

NCER Working Paper Series

Discrete time-series models when counts are unobservable

T M Christensen, A S Hurn, K A Liindsay

Working Paper #35

September 2008

Discrete time-series models when counts are unobservable

T M Christensen, A S Hurn

School of Economics and Finance, Queensland University of Technology

K A Lindsay

Department of Mathematics, University of Glasgow.

Abstract

Count data in economics have traditionally been modeled by means of integer-valued autoregressive models. Consequently, the estimation of the parameters of these models and their asymptotic properties have been well documented in the literature. The models comprise a description of the survival of counts generally in terms of a binomial thinning process and an independent arrivals process usually specified in terms of a Poisson distribution. This paper extends the existing class of models to encompass situations in which counts are latent and all that is observed is the presence or absence of counts. This is a potentially important modification as many interesting economic phenomena may have a natural interpretation as a series of ‘events’ that are driven by an underlying count process which is unobserved. Arrivals of the latent counts are modeled either in terms of the Poisson distribution, where multiple counts may arrive in the sampling interval, or in terms of the Bernoulli distribution, where only one new arrival is allowed in the same sampling interval. The models with latent counts are then applied in two practical illustrations, namely, modeling volatility in financial markets as a function of unobservable ‘news’ and abnormal price spikes in electricity markets being driven by latent ‘stress’.

Keywords

Integer-valued autoregression, Poisson distribution, Bernoulli distribution, latent factors, maximum likelihood estimation.

JEL classification numbers

C13, C25, C32.

Corresponding author

Stan Hurn

School of Economics and Finance

Queensland University of Technology

Brisbane, Qld 4001

Australia

email s.hurn@qut.edu.au

1 Introduction

The integer-valued autoregressive framework to model low-count integer-valued time series was introduced by Al-Osh and Alzaid (1987) and Mackenzie (1988). Essentially these models comprise a model of the survival of counts from the previous period specified in terms of a binomial thinning operation and a model of the arrival of new counts that is independent of the thinning process. An important class of integer-valued time-series model that has received renewed attention in the recent literature is when the arrivals are specified in terms of a Poisson distribution (Freeland and McCabe, 2004a; 2004b; McCabe and Martin, 2005).

The Poisson autoregressive model requires that the values of the counts are observable. In many instances in economics, however, the values of the fundamental underlying driving process of the system are unobservable. All that can be observed is whether ‘nothing has happened’ or ‘something has happened’. For example, the arrival and persistence of significant items of news is usually taken to be the process driving the volatility in the returns to financial assets. The ‘news’ however is generally regarded as being difficult to quantify exactly and volatility is therefore traditionally modeled in a (G)ARCH type framework (Engle, 1982; Bollerslev 1986) in which ‘news’ arrival is inferred from the value of the residual in the model of the conditional mean of the returns to financial assets. Another example concerns electricity markets, in which extreme price events are thought to occur if system stress is present (see, for example, Geman and Roncorni, 2006). From a time series of electricity prices it is impossible to ascertain the number of stresses acting concurrently in the market, but it is possible to infer whether or not at least one stress is acting at each instant in time depending upon whether or not an extreme price event is observed at that time.

The fundamental contribution of this paper is to develop a method for the maximum likelihood estimation of the parameters of count models when the actual counts themselves are unobservable but what is observed is a binary time series indicating whether or not counts are present. It is shown that likelihood is naturally associated with runs of 1s and 0s rather than with individual transitions of the censored process. The analysis includes discussion of the models with Poisson distributed arrivals and Bernoulli distributed arrivals in conjunction with a generic binomial thinning process to model the survival of counts. The results of a simple Monte Carlo simulation exercise illustrate that the maximum likelihood estimators of the models with latent counts have the expected asymptotic properties.

A number of other interesting ancillary results are also developed in the paper. It is shown that the Poisson and Bernoulli autoregressive models are equivalent for low intensities of arrivals. Furthermore, recurrence relations for the computation of the matrix of transitional probabilities are established for both classes of model. These recurrence relations apply equally to models in which counts are observed and provide an efficient mechanism for computing the likelihood function. Finally, two innovative applications of latent count models are provided, one concerning the arrival of news and the volatility of financial asset returns and another modeling system stress in the electricity market.

The rest of the paper is structured as follows. Section 2 sets out the basic integer-valued count

model. Arrivals are specified both in terms of the Poisson distribution and the Bernoulli distribution and the relationship between these two models is established. The asymptotic equivalence of the two models is proved for low arrival processes of low intensity. In Section 3 a set of recurrence relations that govern the transitional probabilities of the counts are derived. These relationships allow the slick computation of the matrix of transitional probabilities. The extension of the integer-valued time-series model to deal with latent counts and maximum likelihood estimation of its parameters is the subject matter of Section 4. A Monte Carlo simulation is presented in Section 5 to illustrate the properties of the maximum likelihood estimators of the latent count models, followed by two practical illustrations of these models in action. Section 6 is a brief conclusion.

2 Integer count models

Let X_0, X_1, \dots be a time series of non-negative integers which are to be interpreted as a series of counts evolving in time through the addition of new arrivals and through the survival of existing counts. The simplest possible time-series model for X_t is the count autoregressive model of order one, specified by the equation

$$X_t = \alpha \circ X_{t-1} + e_t, \quad (1)$$

where $\alpha \circ X_{t-1}$ represents the survival of counts from the previous period and e_t represents the arrival of new counts.

The survival process is generic and is described by a binomial thinning process in which each count at time $(t-1)$ survives to time t with known probability $\alpha \in [0, 1]$, independent of all other counts. Thus if X is the number of counts at one observation then

$$\alpha \circ X = \sum_{n=1}^X B_n \quad (2)$$

denotes the number of counts surviving to the next observation, where B_1, B_2, \dots is a sequence of independent and identically distributed Bernoulli random variables satisfying the property

$$\Pr(B_n = 1) = \alpha, \quad \Pr(B_n = 0) = 1 - \alpha. \quad (3)$$

The operator “ \circ ” is called the *binomial thinning operator*. The binomial thinning operator renders the integer count model nonlinear, but many of the properties of the process are analogous to those of a standard linear autoregressive model of order 1. For example, if $\alpha < 1$ then process (1) has autocorrelation function $\text{Corr}(X_t, X_{t-k}) = \alpha^k$. Thus α acts as a measure of the *persistence* or *memory* exhibited by the time series, with higher values of α corresponding to stronger degrees of persistence. Of course, there is no need to restrict the order of the model to the first order, but only first-order processes are considered in the paper.¹

While the survival process is generic, the integer count model is completely characterized by the assumptions made about the arrival process. There has been recent interest in the literature on

¹A more general model (see, Du and Li, 1991) is the count model of order p

$$X_t = \alpha_1 \circ X_{t-1} + \dots + \alpha_p \circ X_{t-p} + e_t.$$

The stationarity of this model requires that $\alpha_k \in [0, 1)$ for all k and that $\sum_{k=1}^p \alpha_k < 1$.

the model which specifies the arrival process, e_t , to be a Poisson process with parameter λ . If the arrivals process is specified in terms of the Poisson distribution, then the model in equation (1) is called a Poisson autoregressive model of order 1 or PAR(1) model. The key feature of a Poisson model for e_t , the arrival process, is that it allows more than one arrival to occur in any fixed interval of time. The underlying concept driving a Poisson arrival process is that the arrivals are purely random and unstructured but that they occur at a fixed rate. An alternative specification of e_t is that most one arrival to occur in the interval between observations, that is, $\Pr(e_t = 1) = \lambda$ and $\Pr(e_t = 0) = 1 - \lambda$. In this case, the arrival process is Bernoulli distributed and the model in equation (1) is called a Bernoulli autoregressive model of order 1 or BAR(1) model.

There is a subtle difference in interpretation of the parameter λ in the PAR and BAR models. In the PAR model λ is the intensity of arrivals, but in the BAR model λ represents the probability of one arrival between successive observations. The values of λ for both specifications can be connected in an approximate sense by equating the probability of no arrival in each model to obtain $1 - \lambda_{\text{BAR}} = e^{-\lambda_{\text{PAR}}}$, or equivalently, $\lambda_{\text{PAR}} = -\log(1 - \lambda_{\text{BAR}})$.

Of course, there is no need for the survival rate α and arrival rate λ in these count models to be constant, but instead they can be extended to depend on vectors of time-varying covariates x_t and y_t . In the context of a PAR model, Freeland and McCabe (2004a) suggest the use of the representations

$$\alpha_t = \frac{1}{1 + \exp(-y_t' \gamma)}, \quad \lambda_t = \exp(x_t' \delta), \quad (4)$$

for α_t and λ_t , the time-varying specifications of α and λ . The advantage of these forms is that they intrinsically respect the essential constraints $\alpha_t \in (0, 1)$ and $\lambda_t \in (0, \infty)$ irrespective of the values taken by the covariates y_t and z_t . In the BAR model, the corresponding forms would be

$$\alpha_t = \frac{1}{1 + \exp(-y_t' \gamma)}, \quad \lambda_t = \frac{1}{1 + \exp(-x_t' \delta)}, \quad (5)$$

which satisfies the constraint $\lambda_t \in (0, 1)$ irrespective of the value of x_t . Testing whether the rates of the processes are time-varying is simply a matter of testing restricted forms of γ and δ . Without loss of generality, the theoretical results that follow are confined to the case of constant survival and arrival probabilities.

The PAR model has attracted a fair amount of recent interest in the literature (see for example, Al-Osh and Alzaid, 1987; Mackenzie, 1988; Freeland and McCabe, 2004a; 2004b; McCabe and Martin, 2005). By contrast, the BAR model has received little attention. This is not a desirable situation, given that the two models are closely related. Because the rate of arrival is specified exogenously in each of the models, then for suitably short sampling intervals the arrivals process is well approximated by a Bernoulli process even if the true arrivals distribution is Poisson. The theoretical results that follow demonstrate, among other things, that the BAR(1) model is asymptotically identical to the PAR(1) model for small values of λ .

The analysis is based on the concept of a probability generating function. Recall that if X is a random variable taking non-negative integer values such that $\Pr(X = k) = p_k$ where

$p_0 + p_1 + \dots = 1$, then the probability generating function of X is the function defined by

$$G_X(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Probability generating functions for non-negative integer-valued random variables satisfy the following two important properties.

