Efficient Method of Moments Estimators for Integer Time Series Models¹

Vance L. Martin
University of Melbourne

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

Robert C. Jung
Universität Hohenheim

May 2014

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

JEL Classification: C15, C22, C25

¹We would like to thank the editor and three referees for helpful comments on an earlier draft of the paper. Comments by participants at the Australasian Econometric Meetings in Adelaide 2011, and the Singapore Management University econometrics workshop in Singapore 2012, are also gratefully acknowledged. Finally, we thank Christian Weiß for supplying the downloads data.

Abstract

The parameters of integer autoregressive models with equi-dispersed Poisson, or over-dispersed negative binomial innovations 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 efficient method of moment techniques as a means of obtaining practical estimators of relevant parameters using simulation methods. Under appropriate regularity conditions, the resultant estimators are consistent, asymptotically normal and under certain conditions achieve the same efficiency as maximum likelihood estimators. 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 well-known data sets in the time series literature, one of which has hitherto proved difficult to model satisfactorily, where it can be plausibly argued that moving average components should be entertained.

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). This familiar model may provide a useful starting point for modelling real life phenomena. Suppose that a stock-type count variable, y_t , is observable to represent the number of countries that are in some kind of financial crisis at time t . At each time point a country currently in crisis may continue to be in crisis at time $t + 1$, or it may exit crisis and other countries previously not in crisis may be deemed to have entered into crisis. The INAR(1) model may capture the data generating mechanism using a parameter α_1 which controls the strength of autocorrelation in y_t between adjacent time periods together with a ‘thinning’ operator allowing countries to either continue in crisis mode in the next period, or exit it. A further parameter λ controls the number of new countries entering crisis at time $t + 1$.

A steady stream of papers relating to estimation and forecasting with the first order model followed those cited above, 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 mod-

elling 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 (1993). Integer autoregressive specifications are considered, particularly 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 a subsampling approach. Finally, we use EMM to fit models for two real life data sets, each of which may be plausibly regarded as exhibiting moving average behaviour and, therefore, eliminating consideration of MLE. Diagnostic testing of the adequacy of each specification for the two data sets suggest that the method works satisfactorily.

2 Integer Models Based on Binomial Thinning

In what follows we shall study integer valued autoregressive moving average (INARMA) models in the spirit of McKenzie (1988, Sec. 4) which are given in their most general form by

$$\begin{aligned} x_t &= \sum_{i=1}^p \alpha_i \circ x_{t-i} + u_t, \\ y_t &= x_{t-q} + \sum_{j=1}^q \beta_j \circ u_{t+1-j}, \end{aligned} \tag{1}$$

where y_t is a stock-type observable process and x_t is a p th order unobservable latent process carrying the autoregressive dependence. The *iid* innovations u_t drive the moving average aspect of the model parameterized by $\beta_1, \beta_2, \dots, \beta_q$, where each $\beta_j \in [0, 1], j = 1, \dots, q$. The autoregressive part of the model is parameterized by $\alpha_1, \alpha_2, \dots, \alpha_p$ where each $\alpha_i \in [0, 1), i = 1, \dots, p$. For some of the special cases discussed below, we provide additional restrictions for the parameter space of the α 's.

In (1) ‘ $\circ$ ’ denotes the binomial thinning operator introduced by Steutel and van Harn (1979) which preserves the status of the integer random variables w when operated on by a parameter $\theta \in [0, 1)$ via $r = \theta \circ w = \sum_{k=1}^w e_k$, where the e_k , $k = 1, 2, \dots, w$ are *iid* Bernoulli counting sequence random variables $P(e_k = 1) = \theta$ and $P(e_k = 0) = 1 - \theta$. The operator is a random operator and the random variable r (number of ‘survivors’) has a binomial distribution with parameters w and θ and counts the number of ‘survivors’ from the count w remaining after (binomial) thinning. 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’, $\alpha_i, i = 1, \dots, p$ and $\beta_j, j = 1, \dots, q$. 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.

To close the model, a parametric assumption for the *iid* random innovation sequence $\{u_t\}$ is required. In the case where these innovations are equi-dispersed, a Poisson distribution with constant mean λ and variance $\sigma_U^2 \equiv \lambda$ is commonly specified for u_t . Alternatively, in the case of over-dispersion where the volatility of innovations is greater than λ , a negative binomial distribution can be specified.

In the following we briefly present special cases of the general INARMA(p, q) model which may be useful for modelling time series of counts and discuss their main properties. The prototypical model in the class is the pure first order integer autoregressive, INAR(1), model following from (1) if $\alpha_2 = \alpha_3 = \dots = \alpha_p = 0$ and $\beta_1 = \beta_2 = \dots = \beta_q = 0$ and given by

$$\begin{aligned} x_t &= \alpha_1 \circ x_{t-1} + u_t \\ &= \sum_{j=1}^{x_{t-1}} e_{j,t-1} + u_t \\ E[e_{j,t-1}] &= \alpha_1, \end{aligned} \tag{2}$$

where the $\{e_{j,t-1}\}$ are Bernoulli random variables. In contrast to the general process given in (1), the random variable x_t in the INAR(1) process is of course not latent. If equi-dispersed innovations $u_t \sim iid Po(\lambda)$ are assumed, it can be shown that the stationary marginal distribution of x_t is $Po(\lambda/(1 - \alpha_1))$. The process is Markovian and hence depends upon its past history, $\mathcal{F}_{t-1}$, only through x_{t-1} . Moreover, the process

can easily be shown to have a linear conditional mean (regression) function given by

$$E(x_t|\mathcal{F}_{t-1}) = \alpha_1 x_{t-1} + \lambda \quad (3)$$

and an autocorrelation function (ACF) that mirrors that of the standard AR(1) where x_t is a continuous random variable in that the j th autocorrelation is given by $\rho_j = \alpha_1^j, j = 1, 2, \dots$. 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 $x_1, x_2, \dots, x_T$. Estimation methods based on least squares and maximum likelihood are also readily available; see below.

Should one believe that equi-dispersed innovations are undesirable, because over-dispersion is called for, we may specify negative binomial innovations, $u_t \sim iid NB(\lambda, \pi)$, with $\pi = \lambda/(\lambda + \gamma)$ where the mean is λ and the variance is $\lambda(1 + \lambda/\gamma) = \lambda/(1 - \pi)$, whereby over-dispersion, measured by the ratio of the variance to the mean of the innovations, is given by $1 + \lambda/\gamma$. As $\gamma \rightarrow \infty$ (resp. $\pi \rightarrow 0$), the negative binomial distribution approaches the Poisson distribution as a limiting case.

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

$$x_t = \sum_{i=1}^p \alpha_i \circ x_{t-i} + u_t, \quad (4)$$

where the p thinnings are applied independently at each time period and $\mathcal{F}_{t-1} = (x_{t-1}, x_{t-2}, \dots, x_{t-p})$. The particular case on which we shall concentrate here (in simulation experiments at least) is $p = 2$, *i.e.* the INAR(2) model. The parameters of this model must satisfy $\alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 < 1$. This and other models with $p > 2$ of the Du and Li-type do not have the property of closure under convolution enjoyed by the first order counterpart with Poisson innovations; the marginal distribution of x_t is, in fact, over-dispersed. But it does have the same ACF as an AR(2) with Gaussian innovations and retains conditional linearity in that the conditional mean function is

$$E(x_t|\mathcal{F}_{t-1}) = E(x_t|x_{t-1}, x_{t-2}) = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + E(u_t).$$

Available estimation methods include conditional least squares (CLS) estimators (Klimko

and Nelson, 1978) which can be obtained from a minimization of the criterion function

$$S(\alpha_1, \alpha_2, \lambda) = \sum_{t=3}^T (x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2} - \lambda)^2.$$

Conditional maximum likelihood estimators can be based upon maximizing the function

$$\begin{aligned} \ell(\alpha_1, \alpha_2, \lambda; x_1, \dots, x_T) &\propto \sum_{t=3}^T \log p(x_t | x_{t-1}, x_{t-2}) \\ &= \sum_{t=3}^T \log \sum_{r=0}^l p_R(r | x_{t-1}) \left[\sum_{s=0}^n p_S(s | x_{t-2}) p_U(x_t - r - s) \right], \end{aligned}$$

where $l = \min(x_t, x_{t-1})$, $p_U(\cdot)$ denotes the distribution of the innovations, $p_R(r | x_{t-1})$ denotes the binomial distribution of the number of survivors from thinning the count x_{t-1} at time t using parameter α_1 , $p_S(s | x_{t-2})$ is the binomial distribution of the number of survivors from the population count x_{t-2} that survive the binomial thinning parameterized by α_2 and $n = \min(x_{t-1}, x_{t-2})$. The term after the logarithm on the right hand side is, of course, a convolution and the complete expression for the conditional distribution $p(x_t | \mathcal{F}_{t-1})$ is given by Bu, Hadri and McCabe (2008) as their (13).

