Efficient Method of Moments Estimators for Integer Time Series Models

Vance L. Martin
University of Melbourne

A.R.Tremayne
University of New South Wales and University of Liverpool

Robert C. Jung
Universität Hohenheim

September 2012

Keywords: count time series; binomial thinning; EMM; subsampling standard errors

JEL Classification: C15, C22, C25

Abstract

The parameters of Poisson Integer AutoRegressive, or PoINAR, models can be estimated by maximum likelihood where the prediction error decomposition, together with convolution methods, is used to write down the likelihood function. When a moving average component is introduced this is not the case. In this paper we consider the use of techniques of efficient method of moment techniques as a means of obtaining practical estimators of relevant parameters using simulation methods. Under certain regularity conditions, the resultant estimators are consistent, asymptotically normal and achieve the Cramér-Rao bound. Simulation evidence on the efficacy of the approach is provided and it is seen that the method can yield serviceable estimates, even with relatively small samples. Estimated standard errors for parameters are obtained using subsampling methods. Applications are in short supply with these models, though the range is increasing. We provide two examples using real life data: one data set is adequately modelled using a pure integer moving average specification; and the other, a well known data set in the branching process literature that has hitherto proved difficult to model satisfactorily, uses a mixed specification with special features.

1 Introduction

Time series models for counts have been an active area of research for quite a time. Some of those most frequently encountered are integer autoregressive (INAR) models where the integer nature of the data is preserved by using the binomial thinning operator of Steutel and van Harn (1979). Three of the early papers investigating the first order INAR model are those by Al-Osh and Alzaid (1987) and McKenzie (1985, 1988). There then followed a steady stream of papers relating to estimation and forecasting with the first order model, including Brännäs (1994) and Freeland and McCabe (2004a,b). Subsequently, higher order INAR models have been investigated, but the extension from the first order case is not unique. The paper by Du and Li (1991) provides one such specification and is still based on independent binomial thinning operators. Other authors have considered higher order specifications, including Alzaid and Al-Osh (1990) and Bu, Hadri and McCabe (2008). That the general topic remains of interest today is evidenced by the important recent contributions of, *inter alia*, Drost, van den Akker and Wekker (2009) and McCabe, Martin and Harris (2011). An interesting review paper is provided by McKenzie (2003).

Most of these papers, and others in the field not cited here, when they are concerned with parameter estimation, focus solely on autoregressive specifications. A number of methods are available, ranging from moment-based estimators, through least squares procedures to maximum likelihood, be it fully parametric or semi-parametric. There are various reasons for restricting to this type of model, one of which is the intractability of writing down, in a compact manner, the conditional distributions needed to permit likelihood-based procedures when models with moving average components are entertained, a feature which would also complicate forecasting. Indeed, one important reason for seeking models that preserve integer status is that one of the goals of modelling may be forecasting or prediction. The matter of coherent forecasting has also been the subject of attention in the literature, see, *inter alia*, the work of McCabe and Martin (2005) and Jung and Tremayne (2006a).

Few papers consider moving average models based upon binomial thinning, although mention should be made of the early contributions of Al-Osh and Alzaid (1988) and McKenzie (1988) and also more recent work in Brännäs and Hall (2001) and Brännäs and Shahiduzzaman Quoreshi (2010). Feasible parameter estimators in such models include (generalized) methods of moments and conditional least squares procedures.

The purpose of this paper is to advance the use of the efficient method of moments estimator (EMM) of Gallant and Tauchen (1996) and the corresponding indirect inference methods of Gouriéroux, Monfort and Renault (2003). Integer autoregressive specifications are considered, specifically of first and second order types, because there is extant literature available to which the performance of EMM relative to maximum likelihood can be assessed. We also examine the behaviour of EMM in integer time series models featuring moving average components using Monte Carlo methods. Standard errors for the resultant estimators are obtained using subsampling methods. Finally, we use EMM to fit models for two real life data sets, each of which may plausibly regarded as exhibiting moving average behaviour. Diagnostic testing of the adequacy of each specification for the relevant data set suggest that the method works satisfactorily.

The remainder of the paper proceeds as follows. A brief introduction to binomial thinning models and some other basic ideas are introduced in Section 2; both autoregressive and moving average representations are discussed. There then follows in Section 3 a discussion of the EMM estimator for binomial thinning models, with the estimation of standard errors being treated in Section 3.2. The finite sample properties of this estimator are examined in Section 3.3, which is then applied in Section 4 to the modelling of two real life integer time series data sets. Concluding comments are provided in Section 5.

2 Integer Models Based on Binomial Thinning

The binomial thinning operator '$\circ$' introduced by Steutel and van Harn (1979) preserves the status of an integer random variable W when operated on by a parameter $\theta \in [0, 1)$ *via* $R = \theta \circ W = \sum_{s=1}^{W} e_s$, where the e_s, $s = 1, ..., W$ are *iid* Bernoulli counting sequence random variables $P(e_s = 1) = \theta$ and $P(e_s = 0) = 1 - \theta$. The operator is a random operator and the random variable R has a binomial distribution with parameters W and θ and counts the number of 'survivors' from the count W remaining after thinning.

Notice that the thinning operator confers greater dispersion on the number of survivors than does the ordinary multiplication operator. For instance, in integer time series models, W may often be an equi-dispersed Poisson random variable with equal mean and variance λ, say. Suppose W_{t-1} is an integer random variable arising at time $t-1$ and subjected to binomial thinning to produce the number of survivors in the next period, R_t, then, conditional on W_{t-1}, R_t is an integer random variable with variance $\theta(1-\theta)W_{t-1}$, whereas $S_t = \theta W_{t-1}$ has zero conditional variance (the unconditional counterparts are $\theta\lambda$ and $\theta^2\lambda$). In this paper, we shall assume that all thinnings at any given period are performed independently of one another and that each thinning is performed independently at each time period with constant probability of 'success',

θ. These conditions are standard, see, *e.g.* Du and Li (1991, p.129). Although other random operators are sometimes used in integer time series analysis (see, for example, Alzaid and Al-Osh, 1990, who use binomial thinning operations in a somewhat different way and Joe, 1996, who, on occasion, uses a different operator), the binomial thinning operator used in this way is by far the most commonly encountered. It is possible to entertain nonconstancy of thinning parameter(s), but in this paper we shall not do this.

The prototypical model in the class of binomial thinning models for integer time series is the first order integer autoregressive, or INAR(1), model given by

$$\begin{aligned} y_t &= \alpha \circ y_{t-1} + u_t \\ &= \sum_{s=1}^{y_{t-1}} e_{s,t-1} + u_t \\ E\left[e_{s,t-1}\right] &= \alpha, \end{aligned} \tag{1}$$

where u_t is an *iid* random variable with mean μ_U and variance σ_U^2. With an equi-spaced series it is often convenient to specify that $u_t \sim iid\ Po(\lambda)$, an independent Poisson random variable with mean λ. In this case it can be shown that the stationary marginal distribution of y_t is $Po(\lambda/(1-\alpha))$. The process is Markovian and hence depends upon its past history, $\mathcal{F}_{t-1}$, only through y_{t-1}. Moreover, the process can be shown to be a member of the class of conditionally linear autoregressive models (CLAR), see Grunwald, Hyndman, Tedesco and Tweedie (2000), with a conditional mean (regression) function given by

$$E(y_t|\mathcal{F}_{t-1}) = \alpha y_{t-1} + \lambda. \tag{2}$$

The autocorrelation function (ACF) mirrors that of the continuous AR(1) counterpart in that the jth autocorrelation is given by $\rho_j = \alpha^j, j = 1, 2, \ldots$. Notice that only positive autocorrelation can be modelled with an INAR(1), and, indeed, any INAR(p) specification. All these details are discussed in greater depth in Jung, Ronning and Tremayne (2005) and some of the references to be found there.

Even from this limited discussion, it is clear that method of moments, MM, estimators of the model's parameters could be obtained in many ways, including being based on the first order sample autocorrelation coefficient together with the sample mean of the data $y_1, y_2, ..., y_T$. Conditional least squares (CLS) estimators (Klimko and Nelson, 1978) can be obtained from a minimization of the criterion function

$$S(\alpha, \lambda) = \sum_{t=2}^{T} (y_t - \alpha y_{t-1} - \lambda)^2,$$

and conditional maximum likelihood estimators, MLE, can be based upon maximizing the function

$$\begin{aligned}\ell(\alpha, \lambda; y_1, ..., y_T) &\propto \sum_{t=2}^{T} \log p(y_t|y_{t-1}) \\ &= \sum_{t=2}^{T} \log \sum_{r=0}^{m} p_R(r|y_{t-1}) p_U(y_t - r),\end{aligned}$$

where $p_R(r|y_{t-1})$ is the binomial distribution of the number of survivors from the previous population count y_{t-1} that have survived the binomial thinning parameterized by α; $p_U(\cdot)$ denotes the (Poisson) distribution of the innovations and $m = \min(y_t, y_{t-1})$. The term after the logarithm on the right hand side is, of course, a convolution and the complete expression for the conditional distribution $p(y_t|y_{t-1})$ with Poisson innovations is given by, *e.g.*, Jung, Ronning and Tremayne (2005) in their expression (B.7).

The model can be generalized to a pth order one in a variety of ways. Here we confine attention to the extension proposed by Du and Li (1991)

$$y_t = \sum_{i=1}^{p} \alpha_i \circ y_{t-i} + u_t,$$

where the p thinnings are applied independently at each time period and the u_t are *iid* innovations with finite mean and variance independent of all Bernoulli counting sequences. The particular case with which we shall be concerned here is $p = 2$, *i.e.* the INAR(2) model. This model does not have the property of closure under convolution enjoyed by the first order counterpart with Poisson innovations; the marginal distribution of y_t is, in fact, overdispersed. But it does have the same ACF as a continuous Gaussian AR(2) and retains conditional linearity in that the conditional mean function (Poisson innovations assumed)

$$E(y_t|\mathcal{F}_{t-1}) = E(y_t|y_{t-1}, y_{t-2}) = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \lambda.$$

The parameters of the model must satisfy $\alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 < 1$. As before, a variety of approaches to estimating the model's parameters is available. Bu, Hadri and McCabe (2008) provide the MLE for this model and some Monte Carlo evidence relating to its behaviour. The conditional log-likelihood is of the form

$$\begin{aligned}\ell(\alpha_1, \alpha_2, \lambda; y_1, ..., y_T) &\propto \sum_{t=3}^{T} \log p(y_t|y_{t-1}, y_{t-2}) \\ &= \sum_{t=3}^{T} \log \sum_{r=0}^{m} p_R(r|y_{t-1}) \left[\sum_{s=0}^{n} p_S(s|y_{t-2}) p_U(y_t - r - s) \right],\end{aligned}$$

where m and $p_U(\cdot)$ are as before, $p_R(r|y_{t-1})$ denotes the binomial distribution of the number of survivors from thinning the count y_{t-1} at time t using parameter α_1, $p_S(s|y_{t-2})$ is the binomial distribution of the number of survivors from the population count y_{t-2} that survive the binomial thinning parameterized by α_2 and $n = \min(y_{t-1}, y_{t-2})$. The full expression for the conditional distribution $p(y_t|\mathcal{F}_{t-1})$ is given by Bu, Hadri and McCabe (2008) as their (13).