PROPERTY 1 (Sum of independent integer-valued random variables)

If X and Y are independent non-negative integer-valued random variables with respective generating functions $G_X(z)$ and $G_Y(z)$, then the random variable $X + Y$ has probability generating function $G_X(z)G_Y(z)$. ■

PROPERTY 2 (Random sum of independent integer-valued random variables)

If Y is a non-negative integer-valued random variable with probability generating function $G_Y(z)$, and X_1, X_2, \dots are independent and identically distributed non-negative integer-valued random variables independent of Y and with probability generating function $G_X(z)$, then $\sum_{k=0}^Y X_k$ has probability generating function $G_Y(G_X(z))$. ■

These two properties of probability generating functions are now used to establish that at each instant, the process X_t in a PAR(1) model has a Poisson distribution at all finite times, provided the starting value X_0 has a Poisson distribution.

THEOREM 1: *Let X_t be the process defined by the PAR(1) model*

$$X_t = \alpha \circ X_{t-1} + e_t \tag{6}$$

in which $\alpha \circ X_{t-1}$ is a binomial thinning process with parameter $\alpha \in [0, 1]$, e_t is an independent Poisson arrivals process with parameter $\lambda \in (0, 1)$. If X_0 is a Poisson distributed random variable with parameter λ_0 , then X_k is Poisson distributed with parameter

$$\alpha^k \lambda_0 + \lambda \sum_{n=0}^{k-1} \alpha^n. \quad \blacksquare$$

The existence of the stationary distribution now only requires the convergence of the geometric sum $\sum_{n=0}^{k-1} \alpha^n$, which is the case when $\alpha < 1$. The result is stated in the following Corollary.

COROLLARY 1.1: *If the binomial thinning parameter $\alpha < 1$ then the process defined by equation (6) has a stationary distribution which is a Poisson distribution with parameter $\lambda/(1 - \alpha)$. ■*

The analysis of Theorem 1 can now be repeated for the BAR(1) model to demonstrate that this process also has a stationary distribution and that this distribution is asymptotically Poisson for small values of λ . The details, however, are no longer straightforward because the sample space of the BAR process changes from iteration to iteration so that the BAR process does not enjoy the self-similarity property of the PAR process. The primary property of the BAR process is established in Theorem 2.

THEOREM 2: *Let X_t be the process defined by the integer-valued autoregressive process*

$$X_t = \alpha \circ X_{t-1} + e_t \tag{7}$$

in which $\alpha \circ X_{t-1}$ is a binomial thinning process with parameter $\alpha \in [0, 1]$, e_t is an independent Bernoulli arrivals process with parameter $\lambda \in (0, 1)$ and X_0 , the first term in the process (7), is a random draw from a Bernoulli process with parameter p_0 , then the probability generation function of process X_k is

$$G_k(z) = (1 + p_0\alpha^k(z-1)) \prod_{n=0}^{k-1} (1 + \lambda\alpha^n(z-1)), \quad k > 0. \quad \blacksquare$$

The result is for k iterations of the BAR(1) model where k is finite. Unlike the case in Theorem 1 for the PAR model, it is not immediately obvious that the limit of the product on the right hand side of equation (7) exists as $k \rightarrow \infty$. The existence of this limit and therefore the existence of a stationary distribution for the BAR process together with its first and second moments is now established.

COROLLARY 2.1: *If the process X_t has thinning process with parameter $\alpha < 1$ then the process has a stationary state with probability generating function*

$$G(z) = \prod_{n=0}^{\infty} [1 + \lambda\alpha^n(z-1)],$$

with mean value $\lambda/(1-\alpha)$ and variance $\lambda(1+\alpha-\lambda)/(1-\alpha^2)$. ■

The main result of this section now follows from Corollary 2.1.

COROLLARY 2.2: *The stationary distribution of the BAR(1) process, X_t , is asymptotically a Poisson distribution with parameter $\lambda/(1-\alpha)$ as $\lambda \rightarrow 0^+$.* ■

Provided that there is genuine thinning in the integer count process, that is $\alpha < 1$, then both the PAR and BAR model have stationary distributions. In particular, if the arrival rate in the PAR model is low, then the PAR model and the BAR model are asymptotically identical.

The next section develops a series of recurrence relations governing the behavior of the transitional probability distributions of the count processes for the PAR and BAR models. These relationships provide the basis of an efficient recursive procedure for computing the transitional probabilities that are required in the construction of the log-likelihood or score functions.

3 Recurrence Relations

Given a sample X_0, X_1, \dots, X_N of counts, maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}$ of the values of the parameters α and λ are obtained by maximizing the likelihood function

$$\mathcal{L}(\alpha, \lambda) = \prod_{k=1}^N F_{X_k, X_{k-1}},$$

with respect to α and λ , or equivalently, are obtained by minimizing the negative log-likelihood function

$$-\log \mathcal{L}(\alpha, \lambda) = -\sum_{k=1}^N \log (F_{X_k, X_{k-1}}), \quad (8)$$

where $F_{X_t, X_{t-1}}$ is simply the model value of the transitional probability of the observed transition from the state X_{k-1} to the state X_k . Of course, these transitional probabilities will be different for the PAR and the BAR models.

Initially it would seem that maximum likelihood estimation unavoidably requires the computation of a separate sum for each pairing (X_k, X_{k-1}) arising in the expression (8), a procedure which is potentially both numerically intensive and arithmetically error-prone. In what follows, an efficient recursive procedure is developed for calculating the matrix transitional probabilities.

3.1 PAR recurrence relation

In practice it is the value of X_t that is observed in a PAR process rather the component arrival and survival processes. Each observation of X_t can arise from X_{t-1} in a number of mutually exclusive ways leading in turn to a transitional probability $\Pr(X_t = p | X_{t-1} = q)$ which is formed by constructing the weighted sum of probabilities over all the possible routes consistent with the observation $X_t = p$ and $X_{t-1} = q$. The route in which precisely n counts of X_{t-1} survive and are supplemented by $p - n$ arrivals occurs with probability

$$\left[\binom{q}{n} \alpha^n (1 - \alpha)^{q-n} \right] \times \left[\frac{e^{-\lambda} \lambda^{p-n}}{(p-n)!} \right].$$

Consequently, the probability $\Pr(X_t = p | X_{t-1} = q)$ is given by

$$P_{p,q} = e^{-\lambda} \sum_{n=0}^{\min(p,q)} \binom{q}{n} \frac{\alpha^n (1 - \alpha)^{q-n} \lambda^{p-n}}{(p-n)!}. \quad (9)$$

The recurrence relations satisfied by these transitional probabilities are now given in Theorem 3 and proved in the Appendix.

THEOREM 3: *Given parameters $\alpha \in [0, 1)$ and $\lambda \in (0, \infty)$ and non-negative integers p and q , then the function*

$$P_{p,q} = e^{-\lambda} \sum_{n=0}^{\min(p,q)} \binom{q}{n} \frac{\alpha^n (1 - \alpha)^{q-n} \lambda^{p-n}}{(p-n)!},$$

satisfies the recurrence relations

$$\begin{aligned} (i) \quad (p+1)P_{p+1,q} &= \lambda P_{p,q} + \alpha q P_{p,q-1}, & p \geq q \\ (ii) \quad P_{p,q+1} &= (1 - \alpha)P_{p,q} + \alpha P_{p-1,q} & p < q, \\ (iii) \quad P_{q+1,q+1} &= \alpha P_{q,q} + (1 - \alpha)P_{q+1,q} \end{aligned} \quad (10)$$

where $P_{0,0} = e^{-\lambda}$ and it is assumed that $P_{j,k} = 0$ if either j or k is a negative integer. ■

Recall that the value of $P_{p,q}$ is the probability of a transition from the state $X_{t-1} = q$ to the state $X_t = p$. The matrix $F = [P_{p,q}]$ is therefore the matrix of transitional probabilities for the PAR(1) model and gives a complete description of how X_t evolves between observations. The value of Theorem 3 is that it enables F to be constructed in a systematic way with the minimum of numerical effort (and maximum accuracy) because each entry of F is calculated without the need to compute a summation.

The algorithm for the construction of F uses these recurrence relations in the following way. The process is started by initializing $P_{0,0}$ to the value $e^{-\lambda}$ and in so doing seeds the leading entry of the main diagonal of F . The following steps are then repeated in sequence until the matrix F is fully populated.

- (1) Condition (i) is used to fill all the entries of column q of P *below* the main diagonal in increasing row number starting with the entry $P_{q+1,q}$. The successful implementation of this step assumes that column $(q-1)$ of P has been filled previously and that the diagonal entry of column q has been seeded.
- (2) Condition (ii) is used to fill all entries of row q of P to the *right* of the main diagonal in increasing column number starting with the entry $P_{q,q+1}$. The successful implementation of this step assumes that row $(q-1)$ of P has been filled previously and that the diagonal entry of column q has been seeded.
- (3) At this juncture all rows and columns of the matrix P up to and including row q and column q have been filled. Condition (iii) now allows entry $(q+1)$ of the main diagonal of P to be seeded from previously known information.

The matrix F constructed in this way may now be used either to evaluate the log-likelihood function directly from equation (8), or used to compute the score function and related entities from the expressions provided by Freeland and McCabe (2004a).

3.2 BAR recurrence relation

As in the case of the PAR process, useful recurrence relations can be constructed to facilitate the computation of the transition matrix for a BAR process. Let $B_{p,q} = \Pr(X_t = p | X_{t-1} = q)$ denote the probability of a transition from $X_{t-1} = q$ to $X_t = p$ in a BAR process with arrival and survival probabilities λ and α . Because at most one count can arrive in the BAR model in the sampling interval, if $X_{t-1} = q$ then X_t can be at most $(q+1)$. Therefore $B_{p,q} = 0$ when $p > q+1$ which in turn means that the transition matrix of a BAR process necessarily has upper Hessenberg form. However, a transition from q to $p = q+1$ is possible provided each count at time $(t-1)$ survives to time t (which happens with probability α^q) and simultaneously and independently of this, there is an arrival (which happens with probability λ). Consequently, $B_{q+1,q} = \alpha^q \lambda$. Finally, in a BAR process there are precisely two ways in which a transition from $X_{t-1} = q$ to $X_t = p$ can occur. Either $(q-p)$ counts fail to survive and there is no arrival in the period between observations, or alternatively, $(q+1-p)$ counts fail to survive but there is one arrival during this period. Because survival and arrival process behave independently, the possible expressions for $B_{p,q}$ are

$$B_{p,q} = \begin{cases} 0 & p > q+1, \\ \alpha^q \lambda & p = q+1, \\ \binom{q}{p} \alpha^p (1-\alpha)^{q-p} (1-\lambda) + \binom{q}{p-1} \alpha^{p-1} (1-\alpha)^{q+1-p} \lambda & p < q+1. \end{cases} \quad (11)$$

The efficient computation of the matrix B is now described in Theorem 3.

