predictions are based on 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 contrast to the usual applications of this model, in which the arrivals process is fully specified (usually as Poisson), we allow the arrivals to follow any distribution 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 probability mass function (pmf). Since the predictive pmf assigns probabilities to the different possible values of the random variable in its discrete support, forecast coherency is assured. This contrasts with the non-integer point forecasts that would be produced via the use of conditional expectations and the non-integer values covered by the associated prediction intervals. To eliminate unwanted elements from the conditioning set of the predictive pmf, Bayesian methods are used. That is, the mass function on which forecasts are ultimately based averages over uncertainty in the parameter values as well as the uncertainty in 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. An alternative approach is to model the dynamics in the

counts indirectly, via a latent process. For example, West and Harrison (1997), Chib, Greenburg and Winkelmann (1998), Durbin and Koopman (2000) and Chib and Wilkemann (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 tenuous (see McCabe, Martin and Freeland 2003). Moreover, the addition to the set of unknowns of an unobservable latent factor for each observation, although manageable via the techniques of data augmentation, does markedly increase the computational burden. In contrast, analysis of the parsimonious specifications considered in this paper requires only low-dimensional numerical quadrature 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 two sets 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 which can assume only count values $\{0, 1, 2...\}$ at each point in time $t = 1, 2, \ldots, T$. The model generating Y_t is assumed to be any one within a set of K models which are entertained, with those models denoted by M_k, $k = 1, 2, \ldots, K$. Given the vector of observed data, $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, the p-step ahead predictive pmf is defined as

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

where $P(Y_{T+p} = y_{T+p}|M_k, \mathbf{y})$ is the kth 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, $P(M_k|\mathbf{y})$, $k = 1, 2, \ldots, K$, are constructed via the posterior odds ratios for models $M_2, ..., M_K$, relative to reference model M_1. The posterior odds ratio for model M_k relative to M_1, $PO_{k,1}$, is given by the product of the prior odds ratio and the Bayes Factor,

$$PO_{k,1} = \frac{P(M_k|\mathbf{y})}{P(M_1|\mathbf{y})} = \frac{P(M_k)}{P(M_1)} \times \frac{P(\mathbf{y}|M_k)}{P(\mathbf{y}|M_1)}, \qquad k = 2, ..., K, \tag{2}$$

where

$$P(\mathbf{y}|M_k) = \int_{\boldsymbol{\theta}_k} \ell(\boldsymbol{\theta}_k|M_k)p(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k \tag{3}$$

is the marginal likelihood of model M_k. The marginal likelihood 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, $p(\boldsymbol{\theta}_k|M_k)$. The

likelihood function reflects the form of the joint probability mass function for Y_t, $t = 1, 2, \ldots, T$, under M_k,

$$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T | M_k, \boldsymbol{\theta}_k), \tag{4}$$

whilst the prior density, $p(\boldsymbol{\theta}_k|M_k)$, expresses the prior beliefs about the unknown parameters in M_k.

The kth model-specific predictive pmf is defined as

$$P(Y_{T+p} = y_{T+p}|M_k, \mathbf{y}) = \int_{\boldsymbol{\theta}_k} P(Y_{T+p} = y_{T+p}|M_k, \boldsymbol{\theta}_k, \mathbf{y}) p(\boldsymbol{\theta}_k|M_k, \mathbf{y}) d\boldsymbol{\theta}_k, \tag{5}$$

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

$$p(\boldsymbol{\theta}_k|M_k, \mathbf{y}) \propto \ell(\boldsymbol{\theta}_k|M_k) p(\boldsymbol{\theta}_k|M_k). \tag{6}$$

As is clear from the expressions in (5) and (1), the p-step ahead predictive pmf averages over both parameter uncertainty, conditional on an assumed model, as quantified by $p(\boldsymbol{\theta}_k|M_k, \mathbf{y})$, and model uncertainty, as quantified by the distribution of posterior model probabilities $P(M_k|\mathbf{y})$, $k = 1, 2, \ldots, K$. That is, the Bayesian method enables $\boldsymbol{\theta}_k$ and M_k to be eliminated from the conditioning set of the conditional mass function, $P(Y_{T+p} = y_{T+p}|M_k, \boldsymbol{\theta}_k, y)$, in a way which 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, $p(\boldsymbol{\theta}_k|M_k, \mathbf{y})$ can be evaluated using deterministic numerical integration. The kth model-specific p-step ahead predictive pmf in (5) 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. That is, the kth model-specific predictive pmf is estimated as

$$\widehat{P}(Y_{T+p} = y_{T+p}|M_k, \mathbf{y}) = h \sum_{i=1}^{N_{\theta_k}} P(Y_{T+p} = y_{T+p}|M_k, \boldsymbol{\theta}_k^{(i)}, \mathbf{y}) p(\boldsymbol{\theta}_k^{(i)}|M_k, \mathbf{y}), \tag{7}$$

where N_{θ_k} is the number of gridpoints used in evaluating the density $p(\boldsymbol{\theta}_k|M_k, \mathbf{y})$, h is the grid width and $p(\boldsymbol{\theta}_k^{(i)}|M_k, \mathbf{y})$ denotes the ordinate of $p(\boldsymbol{\theta}_k|M_k, \mathbf{y})$ at grid value $\boldsymbol{\theta}_k^{(i)}$. When $\boldsymbol{\theta}_k$ is of low dimension, the marginal likelihood for model M_k in (3) 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 $\widehat{P}(M_k|\mathbf{y})$, the model-averaged predictive pmf in (1) is then estimated as

[1] See Hoeting, Madigan, Raftery and Volinsky (1999) for an overview of the use of Bayesian model averaging in prediction.

$$\widehat{P}(Y_{T+p} = y_{T+p}|\mathbf{y}) = \sum_{k=1}^{K} \widehat{P}(Y_{T+p} = y_{T+p}|M_k, \mathbf{y})\widehat{P}(M_k|\mathbf{y}). \tag{8}$$

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 Markov Chain Monte Carlo sampling could be used to produce a sample from $p(\boldsymbol{\theta}_k|M_k, \mathbf{y})$, with $P(Y_{T+p} = y_{T+p}|M_k, \mathbf{y})$ estimated as an average of the conditional mass functions over the parameter draws. Estimates of the marginal likelihoods and, hence, the model probabilities, would also be based on the simulation output; see, for example, Chib (1995).

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 applied in Freeland and McCabe (2000a,b) and Jung and Tremayne (2003), amongst others. Let $Y_1, Y_2, \cdots, Y_T$ be a series of dependent counts generated according to the following model,

$$Y_t = \alpha \circ Y_{t-1} + \varepsilon_t, \tag{9}$$

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

$$\alpha \circ Y_{t-1} = \sum\nolimits_{i=1}^{Y_{t-1}} \mathsf{B}_{it}, \tag{10}$$

where $\mathsf{B}_{1t}, \mathsf{B}_{2t}, \ldots, \mathsf{B}_{Y_{t-1}t}$ are iid Bernoulli random variables with

$$P(\mathsf{B}_{it} = 1) = 1 - P(\mathsf{B}_{it} = 0) = \alpha. \tag{11}$$

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} 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).[2]

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. The model set, M_k, $k = 1, 2, \ldots, K$, 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

[2] 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.

binomial. These three models are appropriate, respectively, for arrivals which 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 we present in Section 3.1 is, however, appropriate for any arrivals process embedded within the INAR(1) class.[3] 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 (8) requires evaluation of the conditional mass function, $P(Y_{T+p} = y_{T+p}|M_k, \boldsymbol{\theta}_k, \mathbf{y})$, in the integrand in (5). In the case of the INAR(1) model, the form of this function is determined by the structure of the model in (9). 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}, given arrivals model M_k, is thus given by the convolution of the two random components, as follows,