Turning to pure moving average models based on binomial thinning, the basic INMA(q) follows from (1) if $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$ and is given by

$$y_t = \sum_{i=1}^q \beta_i \circ u_{t-i} + u_t. \quad (5)$$

The model was first considered by McKenzie (1988) in the context of $u_t \sim iid Po(\lambda)$. Again, 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 $\lambda / (1 + \sum_{i=1}^q \beta_i)$ under Poisson innovations, is given by $\rho_j = \sum_{i=0}^{q-j} \beta_i \beta_{i+j} / (1 + \sum_{i=1}^q \beta_i)$, $j = 1, \dots, q$ ($\beta_0 = 1$) 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 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 (3) though, of course, with parameter $\beta_1 / (1 + \beta_1)$ replacing α_1 ; see Al-Osh and Alzaid (1988, eq. 3.10). The ACF has only one non-zero ordinate,

$\rho_1 = \beta_1/(1 + \beta_1) \leq 0.5$. 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, whether the innovations are Poisson or negative binomial. 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.

Finally, we shall also discuss here the most simple mixed specification given by the INARMA(1,1) form. This model is obtained as a special case of (1) by setting $\alpha_2 = \alpha_3 = \dots = \alpha_p = 0$ and $\beta_2 = \beta_3 = \dots = \beta_q = 0$, whereby the model is given by

$$\begin{aligned} x_t &= \alpha_1 \circ x_{t-1} + u_t \\ y_t &= x_{t-q} + \beta_1 \circ u_t. \end{aligned} \tag{6}$$

Assuming $u_t \sim iid Po(\lambda)$ McKenzie (1988, Sec. 4) is able to show that the marginal distribution of y_t is Poisson with mean and variance of $\lambda[(1 - \alpha_1)^{-1} + \beta_1]$ and that the ACF is given by

$$\begin{aligned} \rho_1 &= \frac{\alpha_1 - \alpha_1\beta_1 + \beta_1}{1 - \alpha_1\beta_1 + \beta_1} \\ \rho_j &= \alpha_1^{j-1}\rho_1, \quad j = 2, 3, \dots \end{aligned} \tag{7}$$

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 time series models for counts 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 provides an

approximation to 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 suitable approximation to 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 termed just-identified, since the number of equations and parameters to be solved for are equal, thus yielding unique estimates of the parameters. In this setting the elements of the gradient vector are all zero when evaluated using the simulated data. In choosing between alternative auxiliary models, this situation constitutes a lower bound to the number of parameters in the auxiliary. 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 non-zero. A natural choice for the auxiliary model is an $AR(m)$ model. This 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 m , the length of the lag structure.

The INMA model can be viewed as an infinite INAR model in much the same way that, for a time series model with Gaussian innovations, a MA model is represented by an infinite AR model. Hence, increasing m in the auxiliary model improves the quality of the approximation of the model, thereby improving the efficiency of the EMM estimator (Gouriéroux, Monfort and Renault, 1993). There remains the question of

how best to select m in applying the technique. In our applications to real data series in Section 5, we propose to operationalize the suggestions made by Chen and Deo (2006, page 265) for the use of Akaike's Information Criterion (AIC) (or FPE) which was developed specifically for the framework of an approximating model where there may be no true degrees.

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

$$\begin{aligned}x_t &= \sum_{j=1}^p \alpha_j \circ x_{t-j} + u_t \\y_t &= x_{t-1} + \beta_1 \circ u_t \\u_t &\sim iid(\lambda),\end{aligned}\tag{8}$$

adopted in the Monte Carlo experiments and the empirical applications. This is indeed the case when the $p + 1$ thinning operations are all treated independently. In (8) $iid(\lambda)$ indicates *iid* innovations with mean λ , either Poisson or negative binomial as appropriate. The model is simulated by expressing the thinning operations in terms of independent Bernoulli random numbers $e_{j,r,t}$, $j = 0, 1, 2, \dots, p$, $r = 1, \dots, x_{s,t-j}$ and independent draws $u_{s,t}$ from the relevant distribution according to

$$\begin{aligned}x_{s,t} &= \sum_{j=1}^p \sum_{r=1}^{x_{s,t-j}} e_{j,r,t} + u_{s,t} \\y_{s,t} &= x_{s,t-1} + \sum_{r=1}^{u_{s,t}} e_{0,r,t} \\u_{s,t} &\sim iid(\lambda),\end{aligned}\tag{9}$$

where the Bernoulli random numbers have moments

$$\begin{aligned}E[e_{j,r,t}] &= \alpha_j, & j = 1, 2, \dots, p, \\E[e_{0,r,t}] &= \beta_1.\end{aligned}\tag{10}$$

Formally, the EMM estimator is based on solving

$$\hat{\theta} = \arg \min_{\theta} G'_s I^{-1} G_s = \arg \min_{\theta} Q(\theta),\tag{11}$$

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(m) model with a constant

$$y_t = \phi_0 + \sum_{i=1}^m \phi_i y_{t-i} + v_t,\tag{12}$$

where v_t is *iid* $N(0, \sigma_v^2)$. The gradients of this model at time t are

$$\begin{aligned} g_{1,t} &= \frac{v_t}{\sigma_v^2} \\ g_{i,t} &= \frac{v_t y_{t-i+1}}{\sigma_v^2}, \quad i = 2, 3, \dots, m+1 \\ g_{m+2,t} &= \left(\frac{v_t^2}{\sigma_v^2} - 1 \right) \frac{1}{2\sigma_v^2}, \end{aligned} \quad (13)$$

a total of $K = m + 2$ elements. Letting g_t be the column vector with i^{th} component $g_{i,t}$, $i = 1, 2, \dots, m + 2$, and averaging across a sample of size T , gives

$$G = \frac{1}{T} \sum_{t=m+1}^T g_t. \quad (14)$$

Evaluating G at the QMLE $\{\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_v^2\}$, 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 (14) by the simulated data $(y_{s,t})$, evaluated at the parameter estimates of the AR(p) model from using the actual data, $\{\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_v^2\}$. 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 (11) is obtained from (14) according to

$$G_s = \frac{1}{N} \sum_{t=p+1}^N g_{s,t}, \quad (15)$$

where $g_{s,t}$ is the column vector with i^{th} component $g_{s,i,t}$, $i = 1, 2, \dots, m + 2$, given by

$$\begin{aligned} g_{s,1,t} &= \frac{v_{s,t}}{\sigma_v^2} \\ g_{s,i,t} &= \frac{v_{s,t} y_{s,t-i+1}}{\sigma_v^2}, \quad i = 2, 3, \dots, m+1 \\ g_{s,m+2,t} &= \left(\frac{v_{s,t}^2}{\sigma_v^2} - 1 \right) \frac{1}{2\sigma_v^2}, \end{aligned} \quad (16)$$

and

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

Finally, the weighting matrix in (11) is defined as (Gallant and Tauchen, 1996)

$$I = \frac{1}{T} \sum_{t=m+1}^T g_t g_t'. \quad (17)$$

In applying the EMM estimator here we use an auxiliary model in which all $m + 2$ gradient conditions are used. In the case of an INAR(1) model with parameters $\theta = \{\alpha_1, \lambda\}$, the first auxiliary model with $m = 1$ amounts to estimating an AR(1) regression with an intercept. If Poisson innovations are specified, the third gradient condition is overidentifying for λ , whereas if negative binomial innovations are specified, the third gradient is needed to identify the innovation variance. Taking expectations of the auxiliary model gradients in (13) with respect to the true model being an INAR(1) process with Poisson innovations, and using (12) to substitute out v_t , gives