Turning to moving average models based on binomial thinning, the basic INMA(q) model is given by

$$y_t = \sum_{i=1}^{q} \beta_i \circ u_{t-i} + u_t, \tag{3}$$

and was first considered by McKenzie (1988) in the context of $u_t \sim iid\ Po(\lambda)$. Each $\beta_i \in [0, 1], i = 1, ..., q$. In the most basic model, all thinnings are independent, so that an individual (birth) entering the system at time t is available until time $t+q$ but then exits it. The ACF of the process, which has mean and variance equal to $\lambda/(1+\sum_{i=1}^{q} \beta_i)$, is given by $\rho_j = \sum_{i=0}^{q-j} \beta_i \beta_{i+j}/(1 + \sum_{i=1}^{q} \beta_i), i = 1, ..., q$ and zero otherwise. Brännäs and Hall (2001) discuss a range of other dependent thinning specifications that change the ACF of the basic model.

The special case of the INMA(1) model will be used at a number of junctures in this paper and this model has some interesting features. For instance, when the innovations, or births, are Poisson distributed the joint distribution of (y_t, y_{t-1}) is the same as that of the Poisson INAR(1) counterpart. This implies that the conditional mean function is also given by (2) though, of course, with parameter $\beta/(1 + \beta)$ replacing α; see Al-Osh and Alzaid (1988, eq. 3.10). The ACF has only one non-zero ordinate, $\rho_1 = \beta/(1+\beta) \leq 0.5$. Finally, the process is time reversible; see McKenzie (1988, Sec. 3.1) for more details. However, there remain significant differences between the INAR(1) and INMA(1) models, even with Poisson innovations. This is because their distributional properties are dissimilar; the INMA(1), unlike the INAR(1), is not Markovian and there is no sufficiency in the sense of reduction of the data. This makes the obtaining of maximum likelihood estimators intractable.

In what follows we shall also consider mixed specifications, which have received little attention in the literature. The basic specification to be used is due to McKenzie (1988, Sec. 4) and, in its simplest INARMA(1,1) form, is given by

$$\begin{aligned} x_t &= \alpha \circ x_{t-1} + u_t \\ y_t &= x_{t-1} + \beta \circ u_t, \end{aligned} \tag{4}$$

where y_t is the observable process and x_t is an unobservable latent process carrying the autoregressive dependence. The *iid* Poisson random variables u_t drive the moving average aspect of the model parameterized by β and use has been made of the reversibility of the INMA(1) model. Thinnings are performed independently and the presence of the lagged latent autoregressive component in contributing to y_t ensures the independence of the two birth counts implicit in its determination; if the model were specified as an INARMA(1,q), x_{t-q} would have entered in the second line of (4). The model reduces to an INAR(1) if $\beta = 0$ and to an INMA(1) if $\alpha = 0$. The mean and variance of y_t are $\lambda[(1-\alpha)^{-1} + \beta]$ and the ACF is given by

$$\begin{aligned} \rho_1 &= \frac{\alpha - \alpha^2 + \beta}{1 - \alpha + \beta} \\ \rho_j &= \alpha^{j-1}\rho_1, \quad j = 2, 3, \ldots \quad . \end{aligned}$$

3 The EMM Estimator

Given the difficulty of obtaining the maximum likelihood estimator for higher order INAR models and the intractability of it for INMA and mixed INARMA models in general, an alternative estimator is adopted based on the efficient method of moments estimator (EMM) of Gallant and Tauchen (1996). This estimator is also related to the indirect inference approach of Gouriéroux, Monfort, and Renault (1993), the simulated method of moments estimator (SMM) of Duffie and Singleton (1993), and the simulated quasi maximum likelihood estimator (SQML) of Smith (1993). In the next subsection we outline parameter estimation in count models using EMM and follow this with an account of how estimated asymptotic standard errors can be obtained *via* subsampling methods using the proposals described in Politis, Romano and Wolf (1999).

3.1 Parameter Estimation

The approach is to specify an alternative model and corresponding likelihood function, commonly referred to as the auxiliary model, which has the property that it represents a good approximation of the true model and hence the true likelihood function, but nonetheless is simpler to compute than the maximum likelihood estimator of an INARMA model. The estimator of the auxiliary model is also known as a quasi maximum likelihood estimator (QMLE) as it is based on a likelihood that is an approximation of the true likelihood. The approach of the EMM estimator stems from the property that the gradient vector estimator of the auxiliary model is zero when evaluated at the actual data y_t. This suggests that, by choosing parameters of the true INARMA model, simulated data from the INARMA model, $y_{s,t}$, can be generated and used to evaluate the gradient vector of the auxiliary model. The EMM solution is given by the set of parameter values of the true model that generate the smallest value of the gradient vector of the auxiliary model evaluated at the simulated data.

The stipulation that the auxiliary model be a good approximation of the true model is a requirement that there is a mapping between the parameters of the auxiliary model and the INARMA model. This mapping is also known as the binding function (Gouriéroux, Monfort, and Renault,1993). This condition means that Slutsky's theorem can be used to establish the consistency of the EMM estimator as the existence of a consistent estimator of the auxiliary model's parameters also establishes the existence of a consistent estimator of the parameters of the INARMA model.

In the case where the number of gradients of the auxiliary model matches the number of unknown parameters in the INARMA model, the model is just identified and the gradient vector evaluated using the simulated data is zero. If the dimension of the gradient vector exceeds the number of parameters in the INARMA model, the model is over-identified and the gradient vector evaluated at the simulated data, in general, is positive. A natural choice for the auxiliary model is an AR(p) model, which is motivated by the Klimko and Nelson (1978) conditional least squares estimator, which yields a consistent estimator of the parameters of INAR models. The quality of the approximation is then a function of p, the length of the lag structure. The INMA model can be viewed as an infinite INAR model in much the same way that, for continuous time series models, a MA model is represented by an infinite AR model. Hence, increasing p in the auxiliary model improves the quality of the approximation of the model, thereby improving the efficiency of the EMM estimator. In particular, by allowing p to increase at a certain rate as the sample size increases, the approximating model approaches the true model resulting in the EMM estimator approaching the MLE asymptotically.

The last requirement needed to implement the EMM estimator is that the INARMA model can be simulated. For the INARMA(p,1) specification

$$\begin{aligned} x_t &= \sum_{i=1}^{p} \alpha_i \circ x_{t-i} + u_t \\ y_t &= x_{t-1} + \beta_1 \circ u_t \\ u_t &\sim Po\,(\lambda)\,, \end{aligned} \tag{5}$$

adopted in the Monte Carlo experiments and the empirical applications, this is indeed the case when the $p+1$ thinning operations in (5) are all treated independently. The model is simulated by expressing the thinning operations in terms of independent uniform random numbers $e_{s,t-j}$, $j = 0, 1, 2, \cdots, p$, according to

$$\begin{aligned} x_{s,t} &= \sum_{i=1}^{p} \sum_{s=1}^{x_{t-i}} e_{s,t-i} + u_t \\ y_{s,t} &= x_{s,t-1} + \sum_{s=1}^{u_t} e_{s,t} \\ u_t &\sim Po(\lambda), \end{aligned} \tag{6}$$

where the uniform random numbers have moments

$$\begin{aligned} E[e_{s,t-i}] &= \alpha_i, \qquad i = 1, 2, \cdots, p, \\ E[e_{s,t}] &= \beta_1. \end{aligned} \tag{7}$$

Formally, the EMM estimator is based on solving

$$\widehat{\theta} = \arg\min_{\theta} G_s' I^{-1} G_s = \arg\min_{\theta} Q(\theta), \tag{8}$$

where G_s is a $(K \times 1)$ vector of gradients of the specified auxiliary model evaluated at the simulated data and the QMLE estimates of the auxiliary model, and I is a $(K \times K)$ optimal weighting matrix defined below. The auxiliary model specification is an AR(p) model with a constant

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + v_t, \tag{9}$$

where v_t is *iid* $N(0, \sigma_v^2)$. The gradient conditions of the auxiliary model at time t use

$$\begin{aligned} g_{1,t} &= \frac{v_t}{\sigma_v^2} \\ g_{i,t} &= \frac{v_t y_{t-i+1}}{\sigma_v^2}, \qquad i = 2, 3, \cdots, p+1 \\ g_{p+2,t} &= \left(\frac{v_t^2}{\sigma_v^2} - 1 \right) \frac{1}{2\sigma_v^2}, \end{aligned} \tag{10}$$

a total of $K = p + 2$ elements. Stacking these then and averaging across a sample of size T, gives

$$G = \frac{1}{T}\sum_{t=p+1}^{T}\begin{bmatrix} g_{1,t} \\ g_{2,t} \\ g_{3,t} \\ \vdots \\ g_{p+1,t} \\ g_{p+2,t} \end{bmatrix} = \frac{1}{T}\sum_{t=p+1}^{T}\begin{bmatrix} \frac{v_t}{\sigma_v^2} \\ \frac{v_t y_{t-1}}{\sigma_v^2} \\ \frac{v_t y_{t-2}}{\sigma_v^2} \\ \vdots \\ \frac{v_t y_{t-p}}{\sigma_v^2} \\ \left(\frac{v_t^2}{\sigma_v^2} - 1\right)\frac{1}{2\sigma_v^2} \end{bmatrix}. \tag{11}$$

Evaluating G at the QMLE $\{\widehat{\phi}_0, \widehat{\phi}_1, \cdots, \widehat{\phi}_p, \widehat{\sigma}_v^2\}$, which is also equivalent to the Klimko and Nelson (1978) CLS estimator, yields a null vector by construction.