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

where $P_{M_k}(\varepsilon_t = y_t - s)$ denotes the mass function for the arrival ε_t under M_k. Equally, under thinning, the distribution of Y_t given 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

$$P[Y_t = y_t|y_{t-p}, M_k, \boldsymbol{\theta}_k] = \sum_{s=0}^{\min(y_t, y_{t-p})} P\left[Bin(\alpha^p, y_{t-p}) = s\right] P\left[\sum_{j=0}^{p-1} \left(\alpha^j \circ \varepsilon_{t-j} = y_t - s\right)\right]. \qquad (12)$$

Also, since the arrivals are iid,

$$P\left[\sum_{j=0}^{p-1} \left(\alpha^j \circ \varepsilon_{t-j} = y_t - s\right)\right] = \left\{P\left[\left(\alpha^{p-1} \circ \varepsilon_{t-j} = y_t - s\right]\right\} * \ldots * \left\{P\left[\left(\alpha^0 \circ \varepsilon_{t-j} = y_t - s\right]\right\}\right., \qquad (13)$$

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 $p = 2$, $\{a_n\} = \left\{P\left[\left(\alpha^1 \circ \varepsilon_{t-j} = n\right]\right\}\right($and

[3] 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.

$\{b_n\} = \{P[(\alpha^0 \circ \varepsilon_{t-j} = \varepsilon_{t-j} = n]\}$(and

$$
\begin{aligned}
P\left[\sum_{j=0}^{1}(\alpha^j \circ \varepsilon_{t-j} = y_t - s\right] &= P[\alpha \circ \varepsilon_{t-1} + \varepsilon_t = y_t - s] \\
&= \sum_{l=0}^{y_t-s}(P[\alpha \circ \varepsilon_{t-1} = l]P_{M_k}(\varepsilon_t = y_t - s - l) \\
&= \sum_{l=0}^{y_t-s}\left[\sum_{m=l}^{\infty}(P[Bin(\alpha, m) = l]P_{M_k}(\varepsilon_{t-1} = m)\right] \\
\times P_{M_k}(\varepsilon_t &= y_t - s - l), \qquad (14)
\end{aligned}
$$

where the expression in square brackets in the last line in (14) follows from mixing the conditional binomial variate, $Bin(\alpha, m)$, 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 (13) may be carried out in any order.

For $j = 0, .., p-1$, and where the summation limits may be truncated to accommodate restrictions on the support of the arrivals process if necessary,

$$
P\left[\alpha^j \circ \varepsilon_{t-j} = y_t - s\right](= \left[\sum_{m=y_t-s}^{\infty}(\ P[Bin(\alpha^j, m) = y_t - s]P_{M_k}[\varepsilon_{t-j} = m]\right](. \qquad (15)
$$

Hence, the p-step ahead conditional distribution is given by

$$
P[Y_{T+p} = y_T | y_T, M_k, \boldsymbol{\theta}_k] = \sum_{s=0}^{\min(y_{T+p}, y_T)} P[Bin(\alpha^p, y_T) = s] P\left[\sum_{j=0}^{p-1}(\alpha^j \circ \varepsilon_{t-j} = y_{T+p} - s\right], \qquad (16)
$$

with each component in (16) evaluated using (13) and (15).

The precise form of the conditional distribution in (16) depends of course on the distributional assumption, M_k, adopted for the arrivals process. We demonstrate the form of (16) 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

$$
P_{M_1}(\varepsilon_t = w) = \frac{e^{-\lambda}\lambda^w}{w!}, \quad w = 0, 1, 2.... \qquad (17)
$$

for all $t = 1, 2, \ldots, T$, with the ε_t independent. The binomial thinning process in (9) 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, $\alpha^j \circ \varepsilon_{t-j}$ in (16), is also Poisson, as

is the sum of the p such independent variates which appears as the second term on the right hand side of the expression.[4] Thus, in this case, (16) collapses to

$$P(Y_{T+p} = y_{T+p}|M_1, \boldsymbol{\theta}_1, \mathbf{y}) = \sum_{s=0}^{\min(y_{T+p}, y_T)} \Big(P[Bin(\alpha^p, y_T) = s] \times \\ \times \frac{1}{(y_{T+p}-s)!} \exp\left\{ -\lambda \frac{1-\alpha^p}{1-\alpha} \right\} \left(\lambda \frac{1-\alpha^p}{1-\alpha} \right)^{y_{T+p}-s}. \quad (18)$$

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

$$P_{M_2}(\varepsilon_t = w) = \binom{n}{w} q^w (1-q)^{(n-w)}; \quad w = 0, 1, 2, \ldots, n, \quad (19)$$

for all $t = 1, 2, \ldots T$, with the ε_t independent. We refer to the binomial thinning process in (9), 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 (16), with evaluation of that distribution requiring repeated iteration of the relevant convolution formulae, with $P_{M_k}(.)$ being as given in (19). For example, when $p = 2$, (14) becomes

$$P[\varepsilon_t + \alpha \circ \varepsilon_{t-1} = y_{T+2} - s] = \sum_{l=0}^{y_{T+2}-s} \Big(P[Bin(\alpha, \varepsilon_{t-1}) = l] \, P[Bin(q, n) = y_{T+2} - s], \quad (20)$$

where

$$P[Bin(\alpha, \varepsilon_{t-1}) = l] = \sum_{m=l}^{\infty} \Big(P[Bin(\alpha, m) = l] \, P[Bin(q, n) = m]. \quad (21)$$

Substituting (20) and (21) into (16) leads to a 2-step ahead conditional predictive pmf of the form

$$P(Y_{T+2} = y_{T+2}|M_2, \boldsymbol{\theta}_2, \mathbf{y}) = \sum_{s=0}^{\min(y_{T+2}, y_T)} \{ P\left[\Big(Bin(\alpha^2, y_T) = s \right] \Big(\times \\ \sum_{l=0}^{y_{T+2}-s} \sum_{m=l}^{\infty} \Big(P[Bin(\alpha, m) = l] \, P[Bin(q, n) = m] \\ \times P[Bin(q, n) = y_{T+2} - s - l] \}. \quad (22)$$

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

[4] 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 $\lambda/(1-\alpha)$.

as the number of failures, w, prior to the rth success, with mass function given by

$$P_{M_3}(\varepsilon_t = w) = \binom{r+w-1}{w}\pi^r(1-\pi)^w; \quad w = 0,1,2,\ldots, \tag{23}$$

for all $t = 1,2,\ldots T$, with the ε_t independent. We refer to the binomial thinning process in (9), 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 (16), with evaluation of that distribution requiring repeated iteration of the relevant convolution formulae, with $P_{M_k}(.)$ now being as given in (23). For $p = 2$, the conditional predictive pmf is

$$\begin{aligned} P(Y_{T+2} = y_{T+2}|M_3, \boldsymbol{\theta}_3, \mathbf{y}) \quad = \quad & \sum_{s=0}^{\min(y_{T+2}, y_T)} \left\{P\left[Bin(\alpha^2, y_T) = s\right] \times \right. \\ & \sum_{l=0}^{y_{T+2}-s} \sum_{m=l}^{\infty} \left(P\left[Bin(\alpha, m) = l\right] P\left[NBin(\pi, r) = m\right]\right. \\ & \left.\left. \times P\left[NBin(\pi, r) = y_{T+2} - s - l\right]\right\}. \end{aligned} \tag{24}$$

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 (5) 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 which 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 values of Y_t, $t = 1,2,\ldots T$. The draws of Y_{T+1} would then be used to produce an estimate of $P(Y_{T+1} = y_{T+1}|M_k, \boldsymbol{\theta}_k, \mathbf{y})$, 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 conditional on $p-1$ additional out-of-sample values, $Y_{T+p-1}, Y_{T+p-2}, \ldots, Y_{T+1}$, which would, in turn, be produced from the estimated mass functions for the corresponding periods.