$$\begin{aligned}
E[g_{1,t}] &= E\left[\frac{y_t - \phi_0 - \phi_1 y_{t-1}}{\sigma_v^2}\right] = \frac{E[y_t] - \phi_0 - \phi_1 E[y_{t-1}]}{\sigma_v^2} \\
&= \frac{1}{\sigma_v^2} \left(\frac{\lambda(1 - \phi_1)}{1 - \alpha_1} - \phi_0 \right) \\
E[g_{2,t}] &= E\left[\frac{(y_t - \phi_0 - \phi_1 y_{t-1}) y_{t-1}}{\sigma_v^2}\right] = \frac{E[y_t y_{t-1}] - \phi_0 \frac{\lambda}{1 - \alpha_1} - \phi_1 E[y_{t-1}^2]}{\sigma_v^2} \\
&= \frac{1}{\sigma_v^2} \left(\frac{\alpha_1 \lambda}{1 - \alpha_1} + \frac{\alpha_1 \lambda^2}{(1 - \alpha_1)^2} - \phi_0 \frac{\lambda}{1 - \alpha_1} - \phi_1 \left(\frac{\lambda}{1 - \alpha_1} + \frac{\lambda^2}{(1 - \alpha_1)^2} \right) \right) \\
E[g_{3,t}] &= E\left[\frac{(y_t - \phi_0 - \phi_1 y_{t-1})^2}{2\sigma_v^4} - 1\right] \\
&= E\left[\frac{y_t^2 + \phi_0^2 + \phi_1^2 y_{t-1}^2 - 2\phi_0 y_t - 2\phi_1 y_t y_{t-1} + \phi_0 \phi_1 y_{t-1}}{2\sigma_v^4} - 1\right] \\
&= \frac{1}{2\sigma_v^4} \left(\left(\frac{\lambda}{1 - \alpha_1} + \frac{\lambda^2}{(1 - \alpha_1)^2} \right) + \phi_0^2 + \phi_1^2 \left(\frac{\lambda}{1 - \alpha_1} + \frac{\lambda^2}{(1 - \alpha_1)^2} \right) \right. \\
&\quad \left. - 2\phi_0 \frac{\lambda}{1 - \alpha_1} - 2\phi_1 \left(\frac{\alpha_1 \lambda}{1 - \alpha_1} + \frac{\alpha_1 \lambda^2}{(1 - \alpha_1)^2} \right) + \phi_0 \phi_1 \frac{\lambda}{1 - \alpha_1} - 1 \right)
\end{aligned} \tag{18}$$

where the last step uses the analytical solutions of the moments in an INAR(1) model with Poisson innovations. Setting the three moments in (18) to zero shows that the parameter vector $\theta = \{\alpha_1, \lambda\}$ is over-identified as there are three equations and two unknowns. The EMM estimator in this case is obtained by replacing the theoretical expectations in (18) by sample averages of the simulated data, $y_{s,t}$. 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 (15), 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 and resulting in some efficiency loss. This potential efficiency loss is investigated in the Monte Carlo experiments.

To operationalize the EMM algorithm in practice, we propose to implement the following steps.

- Step 1: Use standard Box-Jenkins type model identification methodology to identify a plausible range of potential auxiliary models. For INAR models, the sample autocorrelation function (SACF) is a standard tool. Moreover, the sample partial correlation structure (SPACF) may also be useful with integer models, particularly if a moving average component is suspected; see Al-Osh and Alzaid (1988, Sec. 3.3) for details and Mills and Seneta (1989) for a paper largely devoted to the subject. However, a caveat is in order here. A referee has carefully pointed out that the use of the SPACF for non-Gaussian time series has not been established theoretically and its use, therefore, should be taken only as a heuristic. We restrict attention to continuous $AR(m)$ auxiliary models because they represent suitable approximations to the integer models under consideration. The value of m to be used is informed by low values of $AIC(k)$ across a range of values $k = 1, \dots, M$. Estimate the auxiliary models by least squares and generate the estimates $\hat{\phi}_i, i = 0, 1, 2, \dots, m$, $\hat{\sigma}^2 = T^{-1} \sum_{t=m+1}^T \hat{v}_t^2$, where $\hat{v}_t = y_t - \hat{\phi}_0 - \sum_{i=1}^m \hat{\phi}_i y_{t-i}$, is the least squares residual for the preferred value of m . The innovation distribution should also be specified at this stage. Under negative binomial innovations, innovation mean and variance differ. This is a default position, but the equality of innovation mean and variance can be tested. If this null hypothesis is not rejected, Poisson innovations are confirmed, otherwise the negative binomial distribution is used.
- Step 2: Simulate the true INARMA model for the chosen innovation distribution from Step 1 and some starting parameter values $\theta = \theta_{(0)}$, and let the simulated values be given by $y_{s,t}$. Starting values can be based on either the conditional least squares estimator of Klimko and Nelson (1978), or the sample autocorrelation function (see Jung and Tremayne, 2006b, for details).
- Step 3: Evaluate the gradient vector in (15) and the weighting matrix in (17) for the current values of the parameters.
- Step 4: Across a suitable range of alternative parameter values the EMM estimator is given as the solution of (11).

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, as 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.

It seems prudent to discuss if the INARMA models used here satisfy the conditions for valid use of subsampling. PRW stipulate in Theorem 3.2.1 that, in addition to conditions on block length choice, to be addressed below, the stochastic process being modelled may be dependent but should be strong-mixing. The theory relating to ergodicity, mixing and asymptotic properties with models for integer counts is as yet incomplete, though progress is being made, see, for example, Neumann (2011) in the context of discrete autoregressive conditional mean models. Grunwald, Hyndman, Tedesco and Tweedie (2000, Proposition 3, Case II) establish that the Poisson INAR(1) process is ergodic and strong-mixing. More generally, the INAR(p) processes used in this paper are stationary Markov chains of order p . Since it is easy to see that, with either Poisson or negative binomial innovations, they are both irreducible and aperiodic, it follows that they satisfy β -mixing, which implies α -, or strong-mixing. Full details of these and related features may be found in Bradley (2005, Sec. 3). Moreover, the ACF of an INAR(p) is identical to that of a conventional Gaussian AR(p) and, as pointed out by White and Domowitz (1984, p. 146), for example, strong-mixing still applies for processes with memories much longer than AR(MA) processes. The INMA(q) process is, of course, q -dependent and hence obviously strong-mixing. Consider the INARMA(1,1) process given at (6). Equation (7) gives the ACF of such a process and second and subsequent ordinates decline geometrically, as in an INAR(1). Since the ACF of an INARMA(p ,1) process also mimics that of an INAR(p) in terms of its decline, it seems reasonable to maintain that INARMA(p ,1) (and, more generally,

INARMA(p, q) processes are strong-mixing, so that subsampling methods may be employed with the models of this paper.

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 $y_1, \dots, y_T$ 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 $y_i, \dots, y_{i+B-1}$. Let there be $j = 1, \dots, N_B$ such blocks used and suppose that a (scalar) parameter θ is estimated by $\hat{\theta}_T$ using the full sample and by $\hat{\theta}_{T,i,B}$ 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 $\sqrt{T}(\hat{\theta}_T - \theta)$ by $N(\theta, \sigma_\infty^2)$, where σ_∞^2 is an unknown long-run variance to be estimated. Provided $B \rightarrow \infty$ with T and $B/T \rightarrow 0$ the distribution of $\sqrt{B}(\hat{\theta}_{T,i,B} - \theta)$ is the same Gaussian distribution involving the unknown long-run variance. Then a suitable estimator of the variance of $\hat{\theta}_T$ is given by

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

where $\hat{\theta}_{T,B,\cdot} = N_B^{-1} \sum_{j=1}^{N_B} \hat{\theta}_{T,B,j}$, 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, leaving discussion of choice of block length until it is needed in Section 5.

4 Finite Sample Properties

The finite sample properties of the EMM parameter estimator are investigated in this section 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). In such circumstances one would hope that EMM would not perform significantly worse than CLS, or the asymptotically efficient MLE. However, an important point to bear in mind is that EMM is sometimes available when one (or both) of these alternatives is not and, if it performs adequately when its competitors are available, there is reason to be confident that it represents a useful addition to inference methodology for integer models because of this added flexibility.

4.1 Experimental Design

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

$$\begin{aligned}x_t &= \alpha_1 \circ x_{t-1} + \alpha_2 \circ x_{t-2} + u_t \\y_t &= x_{t-1} + \beta_1 \circ u_t \\u_t &\sim iid(\lambda),\end{aligned}\tag{19}$$

where the three thinning operators in (19) are assumed independent following the approach of Du and Li (1991). To save space in the presentation of results with simulated data, the auxiliary model used to generate them is invariably one where $m \leq 3$ and the variance gradient given as the final element of (13) is exploited. For those simulations where the innovations u_t in (19) are negative binomial, $\lambda/\gamma = 1$ ($\pi = 0.5$) resulting in overdispersion of 2. Experimental results are presented based on the parameterizations given in Table 1.