The EMM estimator replaces the actual data (y_t) in (11) by the simulated data $(y_{s,t})$, evaluated at $\{\widehat{\phi}_0, \widehat{\phi}_1, \cdots, \widehat{\phi}_p, \widehat{\sigma}_v^2\}$. Letting the length of the simulated data be $N = HT$ where $H > 0$ is a constant which is chosen large enough to ensure reliable estimates, the gradient vector in (8) is obtained from (11) according to

$$G_s = \frac{1}{N}\sum_{t=p+1}^{N}\begin{bmatrix} \frac{\widehat{v}_{s,t}}{\widehat{\sigma}_v^2} \\ \frac{\widehat{v}_{s,t} y_{s,t-1}}{\widehat{\sigma}_v^2} \\ \frac{\widehat{v}_{s,t} y_{s,t-2}}{\widehat{\sigma}_v^2} \\ \vdots \\ \frac{\widehat{v}_{s,t} y_{s,t-p}}{\widehat{\sigma}_v^2} \\ \left(\frac{\widehat{v}_{s,t}^2}{\widehat{\sigma}_v^2} - 1\right)\frac{1}{2\widehat{\sigma}_v^2} \end{bmatrix}, \tag{12}$$

where

$$v_{s,t} = y_{s,t} - \phi_0 - \sum_{i=1}^{p}\phi_i y_{s,t-1}.$$

Finally, the weighting matrix in (8) is defined as (Gallant and Tauchen, 1996)

$$I = \frac{1}{T}\sum_{t=p+1}^{T}\begin{bmatrix} g_{1,t}^2 & g_{1,t}g_{2,t} & g_{1,t}g_{3,t} & \cdots & g_{1,t}g_{p+1,t} & g_{1,t}g_{p+2,t} \\ g_{2,t}g_{1,t} & g_{2,t}^2 & g_{2,t}g_{3,t} & \cdots & g_{2,t}g_{p+1,t} & g_{2,t}g_{p+2,t} \\ g_{3,t}g_{1,t} & g_{3,t}g_{2,t} & g_{3,t}^2 & \cdots & g_{3,t}g_{p+1,t} & g_{3,t}g_{p+2,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{p+1,t}g_{1,t} & g_{p+1,t}g_{2,t} & g_{p+1,t}g_{3,t} & \cdots & g_{p+1,t}^2 & g_{p+1,t}g_{p+2,t} \\ g_{p+2,t}g_{1,t} & g_{p+2,t}g_{2,t} & g_{p+2,t}g_{3,t} & \cdots & g_{p+2,t}g_{p+1,t} & g_{p+2,t}^2 \end{bmatrix}. \quad (13)$$

In applying the EMM estimator, two auxiliary models are used here. The first, Aux.1, uses the first $p+1$ moment conditions in (12), whereas the second, Aux.2, uses all $p+2$ of them. In the case of an INAR(1) model with parameters $\theta = \{\alpha_1, \lambda\}$, the first auxiliary model with $p = 1$ amounts to estimating an AR(1) regression with an intercept. The parameter vector $\theta = \{\alpha_1, \lambda\}$ is just identified, with λ being identified by the first moment condition in (12) and α_1 being identified by the second moment condition. From the properties of the INAR(1) model and the relationship between the intercept and the disturbance variance λ in particular, the inclusion of the last moment condition in (12), provides additional information to identify λ. For the INMA(1) model, increasing the number of lags in the auxiliary model is expected to raise the efficiency of the EMM estimator, asymptotically, as the auxiliary model provides a better approximation to the true likelihood. However, in finite samples, the additional moments from increasing the lag length of the auxiliary model may provide less information than lower order lags resulting in some efficiency loss. This potential efficiency loss is investigated in the Monte Carlo experiments.

To implement the EMM algorithm the following steps are followed.

Step 1: Estimate the auxiliary model by least squares and generate the estimates $\widehat{\phi}_i, i = 0, 1, 2, \cdots, p$, $\widehat{\sigma}^2 = T^{-1}\sum_{t=p+1}^{T}\widehat{v}_t^2$, where $\widehat{v}_t = y_t - \widehat{\phi}_0 - \sum_{i=1}^{p}\widehat{\phi}_i y_{t-i}$, is the least squares residual.

Step 2: Simulate the true INARMA model for some starting parameter values $\theta = \theta_{(0)}$; see below for specific details. Let the simulated values be $y_{s,t}$.

Step 3: Evaluate the gradient vector in (12) and the weighting matrix in (13).

Step 4: The EMM estimator is given as the solution of (8).

3.2 Estimating Asymptotic Standard Errors

The EMM estimator does not require the specification of a likelihood function, which would be based on the conditional distribution of an observation, conditional on its past. For the models of specific interest in this paper, such a likelihood function is generally not available and so the routine evaluation of appropriate estimated standard errors is not possible. Since, with any statistical modelling procedure, it is important to have some device for assessing model and parameter uncertainty, a means of overcoming this difficulty must be found. An attractive approach that is available under fairly mild conditions is that of subsampling, discussed in detail in the work of Politis, Romano and Wolf (1999). As pointed out in that book, subsampling is available in certain situations where a bootstrap approach would be invalid and the requirements that must be met for it to provide a valid inference tool are mild, often requiring little more than stationarity of the underlying data generating process. The account presented here draws heavily on the book, which will henceforth be termed PRW, as will its authors.

The types of problems to which subsampling can be applied include: variance estimation; distribution function estimation; bias reduction; interval estimation; and hypothesis testing methods. Moreover, the methods can be applied in the context of both dependent and independent data. Suppose, as elsewhere in the paper, that $y_1, ..., y_T$ represents a sample of T observations on a stationary time series. Generally, any statistical quantity calculated will be based on all observations (apart, possibly, from end effects). Subsampling methods are based on making repeated computations of similar statistics based on subsamples of length B using the observations $y_i, ..., y_{i+B-1}$. Let there be $j = 1, ..., N_B$ such blocks used and suppose that a (scalar) parameter θ is estimated by $\widehat{\theta}_T$ using the full sample and by $\widehat{\theta}_{T,i,B}$ using the block of length B beginning at observation i. Suppose further, as applies in the cases used here for EMM, that we can approximate the distribution of $\sqrt{T}(\widehat{\theta}_T - \theta)$ by $N(\theta, \sigma^2_\infty)$, where σ^2_∞ is an unknown long-run variance to be estimated. Provided $B \to \infty$ with T and $B/T \to 0$ the distribution of $\sqrt{B}(\widehat{\theta}_{T,i,B} - \theta)$ is the same Gaussian distribution involving the unknown long-run variance. Then a suitable estimator of the variance of $\widehat{\theta}_T$ is given by

$$\widehat{Var}_{T,B}(\widehat{\theta}_T) = \frac{B}{TN_B} \sum_{j=1}^{N_B} (\widehat{\theta}_{T,B,j} - \widehat{\theta}_{T,B,.})^2,$$

where $\widehat{\theta}_{T,B,.} = N_B^{-1} \sum_{j=1}^{N_B} \widehat{\theta}_{T,B,j}$, compare PRW, eq. (3.40).

Various difficulties still need to be addressed including: what block length B to use; how many blocks N_B to use; and how to appropriately estimate the unknown long-run variance σ^2_∞, should it be required (which, as we shall see, it is). The first occurrence of variance estimation using subsampling is due to Carlstein (1986), who proposes non-overlapping blocks. Presuming for simplicity that T is an exact multiple of B, it would follow that $N_B = T/B$. But PRW, Section 3.8.2 show that, in the case of a sample mean at least, it is preferable to use the maximum available number of overlapping blocks of size B, *viz.* $N_B = T - B + 1$. This is because it can be shown that the variance estimator based upon subsampling using all available blocks is 50% more efficient, asymptotically, than the Carlstein estimator (see eq. (3.46) in PRW). We adopt the approach recommended by PRW here.

There remains the crucial choice of block length B. There is work, particularly relating to the sample mean, that indicates that B should be $O(T^{1/3})$ and that the asymptotic mean squared error of the estimated variance in this case is $T^{1/3}$ times a complex quantity depending on the long-run variance σ^2_∞ (see PRW eq. (9.4)). In the case of a more general statistic, not much is known about the choice of B. However, in their Remark 10.5.4, PRW provide evidence that a bias-corrected estimator is to be preferred and that its use 'comes with the added bonus of an easy way to estimate the optimal block size in practice' (PRW, page 240). They suggest using two block lengths, say b and B, $b < B$, and associated variance quantities $\widehat{Var}_{T,b}(\widehat{\theta}_T)$ and $\widehat{Var}_{T,B}(\widehat{\theta}_T)$ and plotting these points with a x-axis B^{-1} and determining the intercept in the associated linear relationship obtained by joining the two points as the long-run variance. This follows from the display equation at the top of page 238 in PRW, and the succeeding discussion. As a rule of thumb, PRW propose that b should be determined by looking at the correlogram of the data in question and taking the value of b to be one greater than the last 'significant' ordinate of the sample autocorrelation function with $B = 2b$. We experimented with this suggestion and found that it did not work particularly well; for instance it can be very sensitive to the choice of b and often leads to short block lengths. Moreover, it can lead to quite differing values of the long-run variance estimate obtained for minor variations in b

Of course, the long-run variance estimate can be determined by linearly combining more than two such variance estimates based upon differing block sizes B and estimating the intercept in a regression of variance estimate on the inverse block size. We found this suggestion works better in a case where the desired asymptotic result is known. For instance, one might experiment with a continuous Gaussian AR(1) specification where the desire is to estimate the true asymptotic variance of $\widehat{\alpha}$, the MLE of the true dependence parameter α for which it is well known that $\sqrt{T}(\widehat{\alpha} - \alpha) \xrightarrow{d} N(0, 1 - \alpha^2)$.

To remain within the spirit of integer modelling, we opted to work with a stationary Poisson INAR(1) data generating process. Evaluation of the MLE of $\theta = (\alpha, \lambda)$ and the associated asymptotic variance-covariance matrix is discussed elsewhere, see, for example, Al-Osh and Alzaid (1987). Choosing B too small leads to unreliably large values of the long-run variance, whilst choosing it too big leads to the opposite. We found, after some experimentation, that B, perhaps not surprisingly, needs to be larger the stronger is the dependence in any process and that, for $T = 100$, say, reliable long -run variance estimators are found by combining estimates based on B in a range $kT^{1/3}$ to $8kT^{1/3}$ with some suitable k. Since the count data used in the Applications section are taken to be stationary without near nonstationary behaviour by reference to the sample autocorrelation properties of the two data sets involved, we believe that similar choices made for block length with these applications will provide reliable estimates for asymptotic standard errors.

As a test of the approach to compute standard errors for the EMM estimates based on subsampling, a INAR(1) model is simulated for a sample of size $T = 100$ with parameter values of $\alpha_1 = 0.3$ and $\lambda = 3.5$. The maximum likelihood parameter estimates are $\widehat{\alpha}_1 = 0.217$ and $\widehat{\lambda} = 3.768$, with respective estimated standard errors computed as the square root of the diagonal elements of the inverse of the (negative) Hessian evaluated at the maximum likelihood estimates, given by $se\left(\widehat{\alpha}_1\right) = 0.086$ and $se(\widehat{\lambda}) = 0.454$. Using Aux.1 with one lag, the EMM parameter estimates are $\widehat{\alpha}_1 = 0.242$ and $\widehat{\lambda} = 3.647$. The standard errors of the EMM estimates are computed using 4 block lengths $B = \{8, 16, 32, 64\}$. For each block size B the maximum number of data subsamples is $T - B + 1$. The EMM estimate of $\theta = \{\alpha_1, \lambda\}$ is computed for each subsample using 100 searches, which, in turn, is used to compute an estimate of the variance of $\widehat{\theta}_{100,B}$. In the case of $B = 8$, this amounts to computing the EMM estimates $100 - 8 + 1 = 93$ times with the EMM objective evaluated in each case 100 times in performing the search procedure to minimize this objective function. The long-run variance σ^2_∞ is estimated as the intercept from a regression of the estimated variances corresponding to the four block sizes, on a constant and the regressor $\{1/8, 1/16, 1/32, 1/64\}$. The standard errors of $\widehat{\theta}_T$ are computed as $\widehat{\sigma}_\infty/\sqrt{T}$ where $\widehat{\sigma}^2_\infty$ is the estimate of σ^2_∞. The estimated standard errors are: $se\left(\widehat{\alpha}_1\right) = 0.088$; and $se(\widehat{\lambda}) = 0.489$, which agree closely with the MLE standard errors.