THEOREM 4: *Given parameters $\alpha \in [0, 1)$ and $\lambda \in (0, \infty)$ and non-negative integers p and q , then the function*

$$B_{p,q} = \begin{cases} 0 & p > q + 1, \\ \alpha^q \lambda & p = q + 1, \\ \binom{q}{p} \alpha^p (1 - \alpha)^{q-p} (1 - \lambda) + \binom{q}{p-1} \alpha^{p-1} (1 - \alpha)^{q+1-p} \lambda & p < q + 1. \end{cases}$$

satisfies the recurrence relations

$$\begin{aligned} (i) \quad & B_{p,q} = 0, & p > q + 1 \\ & B_{p,q} = \alpha^q \lambda, & p = q + 1 \\ (ii) \quad & B_{p,q} = (1 - \alpha)B_{p,q-1} + \alpha B_{p-1,q-1}, & p < q + 1 \end{aligned} \tag{12}$$

where $B_{0,0} = (1 - \lambda)$ and it is assumed that $B_{j,k} = 0$ whenever either $j < 0$ or $k < 0$. ■

As with Theorem 3, the focus here is on the practical application of the recurrence relations established in Theorem 4 for the efficient construction of the matrix of transitional probabilities, F , for the BAR model.

The algorithm is as follows. To initialize the process, fill the row $p = 0$ of F with the entries $B_{0,q} = (1 - \alpha)^q (1 - \lambda)$. The remaining rows of F are filled by repeating the following steps for each row starting with $p = 1$.

- (1) Conditions (i) are used to set $B_{p,q} = 0$ for rows $q < p - 1$ and to set $B_{p,p-1} = \alpha^{p-1} (1 - \lambda)$.
- (2) Condition (ii) is now used to complete row p for all values of $q \geq p$. This calculation uses values from the row above which has been filled at the previous iteration.

The recurrence relations established in Theorems 3 and 4 enable the matrix of transitional probabilities, F , to be constructed easily for both the PAR and BAR models given values for α and λ . If counts data X_0, X_1, \dots, X_n are available, then the negative log-likelihood function in equation (8) may be computed directly from the entries of this matrix and subsequently minimized by choice of parameters α and λ . It is often the case in economics and finance, however, that the counts are unobserved or not easily measurable. The next section will demonstrate how maximum likelihood estimation can be adapted to deal with this situation.

4 Estimation of a Model of Latent Counts

Suppose now that it is relatively easy to identify the presence or absence of a latent feature driving the counts, but the counts themselves are unobserved. In this situation, the series of integer counts X_t is replaced by the binary time series

$$Y_t = \begin{cases} 1 & X_t > 0, \\ 0 & X_t = 0. \end{cases}$$

In the latent counts model the fundamental drivers of the process X_t are unchanged in the respect that

$$X_t = \alpha \circ X_{t-1} + e_t, \quad (13)$$

where the survival component $\alpha \circ X_t$ modeled again as a binomial thinning process and e_t is an independent arrival component with distribution to be specified.² The analysis will again consider two specifications of the arrivals process, namely, Poisson arrivals which allow more than one arrival between observations and Bernoulli arrivals which allow at most one arrival between observations.

When counts information is observed then the probability passed to the maximum likelihood procedure for the transition from X_{t-1} to X_t is simply the model value of the transitional probability of the observed transition, namely the appropriate entry in the matrix F . By contrast, in a latent process where Y_t alone is known, what is passed to the maximum likelihood procedure is the probability that $Y_t = 0$ or that $Y_t = 1$ conditioned on the observed value of Y_{t-1} . The computation of the probability that $Y_t = 0$ or that $Y_t = 1$ is calculated by convolving the matrix of transitional probabilities for the model with the estimated state of the unobserved distribution of counts when conditioned on the observed value of Y_{t-1} .

The construction of the log-likelihood function for the latent counts models is best described within the framework of a filtering procedure. The procedure involves a prediction phase, an observation phase in which the contribution to the log-likelihood function is determined and an update phase. Integral to this procedure is F , the matrix of transitional probabilities, in this instance taken to be the $(M + 1) \times (M + 1)$ matrix,

$$F = \left[\begin{array}{c|ccc} F_{0,0} & F_{0,1} & \cdots & F_{0,M} \\ \hline F_{1,0} & F_{1,1} & \cdots & F_{1,M} \\ \vdots & \vdots & \ddots & \vdots \\ F_{M,0} & F_{M,1} & \cdots & F_{M,M} \end{array} \right]. \quad (14)$$

The matrix of transitional probabilities therefore has the set of initial states of a transition along the rows of the matrix and the final states of each transition down the columns. For example, the first row of the transition matrix gives the unconditional probabilities of the transitions from all possible values of X_{t-1} to $X_t = 0$, while the first column of the transition matrix gives the unconditional probabilities of a transition from $X_{t-1} = 0$ to all possible values of X_t . Note that the entries of F are determined solely by the model of the survival and arrival processes, whereas the choice for the value of M is determined by the model in combination with K , the length of the longest run of ones in the observations of Y . In the case of the BAR model, $M = K + 1$ is adequate since each transition can add at most one count to the existing count process making $(K + 1)$ the maximum possible number of latent counts. In the case of the PAR model there is no technical restriction on the number of possible latent states although, of course, high values of counts are associated with negligible probabilities. A subjective decision must be taken for the value of M , and in this work that decision is $M = 2K$.

It is also necessary to introduce two auxiliary vectors to be denoted S^+ and S^- . The vector S^+

²For example, if X_t denotes the number of customers in a queue, then Y_t signifies whether or not there is a queue.

holds the best available estimate of the distribution of states of the latent counts at the current instant in time, whereas S^- is the prediction of the one-step-ahead distribution of states prior to using the next observation and is derived from the convolution of F and S^+ . One iteration of the filtering cycle is now described.

Predict: Prior to transition but after the previous observation at $(t - 1)$, the best estimate of the distribution of states is held in the vector S^+ . The one-step-ahead prediction of the state at time t is then $S^- = FS^+$.

Observe: The observation Y_t now becomes available and takes the value 0 or 1. If $Y_t = 0$ is observed, then $X_t = 0$ and the leading entry of S^- is incorporated into the log-likelihood function, being of course the probability that $X_t = 0$. If $Y_t = 1$ then $X_t > 0$ and the sum of all the entries of S^- below the leading entry is incorporated into the log-likelihood function. However, the column sum of S^- is 1 and so the numerically equivalent value is the difference between 1 and the leading entry of S^- .

Update: If $Y_t = 0$ then all the entries of S^+ are set to zero except the leading entry which is 1. On the other hand, if $Y_t = 1$ then the state $X_t = 0$ is populated with zero probability. The leading entry of S^+ is set to 0 and the remaining elements of S^+ are constructed from the non-leading entries of S^- by scaling these entries with a constant multiplier so as to ensure that column sum of S^+ is 1.

The construction of the log-likelihood function requires that this cycle be repeated for each transition in the observations of Y_t . The remaining practical problem concerns the initialization of S^+ prior to the first transition. If $Y_0 = 0$ then S^+ has leading entry 1 and 0 elsewhere. If $Y_0 = 1$, however, S^+ must be populated with the stationary distribution of the X_t process scaled to reflect the fact that $X_0 > 0$ because the leading entry of S^+ must be 0.

The development of the algorithm for maximum likelihood estimation of the parameters of the latent count model focuses on individual transitions, whereas an alternative, broader view of the observations Y_t would see a series of runs of 1s and 0s. Every transition from $Y_{t-1} = 0$ to $Y_t = 0$ occurs with the probability of the transition from $X_{t-1} = 0$ to $X_t = 0$, namely $F_{0,0}$. The contribution to the negative log-likelihood function of a run of m such transitions is therefore $-m \log F_{0,0}$. The contribution to the negative log-likelihood function of a run of k ones is more difficult to compute. To establish this probability, it is beneficial to partition F into the block matrix form

$$F = \begin{bmatrix} \beta & U' \\ V & P \end{bmatrix}, \quad (15)$$

where $\beta = F_{0,0}$ and U , V and P are respectively

$$V = \begin{bmatrix} F_{1,0} \\ \vdots \\ F_{M,0} \end{bmatrix}, \quad U = \begin{bmatrix} F_{0,1} \\ \vdots \\ F_{0,M} \end{bmatrix}, \quad P = \begin{bmatrix} F_{1,1} & \cdots & F_{1,M} \\ \vdots & \ddots & \vdots \\ F_{M,1} & \cdots & F_{M,M} \end{bmatrix}. \quad (16)$$

Theorem 5 states the general result from which the contribution to the negative log-likelihood function of a run of k ones may be computed directly.

THEOREM 5: Let X_t denote an unobserved integer-valued random variable modeled by the autoregressive process $X_t = \alpha \circ X_{t-1} + e_t$ where the survival process $\alpha \circ X_{t-1}$ is taken to be a binomial thinning process and e_t is an independent arrivals process. Suppose that

$$F = \begin{bmatrix} \beta & U' \\ V & P \end{bmatrix}$$

is the matrix of transitional probabilities of X in which U and V are column vectors of dimension M , P is a square matrix and $F_{p,q}$ denotes the probability of a transition from $X_{t-1} = q$ to $X_t = p$. Let Y_t denote the censored random variable taking the value zero if $X_t = 0$ and one otherwise. The probability of observing an unbroken run of precisely k ones in Y is then

$$Pr(Y_t = 0, Y_{t+1} = 1, \dots, Y_{t+k} = 1, Y_{t+k+1} = 0) = \begin{cases} \beta & k = 0, \\ U' P^{k-1} V & k \geq 1. \end{cases} \quad \blacksquare$$

5 Empirical Applications

Three practical illustrations of the BAR and PAR models in action are presented in this Section. The first is a simple Monte Carlo exercise to demonstrate that the maximum likelihood estimates of the BAR model computed using the recurrence relations developed in this paper behave as predicted by the theory of maximum likelihood.³ The second example applies the BAR and PAR models with constant arrival and survival probabilities to binary time series constructed from the daily volatility of four major stock market indices. The idea here is that news arrival drives volatility but is unobservable. The final illustration estimates a BAR model with time-varying arrival and survival probabilities for binary time series of abnormal price events in electricity prices. The fundamental assumption in this example is that abnormally high electricity prices are due to unobserved system stress.