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 both MLE and CLS; the results of Jung, Ronning and Tremayne (2005) are used to facilitate comparison with the former. The next 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) for MLE. A comparison of the sampling performance of the EMM and CLS estimators of the INMA(1) model (for which MLE is unavailable) determines whether there are any efficiency gains in finite samples from adopting the EMM estimator over the CLS one. The INARMA(1,1) and INARMA(2,1) models represent more complex DGPs and these

Table 1:
Parameterizations of INARMA model specifications used in the Monte Carlo experiments.

Experiment	α_1	α_2	λ	β_1	Unconditional Mean, $E[y_t]$
INAR(1)	0.3	0.0	3.5	0.0	$\frac{\lambda}{1 - \alpha_1} = 5.00$
INAR(2)	0.5	0.3	1.0	0.0	$\frac{\lambda}{1 - \alpha_1 - \alpha_2} = 5.00$
INMA(1)	0.0	0.0	3.5	0.3	$\lambda(1 + \beta_1) = 4.55$
INARMA(1,1)	0.3	0.0	3.5	0.7	$\lambda \left(\beta_1 + \frac{1}{1 - \alpha_1} \right) = 7.45$
INARMA(2,1)	0.5	0.3	1.0	0.7	$\lambda \left(\beta_1 + \frac{1}{1 - \alpha_1 - \alpha_2} \right) = 5.70$

experimental results inform 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 autoregression. The selected parameterizations in Table 1 are based on DGPs that yield relatively “low” counts, with the expected counts ranging from 4.55 to 7.45.

To keep the exposition simple, in implementing the EMM estimator a single auxiliary model is used, *viz.* an approximating autoregressive model (12) where the maximum lag used is $m = 3$. The parameter λ is identified by the intercept with both Poisson and negative binomial innovations and the error variance in the latter case leads to identification of π .

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, of course, 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 (3). 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 in the case of an independent error variance parameter the sample variance of v_t is used in the usual way.

When computing the EMM estimator, a range of simulation runs was initially experimented with. A value of $H = 500$ was settled upon, resulting in simulation runs of length $N = 500T$ in (15), where T is the sample size given by $T = \{50, 100, 200\}$. As the objective function of the EMM estimator $Q(\theta)$ (see (11) above) is not continuous with respect to the parameter vector θ , a grid search is used to minimize the objective function. The number of searches is set to 100 for the INAR(1) and INMA(1) models and 200 for the INARMA(1,1) model with Poisson innovations. For the more complex INARMA(1,1) with negative binomial innovations and the INARMA(2,1) models 500 searches are used. 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 search procedure adopted in the Monte Carlo experiments reported here suffices. Finally, all Monte Carlo experiments are based on 5000 replications for the INAR(1), INMA(1) and INARMA(1,1) specifications, and 1000 replications for the INARMA(2,1) model.

4.2 Simulation Results

4.2.1 INAR(1)

The basic results for the INAR(1) Monte Carlo experiments are given in Table 2. Sampling statistics based on the mean and the root mean squared error, RMSE, from the 5000 replications are presented. In the case of MLE and CLS the finite sample results are comparable to those reported in Jung, Ronning and Tremayne (2005). The results for the EMM estimator are based on an AR auxiliary model with a constant and $m = 1$ lag. Experiments were also run by specifying auxiliary models with lag lengths of 2 and 3, but these results are not reported here to save space. For comparison the results for the MLE and CLS estimators are also reported. Inspection 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 terms of bias there is

very little to choose between CLS and EMM where both estimators exhibit a bias of around 13% for $T = 50$; this is roughly twice as large as the bias of MLE.

The type of results discussed in the last paragraph is similar to those reported after many simulation experiments, yet it is possible to glean much more information by examining the results in closer detail. Since, asymptotically, MLE cannot be outperformed by any other estimator, we use the INAR(1) results to see if there is any marked superiority of MLE at the sample sizes used in the experiments, over and above its lower bias and slightly lower RMSE. We initially decomposed the replications with regard to the EMM and MLE estimates of $\alpha_1 = 0.3$ at each replication into four groups: (i) those cases where both EMM and MLE estimates are less than $\alpha_1 = 0.3$; (ii) those cases where both EMM and MLE estimates are greater than $\alpha_1 = 0.3$; (iii) those cases where EMM is less than $\alpha_1 = 0.3$ while MLE estimates are greater than $\alpha_1 = 0.3$; and finally (iv) those cases where EMM is greater than $\alpha_1 = 0.3$ while MLE estimates are less than $\alpha_1 = 0.3$. To make the analysis comparable, the same random numbers are used to compute the EMM and MLE estimates in each simulation run. For Cases (i) and (iii), the proportion of times that EMM generates estimates that are closer to $\alpha_1 = 0.3$ than MLE are, respectively, 40.4% and 48.6%. For Cases (ii) and (iv) the performance of EMM is better than that of MLE with the respective percentages now 57.5% and 54.1%. Across all cases the correlation between EMM and MLE estimates is 0.905. Similar statistical properties occur for the CLS estimator, although the results are not reported here. A further means of comparison of the finite sample behaviour of three estimation methods for this case (i.e. $\alpha_1 = 0.3$ and $T = 50$) is provided by Figure 1 which presents a boxplot of the ranked values of the calculated statistics. The medians of each statistic show the limited superiority of MLE over the other two estimation methods and the inter-quartile ranges are shown to be very similar.

Our final insight into the properties of the EMM, MLE and CLS estimators of α_1 is given in Figure 2 which contains nonparametric kernel estimates of the sampling distributions for samples of size $T = 50, 100, 200$. The asymptotics appear to work quite well, for the sampling distributions of all three estimators are similar for all sample sizes, especially when $T = 200$. However, the additional downward bias in the EMM and CLS estimators compared to the MLE estimator can be seen for the case of $T = 50$, with the sampling distributions of the EMM and CLS estimators centred to the left of that of the MLE estimator. The overall conclusions from these additional investigations do confirm the slight superiority of MLE over CLS and EMM but offer realistic hope that EMM will be a useful technique, especially when MLE is not available.

In contrast to the bias properties of estimators of α_1 , Table 2 shows that all estima-

Table 2:

Finite sample properties of alternative estimators of the parameters of the INAR(1) model with Poisson innovations. The sampling properties of the EMM estimator are based on 5000 replications where the auxiliary model consists of a constant and $m = 1$ lag.

Estimator	α_1			λ		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
<i>Mean</i>						
MLE	0.281	0.288	0.295	3.587	3.559	3.521
CLS	0.260	0.280	0.291	3.693	3.596	3.543
EMM	0.260	0.278	0.291	3.681	3.605	3.543
TRUE	0.300	0.300	0.300	3.500	3.500	3.500
<i>RMSE</i>						
MLE	0.133	0.095	0.063	0.713	0.500	0.331
CLS	0.144	0.098	0.069	0.766	0.522	0.361
EMM	0.135	0.098	0.068	0.722	0.517	0.359

tors of the innovation mean parameter λ are biased upwards, with the size of the bias decreasing as the sample size increases. This is unsurprising given the negative correlation between estimators of the two parameters. In particular, MLE has the smallest bias of around 2.5% for samples of size $T = 50$, while CLS and EMM exhibit biases of around 5%. A limited exercise of dissecting the sampling distributions of the EMM and MLE estimators of λ , shows that, when both EMM and MLE under-estimate the true parameter value of $\lambda = 3.5$, the size of the bias is less for EMM than it is for MLE, whereas the opposite is true when both estimators over-estimate λ .

While CLS and EMM exhibit similar finite sample biases, EMM not only demonstrates finite sample efficiency gains over CLS with lower RMSEs, it also exhibits comparable efficiency levels to MLE. This is especially true for the case when $T = 50$. For the mean parameter α_1 , the RMSE of MLE is 0.133, whilst for EMM it is just marginally higher at 0.135 (compared to CLS where it is higher still at 0.144). In the case of the parameter λ for $T = 50$, the RMSE of MLE is 0.713; for EMM it is just marginally higher at 0.722, compared to CLS where it is higher still with a RMSE of 0.766.

4.2.2 INAR(2)

The simulation results from extending the INAR model to having 2 lags are given in Table 3. To be able to compare the finite sample properties of the EMM estimator to the MLE and CLS estimators of the parameters of the INAR(2) model reported in Bu,

Figure 1: Boxplot of the EMM, MLE and CLS estimators of α_1 , for samples of size $T = 50$, where the true value is $\alpha_1 = 0.3$.

Figure 2: Sampling distributions of the EMM, MLE and CLS estimators of α_1 , for samples of size $T = 50, 100, 200$, where the true value is $\alpha_1 = 0.3$.