4 EMPIRICAL APPLICATIONS

4.1 Data Description

We apply the proposed methodology to two data sets, both of which have been obtained from the Workers Compensation Board (WCB) of British Columbia, Canada. Both datasets are also

Table 1: Summary Statistics for BURNS and CUTS data

	BURNS	CUTS
Statistic		
Minimum Count	0	1
Maximum Count	2	21
Median	0	5
Mode	0	5
Mean	0.175	6.130
Variance	0.162	11.700

analyzed respectively in Freeland and McCabe (2002a,b), via application of classical inferential procedures to the PAR model. The datasets both comprise 120 monthly counts of workers collecting Wage Loss Benefits for 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. The first dataset relates to claimants who have received burns injuries, whilst the second dataset relates to claimants whose injuries were either cuts or lacerations. We refer to the two datasets as BURNS and CUTS respectively. Both datasets comprise low integer values, with the BURNS data, in particular, assuming only very low counts of 0, 1 and 2. The sample autocorrelation functions of both datasets also indicate significant first-order autocorrelation, indicating that there is indeed dependence to be modelled. Details of each dataset can be found in Freeland and McCabe (2002a,b) respectively. However, for convenience, the main features of each dataset are reported in Table 1.

The CUTS data are clearly overdispersed, whilst the BURNS data are marginally underdispersed. At first glance, these features of the raw data tend to suggest that binomial and negative binomial arrivals respectively may be the appropriate choice for modelling each dataset. 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 complete 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 for both datasets. With all three of these specifications having low-dimensional parameter sets, all analysis is performed using low-dimensional deterministic integration.[5]

4.2 Prior Specification and Bayes Factor Calculation

We have chosen to base the analysis on noninformative 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. Since prior specification is not the focus of this paper, we adopt the simplest possible specification for the binomial thinning parameter α, namely a uniform prior, defined over the interval from zero to one.[6]. We adopt uniform priors for all parameters of the alternative arrivals processes, subject to the obvious restrictions. For example, the values of n and r for the binomial and negative binomial arrivals distributions respectively, cannot be less than one. The parameters describing the probability of success in each case (q and π respectively) are constrained to lie between zero and one. In order to offset the arbitrariness in the Bayes Factors which results from the arbitrary upper bounds on the parameters n and r, the approach of O'Hagan (1995) is adopted.[7] We follow the convention in the literature of specifying equal prior probabilities for the alternative models, in which case the posterior odds ratio in (2) is equivalent to the Bayes Factor.

4.3 Burns Data Results

Tables 2, 3 and 4 summarize the results associated with estimating the three alternative specifications, PAR, BAR and NBAR, using the BURNS data. 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

[5] Freeland and McCabe (2002a) assume Poisson arrivals in their analysis of the BURNS data. The same arrivals distribution is used in Freeland and McCabe (2000b) to analyse the CUTS data, with the overdispersion catered for via the use of covariates. Estimates of parameters which are common to the formulations of the current paper and the Freeland and McCabe papers are quite close.

[6] Note in the case of continuous autoregressive models, the use of such a prior is associated with bias towards stationarity in the case where the true parameter is close to unity. Since neither dataset under study here has been found, via standard preliminary analysis, to exhibit non stationary unit root behaviour, we hypothesize that any such bias, if it exists in this context, would not have a qualitative impact on the results.

[7] Following O'Hagan, the exponent to which the likelihood function is raised in the construction of the modified Bayes Factor, is chosen to be small, but still large enough to ensure some robustness to the precise specification of the prior. In the case of the uniform priors used here for n and r, this robustness relates to the intervals over which the priors are specified. 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).

well as 95% Highest Probability Density (HPD) intervals.[8] The last line in each table records the posterior probability for the model. The marginal posterior densities for the parameters of each model are graphed in Figures 1, 2 and 3 respectively.

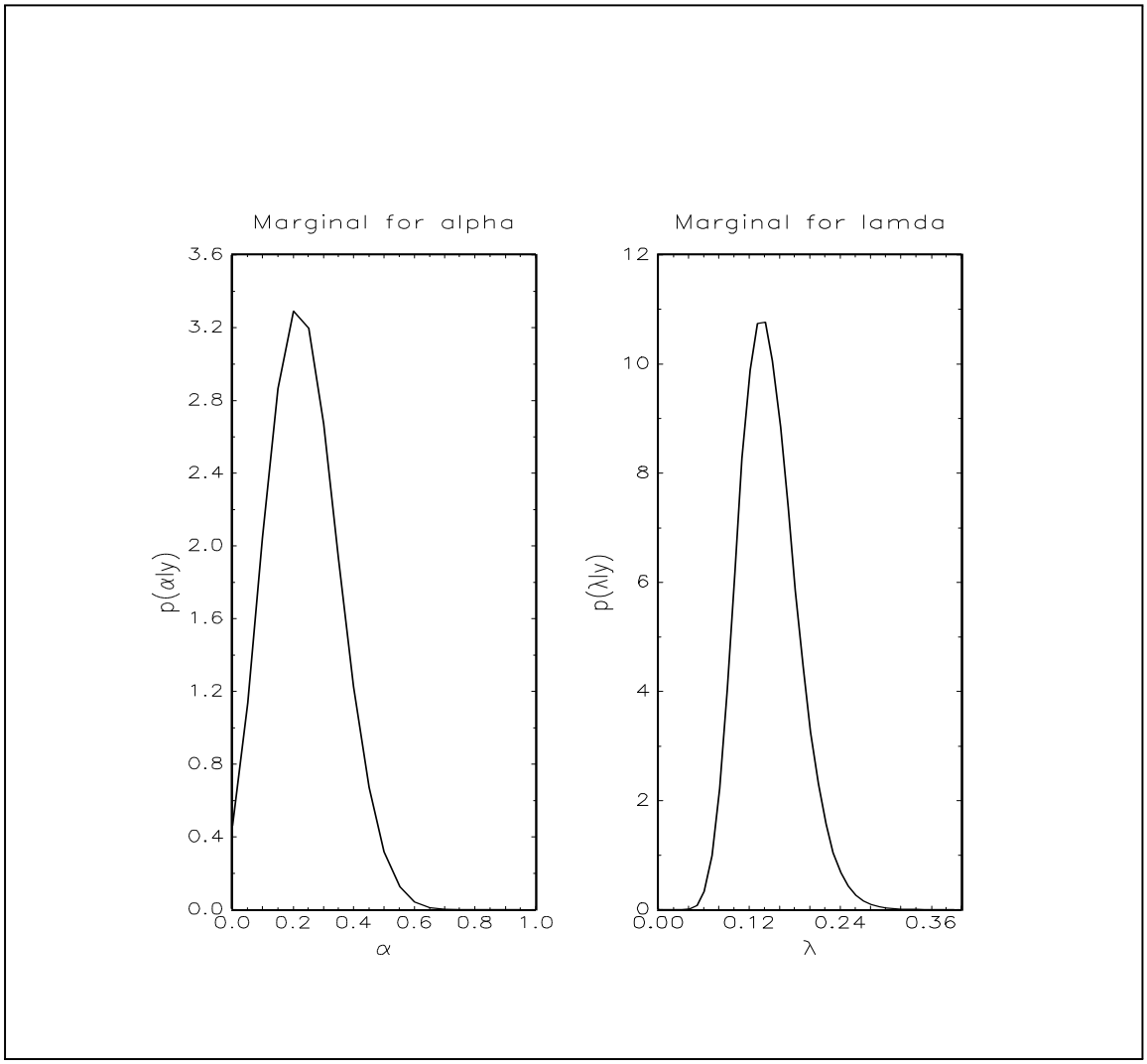


Figure 1: BURNS Data: Poisson Arrivals

The first thing to note regarding the results in Tables 2, 3 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. The modal estimate