3.3 Finite Sample Properties

The finite sample properties of the EMM parameter estimator are investigated using a range of Monte Carlo experiments. Where possible, the sampling properties of the EMM estimator are compared with those obtained from existing estimators, including MLE and CLS (Klimko and Nelson, 1978).

3.4 Experimental Design

The Monte Carlo experiments are based on the INARMA(2,1) model with Poisson innovations

$$\begin{aligned} x_t &= \alpha_1 \circ x_{t-1} + \alpha_2 \circ x_{t-2} + u_t \\ y_t &= x_{t-1} + \beta_1 \circ u_t \\ u_t &\sim P_0(\lambda), \end{aligned} \quad (14)$$

where the three thinning operators in (14) are assumed to be independent following the approach of Du and Li (1991). Four broad DGPs are specified; they are given in Table 1 and all arise from (14) as special cases. The INAR(1) specification serves as a benchmark model, since it enables the EMM estimator to be compared directly to the finite sample properties of the MLE; the results of Jung, Ronning and Tremayne (2005) are used to facilitate the comparison. Here special attention is given to how the performance of the EMM estimator varies according to the lag length of the auxiliary model. A comparison of the sampling performance of the EMM and CLS estimators of the INMA(1) model determines whether there are any efficiency gains in finite samples from adopting the EMM estimator over the CLS one. The INARMA(1,1) model possibly represents a more interesting DGP, as this experiment determines the ability of the EMM estimator to unravel the two types of dynamic components of the model from an auxiliary model that is simply based on an AR(p) regression. The last model specification entertained is an INAR(2) to determine the properties of the EMM estimator in the case of higher order dynamics and results can be compared with those presented by Bu, Hadri and McCabe (2008).

The selected parameterizations in Table 1 are based on DGPs that yield relatively "low" counts, as well as parameterizations adopted in previous studies to aid comparisons. In the case of the INAR(1) model the expected count is

$$E[y_t] = \frac{\lambda}{1-\alpha_1} = \frac{3.5}{1-0.3} = 5.$$

For the INMA(1) model the expected number is

$$E[y_t] = (1+\beta_1)\lambda = (1+0.3) \times 3.5 = 4.55,$$

whereas for the INARMA(1,1) it is

$$E[y_t] = \frac{(1+\beta_1 - \alpha_1\beta_1)\lambda}{1-\alpha_1} = \frac{(1+0.7-0.21)\,3.5}{1-0.3} = 7.45.$$

Finally, for the INAR(2) model the expected number of counts is

$$E[y_t] = \frac{1}{1-\alpha_1-\alpha_2} = \frac{1}{1-0.5-0.3} = 5.0.$$

Table 1:

Choice of DGPs and parameterizations for the Monte Carlo experiments.

Experiment	Type	Parameterization			
		α_1	α_2	λ	β_1
I	INAR(1)	0.3	0.0	3.5	0.0
II	INMA(1)	0.0	0.0	3.5	0.3
III	INARMA(1,1)	0.3	0.0	3.5	0.7
IV	INAR(2)	0.5	0.3	1.0	0.0

In implementing the EMM estimator the two auxiliary models mentioned below (13) are used. Aux.1 involves the first $p+1$ moment conditions in (12) and the maximum lag chosen in $p = 3$. Aux.2 augments Aux.1 with the last element in (12). As the parameter λ is identified by the intercept in both cases and also by the error variance in Aux.2, a comparison of the EMM simulation results based on the two auxiliary models will provide insights into the potential finite sample gains from using both types of identification mechanism.

The MLE is computed using the gradient algorithm MAXLIK in GAUSS Version 10. A numerical differentiation routine is adopted to compute the gradients and the Hessian. As there is no analytical expression for the log-likelihood function of integer models with moving average components, the MLE results are just reported for INAR specifications.

The CLS estimator is computed for the INAR(1) and INMA(1) models, with the latter estimates computed using the estimates of the INAR(1) model by virtue of the fact that they have CLAR(1) conditional mean functions given by (2). This estimator simply involves estimating the AR(1) regression equation

$$y_t = \phi_0 + \phi_1 y_{t-1} + v_t,$$

by least squares with the estimates of λ and α_1 given by the estimates of ϕ_0 and ϕ_1 respectively. The corresponding INMA(1) CLS parameter estimators are obtained as $\phi_1/(1-\phi_1)$ in the case of the moving average parameter β_1, and λ in the case of the error variance parameter.

When computing the EMM estimator, a range of simulation runs were initially experimented with. A value of $H = 500$ was settled upon resulting in simulation runs of length $N = 500T$ in (12) where T is the sample size given by $T = \{50, 100, 200\}$. As the objective function of the EMM estimator $Q(\theta)$ (see (8) above) is not continuous with respect to the parameter vector θ, a grid search is used to minimize the objective function, with the number of searches set equal to 100 for the INAR(1) and INMA(1) models, and 200 for the INARMA(1,1) and INAR(2) models. There do exist more sophisticated search algorithms, including the accelerated search method of Appel, Labarre and Radulovic (2003). In those cases where the results of the EMM estimator can be compared with the MLE, the results suggest that the simple search procedure adopted in the Monte Carlo experiments reported here suffices. Starting values for the EMM algorithm are based on the CLS estimates. In those cases where the CLS estimator yields parameter estimates that violate the stationarity restriction, the true parameter values are chosen as the starting parameter estimates instead. Finally, all Monte Carlo experiments are based on 5000 replications.

3.5 INAR(1) Results

The results of the INAR(1) Monte Carlo experiment are given in Table 2. Sampling statistics based on the mean and the root mean squared error, RMSE, from the 5000 replications are presented. The EMM estimator is computed for six types of auxiliary models: the first three are based on an AR(p) regression with $p = 1, 2, 3$ (Aux.1), and the second three use Aux.2. For comparison the results for MLE and CLS estimator are also reported.

Inspection of the top block of Table 2 shows that all estimators of the autoregressive parameter α_1 are biased downwards, with the size of the bias decreasing as the sample size increases. In the case of MLE and CLS the finite sample results are comparable to those reported in Jung, Ronning and Tremayne (2005). In general the bias of the EMM estimator lies between the MLE (smallest) and the CLS estimator (largest). A consideration of the bias of the EMM estimator for $T = 50$ shows that it decreases as the lag length of the auxiliary model increases with nearly the same level of bias being achieved as the MLE with Aux.2 and $p = 3$. The EMM estimator does no worse than the CLS estimator in the case of Aux.2 with $p = 1$, whereas, for all other auxiliary model specifications, the EMM estimator achieves lower bias than the CLS estimator.

The results in Table 2 demonstrate that the EMM corrects some of the small sample bias of the CLS estimator, even though the EMM estimator uses the CLS framework as an auxiliary model. It can be seen that Aux.1 performs better than Aux.2 for $p = 1$, but the reduction in bias as the number of lags increases is far greater for Aux.2 than it is for Aux.1, *i.e.* the bias quickly becomes relatively lower for Aux.2 when $p > 1$. These results remain qualitatively unchanged for samples of size $T = 100$. However, for larger samples of $T = 200$ the EMM and CLS estimators yield similar biases, with the EMM estimator based on Aux.1 now performing marginally better than it does using

Aux.2.

A comparison of the performance of the three estimators in terms of RMSE shows that the EMM estimator nearly achieves the same level of efficiency as the MLE for a lag of $p = 1$, and, in fact, matches the finite sample efficiency of the MLE if Aux.1 is adopted. In contrast to the performance of the EMM estimator in terms of bias, smaller lags in the auxiliary model produce better results in terms of higher efficiency. In fact, increasing the number of lags of the auxiliary model produces RMSEs that are actually higher than the RMSE of the CLS estimator. This trade-off between bias and efficiency is typical of the finite sample results of most of the models investigated here. Moreover, the additional moment condition used in Aux.2 yields slightly higher RMSEs, suggesting that this moment condition may be partly redundant in terms of identifying the parameters of the model.

Turning to the bottom block of Table 2 it is seen that all estimators of λ are biased upwards, with the size of the bias decreasing as the sample size increases. This is unsurprising given the high negative correlation between estimators of the two parameters. Again, in the case of the MLE and the CLS estimator, the finite sample results are very similar to the results reported in Jung, Ronning and Tremayne (2005). As with the results for α_1, the EMM estimator for $T = 50$, performs best using an auxiliary model with $p = 3$ lags. Interestingly, for Aux.1 and more especially for Aux.2, the EMM estimator yields a lower finite sample bias than the MLE. This result is similar to the finite sample results reported by Gouriéroux, Monfort and Renault (1993), where their simulation based estimator, known as indirect inference, also exhibited better finite sample properties than the MLE. Gouriéroux and Monfort (1994) show that this result reflects the property that the simulation estimator acts as a bootstrap estimator which corrects the second order bias of the MLE.

The RMSE results for λ mirror the RMSE results for α_1, with the EMM estimator yielding relatively lower RMSEs for shorter lag structures in the auxiliary model. In the case of $T = 50$, the EMM estimator again yields an even smaller RMSE than the MLE when $p = 1$ and either of the two auxiliary models is chosen. Increasing the sample size to $T = 100$ and 200 makes little difference in the efficiency of the CLS and EMM estimators, especially where the latter estimator is computed using an auxiliary model with $p = 1$.

3.6 INMA(1) Results

The INMA(1) results given in Table 3 contain just the CLS and EMM results. The EMM estimator of the moving average parameter β_1 exhibits smaller bias than the CLS

Table 2:

Finite sample properties of alternative estimators of the parameters of the INAR(1) model. Population parameters are $\alpha_1 = 0.3$ and $\lambda = 3.5$. The number of draws is 5000.