Table 3:

Finite sample properties of alternative estimators of the parameters of the INAR(2) model with Poisson innovations. The MLE and CLS results are obtained from Bu, Hadri and McCabe (2008) which are based on 1000 replications. The sampling properties of the EMM estimator are based on 5000 replications where the auxiliary model consists of a constant and $m = 2$ lags. The sample size is $T = 100$.

Estimator	α_1	α_2	λ
		<i>Mean</i>	
MLE	0.498	0.271	1.142
CLS	0.481	0.263	1.270
EMM	0.477	0.266	1.262
TRUE	0.500	0.300	1.000
		<i>RMSE</i>	
MLE	0.092	0.111	0.401
CLS	0.109	0.109	0.549
EMM	0.106	0.106	0.515

Hadri and McCabe (2008, Table 1), the sample size is chosen to be $T = 100$. For this choice of parameterization the asymptotic relative efficiency of CLS to MLE is only 0.7 and thus there may be discernible differences, perhaps between all three types of estimators. The auxiliary model of EMM consists of a constant and $m = 2$ lags, which is the minimum needed for identification.

The simulation results for α_1 in Table 3 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 marginally lower bias 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.

4.2.3 INMA(1)

The INMA(1) results given in Table 4 contain just the CLS and EMM simulation results as MLE is no longer feasible. The auxiliary model of the EMM estimator giving the results presented is based on a constant and $m = 1$ lag. The EMM estimator of the moving average parameter, β_1 , exhibits smaller bias than the CLS estimator and this is especially true for $T = 50$. Here the bias for EMM is around 5% and increases to around 7% for CLS. The size of the bias does in fact decrease as the number of lags in the auxiliary model increases. For larger samples, the relative bias of the two

Table 4:

Finite sample properties of alternative estimators of the parameters of the INMA(1) model with Poisson innovations. The sampling properties of the EMM estimator are based on 5000 replications where the auxiliary model consists of a constant and $m = 1$ lag.

Estimator	β_1			λ		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
<i>Mean</i>						
MLE	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS	0.278	0.290	0.292	3.661	3.577	3.547
EMM	0.284	0.290	0.291	3.609	3.556	3.521
TRUE	0.300	0.300	0.300	3.500	3.500	3.500
<i>RMSE</i>						
MLE	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
CLS	0.226	0.156	0.111	0.676	0.463	0.327
EMM	0.219	0.174	0.132	0.633	0.481	0.353

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 exhibits slightly smaller bias than the CLS estimator, in that for samples of size $T = 50$ the bias is 3% for EMM and 5% for CLS. The EMM estimator also exhibits superior finite sample efficiency than the CLS estimator for samples of size $T = 50$. As with the RMSE results for β_1 , the CLS estimator of λ exhibits slightly better efficiency than the EMM estimator for larger sample sizes.

4.2.4 INARMA(1,1)

The finite sample results of the EMM estimator for the INARMA(1,1) model are reported in Table 5. For this type of model, neither MLE nor CLS is feasible. The results are presented for both Poisson innovations and negative binomial innovations. The auxiliary model of the EMM estimator consists of a constant and $m = 3$ lags. In the case of negative binomial innovations the parameters λ and π are jointly identified by the intercept and the disturbance variance of the auxiliary model.

Inspection of the results based on Poisson innovations in Table 5 shows that the EMM estimator performs well in estimating the parameters of the model. The bias in estimating the autoregressive parameter α_1 , is around 2% while for the innovation mean parameter λ , the bias is even smaller. In the case of the moving average parameter the bias is over 10% for samples of size $T = 50$, but halves on doubling the sample size. Given the fairly large bias in estimating the moving average parameter when $T = 50$,

Table 5:

Finite sample properties of the EMM estimator for the INARMA(1,1) model with Poisson and negative binomial innovations. Based on 5000 replications where the auxiliary model consists of a constant and $m = 3$ lags.

	Poisson			negative binomial			
	α_1	λ	β_1	α_1	λ	β_1	π
	<i>Mean</i>			<i>Mean</i>			
$T = 50$	0.306	3.485	0.620	0.282	3.504	0.657	0.287
$T = 100$	0.309	3.487	0.629	0.281	3.510	0.685	0.366
$T = 200$	0.306	3.483	0.651	0.283	3.501	0.702	0.422
TRUE	0.300	3.500	0.700	0.300	3.500	0.700	0.500
	<i>RMSE</i>			<i>RMSE</i>			
$T = 50$	0.190	0.744	0.278	0.169	0.699	0.236	0.286
$T = 100$	0.162	0.545	0.260	0.141	0.521	0.191	0.212
$T = 200$	0.140	0.417	0.218	0.107	0.377	0.142	0.149

it may not be advisable to use the procedure with samples as small as this, though increasing sample size, quite plausibly, attenuates the problem. Moreover, the results also show that progressively increasing the sample size results in progressive falls in the RMSE.

Table 5 shows that by extending the INARMA model to allow for negative binomial innovations the EMM displays similar finite sample properties for α_1 , β_1 and λ , to the case where the innovations are Poisson. In the case of the additional parameter π from the negative binomial distribution the EMM estimates are biased downwards by more than 40%. Increasing the sample size to $T = 100$ reduces to the bias to around 27% which decreases even further for samples of size $T = 200$ where the bias is about 16%.

4.2.5 INARMA(2,1)

The final model structure considered is an INARMA(2,1) model containing either Poisson or negative binomial innovations. Again the finite sample results are only presented for the EMM estimator as neither MLE nor CLS is feasible. The auxiliary model of the EMM estimator consists of a constant and $m = 3$ lags, so the auxiliary model is just identified with negative binomial innovations, while there is one over-identifying restriction in the case of Poisson ones.

The finite sample results of the EMM estimator are presented in Table 6 for the parameters of the INARMA(2,1) model. Qualitatively, the performance of the EMM estimator for this model is similar to its performance for the previous model specifica-

Table 6:

Finite sample properties of the EMM estimator for the INARMA(2,1) model with Poisson and negative binomial innovations. Based on 1000 replications where the auxiliary model consists of a constant and $m = 3$ lags.

	Poisson				negative binomial				
	α_1	α_2	λ	β_1	α_1	α_2	λ	β_1	π
	<i>Mean</i>				<i>Mean</i>				
$T = 50$	0.484	0.167	1.491	0.522	0.460	0.165	1.579	0.509	0.224
$T = 100$	0.498	0.199	1.357	0.514	0.481	0.192	1.426	0.525	0.273
$T = 200$	0.497	0.240	1.243	0.539	0.495	0.226	1.288	0.545	0.334
TRUE	0.500	0.300	1.000	0.700	0.500	0.300	1.000	0.700	0.500
	<i>RMSE</i>				<i>RMSE</i>				
$T = 50$	0.203	0.193	0.786	0.329	0.195	0.182	0.917	0.339	0.330
$T = 100$	0.140	0.157	0.570	0.336	0.148	0.157	0.660	0.324	0.292
$T = 200$	0.101	0.115	0.419	0.305	0.107	0.123	0.506	0.305	0.246

tions with the parameters controlling the dynamics of the model, α_1, α_2 and β_1 , being under-estimated while the parameter λ is over-estimated. This result is the same regardless of the innovations being equi-dispersed (Poisson) or over-dispersed (negative binomial). For samples of size $T = 50$, the size of the bias for the estimator of α_1 is 8% whereas for the estimator of α_2 it is much larger at 45%. Increasing the sample size quickly reduces the bias. For example, for samples of size $T = 200$, the bias is less than 1% for the EMM estimator of α_1 and down to 2% for the corresponding estimator of α_2 . As with the performance of the EMM estimator for the moving average parameter of the INARMA(1,1) model the EMM estimator exhibits relatively large bias even for samples of size $T = 200$.

The results in Table 6 show that the EMM estimator of the negative binomial distribution parameter π is biased downwards by around 50% for samples of size $T = 50$; this decreases to just over 30% for samples of $T = 200$. As with the INARMA(1,1) finite sample results, these results suggest that reasonable sized samples are needed to be able to estimate the parameters of the negative binomial distribution accurately. Finally, the efficiency of EMM is demonstrated by observing that the RMSE for all parameter estimators progressively decrease as the sample size is increased.

5 Applications

The EMM estimator is now applied to estimating INARMA models for two data sets. The first consists of data on software downloads every 2 minutes from a server at

the University of Würzburg, Germany initially considered by Weiß (2007, 2008) who specified an INAR(1) model; henceforth these will be termed the downloads data. For comparison, MLE and CLS estimation results are also reported, whenever available. The second 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 unifying feature of these two data sets is that each plausibly supports the notion that a moving average component might be entertained for integer time series modelling of them. In such cases MLE will not be a viable option for parameter estimation.