[8]For the densities which are unimodal and not truncated from below, the ordinates of all points within the HPD interval exceed the ordinates of all points in the support outside of the interval. For densities which are multimodal, we construct the 95% probability interval in such a way that the ordinates of the upper and lower bounds are equivalent, subject to the restriction that the tail probabilities add to 5%. This means that there may be ordinates within the interval which are smaller than ordinates beyond the interval. In this sense the interval is only an approximate HPD interval. For those densities which are truncated from below, the interval is chosen to ensure that the upper and lower bounds are as close as possible to being equal, subject to the tail probability restriction.

Table 2: Parameter Estimates for the PAR model: BURNS data

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)
$P(M_k\|\mathsf{y})$		0.322

Table 3: Parameter Estimates for the BAR model: BURNS data

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[b]	Mode	0.093
Arrivals variance[b]	Mode	0.083
$P(M_k\|\mathsf{y})$		0.432

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

(b) See Footnote 8 in the text.

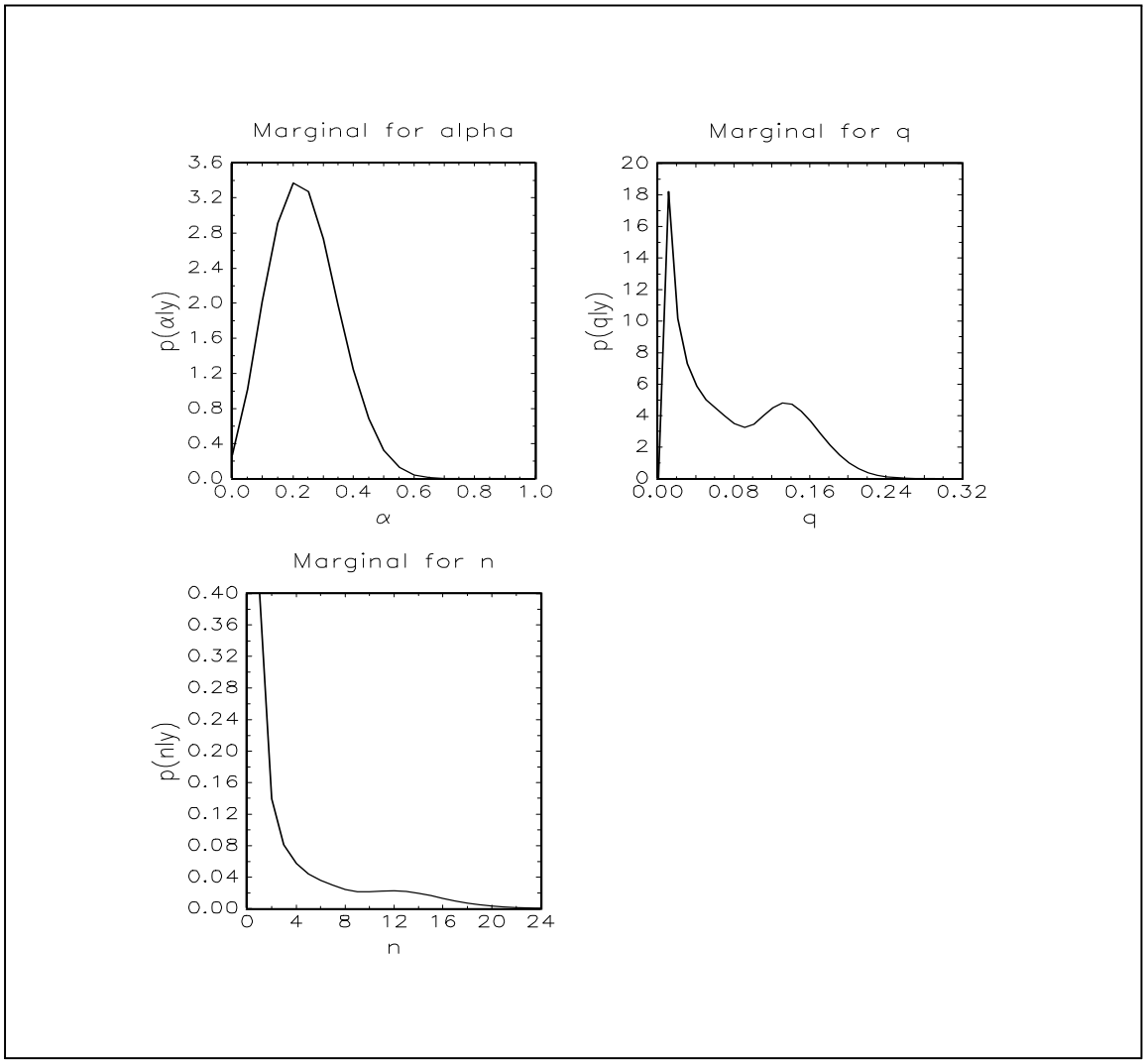


Figure 2: BURNS Data: Binomial Arrivals

of the Poisson mean (and variance), λ, is equal to 0.141, with an HPD interval of (0.081,0.221). The binomial parameters imply a point estimate for the mean of the arrivals process of 0.093 and a variance which is slightly smaller, at 0.083[9]. As 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

[9]The estimates of the mean and variance of both the binomial and negative binomial arrival 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 the 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.