5.1 Monte Carlo experiment

A straightforward Monte Carlo exercise is now undertaken to investigate the properties $\hat{\lambda}$ and $\hat{\alpha}$ the maximum likelihood estimators of λ and α . Specifically, the objective of the simulation exercise is to illustrate that in samples of size T , the distributions of $\hat{\lambda}$ and $\hat{\alpha}$ when scaled by \sqrt{T} converge to a normal distribution with finite variance as $T \rightarrow \infty$. The simulation of the latent BAR processes is accomplished in two stages; the count process X_t is simulated directly using a conventional integer count model as outlined in Section 2 and the binary series Y_t constructed treating X_t as observable. The parameters of the BAR model are then estimated from the simulated binary series Y_t alone.

For the latent BAR model 50,000 independent samples of lengths 500, 1,000, 5,000 and 10,000 were generated using the parameter values $\{\alpha = 0.20, \lambda = 0.50\}$, $\{\alpha = 0.30, \lambda = 0.30\}$ and $\{\alpha = 0.50, \lambda = 0.10\}$. These were taken to be representative of the combinations of arrival and survival rates likely to be encountered in practice. Note that in the BAR model there is

³Simulation results for the PAR model were substantially the same as those reported here for the BAR model.

at most one arrival between successive observations and so the maximum possible number of latent counts is exactly one more than the longest run of consecutive 1s observed in the actual sample.⁴ Table 1 shows the bias and root mean square error (RMSE) of the parameter estimates for the BAR model for all the combinations of parameter values and sample sizes considered.

| T | $\alpha = 0.2$ | $\lambda = 0.5$ | $\alpha = 0.3$ | $\lambda = 0.3$ | $\alpha = 0.5$ | $\lambda = 0.1$ |
|--------|---------------------|---------------------|---------------------|--------------------|---------------------|--------------------|
| 500 | -0.0025 (0.0659) | -0.0015 (0.0326) | -0.0035 (0.0530) | 0.0000 (0.0261) | -0.0063 (0.0577) | 0.0002 (0.0149) |
| 1,000 | -0.0023 (0.0482) | -0.0004 (0.0235) | -0.0016 (0.0372) | 0.0000 (0.0185) | -0.0030 (0.0405) | 0.0001 (0.0106) |
| 5,000 | -0.0004 (0.0214) | 0.0000 (0.0106) | -0.0005 (0.0166) | 0.0001 (0.0083) | -0.0006 (0.0180) | 0.0000 (0.0047) |
| 10,000 | -0.0002 (0.0152) | 0.0000 (0.0075) | -0.0003 (0.0117) | 0.0000 (0.0058) | -0.0003 (0.0127) | 0.0000 (0.0033) |

Table 1: Bias and RMSE (in parentheses) of the estimated parameters of the BAR model. Results are obtained from 50,000 replications of the experiment with samples of size T and for the parameter combinations shown.

It is clear from Table 1 that estimates of α exhibit a negative bias in small samples, but that this bias is not significant and becomes negligible as the sample size increases. By contrast the estimates of λ appear unbiased and extremely well resolved. As required, the RMSE of the estimates of both α and λ reduce as the sample size increases, but note that the RMSE of α is always larger than that of λ . This pattern of bias and RMSE is to be expected. The thinning parameter α can only be estimated from runs of 1s, while λ influences both the length of runs of 1s and 0s. It is therefore not surprising that λ is more precisely resolved than α .

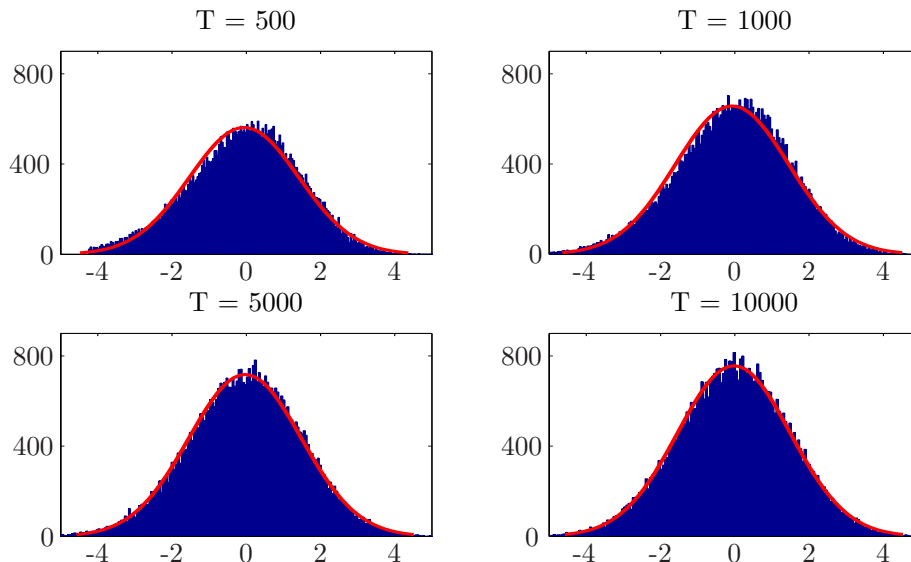


Figure 1: Histogram of parameter estimates with superimposed normal density for $\sqrt{T}(\hat{\alpha} - \alpha)$. Distributions are obtained from 50,000 replications of the experiment with samples of size T and with $\alpha = 0.20$.

⁴Of course, in a PAR model the maximum possible number of arrivals must be truncated at some reasonable number. The convention used in the empirical work in this paper is to set this value at twice the maximum number of consecutive 1s observed in the sample.

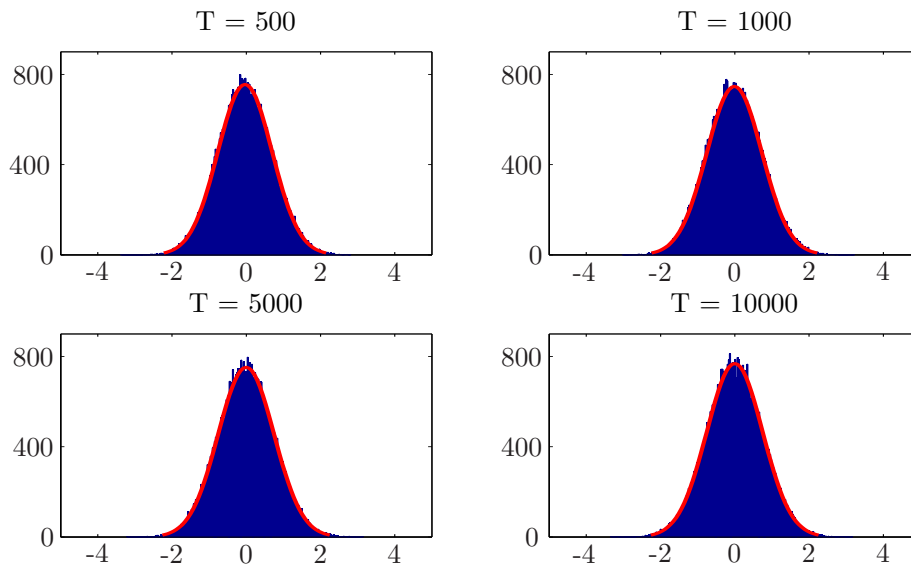


Figure 2: Histogram of parameter estimates with superimposed normal density for $\sqrt{T}(\hat{\lambda} - \lambda)$. Distributions are obtained from 50,000 replications of the experiment with samples of size T and with $\lambda = 0.50$.

Figures 1 and 2 readily demonstrate the asymptotic normality of the maximum likelihood estimators and confirm that the estimators exhibit \sqrt{T} convergence. Similar results are obtained for the other parameter combinations considered in this Monte Carlo exercise but are illustrated here. The robust conclusion of this simulation exercise is that the behavior of the parameter estimates in the latent count BAR model are not consistent with the predictions of the theory of maximum likelihood.

5.2 Volatility in financial markets

Since the seminal work of Clark (1973) the arrival and persistence of news is generally accepted as the fundamental mechanism driving volatility in the returns to financial assets. The difficulty with implementing this view in practice, however, is that ‘news’ is unobservable, or at best, difficult to quantify precisely. Modeling news in financial markets and the effect on volatility is therefore an intuitively appealing application of the models for latent counts developed in this paper.⁵ The proposed procedure is to construct a binary time series by imposing a threshold on the daily volatility of returns to stock market indices and assigning the value 1 to all exceedences. Underlying this binary process is the latent arrival and survival of news with respective intensities to be estimated using the models developed previously.

Most practical applications that deal with modeling volatility adopt a (G)ARCH class of model (Engle, 1982; Bollerslev, 1986) or a model of stochastic volatility (Taylor, 1982; 1986; see also Shephard, 2005). In their most simple forms, these traditional models are fundamentally different to the event-driven models developed in this paper because they focus on the entire trajectory of volatility whereas count models focus on regions of high volatility only. There are precedents in the literature to support the view that the averaging of the persistence of volatility over regions

⁵The use of models of point processes for volatility is a subject of current interest in the literature (see, for example, Brillinger, 2008).

of high and low volatility can have undesirable side effects. Friedman and Laibson (1989), for example, argue that the effects of large shocks to stock market returns do not persist as long as moderate ones. Gray (1996) documents this effect more precisely in a Markov-switching model of volatility. He concludes that volatility is less persistent in regions of high volatility and more persistent in regions of low volatility. It is also found that the values of the persistence parameters in both regimes are significantly lower than the estimates found in traditional (G)ARCH models.