5.1 Downloads Data

Figure 3 presents the downloads data to separate IP addresses in two minute intervals for one working day, 10am to 6pm, on the 29 of November 2005, from a server at the University of Würzburg. 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 over-dispersion. Hence, if attention is restricted to first order models (INAR(1) or INMA(1) specifications), it is likely that Poisson innovations will suffice to model the data, obviating the need to entertain negative binomial ones. As preliminary and tentative model identification devices we turn to the SACF and SPACF plots given in Figure 4, remembering that the latter should be treated as a heuristic device only with non-Gaussian count data (see also Step 1 of the implementation of the EMM procedure at the end of Section 3.1). Both display spikes at lag 1, suggesting the possibility of an INAR(1) model, though the SACF has no significant higher order SACF ordinates, or, perhaps, an INMA(1) though the SPACF has no discernible higher order ordinates. One might also consider an INARMA(1,1) model and, possibly, higher order lag effects in a mixed model.

Before using the EMM approach to parameter estimation, a suitable auxiliary model must be ascertained, as different approximating models will yield different parameter estimates and associated estimated asymptotic standard errors. For the downloads data we entertained continuous autoregressions, $AR(m)$ with $m = 1, \dots, 8$ (see Step 1 at the end of Sec. 3.1 and the discussion above (8)). The value of AIC for each is provided in Figure 5; on the basis of this information, since the minimum of AIC corresponds to a single lag, we used auxiliary models with 1 and 2 lags to provide illustrative results.

Figure 3: Software downloads to different IP addresses every 2 minutes during a working day, on the 29 of November 2005, at a computer at the University of Würzburg, Germany, $T = 241$.

Figure 4: Sample autocorrelation function (SACF) and partial autocorrelation function (SPACF) of downloads data. Lags expressed in 2 minute intervals.

Table 7:

EMM estimates of INARMA models for the downloads data with Poisson (Po) and negative binomial (NB) innovations. The auxiliary model is an $AR(m)$ with the lag length given in the table. Standard errors are in parentheses.

Model	Aux. Model Lag Length	Dist.	Parameter			
			α_1	λ	β_1	π
INAR(1)	$m = 1$	NB	0.228	1.030		0.033
			(0.059)	(0.110)		(0.055)
	$m = 2$	NB	0.199	1.111		0.049
			(0.090)	(1.111)		(0.049)
	$m = 1$	Po	0.206	1.052		
			(0.070)	(0.123)		
	$m = 2$	Po	0.225	1.030		
			(0.083)	(0.133)		
INMA(1)	$m = 1$	NB		1.091	0.236	0.056
				(0.123)	(0.153)	(0.065)
	$m = 2$	NB		1.023	0.320	0.057
				(0.076)	(0.075)	(0.035)
	$m = 1$	Po		1.001	0.286	
				(0.131)	(0.150)	
	$m = 2$	Po		1.088	0.233	
				(0.116)	(0.106)	
INARMA(1,1)	$m = 2$	NB	0.134	1.064	0.077	0.037
			(0.131)	(0.199)	(0.199)	(0.118)
		Po	0.124	1.078	0.098	
			(0.108)	(0.151)	(0.123)

Figure 5: AIC for auxiliary model lag length for the downloads data.

The EMM parameter estimates of various INARMA models of the downloads data are presented in Table 7, together with their estimated asymptotic standard errors. Parameter estimates based on both negative binomial and Poisson innovations are presented. The standard errors are computed as described in Section 3.2; the block lengths used are now discussed. We adopted a regression-based procedure (fuller details available from the authors on request), noting that, 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). An outline of the approach can be gleaned 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 $B = 2b$. But 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 .

We adopted an approach of 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 satisfactorily in

a case where the desired asymptotic result is known (e.g. a continuous AR(1) specification where the true variance of the limiting distribution of $\hat{\alpha}_1$, the MLE of the dependence parameter α_1 , is known to be $1 - \alpha^2$). We feel this approach to estimating standard errors by subsampling methods has merit, but more work is required to adduce further evidence to this. Briefly, 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 estimates are found by combining estimates based on B in a range $kT^{1/3}$ to $8kT^{1/3}$ with some suitable k . Since the underlying processes of the count data used in Section 4 are taken to be stationary, without near nonstationary behaviour apparent by reference to the sample autocorrelation properties of the data sets involved, we believe that a similar choice made for the block length with the applications will provide reliable estimates for asymptotic standard errors.

To test out the approach, estimated standard errors based on EMM with a set of 100 observations simulated from an INAR(1) model were obtained. The long-run variances are estimated as the intercepts from a regression of the estimated variances corresponding to four block sizes, on a constant and the regressor $\{1/8, 1/16, 1/32, 1/64\}$. The standard errors of the elements of $\hat{\theta}_T$ are computed as $\hat{\sigma}_\infty/\sqrt{T}$ where $\hat{\sigma}_\infty^2$ is the appropriate estimate of σ_∞^2 and are: $se(\hat{\alpha}_1) = 0.088$; and $se(\hat{\lambda}) = 0.489$; these agree closely with the MLE standard errors for the simulated data. With the downloads data, since the sample size is somewhat larger, the block sizes used for subsampling standard error estimates are $B = \{16, 32, 64, 128\}$.

A comparison of the parameter estimates of the over-dispersion parameter π in Table 7 shows that the estimates range from 0.033 to 0.057 for the models entertained. These estimates are numerically small and, statistically insignificant on conventional criteria in every case, suggesting little evidence of over-dispersion, a result which is consistent with the observation that the mean and the variance of the downloads data are very similar in value. Hence the remainder of the discussion will focus on the results based on the Poisson distribution.

The first set of results in Table 7 is for the INAR(1) model advocated by Weiß (2007, 2008), who compiled the data. The estimates of the autoregressive parameter α_1 , in the case of Poisson innovations are 0.206 ($m=1$) and 0.225 ($m=2$). These estimates are similar to the maximum likelihood estimate of 0.236 and the CLS estimate of 0.221. Using an INMA(1) specification instead yields EMM estimates of the moving average parameter β_1 , of 0.286 ($m=1$) and 0.233 ($m=2$), the former of which is very similar to the CLS estimate of 0.283. Table 7 also gives the results of estimating a combined model

based on an INARMA(1,1) specification. Both the autoregressive and moving average parameter estimates are now statistically insignificant suggesting that the dynamics may best be modelled either by an autoregressive model or a moving average model, but that a mixed specification is not needed.

In standard ARMA models a common approach to compare between AR and MA models is to use information criteria such as the AIC or the BIC. However, this requires being able to evaluate the likelihood function which is not feasible in the case of integer MA models. An alternative strategy to distinguish between the INAR(1) and INMA(1) model specifications is to simulate the two estimated models and compare the SACF (and, possibly, the SPACF) of the simulated data with the corresponding SACF and SPACF of the actual data to see if the specified model captures the memory characteristics observed in the data. These results are given in Table 8 for the downloads data, where EMM parameter estimates based on Table 7 and the length of the simulated time series equals 1,000,000 observations; the INARMA(1,1) model is included for completeness. The INMA(1) model does a better job in capturing the first order autocorrelation coefficient in the actual data, while the INAR(1) and INARMA(1,1) models do marginally better in explaining the (insignificant) second autocorrelation coefficient calculated for the downloads data.

Figure 6 presents the results of a parametric bootstrap exercise following the approach of Tsay (1992) and gives 95% acceptance bounds for the INMA(1) model. He argues that a model may be deemed adequate if it can successfully reproduce important features of the data; we opt to use the first 15 observed SACF and SPACF ordinates of the data series given in Figure 2. To implement the idea, we generate 5000 artificial data sets of length $T = 241$ using the preferred fitted model and *iid* Poisson innovations to yield 5000 SACFs and SPACFs. For each ordinate of these, the $100(1 - \delta/2)\%$ and $100\delta/2\%$ quantiles are used as acceptance bounds; we use $\delta = 0.05$. The model is adjudged to fit any particular feature adequately if the actual value lies within the acceptance bounds. The figure shows that the dynamic properties of the data are well captured by the INMA(1) model. Evidently there is little to choose between the INAR(1) and INMA(1) specifications for these data.

5.2 Fürth Data

Figure 7 gives the number of pedestrians on a city block observed every 5 seconds. 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, possibly conforming with the assumption of Poisson

Table 8:

The sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF) of alternative models for the downloads data, based on the EMM parameter estimates in Table 7 assuming Poisson innovations. The auxiliary model contains one lag in the case of the INAR(1) and INMA(1) models, and two lags in the case of the INARMA(1,1) model. The autocorrelations and partial autocorrelations are computed by simulating each model for 1,000,000 observations and computing the ACF and the SPACF of the simulated data.