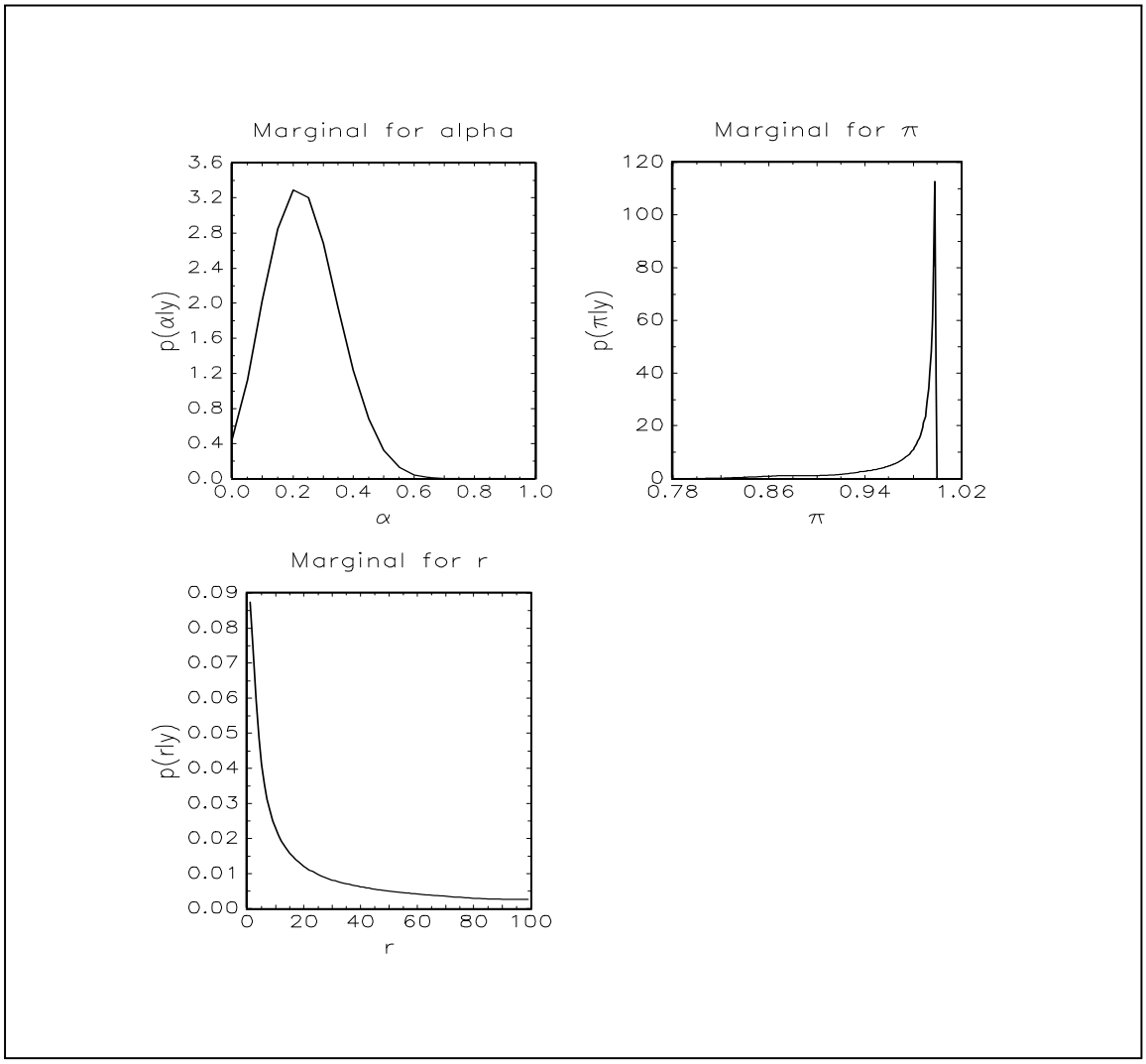


Figure 3: BURNS Data: Negative Binomial Arrivals

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 rth success, for any value of $r > 0$.

The posterior model probabilities favour the BAR model, a result which 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.

Table 5 reports the predictive results associated with each model, as well as the model-averaged predictive results. The predictive output for each model is an estimated probability associated

Table 4: Parameter Estimates for the NBAR model: BURNS data

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[b]	Mode	0.141
Arrivals variance[b]	Mode	0.151
$P(M_k \mid \mathsf{y})$		0.246

(a) The noninteger mean of the marginal density of r is reported for descriptive purposes only.
(b) See Footnote 8 in the text.

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 BURNS 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.

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.

4.4 Cuts Data Results

Once again we estimate the models using the first 118 of the 120 observations of the dataset, with the one-step and two-step ahead predictions compared with the actual values in periods 119 and 120. Tables 6 to 8 summarize the marginal posterior parameter estimates for each model. The last line in each table reports the posterior probability for the model. Figures 4, 5 and 6 graph the densities for each of the three models respectively.

As is the case with the BURNS data, the estimate of the binomial thinning parameter, α, is only very slightly affected by the specification of the arrivals process, with all point estimates of α ranging between approximately 0.44 and 0.46. All of the point and interval estimates of α indicate that the CUTS data is more persistent than the BURNS data, although still clearly stationary.[10] The (modal) estimate of the mean (and variance) of the Poisson arrivals process is 3.400. The BAR model, on the other hand imposes underdispersion in the arrivals process, with an estimated mean of 3.400 and variance of 3.210. The modal estimate of n is equal to 66, and the modal estimate of q equal to 0.051. The spread of the n marginal is quite large, with the HPD interval for n being

[10] The *kth* order autocorrelation of the INAR(1) model is proportional to α^k, for $\alpha < 1$; see Grunwald, Hyndman, Tedesco and Tweedie (2000).

Table 5: Predictive Mass Functions estimated for the BURNS data

Panel A

	Model-Specific Predictive Probabilities			Model-Averaged Predictive Probabilities
	PAR	BAR	NBAR	
Count Value				
0	0.662	0.654	0.663	0.659
1	0.300	0.310	0.298	0.303
2	0.036	0.035	0.036	0.036
3	0.002	0.001	0.003	0.002
Actual Value = 1				

Panel B

	Model-Specific Predictive Probabilities			Model-Averaged Predictive Probabilities
	PAR	BAR	NBAR	
Count Value				
0	0.780	0.775	0.777	0.777
1	0.195	0.206	0.195	0.199
2	0.023	0.018	0.025	0.022
3	0.002	0.001	0.002	0.002
Actual Value = 0				

Table 6: Parameter Estimates for the PAR model: CUTS data

Parameter		
α	Mode	0.450
	Mean	0.442
	95% HPD	(0.325, 0.550)
λ	Mode	3.400
	Mean	3.409
	95% HPD	(2.800, 4.100)
$P(M_k\|\mathsf{y})$		0.0013

Table 7: Parameter Estimates for the BAR model: CUTS data

Parameter		
α	Mode	0.450
	Mean	0.439
	95% HPD	(0.325, 0.550)
n	Mode	66.000
	Mean	64.355
	95% HPD	(32.000, 84.000)
q	Mode	0.051
	Mean	0.056
	95% HPD	(0.010, 0.120)
Arrivals Mean[(a)]	Mode	3.400
Arrivals Variance[(a)]	Mode	3.210
$P(M_k\|\mathsf{y})$		0.0002

(a) See footnote 8 in the text.

Table 8: Parameter Estimates for the NBAR model: CUTS data

Parameter		
α	Mode	0.500
	Mean	0.489
	95% HPD	(0.350, 0.600)
r	Mode	3.000
	Mean	4.261
	95% HPD	(1.000, 12.000)
π	Mode	0.510
	Mean	0.548
	95% HPD	(0.360, 0.760)
Arrivals mean[(a)]	Mode	2.201
Arrivals Variance[(a)]	Mode	3.900
$P(M_k\|\mathsf{y})$		0.9985

(a) See Footnote 10 in the text.

(32.000, 84.000). The NBAR model allows for overdispersion in the arrivals process, with modal point estimates of the mean and variance of the arrivals process being 2.201 and 3.900 respectively.[11] The modal estimate of the success probability, π, is equal to 0.51, with an HPD interval of (0.220, 0.380). The distribution of r is postitively skewed, with a modal and mean values of 3.000 and 4.261 respectively. As indicated by the densities graphed in Figures 4, 5 and 6, skewness is less of a feature of the densities based on the CUTS data, apart from the density for r in the NBAR model. A notable feature of the densities plotted in these figures is the clear bimodality of the marginal for the binomial parameter n.

In contrast with the BURNS data, in which there is a fair degree of spread in the posterior probabilities across the three alternative models, the posterior model probabilities for the CUTS data clearly favour the overdispersed NBAR model, with negligible probability assigned to the PAR and BAR alternatives.