Estimator			Mean (α_1)			RMSE (α_1)		
		$T =$	50	100	200	50	100	200
MLE			0.281	0.288	0.295	0.133	0.095	0.063
CLS			0.260	0.280	0.291	0.144	0.098	0.069
EMM	Aux.1	Lag = 1	0.265	0.279	0.291	0.133	0.098	0.069
	Aux.1	Lag = 2	0.268	0.280	0.293	0.140	0.102	0.070
	Aux.1	Lag = 3	0.274	0.283	0.293	0.149	0.106	0.071
	Aux.2	Lag = 1	0.260	0.278	0.291	0.135	0.098	0.068
	Aux.2	Lag = 2	0.272	0.284	0.291	0.139	0.101	0.070
	Aux.2	Lag = 3	0.280	0.287	0.292	0.150	0.107	0.071
			Mean (λ)			RMSE (λ)		
		$T =$	50	100	200	50	100	200
MLE			3.587	3.559	3.521	0.713	0.500	0.331
CLS			3.693	3.596	3.543	0.766	0.522	0.361
EMM	Aux.1	Lag = 1	3.671	3.605	3.544	0.705	0.525	0.362
	Aux.1	Lag = 2	3.634	3.589	3.533	0.736	0.532	0.366
	Aux.1	Lag = 3	3.584	3.560	3.532	0.772	0.547	0.369
	Aux.2	Lag = 1	3.681	3.605	3.543	0.722	0.517	0.359
	Aux.2	Lag = 2	3.606	3.564	3.538	0.735	0.530	0.367
	Aux.2	Lag = 3	3.518	3.536	3.525	0.774	0.555	0.374

Table 3:

Finite sample properties of alternative estimators of the parameters of the INMA(1) model. Population parameters are $\beta_1 = 0.3$ and $\lambda = 3.5$. The number of draws is 5000.

Estimator			Mean (β_1)			RMSE (β_1)		
		$T =$	50	100	200	50	100	200
MLE			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS			0.278	0.290	0.292	0.226	0.156	0.111
EMM	Aux. 1	Lag = 1	0.288	0.288	0.292	0.217	0.175	0.137
	Aux. 1	Lag = 2	0.280	0.274	0.288	0.206	0.160	0.126
	Aux. 1	Lag = 3	0.283	0.280	0.288	0.204	0.157	0.123
	Aux. 2	Lag = 1	0.284	0.290	0.291	0.219	0.174	0.132
	Aux. 2	Lag = 2	0.298	0.286	0.288	0.213	0.163	0.123
	Aux. 2	Lag = 3	0.301	0.293	0.288	0.215	0.161	0.122
			Mean (λ)			RMSE (λ)		
		$T =$	50	100	200	50	100	200
MLE			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS			3.661	3.577	3.547	0.676	0.463	0.327
EMM	Aux. 1	Lag = 1	3.609	3.576	3.531	0.621	0.482	0.364
	Aux. 1	Lag = 2	3.597	3.585	3.526	0.620	0.472	0.345
	Aux. 1	Lag = 3	3.564	3.562	3.516	0.618	0.472	0.344
	Aux. 2	Lag = 1	3.609	3.556	3.521	0.633	0.481	0.353
	Aux. 2	Lag = 2	3.531	3.541	3.516	0.618	0.479	0.344
	Aux. 2	Lag = 3	3.492	3.510	3.498	0.634	0.466	0.342

estimator, with the size of the bias decreasing as the number of lags in the auxiliary model increases. This is especially true for the smallest sample size investigated, namely $T = 50$. For larger samples, the relative bias of the two estimators diminishes. The EMM estimator also dominates the CLS estimator when $T = 50$ in terms of RMSE, but with the roles reversing as T increases.

The EMM estimator of λ generally exhibits smaller bias than the CLS estimator. The EMM estimator also exhibits superior finite sample efficiency than the CLS estimator for samples of size $T = 50$. As with the β_1 RMSE results, the CLS estimator of λ exhibits slightly better efficiency than the EMM estimator for larger sample sizes.

3.7 INARMA(1,1) Results

The finite sample results of the EMM estimator for the INARMA(1,1) model are reported in Table 4. The auxiliary model of the EMM estimator must be based on a minimum of $p = 2$ lags for identification. Inspection of the results shows that choosing Aux.2 with $p = 3$ lags tends to perform the best for all sample sizes in terms of smaller bias and lower RMSE. The percentage bias of $\widehat{\alpha}_1$ for this model for samples of size $T = 50$, is just $100 \times 0.006/0.3 = 2\%$. Similar results occur for $\widehat{\lambda}$ where the percentage bias is just -0.429%. The EMM estimator of the moving average parameter β_1 is biased downwards, with a percentage bias of -11.429%, which reduces to -7.0%, for $T = 200$. Given the fairly large bias in estimating the moving average parameter with a sample size as small as 50, it may not be advisable to use the procedure with samples as small as this, though increasing sample size, quite plausibly, attenuates the problem.

A comparison of the RMSE for the alternative auxiliary models used to calculate the EMM estimator suggests that, for smaller samples of size $T = 50$, choosing a shorter lag length in the auxiliary model yields marginally more efficient parameter estimates, whereas the opposite is true for larger samples of size $T = 200$.

3.8 INAR(2) Results

The results of the EMM estimator for the INAR(2) model are reported in Table 5. The parameter constellation $(\alpha_1, \alpha_2) = (0.5, 0.3)$ was chosen on the basis of the evidence in Bu, Hadri and McCabe (2008, Table 1) that the asymptotic relative efficiency of CLS to MLE is only 0.7, which means that the former is some 30% less efficient, asymptotically, than the latter and thus there may be discernible differences, perhaps between all three types of estimators. To enable these EMM results to be compared to the MLE and CLS Monte Carlo results reported in Bu, Hadri and McCabe (2008), the experiments are based on a sample of size $T = 100$ only. The EMM estimator is based on Aux.1 with $p = 2$ lags, the minimum required for identification.

The simulation results for α_1 in Table 5 show that the bias is marginally smaller for the CLS estimator than it is for the EMM estimator, whereas EMM displays slightly smaller RMSE. In the case of α_2, EMM yields lower bias and smaller RMSE than CLS. In fact, EMM as well as CLS actually display slightly better finite sample efficiency than MLE, whilst the bias of the MLE of α_2 is marginally smaller than it is for EMM. For the parameter λ, MLE dominates all estimators in terms of bias and RMSE whilst, in turn, EMM dominates CLS.

Table 4:

Finite sample properties of alternative estimators of the parameters of the INARMA(1,1) model. Population parameters are $\alpha_1 = 0.3$, $\lambda = 3.5$ and $\beta_1 = 0.7$. The number of draws is 5000.

Estimator			Mean (α_1)			RMSE (α_1)		
		$T =$	50	100	200	50	100	200
MLE			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
EMM	Aux.1	Lag = 2	0.271	0.285	0.288	0.185	0.165	0.145
	Aux.1	Lag = 3	0.287	0.298	0.303	0.187	0.162	0.142
	Aux.2	Lag = 2	0.280	0.287	0.290	0.183	0.162	0.145
	Aux.2	Lag = 3	0.306	0.309	0.306	0.190	0.162	0.140
			Mean (λ)			RMSE (λ)		
		$T =$	50	100	200	50	100	200
MLE			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
EMM	Aux.1	Lag = 2	3.706	3.618	3.565	0.766	0.581	0.456
	Aux.1	Lag = 3	3.669	3.597	3.541	0.813	0.596	0.470
	Aux.2	Lag = 2	3.561	3.537	3.519	0.673	0.508	0.405
	Aux.2	Lag = 3	3.485	3.487	3.483	0.744	0.545	0.417
			Mean (β_1)			RMSE (β_1)		
		$T =$	50	100	200	50	100	200
MLE			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS			n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
EMM	Aux. 1	Lag = 2	0.603	0.625	0.658	0.302	0.272	0.223
	Aux. 1	Lag = 3	0.579	0.604	0.636	0.301	0.270	0.228
	Aux. 2	Lag = 2	0.642	0.653	0.663	0.276	0.254	0.232
	Aux. 2	Lag = 3	0.620	0.629	0.651	0.278	0.260	0.218

Table 5:

Finite sample properties of alternative estimators of the parameters of the INAR(2) model, $T = 100$ in all cases. Population parameters are $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, and $\lambda = 1.0$. The number of draws used in the computation of the EMM estimator is 5000. Monte Carlo results for the MLE and CLS estimator are obtained from Bu, Hadri and McCabe (2008) which are based on 1000 replications. The EMM estimator is based on Aux.1 with $p = 2$.

Estimator	Mean (α_1)	RMSE (α_1)
MLE	0.498	0.092
CLS	0.481	0.109
EMM	0.477	0.106
	Mean (α_2)	RMSE (α_2)
MLE	0.271	0.111
CLS	0.263	0.109
EMM	0.266	0.106
	Mean (λ)	RMSE (λ)
MLE	1.142	0.401
CLS	1.270	0.549
EMM	1.262	0.515

4 Applications

The EMM estimator is now applied to estimating INARMA models for two data sets. The first stems from the work of Fürth (1918). This is a set of observations of counts taken every 5 seconds of the number of pedestrians present on a city block; these data have been considered by, *inter alia*, Mills and Seneta (1989) and Jung and Tremayne (2006b) but no fully satisfactory time series model seems to have been unearthed because of the complex nature of the autocorrelation structure of these data. The second consists of observations of internet download counts every 2 minutes introduced by Weiß (2007, 2008). For comparison, MLE and CLS estimation results are also reported, where available. For both data sets, the innovation term is envisioned as being Poisson, thereby implying that they are equi-dispersed and this condition is exploited in Aux.2.

4.1 Fürth Data

The Fürth data are depicted in Figure 1. The total number of observations is $T = 505$, with a minimum of 0 and a maximum of 7 pedestrians observed in each 5 second time period. The mean is 1.592 and the variance is 1.508, suggesting equi-dispersion and conforming with the assumption of Poisson innovations. Inspection of the sample ACF and PACF of the Fürth data in Figure 2 suggests that there are higher order dynamics that will not be captured by either an INAR(1) model or an INMA(1) specification. The PACF displays spikes at lags 1 and 2 with positive and negative values respectively, with the ACF displaying a cyclical pattern for higher order lags with a 30 second cycle with peaks at lag 6 (30 seconds), lag 12 (1 minute) etc.

The *prima facie* difficulty of specifying an INMA(1) model is highlighted by the fact that the first sample autocorrelation coefficient is $r_1 = 0.665$ from Figure 2, which violates the restriction $\rho_1 < 0.5$ needed for both continuous MA(1) and INMA(1) specifications to be appropriate. Even allowing for a higher order lag by way of an INAR(2) specification still incurs problems as the first two ordinates of the ACF are $r_1 = 0.665$ and $r_2 = 0.323$, thereby failing to satisfy the restriction $r_2 - r_1^2 > 0$ that is needed for estimated AR parameters (based on autocorrelations, at any rate) to satisfy $\widehat{\alpha}_i \in [0, 1)$ (see Jung and Tremayne, 2006b, eq. 4.5 for more details).