The data used in this application are the daily volatilities (squared returns) to the S&P500 (4,697 observations), DJIA (4,697 observations), FTSE100 (4,709 observations) and NIKKEI225 (4,590 observations) equity indices. The data span the time interval from 2 January 1990 to 19 August 2008 for the S&P500, the DJIA and the FTSE100, with the difference in numbers of observations between the US indices and the UK index being attributable to different numbers of trading days. For the NIKKEI225, the data span the period from 4 January 1990 to 20 August 2008. The average daily trading volume for the indices for the full sample period are respectively 915m, 165m, 953m and 579m shares and the unconditional variances of daily returns ($\times 10^4$) are 1.020, 0.984, 1.081 and 2.175. Given the observed differences in volume and the fact that the unconditional volatility of the NIKKEI225 is substantially higher than the other markets, the thresholds for the volatilities were computed independently for each market. The binary time series for each market for the years 1998 to 2002, constructed by setting the thresholds at the 85th percentile of each volatility distribution, is illustrated in Figure 3. Not much can be read into these plots other than to observe that the behavior of the US and UK indices appears remarkably similar, while the NIKKEI225 appears to show less persistence in the binary process.



Figure 3: Binary time series constructed from daily volatilities using the 85th percentile of each series as the threshold.

Estimates of the PAR and BAR models are presented in Table 2 for thresholds set at the 75th (upper panel) and 85th percentiles (lower panel). A striking result is that the estimated intensity of the news arrival process, λ , is not significantly different across the markets for either choice of threshold. This may be taken as *a priori* evidence that intensity of news arrival is a global phenomenon. Note that BAR and PAR model have slightly different interpretations for λ . In the BAR model with only one possible new arrival, λ is the probability of this arrival. In the PAR model, where multiple arrivals are possible, λ is the *rate* of arrival and $1 - e^{-\lambda}$ is the probability that there is at least one arrival during the same period. Simple calculation will verify that λ for the BAR model and $1 - e^{-\lambda}$ for the PAR model are virtually identical for the results presented here, indicating that each model predicts the same probability for the arrival of news. Since the PAR model allows more than one item of news to arrive, a natural consequence of this property is that the estimated survival rate for the PAR model must necessarily be lower than that for the BAR model. This condition is also satisfied by the results reported in Table 2, indicating that the two models are entirely self-consistent.

| 75 th percentile | PAR | | BAR | |
|-----------------------------|--------------------|--------------------|--------------------|--------------------|
| | α | λ | α | λ |
| S&P500 | 0.0929 (0.0198) | 0.2610 (0.0089) | 0.1052 (0.0177) | 0.2294 (0.0069) |
| DJIA | 0.0732 (0.0187) | 0.2666 (0.0090) | 0.0835 (0.0176) | 0.2338 (0.0070) |
| FTSE100 | 0.1426 (0.0228) | 0.2469 (0.0087) | 0.1583 (0.0177) | 0.2185 (0.0068) |
| NIKKEI225 | 0.1148 (0.0213) | 0.2549 (0.0089) | 0.1283 (0.0179) | 0.2248 (0.0070) |
| 85 th percentile | PAR | | BAR | |
| | α | λ | α | λ |
| S&P500 | 0.1113 (0.0220) | 0.1446 (0.0061) | 0.1188 (0.0182) | 0.1346 (0.0053) |
| DJIA | 0.1165 (0.0224) | 0.1438 (0.0060) | 0.1241 (0.0183) | 0.1338 (0.0053) |
| FTSE100 | 0.1590 (0.0259) | 0.1367 (0.0059) | 0.1675 (0.0189) | 0.1278 (0.0052) |
| NIKKEI225 | 0.0871 (0.0204) | 0.1485 (0.0062) | 0.0937 (0.0179) | 0.1379 (0.0054) |

Table 2: Estimated survival and arrival parameters for the latent PAR and BAR models for the daily abnormal volatility series.

Interestingly enough, the estimated values of the survival parameter, α , are consistent with the observations of Gray (1996) that persistence in higher volatility regimes is significantly lower than the estimates typically reported by simple (G)ARCH models. Not surprisingly, when viewed from the perspective of both models and thresholds, it is difficult to differentiate between estimated persistence for the two US indices, although a marginal case could be made that the persistence in the DJIA is slightly lower than that of the S&P500. It is evident, however, that the estimates of α for the FTSE100 and the NIKKEI225 are significantly different from the

value of α in the US markets. The FTSE100 demonstrates significantly higher persistence than any of the other markets. It is difficult to provide a rational explanation for this behavioral phenomenon, but it may relate to the UK's geographical location between the markets of the US on the one hand and Europe and Asia on the other.

5.3 Abnormal price events in electricity markets

Since electricity is, for all practical purposes, a non-storable commodity, unexpectedly large increases in demand, perhaps due to extreme weather conditions, or bottlenecks in supply due to generator failures, cannot be smoothed by maintaining an inventory. Consequently, a generic feature of deregulated electricity markets worldwide is the periodic occurrence of abnormally high prices or price spikes (Barlow, 2002; de Jong and Huisman, 2003; Escribano, *et al.*, 2002; Lucia and Schwartz, 2002; Burger *et al.*, 2003; Byström, 2005; Cartea and Figueroa, 2005). Retail price regulation, however, prevents electricity retailers from passing the full extent of spot price fluctuations on to their customers thus leaving the retailers bearing a significant proportion of the price risk.

An abnormal price event is defined to be a situation in which the spot price of electricity exceeds a given threshold. The value of this threshold relates to the marginal cost of electricity generation in the market place. At this threshold, it becomes cost effective for more expensive electricity generators (usually gas-fired and diesel generators) to compete with the traditional low-cost generators (usually coal-fired and nuclear generators) and therefore this threshold represents the lowest price that will prevail in a stressed market.⁶

From a series of historical prices, however, all that can be observed is whether or not an abnormal price event occurred at a particular instant in time. The number of stresses acting on the market is not observable, making the analysis of price spikes in electricity markets ideally suited to the models of latent counts developed in this paper. An important distinguishing feature of this problem, by contrast with the previous empirical application to the volatility of returns to equity indices, is that the intensity of the arrivals process and the rate of survival cannot be assumed to be constant. At the very least, it may be conjectured that these parameters will have a temporal dependence driven by daily variations in electricity demand and longer-term seasonal effects. In other words, the assumption of constant arrival and survival probabilities must be relaxed, and these parameters allowed to depend on a set of time-varying covariates along the lines suggested in equations (4).

The data used in this empirical application is drawn from the Australian national electricity market that has been in operation since 1998. Five market regions are analyzed here, namely, New South Wales (NSW), Queensland (Qld), South Australia (SA), the Snowy Mountains (Snowy) and Victoria (Vic). The behavior of prices is examined for the period from the opening of the market on December 13, 1998 to May 1, 2007, a data set spanning 3,061 days or 146,928 half-hours for each region. A binary time series of *daily* data is constructed from the series of

⁶In the Australian market the value of the threshold is generally regarded as approximately A\$100/MWh but the argument concerning the setting of the threshold is a generic one. This choice of threshold lies well outside the 90th percentile of half-hourly prices in the Australian market.

half-hourly spot prices by assigning the value 1 to any day on which the spot price exceeds the threshold value of A\$100/MWh.

Temperature and load are exogenous variables that potentially characterize the intensity of price spikes. The effect of abnormal load on system stress is self-evident. The influence of daily temperature is captured by two variables, namely, T_{\max} , denoting the absolute deviation of the maximum daily temperature from the expected maximum for that day, and T_{\min} denoting the absolute deviation of the minimum daily temperature from the expected minimum temperature for that day. The advantage of using absolute deviations is that these correct for the positive correlation between temperature and price in summer months and the negative correlation between temperature and price in winter months. Temperature is included in the specification merely to search for increased resolution, particularly in respect of the probability of arrival of stress as generator failure may be accentuated by extreme temperatures.

| Parameter | NSW | Qld | SA | Snowy | Vic |
|-------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| δ_0 | -3.064* (0.102) | -2.215* (0.066) | -1.965* (0.063) | -3.112* (0.102) | -3.024* (0.100) |
| δ_1 | 0.341* (0.025) | 0.216* (0.027) | 0.181* (0.013) | 0.236* (0.017) | 0.173* (0.017) |
| δ_2 | 0.199* (0.038) | 0.025 (0.024) | 0.116* (0.020) | 0.058* (0.023) | 0.111* (0.033) |
| δ_3 | 0.740* (0.110) | 0.363* (0.057) | 0.531* (0.082) | 0.053 (0.084) | 0.770* (0.122) |
| λ_t | 0.046 | 0.103 | 0.131 | 0.044 | 0.048 |
| γ_0 | -0.776* (0.113) | -0.423* (0.071) | -0.510* (0.069) | -0.512* (0.101) | -0.816* (0.122) |
| γ_1 | 0.073 (0.037) | 0.038 (0.031) | 0.043* (0.018) | 0.024 (0.024) | 0.116* (0.025) |
| γ_2 | 0.008 (0.043) | 0.053* (0.023) | -0.040 (0.024) | -0.015 (0.024) | -0.037 (0.039) |
| γ_3 | 0.344* (0.131) | 0.121 (0.066) | 0.351* (0.109) | -0.020 (0.072) | 0.431* (0.136) |
| α_t | 0.369 | 0.481 | 0.451 | 0.451 | 0.357 |
| Log-L | -790.2 | -1205.2 | -1279.1 | -765.2 | -781.3 |

Table 3: Coefficient estimates and corresponding implied arrival probabilities (upper panel) and survival probabilities (lower panel) at the sample mean of the exogenous factors using the full set of regressors as drivers for the processes. An asterisk indicates significance at the 5% level. Standard errors are shown in parentheses.

To model the daily series of binary event data constructed from the spot electricity prices in each region, a BAR model with time-varying arrival and survival probabilities is estimated. The

arrivals and survival processes are specified by

$$\lambda_t = \frac{1}{1 + \exp(-\delta_0 - \delta_1 T_{\max_t} - \delta_2 T_{\min_t} - \delta_3 \text{Load}_t)},$$

$$\alpha_t = \frac{1}{1 + \exp(-\gamma_0 - \gamma_1 T_{\max_t} - \gamma_2 T_{\min_t} - \gamma_3 \text{Load}_t)}.$$

The implicit assumption underlying this choice of model is that at most one market stress can arrive each day. The results of the estimation are reported in Table 3. In the upper panel of Table 3, a higher coefficient value indicates a higher arrival probability, whereas in the lower panel a higher coefficient indicates a higher probability that existing stresses will persist.