Lag (2 min .)	SACF			
	Actual	INAR(1)	INMA(1)	INARMA(1,1)
1	0.221	0.204	0.222	0.192
2	0.036	0.041	-0.001	0.023
3	-0.003	0.008	-0.001	0.002
4	0.059	0.000	-0.001	0.000
5	0.063	-0.001	0.000	-0.001
Lag (2 min .)	SPACF			
	Actual	INAR(1)	INMA(1)	INARMA(1,1)
1	0.221	0.204	0.222	0.192
2	-0.012	-0.001	-0.053	-0.015
3	-0.009	-0.001	0.011	0.000
4	0.067	-0.001	-0.003	0.000
5	0.037	-0.001	0.001	-0.001

Figure 6: The results of the Tsay bootstrap exercise for the sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF) for the INMA(1) model for the downloads data. 95% acceptance bounds are based on a parametric bootstrap with 5000 replications.

Figure 7: Fürth data representing the number of pedestrians on a city block observed every 5 seconds, $T = 505$.

innovations. Nevertheless, we shall continue to entertain the possibility of the need for negative binomial innovations and test the mean-variance equality restriction in the normal way. These data have been considered by other authors, *inter alia*, Mills and Seneta (1989, Sec. 5). These authors consider a simple branching process model with immigration that amounts to a Poisson INAR(1) specification and admit that the fit is inadequate. Inspection of the sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF) of the Fürth data in Figure 8 suggest that there are higher order dynamics that will not be captured by either an INAR(1) model or an INMA(1) specification. The SPACF (again used heuristically) displays spikes at lags 1 and 2 with positive and negative values respectively, with the SACF 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 8, which violates the restriction $\rho_1 < 0.5$ needed for both Gaussian MA(1) and INMA(1) specifications to be appropriate. Even allowing for a higher order lag by way of an INAR(2) specification still encounters problems as the first two ordinates of the SACF 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 $\hat{\alpha}_i \in [0, 1)$ (see Jung and Tremayne, 2006b, eq. 4.5 for more details).

Figure 8: Sample autocorrelation function (SACF) and partial autocorrelation function (SPACF) of the Fürth data. Lags expressed in 5-second intervals.

It is, of course, necessary to choose a suitable range of auxiliary (continuous autoregressive) models to entertain. We again followed the suggestions of Deo and Chen (2006) as implemented in the previous subsection, but in view of the complex observed serial correlation in the data we used $AR(m)$, $m = 1, \dots, 15$. Relative local minima for the values of AIC can be seen at lags 3 and 12 (among others) in Figure 9 and so we opted to use an $AR(12)$ with a constant as our auxiliary model. However, to control for the number of moment conditions the lags in the auxiliary model are restricted to lags 1, 2, 3, 6, 12. As the innovations are allowed to be negative binomial in the general specification of the model, the variance moment condition in (16) is also included to identify the negative binomial parameter π , resulting in the total number of moment conditions being 7.

Table 9 gives the EMM parameter estimates for a range of INARMA models for the data. The most general specification is based on the following INARMA(12,1) model

$$\begin{aligned}
 x_t &= \alpha_1 \circ x_{t-1} + \alpha_6 \circ x_{t-6} + \alpha_{12} \circ x_{t-12} + u_t \\
 y_t &= x_{t-1} + \beta_1 \circ u_t \\
 u_t &\sim NB(\lambda, \pi),
 \end{aligned} \tag{20}$$

where u_t is a negative binomial innovation with parameters λ, π . For the special case where $\pi \rightarrow 0$, the innovations become equi-dispersed based on a Poisson distribution. The choice of lags at 6 and 12 is based on the higher order lags observed in the SACF and SPACF given in Figure 8, although the t-statistic associated with the 12th lag reported in Table 9 is statistically insignificant. Some intermediate lags at 2 and 5, were also tried but ultimately discarded, as they did not contribute significantly to the estimated model. For comparison, the results from estimating lower order models

Figure 9: AIC for auxiliary model lag length for the Fürth data.

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 $0 \leq \beta_1 \leq 1$.

Asymptotic standard errors for the EMM parameter estimates in Table 9 are obtained using the approach outlined in Section 3.2. 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 estimates of the long-run variances based on the intercepts. Dividing the relevant long-run estimate by the sample size T and taking square roots yields the estimated asymptotic standard error for each coefficient estimate.

The estimate of the negative binomial parameter π in Table 9 for the most general model specification (first model reported) is 0.099 with a standard error of 0.097, suggesting that the innovations are equally dispersed, a result that is also consistent with the descriptive statistics reported above. Imposing the restriction $\pi = 0$ (second model reported) so the innovations are Poisson, results in qualitatively similar parameter estimates for the remaining parameters.

An important feature of the parameter estimates of the INARMA(12,1) model in

Table 9, including the INARMA(1,1) and INMA(1) specifications, 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. There he states 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 period. Hence, a new arrival 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 $\beta_1 = 1$. Imposing a restriction of unity on the moving average parameter β_1 still produces qualitatively similar estimates for the remaining parameters of the model in the case of both negative binomial innovations (third model reported) and Poisson innovations (fourth model reported).

Excluding the higher order lags by setting $\alpha_6 = \alpha_{12} = 0$ in (20), results in an INARMA(1,1) model with higher estimates of the innovation mean parameter λ , with a value of 0.448 in the case of negative binomial innovations (fifth model reported) and 0.471 in the case of Poisson innovations (sixth model reported). Imposing the additional restriction of $\beta_1 = 0$ and fitting a simple INAR(1) (seventh and eighth models reported) inflates the estimate of the first order autocorrelation parameter α_1 with estimates of 0.664 (negative binomial) and 0.681 (Poisson). These estimates are also comparable to estimates based on maximum likelihood with an estimate of 0.688 in the case of Poisson innovations and CLS with an estimate of 0.665. Alternatively, imposing the restriction $\alpha_1 = 0$ on the INARMA(1,1) specification and fitting a simple INMA(1) model gives the ninth and tenth models reported.

To gauge the ability of the alternative models to capture the dynamics in the data, the SACF and SPACF corresponding to the estimated models are given in Tables 10 and 11 respectively based on EMM parameter estimates in Table 9. The INARMA(12,1) specifications presented are based on the parameterizations where the moving average parameter β_1 is not restricted to unity. Apart from the most general INARMA specification where innovations are negative binomial, the remaining specifications are based on Poisson innovations given the strong evidence of equi-dispersed innovations. The autocorrelations and partial autocorrelations of each model are computed by simulating each model for 1,000,000 observations and computing the SACF and SPACF of the simulated data. For the simple model specifications, such as the INAR(1) and INMA(1) models where analytic expressions of the SACF and SPACF are easily derived, the simulated estimates mimic the expected patterns well.

The INARMA(12,1) model captures the dynamics well for short lags as well as

Table 9:

EMM estimates of INARMA models for the Fürth data with Poisson (Po) and negative binomial (NB) innovations. The auxiliary model is an AR(12) with a constant which is the same for all INARMA model specifications reported. Standard errors are in parentheses.

Model	Dist.	Parameter					
		α_1	α_6	α_{12}	λ	β_1	π
INARMA(12,1)	NB	0.643	0.042	0.019	0.372	0.974	0.099
		(0.063)	(0.052)	(0.039)	(0.056)	(0.133)	(0.097)
	Po	0.575	0.073	0.043	0.395	0.974	0.000
		(0.062)	(0.044)	(0.037)	(0.063)	(0.138)	
	NB	0.584	0.025	0.103	0.286	1.000	0.172
		(0.074)	(0.070)	(0.036)	(0.064)		(0.100)
	Po	0.512	0.100	0.015	0.409	1.000	0.000
		(0.064)	(0.043)	(0.040)	(0.073)		
INARMA(1,1)	NB	0.481			0.448	0.922	0.071
		(0.087)			(0.101)	(0.160)	(0.101)
	Po	0.429			0.471	0.950	0.000
		(0.094)			(0.100)	(0.158)	
INAR(1)	NB	0.664			0.427		0.019
		(0.069)			(0.070)		(0.102)
	Po	0.681			0.395		0.000
		(0.053)			(0.054)		
INMA(1)	NB				0.565	0.976	0.047
					(0.176)	(0.169)	(0.134)
	Po				0.559	0.906	0.000
					(0.206)	(0.170)	

Table 10:

The sample autocorrelation function (SACF) of alternative models for the Fürth data, based on EMM parameter estimates in Table 9. The autocorrelations are computed by simulating each model for 1,000,000 observations and computing the SACF of the simulated data.