To circumvent the problems inherent in specifying either pure INAR or INMA models, Table 6 gives the EMM parameter estimates for a range of INARMA models for the Fürth data. The most general specification is based on the following INARMA(12,1)

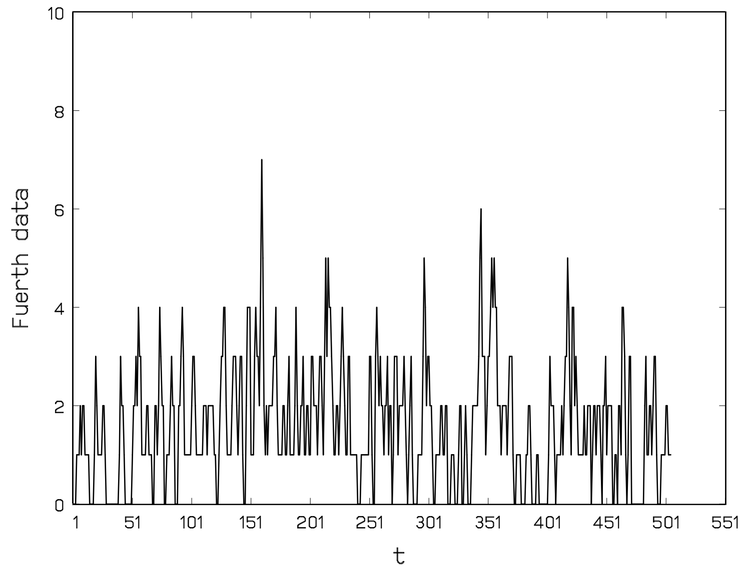

Figure 1: Fürth data representing the number of pedestrians on a city block observed every 5 seconds, $T = 505$.

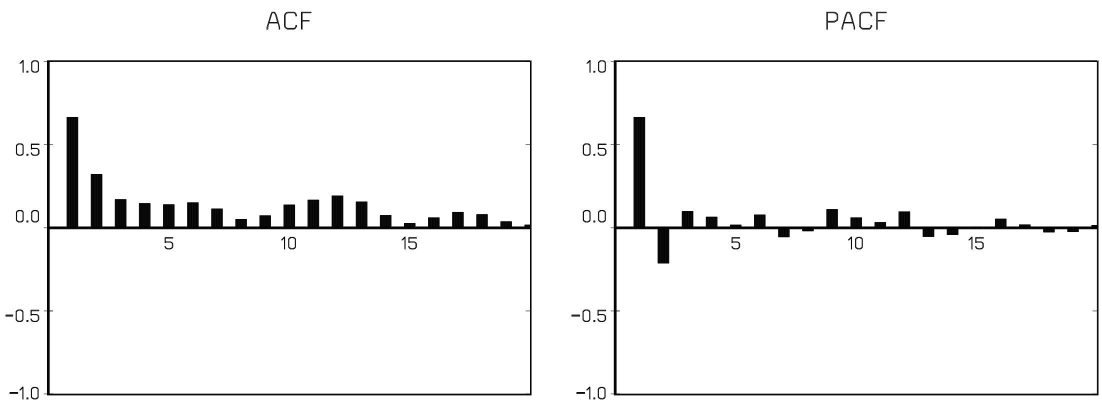

Figure 2: Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the Fürth data. Lags expressed in 5-second intervals.

model

$$
\begin{aligned}
x_t &= \alpha_1 \circ x_{t-1} + \alpha_6 \circ x_{t-6} + \alpha_{12} \circ x_{t-12} + u_t \\
y_t &= x_{t-1} + \beta_1 \circ u_t \\
u_t &\sim Po(\lambda) .
\end{aligned}
\tag{15}
$$

The choice of lags at 6 and 12 is based on the higher order lags observed in the ACF and PACF given in Figure 2, although the t-statistic associated with the 12th lag reported in Table 6 is statistically insignificant. Some intermediate lags at 2 and 5, where also tried but ultimately discarded, as they did not contribute significantly to the estimated model. For comparison, the results from estimating lower order models based on the INARMA(1,1), INAR(1) and INMA(1) models are also given; in the case of the INAR(1) model, the MLE and CLS estimates are also presented. The CLS estimates are not reported for the INMA(1) model as the estimates do not satisfy the restriction $0 \leq \beta_1 \leq 1$. Following the approach outlined in Section 3.2, in computing the estimated asymptotic standard errors of the EMM parameter estimates, the number of searches used to compute the EMM estimates for each block size is set at 100. The block sizes used for all model specifications are $B = \{32, 64, 128\}$ which yield three estimates of the variance corresponding to each of the EMM parameter estimates for each model specification. A regression of the three estimated variances on a constant and the regressor $\{1/32, 1/64, 1/128\}$, produces an estimate of the long-run variance σ^2_∞ based on the intercept. Dividing the long-run estimate by the sample size T and taking the square root of this expression yields the pertinent estimated asymptotic standard error.

An important feature of the parameter estimates of the INARMA(12,1) model, as well as the INARMA(1,1) and INMA(1) models, is that the estimate of the moving average parameter, β_1, is always above 0.9, with estimates that are often larger than 0.95. For the most general model the estimate equals the upper bound of unity. In point of fact, this may not be surprising as a careful read of the original paper of Fürth (1918, p.422) shows. In his syntax "Das Intervall [5 sec.] war so gewählt, daß ein in einem bestimmten Intervall eintretender Mensch mit grosser Wahrscheinlichkeit auch im nächsten Intervall sich noch in dem Beobachtungsraum aufhielt" he indicates that observations were made at intervals of length 5 seconds so that a person entering the observation site in a given interval will, with high probability, remain there in the next interval. Hence, a new arrival (birth, or innovation) entering the count at time t has very little chance of being thinned at time $t + 1$, thereby suggesting a moving average parameter value of $\beta_1 = 1$.

The parameter estimates in Table 6 also show the gains from estimating the general specification of the INARMA(12,1) model as it yields the smallest estimate of the residual variance of $\widehat{\lambda} = 0.395$, across all models. Excluding the higher order lags by setting $\alpha_6 = \alpha_{12} = 0$ in (15), results in an INARMA(1,1) model which yields higher estimates of the residual variance, with values at around 0.473 in general. Imposing the additional restriction of $\beta_1 = 0$ and fitting an INAR(1) results in even higher estimates of λ, with the smallest estimate, based on MLE, being 0.5.

To gauge the ability of the alternative models to capture the dynamics in the data, the ACF and PACF corresponding to the estimated models are given in Tables 7 and 8 respectively based on EMM parameter estimates in Table 6. The autocorrelations and partial autocorrelations of each model are computed by simulating that model for $1,000,000$ observations and computing the ACF and PACF of the simulated data. For the simple model specifications, such as the INAR(1) and INMA(1) models where analytical expressions of the ACF and PACF are easily derived, the simulated estimates are consistent with the theoretical values.

The INARMA(12,1) model captures the dynamics well for the short lags as well as longer lags corresponding to the 30 second cycle. These statistics also demonstrate the failure of the INAR(1) and the INMA(1) models to capture the dynamic structure of the Fürth data. The INAR(1) model does a relatively poor job of capturing lags of the ACF after the first. Unsurprisingly, it models the first lag of the PACF but not the second and higher order lags of the PACF, while the INMA(1) yields the opposite results. By comparison, the advantages in combining the autoregressive and moving average structures in the case of the INARMA(1,1) model are evident as this model is able to model the first few lags of both the ACF and PACF reasonably successfully.

The preferred model is the one with parameter estimates given in the top row of Table 6 and estimated SACF and SPACF in the third columns of Tables 7 and 8, respectively. The adequacy of this model is further investigated by using the parametric resampling method suggested by Tsay (1992), who argues that a model may be deemed adequate if it can successfully reproduce important features of the data; we opt to use the first 20 observed SACF and SPACF ordinates of the Fürth data series given in the leftmost columns of Tables 7 and 8, or, alternatively, in Figure 2. To implement the idea, we generate 5000 artificial data sets of length $T = 505$ using the preferred fitted model and *iid* Poisson innovations to yield 5000 SACFs and SPACFs. For each ordinate of these the $100(1 - \alpha/2)\%$ and $100\alpha/2\%$ quantiles are used as acceptance bounds. The model is adjudged to fit any particular feature adequately if the actual value lies within the acceptance bounds. The results of this exercise for the Fürth data

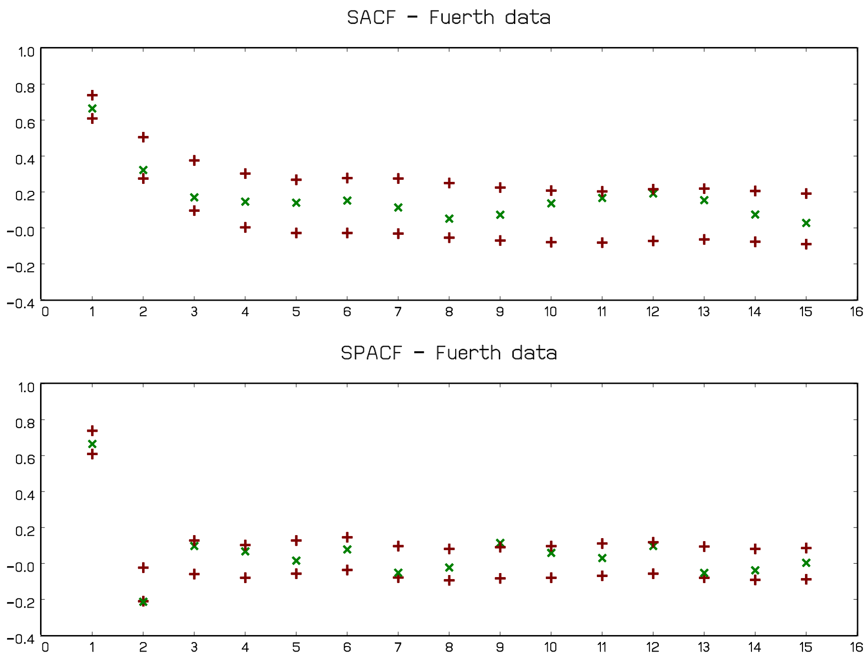

Figure 3: Estimated sample ACF and PACF of the INARMA(12,1) model with $\beta_1 = 1$, for the Fürth data. The 95% confidence bounds are based on a parametric bootstrap with 5000 replications.

are presented in Figure 3 using $\alpha = 0.05$. It is seen that the fitted model faithfully reproduces the main features of the sample autocorrelation properties of the data well.