Turning first to the results in the upper panel of Table 3, the expectation is that all explanatory variables will have positive coefficient estimates. In other words, the probability of a price spike should increase in the presence of extreme temperatures or abnormally high load. This is the case for all variables in all regions, with two notable exceptions. *First*, extreme minimum temperatures are not significant in Qld. This is not unexpected given Qld’s comparatively warm climate. *Second*, load is not significant in explaining arrivals in the Snowy region, a feature which may well be explained by its small load compared with other regions. The lower arrival parameter estimates for NSW, Snowy and Vic are consistent with fewer spikes observed in these regions relative to Qld and SA.

The lower panel of Table 3 prompts the conclusion that the persistence of market stress as estimated by the time-varying survival rate is less dependent on extreme temperatures than the arrivals process.⁷ It appears that the constant term carries most of the weight of the behavior of the survival of stresses, although the load variable is also significant in three of the six regions. What these results suggest is that persistence of stress is less influenced by temperature, but that load and other region-specific factors (embedded in the constant term) are important in modeling persistence. The higher survival parameters for Qld and SA are consistent with these regions having comparatively longer runs of consecutive daily price spikes.

In summary, there is clear evidence in these results that generalizing the latent event models to accommodate time-varying arrival and survival probabilities is a valuable exercise. Furthermore, the results obtained from this empirical application accord strongly with intuition. The arrival of system stress depends strongly on measurable exogenous factors, such as load and temperature, but that persistence is more closely related to region specific factors.

6 Conclusion

Integer-valued autoregressions specify the evolution of a series of non-negative integers, normally interpreted as a series of counts, in terms of the addition of new arrivals and the persistence of existing counts. This paper demonstrates how to estimate the parameters of these models when the counts are latent and the observable series is a binary time series derived from the evolution of the latent count process. It is demonstrated that the specification of the arrivals of new counts

⁷This is consistent with the conjecture made previously, that extreme temperatures may be more important in modeling arrivals.

in terms of a Poisson distributed process or a Bernoulli distributed process is fundamentally equivalent for low arrival intensities. The Poisson and Bernoulli autoregressive models both allow the derivation of a set of recurrence relationships governing the transitions of the counts from one state to another. These relationships are particularly useful in the computation of maximum likelihood estimators of the parameters of the models when the counts are observable and when they are latent. The construction of the log-likelihood function for the latent counts models may be interpreted within a filtering cycle, requiring the prediction of the distribution of latent counts, the evaluation of the probability of the current transition in the observed binary time series and the updating of the distribution of the counts based on the current observation. Alternatively, it is demonstrated that the log-likelihood function may be expressed in terms of the probabilities of observing the lengths of runs of zeros and ones in the observed binary time series. A simple simulation experiment demonstrates that the suggested maximum likelihood procedure yields estimators with the desired asymptotic properties.

Two applications are provided illustrating the latent count models in action. The first example treats the arrival of news in financial markets as a latent count variable and binary time series indicating the presence of excessive volatility in the returns to four major stock indices are constructed. It is found that the intensity of ‘news’ arrival is similar in all markets. The persistence of volatility, however, is a regional phenomenon with the FTSE100 index yielding the highest estimate of persistence. The second example models abnormal price spikes in the Australian electricity market as a function of unobservable system ‘stress’. The importance of this model stems from the fact that the survival and arrival probabilities are allowed to be time varying.

References

- Al-Osh, M. A. and Alzaid, A. A. (1987). 'First-order integer valued autoregressive (INAR(1)) process', *Journal of Time Series Analysis*, 8, 261–275.
- Barlow, M. T. (2002). 'A diffusion model for electricity prices', *Mathematical Finance*, 12, 287–298.
- Bollerslev, Y. (1986). 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, 31, 307–327.
- Brilinger, D. (2008). 'Extending the volatility concept to point processes', *Journal of Statistical Planning and Inference*, 138, 2607–2614.
- Burger, M., Klar, B., Mueller, A. and Schindlmayr, G. (2003). 'A spot market model for pricing derivatives in electricity markets', *Quantitative Finance*, 4, 109–122.
- Byström, H. N. E. (2005). 'Extreme value theory and extremely large electricity price changes', *International Review of Economics & Finance*, 14, 41–55.
- Cartea, Á. and Figueroa, M. G. (2005). 'Pricing in electricity markets: A mean reverting jump diffusion model with seasonality', *Applied Mathematical Finance*, 12, 313–335.
- Clark, P. (1973) 'A subordinated stochastic process model with finite variance for speculative prices', *Econometrica*, 41, 135–155.
- de Jong, C. and Huisman, R. (2003). 'Option prices for power prices with spikes', *Energy and Power Risk Management*, 7, 12–16.
- Du, J. G. and Li, Y. (1991). 'The integer-valued autoregressive (INAR(p)) model', *Journal of Time Series Analysis*, 12, 129–142.
- Engle, R.F. (1982). 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, 50, 987 - 1008.
- Escribano, Á, Peña, J. I. and Villaplana, P. (2002). 'Modelling electricity prices: International evidence', *Working Paper 02-27*, Universidad Carlos III de Madrid.
- Freeland, R. K. and McCabe, B. P. M. (2004a). 'Analysis of low count time series data by Poisson autoregression', *Journal of Time Series Analysis*, 25, 701–722.
- Freeland, R. K. and McCabe, B. P. M. (2004b). 'Forecasting discrete valued low count time series', *International Journal of Forecasting*, 20, 427–434.
- Friedman, B.M. and Laibson, D.I. (1989). 'Economic implications of extraordinary movements in stock prices', *Brookings Papers on Economic Activity*, 2, 137–189.
- Geman, H. and Roncorni, A. (2006). 'Understanding the fine structure of electricity prices', *Journal of Business*, 79, 1225–1261.
- Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Tables of Integrals, Series and Products*, 6th edition. Academic Press: London.

Gray, S.F. (1996). ‘Modeling the conditional distribution of interest rates as a regime-switching process’, *Journal of Financial Economics*, **42**, 27-62.

Lucia, J. J. and Schwartz, E. S. (2002). ‘Electricity prices and power derivatives: Evidence from the Nordic power exchange’, *Review of Derivatives Research*, **5**, 5–50.

McCabe, B. P. M. and Martin, G. M. (2005). ‘Bayesian predictions of low count time series’, *International Journal of Forecasting*, **21**, 315—330.

McKenzie, E. (1988). ‘Some ARMA models for dependent sequences of Poisson counts’, *Advances in Applied Probability*, **20**, 822–35.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press: Oxford.

Taylor, S. (1982). ‘Financial returns modelled by the product of two stochastic processes - A study of daily sugar prices 1961 - 79’. In Anderson, O.D. (ed.) *Time Series Analysis: Theory and Practice*, I, 203-226. North Holland: Amsterdam.

Taylor, S. (1986). *Modelling Financial Time Series*. John Wiley and Sons: New York.

7 Appendix

Proof of THEOREM 1

Suppose that X_0 is a Poisson distributed random variable with parameter λ_0 , then the probability generating function of X_0 is $G_0(z) = \exp(\lambda_0(z - 1))$. Furthermore, the independence of the random variables $\alpha \circ X_0$ and e_t means that the probability generating function of $X_1 = \alpha \circ X_0 + e_t$ is the product of the probability generating function of $\alpha \circ X_0$ and the probability generating function e_t , namely $\exp(\lambda(z - 1))$. However it follows directly from the definition of $\alpha \circ X_0$ in equation (2) and the second property of probability generating functions that the random variable $\alpha \circ X_0$ has probability generating function $G_0(1 - \alpha + \alpha z)$. Consequently

$$G_1(z) = G_0(1 - \alpha + \alpha z) e^{\lambda(z-1)} = e^{(\lambda_0\alpha + \lambda)(z-1)} \quad (17)$$

and therefore X_1 is Poisson distributed with parameter $(\lambda_0\alpha + \lambda)$. The procedure is iterative. The probability generating function $G_2(z)$ is computed by repeating the previous argument with λ_0 replaced by $\lambda_0\alpha + \lambda$ to get $G_2(z) = e^{(\lambda_0\alpha^2 + \lambda(1+\alpha))(z-1)}$ and the probability generating function $G_3(z)$ is computed from the expression for $G_1(z)$ by replacing λ_0 with $(\lambda_0\alpha^2 + \lambda(1+\alpha))$ and so on to deduce that in general

$$G_k(z) = \exp \left[\left(\alpha^k \lambda_0 + \lambda \sum_{n=0}^{k-1} \alpha^n \right) (z - 1) \right]. \quad (18)$$

By comparison with the probability generating function for the Poisson distribution, X_k is a Poisson distributed random variable with parameter

$$\alpha^k \lambda_0 + \lambda \sum_{n=0}^{k-1} \alpha^n. \quad (19)$$

Proof of COROLLARY 1.1

If $\alpha < 1$ then by recognizing that $\alpha^k \rightarrow 0$ as $k \rightarrow \infty$, it follows immediately that the limit of expression (19) as $k \rightarrow \infty$ is $\lambda/(1 - \alpha)$. Thus the process (2) has a stationary distribution with probability generating function

$$G(z) = \exp \left[\frac{\lambda}{1 - \alpha} (z - 1) \right] \quad (20)$$

This is the probability generating function of a Poisson process with parameter $\lambda/(1 - \alpha)$. Note that if $\lambda_0 = \lambda/(1 - \alpha)$ then process (2) is initialised by choosing X_0 to be a realisation of the stationary process, and subsequent deviates X_1, X_2, \dots are themselves realisations of the stationary process.