Tables 9 and 10 report respectively the one-step and two-step ahead predictions associated with the alterative models. For comparison, we also include the frequency distribution of the sample data. As is clear, the data is less than or equal to 14 in all but one case, in which a value of 21 occurs. The last observation in the sample is 2. We have produced predictions for integers ranging

[11] The mean estimates of the arrivals mean and variance allow indicate more overdispersion, with respective values of 3.779 and 7.814.

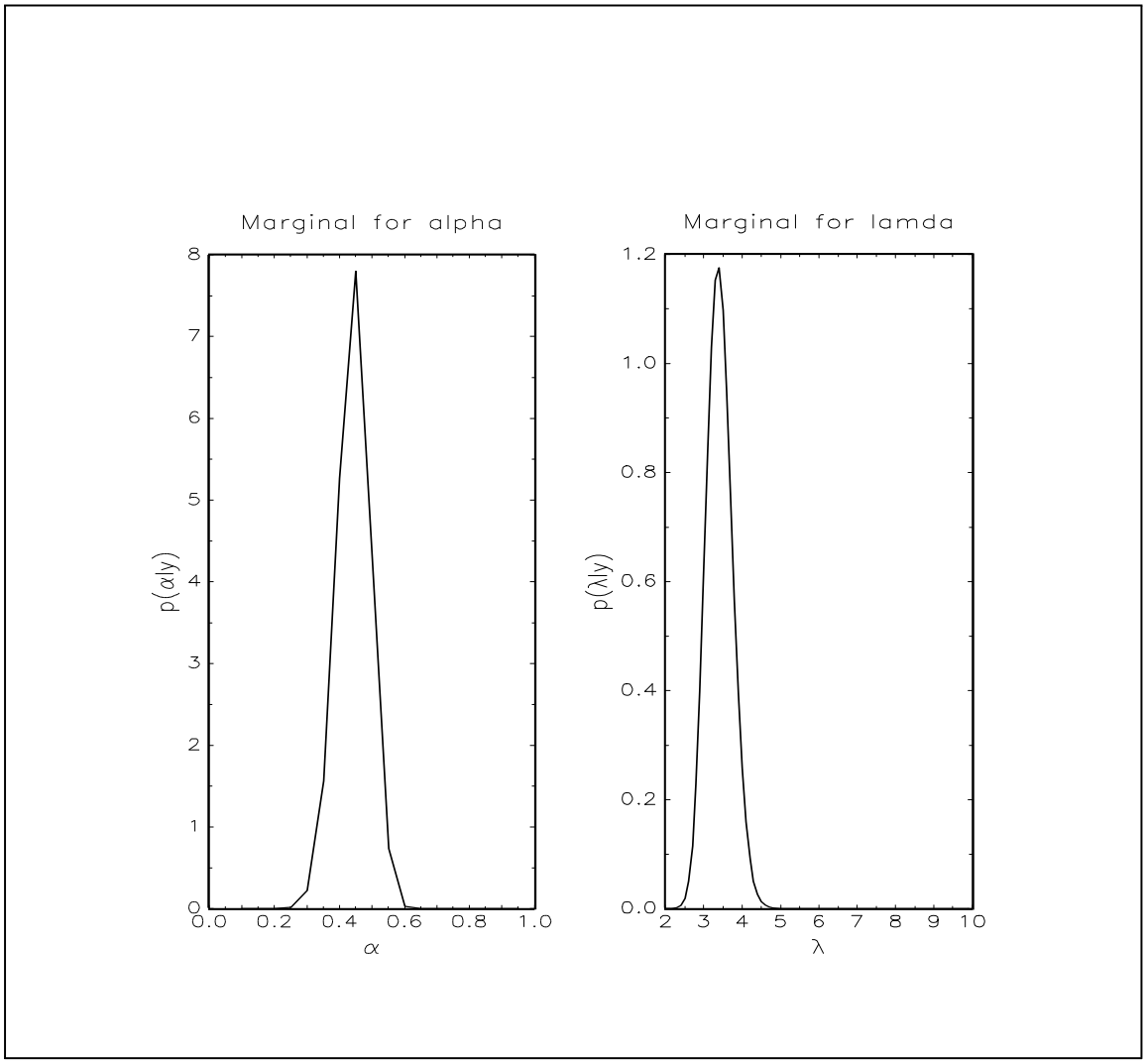


Figure 4: CUTS Data: Poisson Arrivals

from 0 to 22 inclusive.

Since the posterior densities for α in all models assign probabilities to values which are associated with reasonable correlation, the predictions for all models are quite influenced by the low observation at the end of the sample. With the arrivals mean for all models also estimated as a low value, all of the models assign low predictive probability to values exceeding 6. As such, none are particularly accurate in predicting the value of 9 which occurs in period 119. Of the three models, the NBAR is most accurate, since its higher level of dispersion is associated with a higher degree of positive skewness in the predictive pmf than is the case with the other two models. All models are more accurate in period 120, assigning more than 10% predictive probability to the actual value of 5 observed in that period. In this case, the overdispersion of the NBAR model causes it to overpredict values well in excess of the actual value and, hence, to underpredict the true value. Given the high