4.2 Downloads

Figure 4 gives the number of downloads of IP addresses registered every 2 minutes during a working day, 10am to 6pm, on the 29 of November 2005, at a computer at the University of Würzburg, Germany. The total number of observations is $T = 241$, with a range from 0 to a maximum of 8 downloads every 2 minutes. The sample mean and variance are 1.315 and 1.392, respectively, suggesting little evidence of overdispersion. The ACF and PACF given in Figure 5 both display spikes at lag 1, suggesting the possibility of an INAR(1) model, though the ACF has no significant higher order ACF ordinates, or, perhaps, an INMA(1) though the PACF has no discernible higher order ordinates. One might also consider an INARMA(1,1) model and even, possibly, higher order lag effects in a mixed model.

Parameter estimates of various INARMA models of the downloads data are presented in Table 9. The standard errors of the EMM parameter estimates are computed in the same way that they are for the Fürth data with the exception that the number

Table 6:

Estimates of INARMA models for the Fürth data. Standard errors in parentheses.

Model	Estimator			α_1	α_6	α_{12}	λ	β_1
INARMA(12,1)	EMM	Aux.1		0.575	0.073	0.043	0.395	1.000
				(0.080)	(0.042)	(0.031)	(0.085)	
INARMA(1,1)	EMM	Aux.1	Lag = 2	0.504			0.473	0.955
				(0.085)			(0.098)	(0.134)
			Lag = 3	0.354			0.642	0.987
				(0.086)			(0.105)	(0.120)
		Aux.2	Lag = 2	0.504			0.473	0.955
				(0.061)			(0.067)	(0.104)
			Lag = 3	0.504			0.473	0.955
				(0.058)			(0.058)	(0.149)
INAR(1)	EMM	Aux.1	Lag = 1	0.672			0.543	
				(0.061)			(0.179)	
			Lag = 2	0.688			0.533	
				(0.030)			(0.050)	
			Lag = 3	0.699			0.534	
				(0.036)			(0.061)	
		Aux.2	Lag = 1	0.628			0.540	
				(0.016)			(0.035)	
			Lag = 2	0.689			0.484	
				(0.020)			(0.040)	
			Lag = 3	0.670			0.516	
				(0.021)			(0.044)	
INAR(1)	MLE			0.688			0.500	
				(0.023)			(0.041)	
INAR(1)	CLS			0.665			0.536	
INMA(1)	EMM	Aux.1	Lag = 1				0.636	0.992
							(0.156)	(0.072)
			Lag = 2				0.600	0.951
							(0.174)	(0.075)
			Lag = 3				0.631	0.910
							(0.226)	(0.109)
		Aux.2	Lag = 1				0.585	0.988
							(0.156)	(0.055)
			Lag = 2				0.646	0.999
							(0.214)	(0.159)
			Lag = 3				0.662	0.916
							(0.198)	(0.156)

Table 7:

Estimated ACF of alternative models for the Fürth data, based on EMM parameter estimates in Table 6. The autocorrelations are computed by simulating each model for $1,000,000$ observations and computing the ACF of the simulated data.

Lag (5 sec.)	Actual	INARMA(12,1) ($0 < \beta_1 < 1$)	INARMA(12,1) ($\beta_1 = 1$)	INARMA(1,1)	INAR(1)	INMA(1)
1	0.665	0.666	0.684	0.663	0.671	0.479
2	0.323	0.396	0.404	0.334	0.449	0.001
3	0.170	0.244	0.249	0.167	0.302	0.000
4	0.147	0.165	0.167	0.083	0.202	-0.001
5	0.141	0.135	0.134	0.040	0.135	-0.003
6	0.152	0.144	0.139	0.019	0.090	-0.003
7	0.114	0.138	0.135	0.010	0.061	-0.002
8	0.052	0.116	0.113	0.007	0.041	-0.001
9	0.073	0.094	0.093	0.005	0.028	0.001
10	0.138	0.081	0.079	0.004	0.019	0.004
11	0.168	0.079	0.076	0.003	0.013	0.003
12	0.194	0.091	0.087	0.002	0.010	0.001
13	0.157	0.092	0.089	0.001	0.008	0.000
14	0.076	0.079	0.076	0.001	0.006	0.000
15	0.028	0.063	0.061	0.001	0.005	0.000
16	0.061	0.049	0.049	0.002	0.004	0.000
17	0.095	0.040	0.040	0.002	0.003	0.001
18	0.079	0.037	0.036	0.003	0.004	0.002
19	0.038	0.035	0.034	0.004	0.005	0.002
20	0.016	0.033	0.031	0.004	0.005	0.003

Table 8:

Estimated PACF of alternative models for the Fürth data, based on EMM parameter estimates in Table 6. The partial autocorrelations are computed by simulating each model for 1, 000, 000 observations and computing the PACF of the simulated data.

Lag (5 sec.)	Actual	INARMA(12,1) $(0 < \beta_1 < 1)$	INARMA(12,1) $(\beta_1 = 1)$	INARMA(1,1)	INAR(1)	INMA(1)
1	0.665	0.666	0.684	0.663	0.671	0.479
2	-0.213	-0.087	-0.118	-0.189	-0.001	-0.296
3	0.099	0.027	0.040	0.055	0.001	0.201
4	0.066	0.021	0.016	-0.018	-0.001	-0.145
5	0.017	0.040	0.037	0.003	-0.001	0.101
6	0.079	0.065	0.061	-0.001	0.001	-0.076
7	-0.054	0.010	0.012	0.002	0.001	0.055
8	-0.018	0.002	-0.001	0.001	0.000	-0.041
9	0.111	0.009	0.011	0.001	0.000	0.033
10	0.061	0.015	0.013	0.001	0.000	-0.020
11	0.033	0.021	0.021	-0.001	0.001	0.016
12	0.097	0.037	0.034	0.000	0.001	-0.011
13	-0.053	0.004	0.004	0.000	0.001	0.008
14	-0.040	-0.001	-0.003	0.000	0.000	-0.006
15	0.006	0.001	0.003	0.000	0.001	0.003
16	0.053	0.000	0.000	0.001	-0.001	-0.003
17	0.019	0.002	0.002	0.000	0.002	0.004
18	-0.025	0.003	0.003	0.002	0.001	-0.001
19	-0.024	0.002	0.001	0.001	0.002	0.002
20	0.016	0.001	0.001	0.001	0.000	0.001

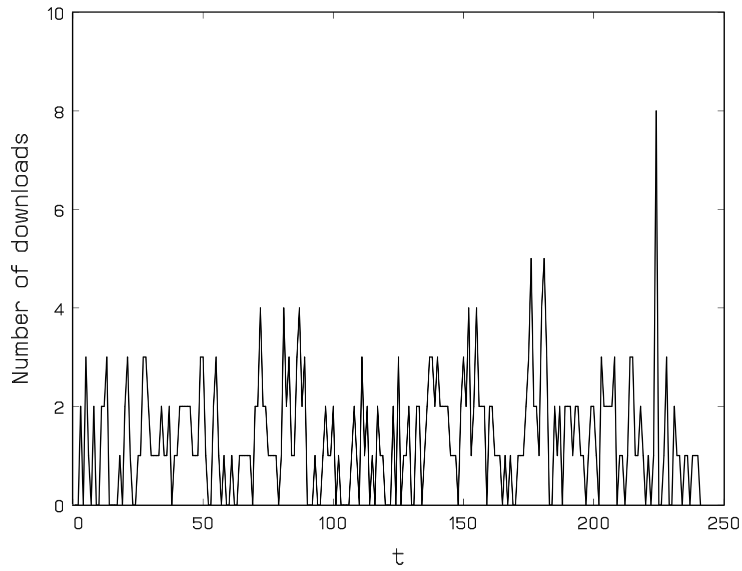

Figure 4: The number of downloads of web addresses every 2 minutes during a working day, on the 29 of November 2005, at a computer at the University of Würzburg, Germany, $T = 241$. We thank Christian Weiß for providing us the data.

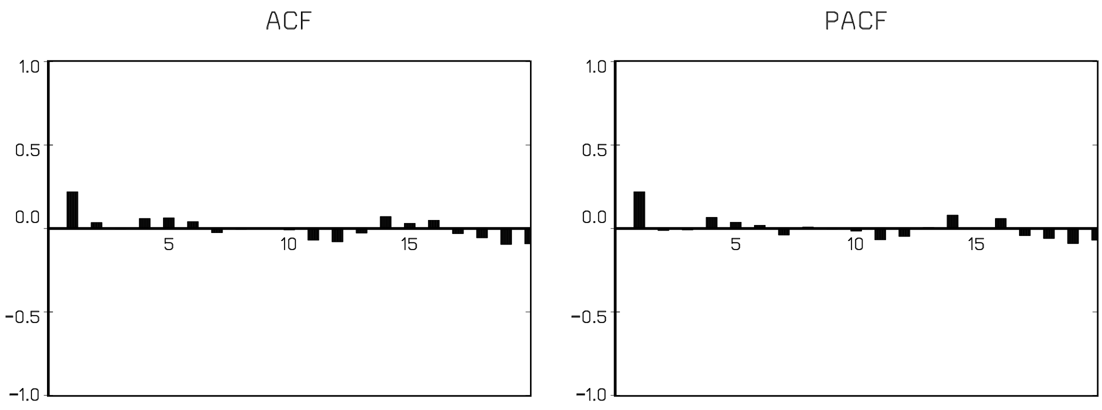

Figure 5: Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of downloads data. Lags expressed in 2 minute intervals.

Table 9:

Estimates of INARMA models for the downloads data. Standard errors in parentheses.

Model	Estimator			α_1	λ	β_1
INAR(1)	MLE			0.236	1.009	
				(0.063)	(0.100)	
INAR(1)	CLS			0.221	1.029	
INMA(1)	CLS				1.029	0.283
INAR(1)	EMM	Aux.1	Lag = 1	0.264	1.030	
				(0.073)	(0.148)	
			Lag = 2	0.198	1.134	
				(0.085)	(0.166)	
			Lag = 3	0.222	1.008	
				(0.112)	(0.211)	
		Aux.2	Lag = 1	0.247	1.012	
				(0.082)	(0.161)	
			Lag = 2	0.153	1.052	
				(0.092)	(0.165)	
			Lag = 3	0.191	1.015	
				(0.109)	(0.202)	
INMA(1)	EMM	Aux.1	Lag = 1		1.009	0.297
					(0.119)	(0.128)
			Lag = 2		1.085	0.234
					(0.123)	(0.098)
			Lag = 3		1.033	0.306
					(0.168)	(0.153)
		Aux.2	Lag = 1		1.026	0.319
					(0.122)	(0.135)
			Lag = 2		1.092	0.258
					(0.107)	(0.101)
			Lag = 3		1.131	0.170
					(0.145)	(0.127)
INARMA(1,1)	EMM	Aux.1	Lag = 1	n.a.	n.a.	n.a.
			Lag = 2	0.007	1.056	0.282
				(0.070)	(0.134)	(0.116)
			Lag = 3	0.007	1.056	0.282
				(0.085)	(0.157)	(0.140)
		Aux.2	Lag = 1	n.a.	n.a.	n.a.
			Lag = 2	0.007	1.056	0.282
				(0.066)	(0.124)	(0.121)
			Lag = 3	0.007	1.056	0.282
				(0.095)	(0.164)	(0.137)

Table 10:

Estimated ACF and PACF of alternative models for the downloads data, based on EMM parameter estimates in Table 9 using Aux.1 with 3 lags. The autocorrelations and partial autocorrelations are computed by simulating each model for $1,000,000$ observations and computing the ACF and the PACF of the simulated data.