Proof of THEOREM 2

The proof begins by noting that for each integer $k > 1$, the independence of the processes $\alpha \circ X_{k-1}$ and e_k requires that $G_k(z)$, the probability generating function of X_k , is the product of the probability generating functions of the processes $\alpha \circ X_{k-1}$ and e_k , that is,

$$G_k(z) = G_{\alpha \circ X_{k-1}}(z)(1 + \lambda(z - 1)).$$

A second property of probability generating functions indicates that

$$G_{\alpha \circ X_{k-1}}(z) = G_{k-1}(1 + \alpha(z - 1)),$$

and therefore the sequence of probability generating functions describing the distribution of states of the integer-valued autoregressive process (6) satisfies

$$\begin{aligned} G_k(z) &= G_{k-1}(1 + \alpha(z - 1))(1 + \lambda(z - 1)), & k \geq 1 \\ G_0(z) &= 1 + p_0(z - 1). \end{aligned} \tag{21}$$

Equation (21) provides the basis for an induction argument. First, property (21) guarantees that the statement of the theorem is true when $k = 1$. The induction procedure now assumes that the statement of the theorem is true for integer $k > 1$ and employs property (21) in the form $G_{k+1}(z) = G_k(1 + \alpha(z - 1))(1 + \lambda(z - 1))$ to show by means of the induction assumption that

$$\begin{aligned} G_{k+1}(z) &= [1 + p_0\alpha^k(1 + \alpha(z - 1) - 1)] \prod_{n=0}^{k-1} [1 + \lambda\alpha^n(1 + \alpha(z - 1) - 1)](1 + \lambda(z - 1)) \\ &= [1 + p_0\alpha^{k+1}(z - 1)] \prod_{n=0}^{k-1} [1 + \lambda\alpha^{n+1}(z - 1)](1 + \lambda(z - 1)). \end{aligned}$$

The product is now re-indexed by replacing $(n + 1)$ with m to obtain

$$\begin{aligned} G_{k+1}(z) &= [1 + p_0\alpha^{k+1}(z - 1)] \prod_{m=1}^k [1 + \lambda\alpha^m(z - 1)](1 + \lambda(z - 1)) \\ &= [1 + p_0\alpha^{k+1}(z - 1)] \prod_{m=0}^k [1 + \lambda\alpha^m(z - 1)]. \end{aligned}$$

The result is thus proved using the principle of induction.

Proof of COROLLARY 2.1

Because $\alpha^k \rightarrow 0$ as $k \rightarrow \infty$ then $[1 + p_0\alpha^k(z - 1)] \rightarrow 1$ as $k \rightarrow \infty$. Moreover, a necessary and sufficient condition for the infinite product $\prod_{n=1}^{\infty} (1 + a_n)$ to converge is that $\sum_{n=1}^{\infty} a_n$ converges absolutely (see, for example, result 0.255 of Gradshteyn and Ryzhik, 2000). In this instance $a_n = \lambda(z - 1)\alpha^n$ and so $\sum_{n=1}^{\infty} a_n$ is absolutely convergent with sum to infinity $(z - 1)\lambda/(1 - \alpha)$. Therefore process (6) has a stationary distribution with probability generating function

$$G(z) = \lim_{k \rightarrow \infty} [1 + p_0\alpha^k(z - 1)] \prod_{n=0}^{k-1} [1 + \lambda\alpha^n(z - 1)] = \prod_{n=0}^{\infty} [1 + \lambda\alpha^n(z - 1)]. \tag{22}$$

However, unlike a Poisson autoregressive model, the probability generating function of the stationary distribution of a Bernoulli autoregressive process does not in general seem to have a closed form expression in terms of conventional functions.

Since $G(z) = \lim_{k \rightarrow \infty} G_k(z)$ has now been shown to exist, then by taking the limit of property (21) as $k \rightarrow \infty$ it is seen that $G(z)$ also satisfies the identity

$$G(z) = G(1 + \alpha(z - 1))(1 + \lambda(z - 1)). \quad (23)$$

Straightforward differentiation of this identity yields

$$\left. \begin{aligned} G'(z) &= \lambda G(\xi) + \alpha G'(\xi)(1 + \lambda(z - 1)), \\ G''(z) &= 2\lambda\alpha G'(\xi) + \alpha^2 G''(\xi)(1 + \lambda(z - 1)), \end{aligned} \right] \quad \xi(z) = 1 + \alpha(z - 1).$$

The values of $G'(1)$ and $G''(1)$ can be recovered from these equations via the intermediary result

$$\left. \begin{aligned} G'(1) &= \lambda G(1) + \alpha G'(1), \\ G''(1) &= 2\lambda\alpha G'(1) + \alpha^2 G''(1), \end{aligned} \right] \rightarrow \begin{aligned} G'(1) &= \frac{\lambda G(1)}{1 - \alpha}, \\ G''(1) &= \frac{2\lambda^2\alpha G(1)}{(1 - \alpha)(1 - \alpha^2)}. \end{aligned}$$

Of course, $G(1) = 1$ since it is by definition the sum of the probabilities of all the possible states of X_t , and therefore the expected value and variance⁸ of the stationary BAR process X_t underpinning the behavior of Y_t are respectively

$$\mathbb{E}[X] = \frac{\lambda}{1 - \alpha}, \quad \mathbb{E}[(X - \mathbb{E}[X])^2] = \frac{\lambda(1 + \alpha - \lambda)}{1 - \alpha^2}. \quad (24)$$

Thus the stationary distributions of the PAR and BAR processes have precisely the same expression for their first moment, namely $\lambda/(1 - \alpha)$, although of course the meaning of λ is different in each case. However this functional similarity fails for second moments since the stationary distribution of the PAR process is a Poisson process which necessarily has identical expressions of the mean and variance.

Proof of COROLLARY 2.2

Although the infinite product in expression (2) does not appear to be expressible in terms of the standard function of mathematics (except in the trivial case $\lambda = 0$), nevertheless the asymptotic behavior of this product can be deduced in the limit $\lambda \rightarrow 0^+$. The analysis begins by taking logarithms of equation (2) to get

$$\log G(z) = \sum_{n=0}^{\infty} \log(1 + \lambda\alpha^n(z - 1)).$$

By noting that $\log(1 + x) = x + O(x^2)$ when x is small, then with $x = \lambda(z - 1)\alpha^n$ and small values of λ the equivalent result is $\log(1 + \lambda\alpha^n(z - 1)) = \lambda\alpha^n(z - 1) + O(\lambda^2)$. The previous equation now gives

$$\log G(z) = \lambda(z - 1) \sum_{n=0}^{\infty} \alpha^n + O(\lambda^2) = \frac{\lambda}{1 - \alpha}(z - 1) + O(\lambda^2).$$

⁸The expected value and variance of a random process with probability generating function $G(z)$ are respectively $G'(1)$ and $G''(1) + G'(1) - [G'(1)]^2$.

from which it follows that

$$G(z) = \exp\left(\frac{\lambda}{1-\alpha}(z-1) + O(\lambda^2)\right).$$

Thus the stationary distribution of the BAR process is asymptotically a Poisson process with parameter $\lambda/(1-\alpha)$ as $\lambda \rightarrow 0^+$, and indeed the BAR process and PAR process are asymptotically identical in this limit. Moreover, the observed binary series Y_t has a stationary distribution for small λ that is (asymptotically) Bernoulli with parameter $\lambda/(1-\alpha)$ irrespective of whether the latent process has Poisson- or Bernoulli-distributed arrivals.

Proof of THEOREM 3

Note first that the restriction on the index of summation in the definition of $P_{p,q}$ is due to the fact that at most q counts of X_{t-1} can survive and that at most p arrivals can contribute to X_t . Each property of the theorem is proved in turn.

Property (i) - $p < q$ The derivation of (i) begins by considering $P_{p,q+1} - (1-\alpha)P_{p,q}$ as the difference of two summations to get

$$P_{p,q+1} - (1-\alpha)P_{p,q} = e^{-\lambda} \sum_{n=0}^p \alpha^n (1-\alpha)^{q+1-n} \frac{\lambda^{p-n}}{(p-n)!} \left[\binom{q+1}{n} - \binom{q}{n} \right].$$

First, the term arising from $n = 0$ is zero in this summation, and second, the difference of binomial coefficients in squared brackets takes the simplified form $\binom{q}{n-1}$. These observations are now introduced into the previous equation to obtain the simplified form

$$P_{p,q+1} - (1-\alpha)P_{p,q} = e^{-\lambda} \sum_{n=1}^p \binom{q}{n-1} \alpha^n (1-\alpha)^{q+1-n} \frac{\lambda^{p-n}}{(p-n)!}. \quad (25)$$

The summation in equation (25) is now re-indexed with $k = n - 1$ giving the final result

$$P_{p,q+1} - (1-\alpha)P_{p,q} = e^{-\lambda} \sum_{k=0}^{p-1} \binom{q}{k} \alpha^{k+1} (1-\alpha)^{q-k} \frac{\lambda^{p-1-k}}{(p-1-k)!} = \alpha P_{p-1,q}. \quad (26)$$

Property (ii) - $p > q$ When $q = 0$, no summation is required in the computation of $P_{p,0}$ which takes the value $\lambda^p e^{-\lambda}/p!$. Thus condition (ii), which asserts that $(p+1)P_{p+1,0} = \lambda P_{p,0}$ for $p \geq 0$, recovers the known specification for $P_{p,0}$. Suppose now that $p \geq q > 0$ and consider

$$(p+1)P_{p+1,q} - \lambda P_{p,q} = e^{-\lambda} \sum_{n=0}^q \binom{q}{n} \alpha^n (1-\alpha)^{q-n} \left[\frac{\lambda^{p+1-n}(p+1)}{(p+1-n)!} - \frac{\lambda^{p+1-n}}{(p-n)!} \right]. \quad (27)$$

The first term ($n = 0$) in this summation is zero and therefore the lower bound of this summation is effectively $n = 1$. When the summation on the right hand side of equation (27) is re-indexed

with $k = n - 1$, the result is

$$\begin{aligned}
(p+1)P_{p+1,q} - \lambda P_{p,q} &= \alpha e^{-\lambda} \sum_{k=0}^{q-1} \binom{q}{k+1} \alpha^k (1-\alpha)^{q-1-k} \left[\frac{(p+1)\lambda^{p-k}}{(p-k)!} - \frac{\lambda^{p-k}}{(p-k-1)!} \right] \\
&= \alpha e^{-\lambda} \sum_{k=0}^{q-1} \binom{q}{k+1} \alpha^k (1-\alpha)^{q-1-k} \frac{\lambda^{p-k}}{(p-k)!} [(p+1) - (p-k)] \\
&= \alpha e^{-\lambda} \sum_{k=0}^{q-1} \frac{q!(k+1)}{(q-1-k)!(k+1)!} \alpha^k (1-\alpha)^{q-1-k} \frac{\lambda^{p-k}}{(p-k)!} \\
&= q\alpha e^{-\lambda} \sum_{k=0}^{q-1} \frac{(q-1)!}{(q-1-k)!k!} \alpha^k (1-\alpha)^{q-1-k} \frac{\lambda^{p-k}}{(p-k)!} \\
&= q\alpha e^{-\lambda} \sum_{k=0}^{q-1} \binom{q-1}{k} \alpha^k (1-\alpha)^{q-1-k} \frac{\lambda^{p-k}}{(p-k)!} \\
&= q\alpha P_{p,q-1}.
\end{aligned}$$