Lag (5 sec.)	Actual	INARMA(12,1)		INARMA(1,1)	INAR(1)	INMA(1)
		NB	Po	Po	Po	Po
1	0.665	0.731	0.682	0.629	0.680	0.475
2	0.323	0.477	0.403	0.269	0.462	0.000
3	0.170	0.316	0.247	0.115	0.314	0.001
4	0.147	0.219	0.165	0.047	0.213	-0.001
5	0.141	0.163	0.132	0.017	0.144	0.001
6	0.152	0.139	0.137	0.004	0.097	-0.002
7	0.114	0.123	0.134	0.000	0.067	-0.004
8	0.052	0.102	0.113	-0.001	0.046	-0.002
9	0.073	0.084	0.092	0.000	0.032	0.000
10	0.138	0.070	0.076	0.001	0.021	0.002
11	0.168	0.061	0.073	0.001	0.014	0.003
12	0.194	0.060	0.085	0.001	0.009	0.002
13	0.157	0.057	0.088	0.001	0.006	0.001
14	0.076	0.050	0.077	0.000	0.003	0.002
15	0.028	0.041	0.061	-0.001	0.002	0.001

Figure 10: Estimated sample SACF and SPACF of the INARMA(12,1) model with Poisson innovations, for the FÜRTH data. The 95% acceptance bounds are based on a parametric bootstrap with 5000 replications.

longer lags corresponding to the 30 second cycle. Briefly, we add that the rightmost two columns demonstrate the failure of the simple INAR(1) and INMA(1) models to capture the dynamic structure of the data. The INAR(1) model does a poor job of capturing lags of the SACF after the first. Unsurprisingly from Table 11, it models the first lag of the SPACF well but not the second and higher order lags of the SPACF, 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 SACF and SPACF reasonably successfully.

The adequacy of INARMA(12,1) model based on Poisson innovations is again further investigated by using the parametric resampling method suggested by Tsay (1992). We opt to use the first 15 observed SACF and SPACF ordinates of the FÜRTH data series given in the leftmost columns of Tables 10 and 11, or, alternatively, in Figure 8. The idea is implemented as described in Section 5.1 using 5000 artificial data sets of length $T = 505$ using the preferred fitted model and *iid* Poisson innovations. The results of this exercise for the FÜRTH data are presented in Figure 10 using a 95% acceptance envelope as before. It is seen that the fitted model faithfully reproduces the main features of the sample autocorrelation properties of the data well. Although the results are not presented to save space, this is not the case for any of the simpler models reported in the lower rows of the table.

Table 11:

The sample partial autocorrelation function (SPACF) of alternative models for the Fürth data, based on EMM parameter estimates in Table 9. The partial autocorrelations are computed by simulating each model for 1,000,000 observations and computing the SPACF of the simulated data.

Lag (5 sec.)	Actual	INARMA(12,1)		INARMA(1,1)	INAR(1)	INMA(1)
		NB	Po	Po	Po	Po
1	0.665	0.731	0.682	0.629	0.680	0.475
2	-0.213	-0.124	-0.115	-0.210	-0.001	-0.292
3	0.099	0.033	0.036	0.073	0.000	0.196
4	0.066	0.011	0.017	-0.029	-0.001	-0.136
5	0.017	0.022	0.037	0.007	0.000	0.094
6	0.079	0.034	0.061	-0.005	-0.001	-0.071
7	-0.054	0.012	0.012	0.003	0.002	0.050
8	-0.018	-0.002	0.001	-0.001	0.000	-0.036
9	0.111	0.006	0.008	0.002	0.000	0.029
10	0.061	0.006	0.010	0.001	-0.002	-0.018
11	0.033	0.010	0.021	-0.001	0.000	0.015
12	0.097	0.015	0.036	0.001	0.000	-0.009
13	-0.053	0.003	0.006	0.000	0.000	0.008
14	-0.040	-0.002	-0.003	-0.001	-0.001	-0.005
15	0.006	0.000	0.001	-0.002	0.000	0.003

6 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 due to Weiß (2007, 2008) (Sec. 5.1) and Fürth (1918) (Sec. 5.2). Certainly, no entirely satisfactory data coherent model for the latter has been provided hitherto. The data sets each prove plausibly to evidence a moving average component, thereby exemplifying the usefulness of the EMM approach, since such the 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] Bradley, R.C. (2005), “Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions”, *Probability Surveys*, 2, 107-144.
- [6] Brännäs, K., (1994), “Estimation and Testing in Integer-valued AR(1) Models”, *University of Umea Discussion Paper No. 335*.
- [7] Brännäs, K. and Hall, A. (2001), “Estimation in Integer-valued Moving Average Models”, *Applied Stochastic Models in Business and Industry*, 17, 277-291.

- [8] 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.
- [9] 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.
- [10] 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.
- [11] Chen, W.W. and Deo, R.S. (2006), “Estimation of Mis-specified Long Memory Models”, *Journal of Econometrics*, 134, 257-281.
- [12] 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.
- [13] Du, J.G and Li, Y. (1991), “The Integer-valued Autoregressive (INAR(p)) Model”, *Journal of Time Series Analysis*, 12, 129-142.
- [14] Duffie, D. and K. Singleton (1993), “Simulated Moments Estimator of Markov Models of Asset Prices”, *Econometrica*, 61, 929-62.
- [15] Freeland, K. and McCabe, B.P.M. (2004a), “Forecasting Discrete Valued Low Count Time Series”, *International Journal of Forecasting*, 20, 427-434.
- [16] 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.
- [17] Fürth, R. (1918), “Statistik und Wahrscheinlichkeitsnachwirkung”, *Physikalische Zeitschrift*, 19, 421-426.
- [18] Gallant A.R. and Tauchen, G. (1996), “Which Moments to Match?”, *Econometric Theory*, 12 (4), 657-81.
- [19] Gouriéroux, C. and Monfort, A. (1994), *Simulation Based Econometric Methods*, CORE Discussion Paper.

- [20] Gouriéroux, C., Monfort, A. and Renault, E. (1993), “Indirect Inference”, *Journal of Applied Econometrics*, 8, S85-S118.
- [21] Joe, H. (1996), “Time Series Models with Univariate Margins in the Convolution-closed Infinitely Divisible Class”, *Journal of Applied Probability*, 33, 664-677.
- [22] Jung, R.C., Ronning, G. and Tremayne, A.R. (2005), “Estimation in Conditional First Order Autoregression with Discrete Support”, *Statistical Papers*, 46, 195-224.
- [23] Jung, R.C. and Tremayne, A.R. (2006a), “Coherent Forecasting in Integer Time Series Models”, *International Journal of Forecasting*, 22, 223-238.
- [24] Jung, R.C. and Tremayne, A.R. (2006b), “Binomial Thinning Models for Integer Time Series”, *Statistical Modelling*, 6, 81-96.
- [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] Neumann, M.H. (2011), “Absolute Regularity and Ergodicity of Poisson Count Processes”, *Bernoulli*, 17, 1268-1284.
- [33] Politis, D.N., Romano, J.P. and Wolf, M. (1999), “Subsampling”, New York: Springer-Verlag.

- [34] Smith, A.A. (1993), “Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions”, *Journal of Applied Econometrics*, 8, S63-S84.
- [35] Steutel, F. and van Harn, K. (1979), “Discrete Analogues of Self-decomposability and Stability”, *The Annals of Probability*, 7, 893-899.
- [36] Tsay, R.S. (1992), “Model Checking via Parametric Bootstraps in Time Series Analysis”, *Applied Statistics*, 41, 1-15.
- [37] Weiß, C.H. (2007), “Controlling Correlated Processes of Poisson Counts”, *Quality and Reliability Engineering International*, 23, 741-754.
- [38] Weiß, C.H. (2008), “Serial Dependence and Regression of INARMA Models”, *Journal of Statistical Planning and Inference*, 138, 2975-2990.
- [39] White, H. and Domowitz, I. (1984), “Nonlinear Regression with Dependent Observations”, *Econometrica*, 52, 143-162.

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Martin, VL;Tremayne, AR;Jung, RC

Title:

Efficient method of moments estimators for integer time series models

Date:

2014-11

Citation:

Martin, V. L., Tremayne, A. R. & Jung, R. C. (2014). Efficient method of moments estimators for integer time series models. *Journal of Time Series Analysis*, 35 (6), pp.491-516. <https://doi.org/10.1111/jtsa.12078>.

Persistent Link:

<http://hdl.handle.net/11343/128253>