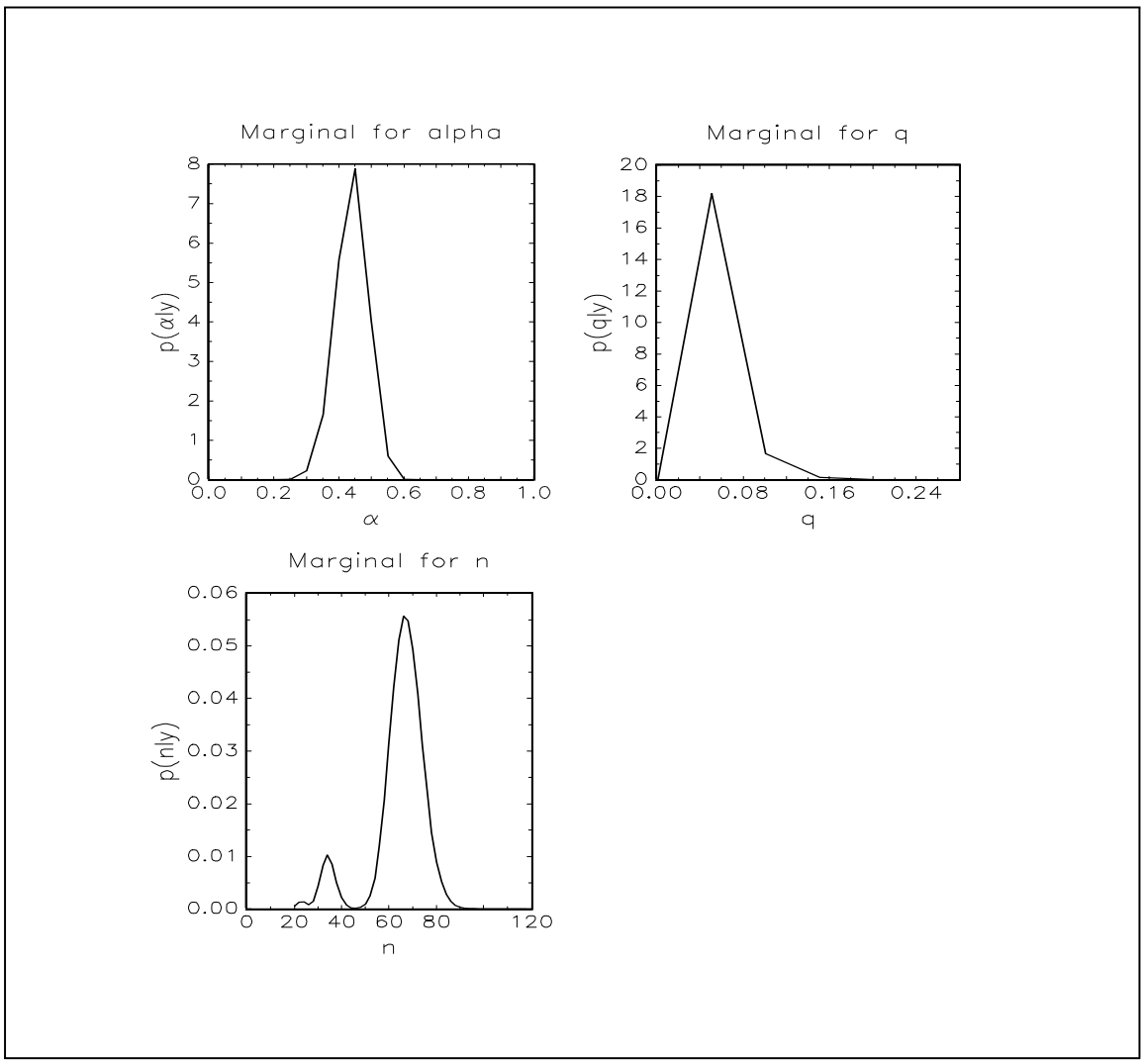


Figure 5: CUTS Data: Binomial Arrivals

posterior weight assigned to the NBAR model, the model-averaged predictive pmf's graphed in Figure 7 reflect almost exclusively the mass functions associated with this particular model.

5 CONCLUSIONS

The paper has presented a new 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.

Table 9: Predictive Mass Functions estimated for the CUTS data

One-Step Ahead Predictions

	Sample Proportions	Model-Specific Predictive Probabilities		
		PAR	BAR	NBAR
Count Value				
0	0.000	0.011	0.009	0.028
1	0.034	0.052	0.048	0.100
2	0.085	0.123	0.118	0.164
3	0.127	0.185	0.184	0.179
4	0.136	0.202	0.206	0.159
5	0.127	0.172	0.178	0.124
6	0.110	0.120	0.124	0.089
7	0.076	0.071	0.072	0.060
8	0.076	0.037	0.036	0.038
9	0.068	0.017	0.016	0.024
10	0.025	0.007	0.006	0.014
11	0.059	0.002	0.002	0.009
12	0.042	0.001	0.001	0.005
13	0.008	0.000	0.000	0.003
14	0.017	0.000	0.000	0.002
15	0.000	0.000	0.000	0.001
16	0.000	0.000	0.000	0.001
17	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000
21	0.008	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000

Actual Value = 9

Table 10: Predictive Mass Functions estimated for the CUTS data

Two-Step Ahead Predictions

	Sample Proportions	Model-Specific Predictive Probabilities		
		PAR	BAR	NBAR
Count Value				
0	0.000	0.005	0.004	0.015
1	0.034	0.027	0.024	0.052
2	0.085	0.070	0.066	0.098
3	0.127	0.124	0.121	0.133
4	0.136	0.164	0.164	0.146
5	0.127	0.174	0.177	0.139
6	0.110	0.153	0.158	0.119
7	0.076	0.116	0.119	0.093
8	0.076	0.077	0.078	0.069
9	0.068	0.045	0.045	0.048
10	0.025	0.024	0.024	0.032
11	0.059	0.012	0.011	0.021
12	0.042	0.005	0.005	0.013
13	0.008	0.002	0.002	0.008
14	0.017	0.001	0.001	0.005
15	0.000	0.000	0.000	0.003
16	0.000	0.000	0.000	0.002
17	0.000	0.000	0.000	0.001
18	0.000	0.000	0.000	0.001
19	0.000	0.000	0.000	0.001
20	0.000	0.000	0.000	0.001
21	0.008	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000

Actual Value = 5

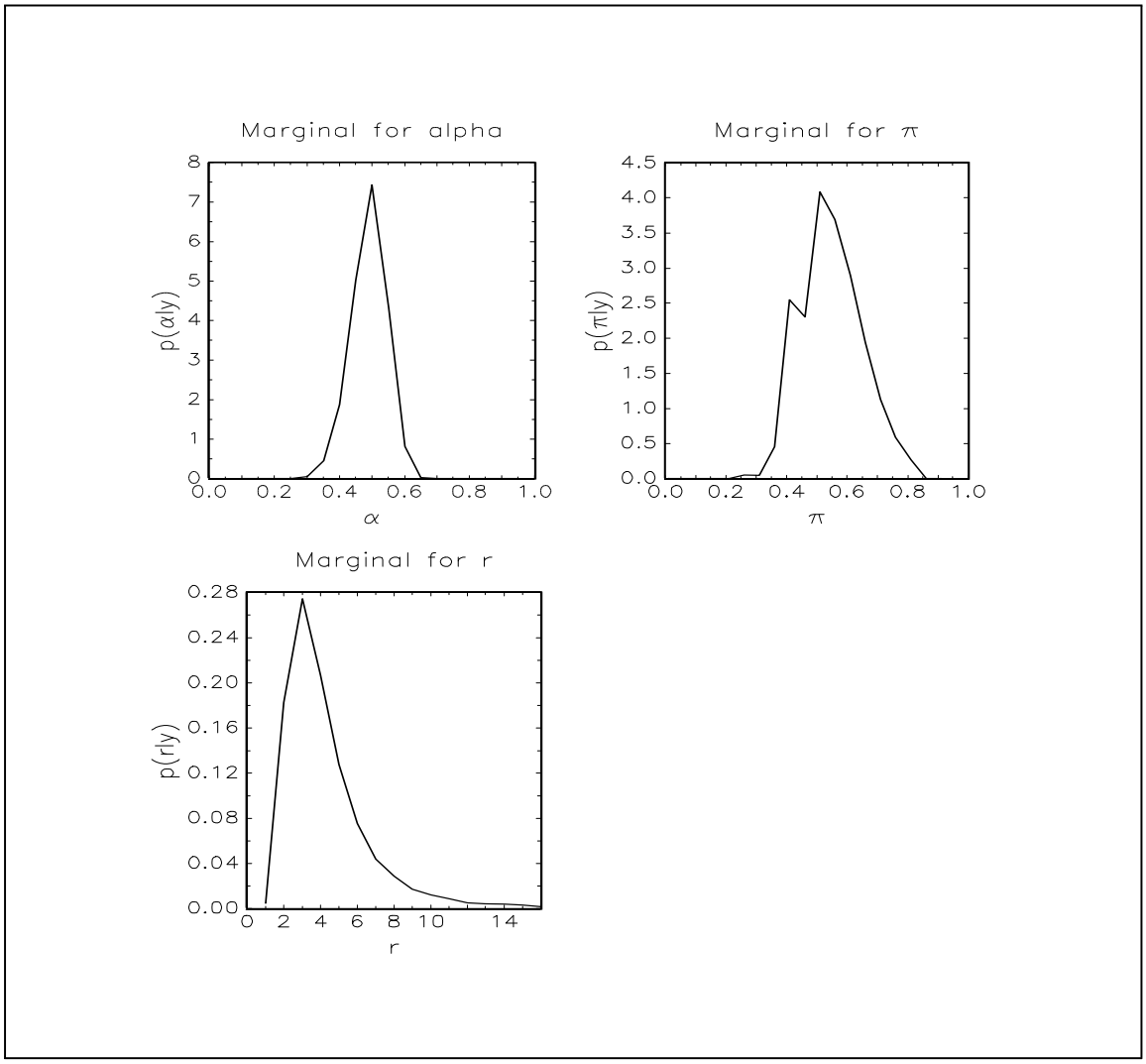


Figure 6: CUTS Data: Negative Binomial Arrivals

The approach demonstrated in the paper has exploited new 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 two sets 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