Property (iii) - $p = q$ The derivation of (iii) begins by expressing $P_{q+1,q+1} - (1-\alpha)P_{q+1,q}$ as the difference of two summations to get

$$e^{-\lambda} \sum_{n=0}^{q+1} \binom{q+1}{n} \alpha^n (1-\alpha)^{q+1-n} \frac{\lambda^{q+1-n}}{(q+1-n)!} - e^{-\lambda} \sum_{n=0}^q \binom{q}{n} \alpha^n (1-\alpha)^{q+1-n} \frac{\lambda^{q+1-n}}{(q+1-n)!}.$$

The term with $n = q + 1$ is separated off in the left hand summation and the summations then collected together to give

$$e^{-\lambda} \alpha^{q+1} + e^{-\lambda} \sum_{n=0}^q \alpha^n (1-\alpha)^{q+1-n} \frac{\lambda^{q+1-n}}{(q+1-n)!} \left[\binom{q+1}{n} - \binom{q}{n} \right]. \quad (28)$$

However, the term with $n = 0$ makes no contribution to the summation in expression (28), and the bracketed difference in binomial coefficients in this summation can be simplified to the binomial coefficient $\binom{q}{n-1}$. When these observations are introduced into expression (28), the outcome is that

$$P_{q+1,q+1} - (1-\alpha)P_{q+1,q} = e^{-\lambda} \alpha^{q+1} + e^{-\lambda} \sum_{n=1}^q \binom{q}{n-1} \alpha^n (1-\alpha)^{q+1-n} \frac{\lambda^{q+1-n}}{(q+1-n)!}. \quad (29)$$

The summation in formula (29) is re-indexed with the change $k = n - 1$ and the first term incorporated into the existing summation to obtain the final result

$$\begin{aligned}
P_{q+1,q+1} - (1-\alpha)P_{q+1,q} &= e^{-\lambda} \alpha^{q+1} + e^{-\lambda} \sum_{k=0}^{q-1} \binom{q}{k} \alpha^{k+1} (1-\alpha)^{q-k} \frac{\lambda^{q-k}}{(q-k)!} \\
&= \alpha e^{-\lambda} \sum_{k=0}^q \binom{q}{k} \alpha^{k+1} (1-\alpha)^{q-k} \frac{\lambda^{q-k}}{(q-k)!} = \alpha P_{q,q}.
\end{aligned}$$

This completes the proof of Theorem 3.

Proof of THEOREM 4

Part (i) of the theorem has been argued previously but is present in the statement of the theorem for completeness. The meat of the theorem lies in result (ii) which in fact combines two different cases, namely the case in which $p = q$ and the case in which $p < q$. The proof of each case begins by considering the expression $(1 - \alpha)B_{p,q-1} + \alpha B_{p-1,q-1}$ assuming that $p, q \geq 1$.

Property (i) - $p = q$ The proof of the theorem in this special case hinges on the observation that $B_{q,q} = \alpha^q(1 - \lambda) + q\alpha^{q-1}(1 - \alpha)\lambda$. When $p = q$ expression $(1 - \alpha)B_{p,q-1} + \alpha B_{p-1,q-1}$ becomes

$$\begin{aligned} (1 - \alpha)B_{q,q-1} + \alpha B_{q-1,q-1} &= (1 - \alpha)\alpha^{q-1}\lambda + \alpha \left[\alpha^{q-1}(1 - \lambda) + (q - 1)\alpha^{q-2}(1 - \alpha)\lambda \right] \\ &= (1 - \alpha)\alpha^{q-1}\lambda + \alpha^q(1 - \lambda) + (q - 1)\alpha^{q-1}(1 - \alpha)\lambda \\ &= \alpha^q(1 - \lambda) + q\alpha^{q-1}(1 - \alpha)\lambda = B_{q,q}. \end{aligned}$$

Property (ii) - $p < q$ In this case each expression in $(1 - \alpha)B_{p,q-1} + \alpha B_{p-1,q-1}$ is first replaced by its definition to obtain

$$\begin{aligned} (1 - \alpha) \left[\binom{q-1}{p} \alpha^p (1 - \alpha)^{q-1-p} (1 - \lambda) + \binom{q-1}{p-1} \alpha^{p-1} (1 - \alpha)^{q-p} \lambda \right] \\ + \alpha \left[\binom{q-1}{p-1} \alpha^{p-1} (1 - \alpha)^{q-p} (1 - \lambda) + \binom{q-1}{p-2} \alpha^{p-2} (1 - \alpha)^{q+1-p} \lambda \right]. \end{aligned}$$

The terms are now regrouped to give the simplified expression

$$\left[\binom{q-1}{p} + \binom{q-1}{p-1} \right] \alpha^p (1 - \alpha)^{q-p} (1 - \lambda) + \left[\binom{q-1}{p-1} + \binom{q-1}{p-2} \right] \alpha^{p-1} (1 - \alpha)^{q+1-p} \lambda,$$

which can be further simplified using the properties of binomial coefficients to obtain

$$\binom{q}{p} \alpha^p (1 - \alpha)^{q-p} (1 - \lambda) + \binom{q}{p-1} \alpha^{p-1} (1 - \alpha)^{q+1-p} \lambda.$$

Thus it has been demonstrated that $B_{p,q} = (1 - \alpha)B_{p,q-1} + \alpha B_{p-1,q-1}$ provided $p, q \geq 1$. The case $p = 0$ is the special case of property (ii) in which $B_{-1,q-1} = 0$. The recursive formula now takes the simplified form $B_{0,q} = (1 - \alpha)B_{0,q-1}$ with solution $B_{0,q} = (1 - \alpha)^q B_{0,0} = (1 - \alpha)^q (1 - \lambda)$. This completes the proof of Theorem 4.

Proof of THEOREM 5

Let $S_t, S_{t+1}, \dots, S_{t+k}$ be a sequence of column vectors, each of dimension $(M + 1)$, with S_{t+j} storing the conditional distribution of the censored process X after a run of j consecutive ones has been observed. Since $Y_t = 0$ then $X_t = 0$ and $S_t = [1, 0, \dots]'$. The distribution of states of X_{t+1} prior to observing Y_{t+1} is given by

$$FS_t = \begin{bmatrix} \beta & U' \\ V & P \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \beta \\ V \end{bmatrix},$$

The probability that $X_{t+1} = 0$ is given by the leading entry of FS_t , namely β .

Suppose now that $Y_{t+1} = 1$ is observed, then $X_{t+1} > 0$ and the conditional distribution of X_{t+1} becomes

$$S_{t+1} = \frac{1}{(1 - \beta_{t+1})} \begin{bmatrix} 0 \\ V \end{bmatrix}, \quad \beta_{t+1} = \beta.$$

The previous procedure is repeated, in this instance treating the initial state as S_{t+1} . Prior to observing the value of Y_{t+2} , the anticipated distribution of X_{t+2} is therefore

$$FS_{t+1} = \frac{1}{(1 - \beta_{t+1})} \begin{bmatrix} \beta & U' \\ V & P \end{bmatrix} \begin{bmatrix} 0 \\ V \end{bmatrix} = \frac{1}{(1 - \beta_{t+1})} \begin{bmatrix} U'V \\ PV \end{bmatrix},$$

The leading entry of FS_{t+1} , namely $U'V/(1 - \beta_{t+1})$, gives the probability of observing $Y_{t+2} = 0$, and therefore the probability of observing a single isolated one is

$$(1 - \beta_{t+1}) \times \frac{U'V}{(1 - \beta_{t+1})} = U'V.$$

On the other hand if $Y_{t+2} = 1$ is observed, then $X_{t+2} > 0$ and the conditional distribution of X_{t+2} becomes

$$S_{t+2} = \frac{1}{(1 - \beta_{t+1})(1 - \beta_{t+2})} \begin{bmatrix} 0 \\ PV \end{bmatrix}, \quad \beta_{t+2} = \frac{U'V}{(1 - \beta_{t+1})}.$$

Suppose now that S_j ($j \geq 2$) is the conditional distribution of X_{t+j} after a run of j consecutive ones has been observed. Mathematical induction will be used to show that

$$S_{t+j} = \frac{1}{\prod_{n=1}^j (1 - \beta_{t+n})} \begin{bmatrix} 0 \\ P^{j-1}V \end{bmatrix}, \quad \beta_{t+j} = \frac{U'P^{j-2}V}{\prod_{n=1}^{j-1} (1 - \beta_{t+n})}. \quad (30)$$

The result is true by inspection for $j = 2$. Assume now that the result is true for j consecutive ones then

$$FS_{t+j} = \frac{1}{\prod_{n=1}^j (1 - \beta_{t+n})} \begin{bmatrix} \beta & U' \\ V & P \end{bmatrix} \begin{bmatrix} 0 \\ P^{j-1}V \end{bmatrix} = \frac{1}{\prod_{n=1}^j (1 - \beta_{t+n})} \begin{bmatrix} U'P^{j-1}V \\ P^jV \end{bmatrix},$$

The model probability associated with the observation $Y_{t+j+1} = 0$ is therefore

$$\beta_{t+j+1} = \frac{U'P^{j-1}V}{\prod_{n=1}^j (1 - \beta_{t+n})}$$

while the model probability associated with the observation $Y_{t+j+1} = 1$ is $(1 - \beta_{t+j+1})$ and the associate distribution of censored counts is given by the vector

$$\frac{P^jV}{\prod_{n=1}^{j+1} (1 - \beta_{t+n})}.$$

The expressions obtained for β_{t+j+1} and S_{t+j+1} are precisely those which would have been obtained by replacing j by $(j + 1)$ in the induction assumption, and so the conjecture of the induction assumption is substantiated. In particular, the condition satisfied by the coefficients $\beta_{t+1}, \dots, \beta_{t+k}$ can be re-expressed in the equivalent form

$$\beta_{t+j+1} \prod_{n=1}^j (1 - \beta_{t+n}) = U'P^{j-1}V.$$

If a run of k consecutive ones is observed then the model estimate of the probability of this event is

$$(1 - \beta_{t+1}) \times \cdots \times (1 - \beta_{t+k}) \times \beta_{t+k+1} = \beta_{t+k+1} \prod_{n=1}^k (1 - \beta_{t+n}) = U' P^{k-1} V.$$

This completes the proof of Theorem 5.