Lag	ACF			
(2 min .)	Actual	INAR(1)	INMA(1)	INARMA(1,1)
1	0.221	0.223	0.236	0.223
2	0.036	0.049	0.000	0.000
3	-0.003	0.011	-0.001	0.000
4	0.059	0.002	-0.001	-0.001
5	0.063	0.001	-0.001	0.000
Lag	PACF			
(2 min .)	Actual	INAR(1)	INMA(1)	INARMA(1,1)
1	0.221	0.223	0.236	0.223
2	-0.012	-0.001	-0.059	-0.052
3	-0.009	0.000	0.014	0.013
4	0.067	0.000	-0.004	-0.004
5	0.037	0.000	0.001	0.001

of blocks chosen is now $B = \{16, 32, 64, 128\}$, with the smallest block size of 16 chosen as of result of the relatively simple correlation structure of the downloads data.

The first set of results in Table 9 is for the INAR(1) model advocated by Weiß (2007, 2008), who compiled the data; MLE, CLS and EMM generally yield similar inferences. Using an INMA(1) specification instead yields CLS and EMM estimators of the moving average parameter, β_1, around 0.3. The EMM results from extending the model to an INARMA(1,1) specification provide strong support for the INMA(1), as the moving average estimate (0.282) dominates the autoregressive estimate (0.007) with the latter being statistically insignificant. This finding is further supported by the sample ACF and PACF of the alternative estimated models given in Table 10, where EMM parameter estimates from Table 9 are used. Finally, a parametric bootstrap exercise similar to that reported in the previous sub-section for the Fürth data is also conducted for these downloads data; the results are portrayed in Figure 6 and indicate the adequacy of the fitted INMA(1) specification. The strong evidence in favour of the INMA(1) model for the downloads data contrasts with the INAR(1) specification discussed by Weiß.

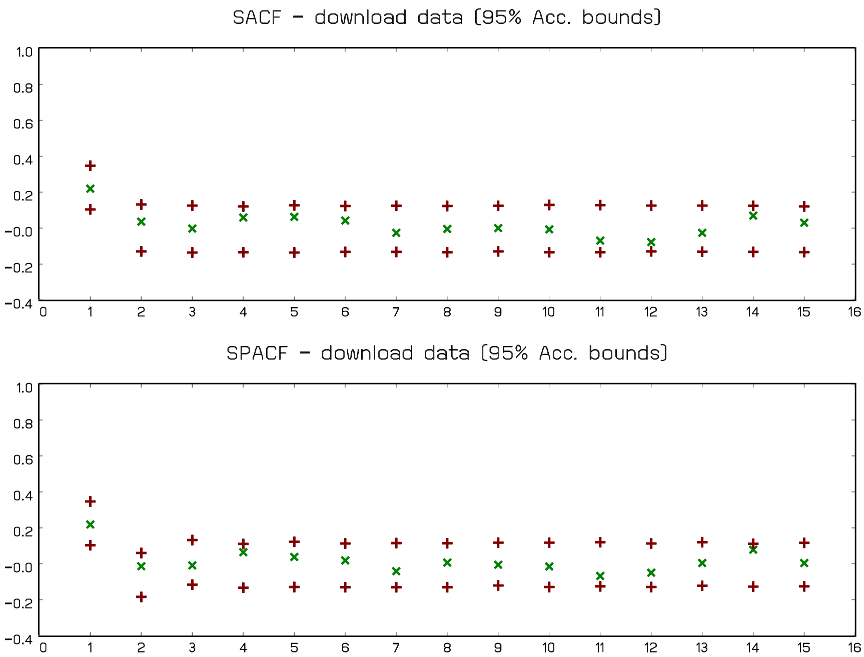

Figure 6: Estimated sample ACF and PACF of the INMA(1) model for the downloads data. The 95% confidence bounds are based on a parametric bootstrap with 5000 replications.

5 Conclusions

This paper advocates the use of Efficient Method of Moments estimators in binomial thinning models for count time series. The methods are easy to implement and especially useful when moving average components are specified. Where comparison with other estimators, such as CLS and MLE, is feasible the EMM procedure is generally superior to the former and rarely noticeably inferior to the latter on the basis of bias and mean squared error computations from Monte Carlo experimentation. In the last substantive section of the paper, we apply the foregoing ideas to two data sets for which satisfactory data generating mechanisms do not seem to have been provided hitherto. Each proves to have a significant moving average component, thereby exemplifying the usefulness of the EMM approach, since such specifications cannot be readily fitted by maximum likelihood methods.

References

[1] Al-Osh, M. and Alzaid, A.A. (1987), "First-order Integer-valued Autoregression (INAR(1)) Process", *Journal of Time Series Analysis*, 8, 261-275.

[2] Al-Osh, M. and Alzaid, A.A. (1988), "Integer-valued Moving Average (INMA) Process", *Statistical Papers*, 29, 281-300.

[3] Alzaid, A.A. and Al-Osh, M. (1990), "An Integer-valued pth Autoregressive Structure (INAR(p)) Process", *Journal of Applied Probability*, 27, 314-324.

[4] Appel, M. J., Labarre, R. and Radulovic, D. (2003), "On Accelerated Random Search", *SIAM Journal on Optimization*, 14, 708-731.

[5] Brännäs, K., (1994), "Estimation and Testing in Integer-valued AR(1) Models", *University of Umea Discussion Paper No. 335.*

[6] Brännäs, K. and Hall, A. (2001), "Estimation in Integer-valued Moving Average Models", *Applied Stochastic Models in Business and Industry*, 17,277-291.

[7] Brännäs, K. and Shahiduzzaman Quoreshi, A.M.M. (2010), "Integer-valued Moving Average Modelling of the Number of Transactions in Stocks", *Applied Financial Economics*, 20, 1429-1440.

[8] Bu, R., Hadri, K. and McCabe, B.P.M. (2008), "Conditional Maximum Likelihood Estimation of Higher-Order Integer-valued Autoregressive Processes", *Journal of Time Series Analysis*, 29, 973-994.

[9] Carlstein, E., (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Time Series", *Annals of Statistics*,14, 1171-1179.

[10] Drost, F.C., van den Akker, R. and Wekker, B.J.M. (2009), "Efficient Estimation of Auto-Regression Parameters and Innovation Distributions for Semiparametric Integer-valued AR(p) Models", *Journal of the Royal Statistical Society B*, 71, 467-485.

[11] Du, J.G and Li, Y. (1991), "The Integer-valued Autoregressive (INAR(p)) Model", *Journal of Time Series Analysis*, 12, 129-142.

[12] Duffie, D. and K. Singleton (1993), "Simulated Moments Estimator of Markov Models of Asset Prices", *Econometrica*, 61, 929-62.

[13] Freeland, K. and McCabe, B.P.M. (2004a), "Forecasting Discrete Valued Low Count Time Series", *International Journal of Forecasting*, 20, 427-434.

[14] Freeland K. and McCabe, B.P.M. (2004b), "Analysis of Count Data by Means of the Poisson Autoregressive Model", *Journal of Time Series Analysis*, 25, 701-722.

[15] Fürth, R. (1918), "Statistik and Wahrscheinlichkeitsnachwirkung", *Physikalische Zeitschrift*, 19, 421-426.

[16] Gallant A.R. and Tauchen, G. (1996), "Which Moments to Match?", *Econometric Theory*, 12 (4), 657-81.

[17] Gouriéroux, C. and Monfort, A. (1994), *Simulation Based Econometric Methods*, CORE Discussion Paper.

[18] Gouriéroux, C., Monfort, A. and Renault, E. (1993), "Indirect Inference", *Journal of Applied Econometrics*, 8, S85-S118.

[19] 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, 479-495.

[20] Joe, H. (1996), "Time Series Models with Univariate Margins in the Convolution-closed Infinitely Divisible Class", *Journal of Applied Probability,* 33, 664-677.

[21] Jung, R.C., Ronning, G. and Tremayne, A.R. (2005), "Estimation in Conditional First Order Autoregression with Discrete Support", *Statistical Papers*, 46, 195-224.

[22] Jung, R.C. and Tremayne, A.R. (2006a), "Coherent Forecasting in Integer Time Series Models", *International Journal of Forecasting*, 22, 223-238.

[23] Jung, R.C and Tremayne, A.R. (2006b), "Binomial Thinning Models for Integer Time Series", *Statistical Modelling*, 6, 81-96.

[24] Jung, R.C. and Tremayne, A.R. (2011), "Useful Models for Time Series of Counts or Simply Wrong Ones", *Advances in Statistical Analysis*, 95, 59-91.

[25] Klimko, L.A. and Nelson, P.I. (1978), "On Conditional Least Squares Estimation for Stochastic Processes", *Annals of Statistics*, 6, 629-642.

[26] McCabe, B.P.M. and Martin, G.M. (2005), "Bayesian Predictions of Low Count Time Series", *International Journal of Forecasting*, 21, 315-330.

[27] McCabe, B.P.M., Martin, G.M. and Harris, D. (2011), "Efficient Probabilistic Forecasts for Counts", *Journal of the Royal Statistical Society B*, 73, 253-272.

[28] McKenzie, E. (1985), "Some Simple Models for Discrete Variate Time Series", *Water Resources Bulletin*, 21, 645-650.

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

[30] McKenzie, E. (2003), "Discrete Variate Time Series", in *Handbook of Statistics* Vol. 20, eds. Shanbhag, D.N. and Rao, C.R., Amsterdam; Elsevier, 573-606.

[31] Mills, T.M. and Seneta, E. (1989), "Goodness-of-fit for a Branching Process with Immigration", *Stochastic Processes and Their Applications*, 33, 151-161.

[32] Politis, D.N., Romano, J.P. and Wolf, M. (1999), "Subsampling", New York: Springer-Verlag.

[33] Smith, A.A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, 8, S63-S84.

[34] Steutel, F. and van Harn, K. (1979), "Discrete Analogues of Self-decomposability and Stability", *The Annals of Probability*, 7, 893-899.

[35] Tsay, R.S. (1992), "Model Checking via Parametric Bootstraps in Time Series Analysis", *Applied Statistics*, 41, 1-15.

[36] Weiß, C.H. (2007), "Controlling Correlated Processes of Poisson Counts", *Quality and Reliability Engineering International*, 23, 741-754.

[37] Weiß, C.H. (2008), "Serial Dependence and Regression of INARMA Models", *Journal of Statistical Planning and Inference*, 138, 2975-2990.