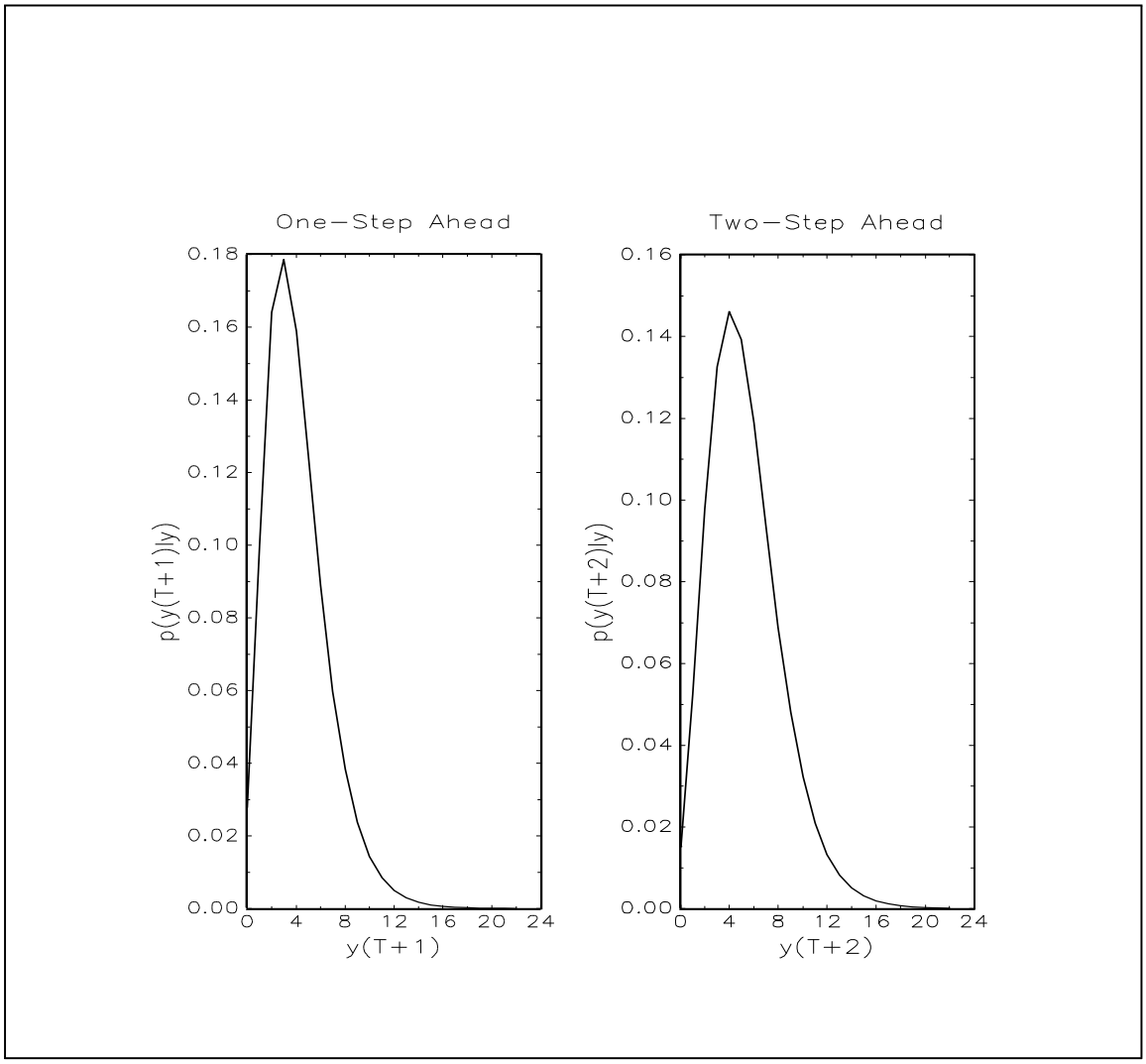


Figure 7: CUTS Data: One-Step and Two-Step Ahead Predictives

the presence of short-run dependence in both sets of data. The posterior probabilities associated with alternative arrivals processes overwhelmingly favour the negative binomial in the case of the overdispersed CUTS data, whilst only slightly favouring the binomial in the case of the underdispersed BURNS data. 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.

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. and 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. and Mortera, J. (1999), "Default Bayes Factors for Nonnested Hypothesis Testing", Journal of the American Statistical Society, 94, 542-554.

[4] Chib, S. (1995), "Marginal Output from the Gibbs Output", Journal of the American Statistical Association, 90, 1313-1321.

[5] Chib, S., Greenburg, E. and Winkelmann, R. (1998), "Posterior Simulation and Bayes Factors in Panel Count Data Models", Journal of Econometrics, 86, 33-54.

[6] Chib, S. and Winkelmann, R. (2001), "Markov Chain Monte Carlo Analysis of Correlated Count Data", Journal of Business and Economic Statistics, 19, 428-435.

[7] Durbin, J. and 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.

[8] Cameron, A.C. and Trevedi, P.K. (1998), Regression Analysis of Count Data. Cambridge University Press.

[9] Fahrmeir, L. and Tutz, G. (1994), Multivariate Statistical Modeling Based on Generalized Linear Models, Springer-Verlag, New York.

[10] 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.

[11] Freeland, R.K. & McCabe, B.P.M. (2002a), "Forecasting Discrete Valued Low Count Time Series", forthcoming International Journal of Forecasting.

[12] Freeland, R.K. & McCabe, B.P.M. (2002b), "Analysis of Low Count Time Series Data by Poisson Autoregression", mimeo, University of Liverpool.

[13] Grunwald, G.K., Hyndman, R.J., Tedesco, L. and Tweedie, R.L. (2000), "Non-Gaussian Conditional Linear AR(1) Models", Australian and New Zealand Journal of Statistics, 42, 478-495.

[14] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999), "Bayesian Model Averaging: A Tutorial", Statistical Science, 14, 382-417.

[15] Jung, R.C. and Tremayne, A.R., (2003), "Testing for Serial Dependence in Time Series Models of Counts", Journal of Time Series Analysis, 24, 65-84.

[16] Kleibergen, F. (2003), "Invariant Bayesian Inference in Regression Models that is Robust Against the Jeffreys-Lindley's Paradox", mimeo, University of Amsterdam.

[17] McCabe, B.P.M., Martin, G.M. and Freeland, R.K. (2003), "Specification Analysis for Models of Non-Gaussian Data", mimeo, University of Liverpool.

[18] McDonald, I.L. & Zucchini, W. (1997), Hidden Markov and Other Models for Discrete-Valued Time Series, Chapman and Hall, London.

[19] McKenzie, E. (1988), "Some ARMA Models for Dependent Sequences of Poison Counts", Advances in Applied Probability, 20, 822-835.

[20] O'Hagan, (1995), "Fractional Bayes Factors for Model Comparison", Journal of the Royal Statistical Society, B, 57, 99-138.

[21] West, M., and Harrison, J. (1997), Bayesian Forecasting and Dynamic Models, Second Edition, Springer-Verlag, New York.