Statistical Papers 46, 195-224 (2005)

Statistical Papers

Estimation in conditional first order autoregression with discrete support

Robert C. Jung[1], Gerd Ronning[1], A. R. Tremayne[2]

[1] Wirtschaftswissenschaftliche Fakultät, Universität Tübingen, Mohlstr. 36, 72074 Tübingen, Germany (e-mail: Robert.Jung@uni-tuebingen.de)
[2] University of Sydney, Australia and University of York, UK

Received: June 25, 2003; revised: September 10, 2003

We consider estimation in the class of first order conditional linear autoregressive models with discrete support that are routinely used to model time series of counts. Various groups of estimators proposed in the literature are discussed: moment-based estimators; regression-based estimators; and likelihood-based estimators. Some of these have been used previously and others not. In particular, we address the performance of new types of generalized method of moments estimators and propose an exact maximum likelihood procedure valid for a Poisson marginal model using backcasting. The small sample properties of all estimators are comprehensively analyzed using simulation. Three situations are considered using data generated with: a fixed autoregressive parameter and equidispersed Poisson innovations; negative binomial innovations; and, additionally, a random autoregressive coefficient. The first set of experiments indicates that bias correction methods, not hitherto used in this context to our knowledge, are sometimes needed and that likelihood-based estimators, as might be expected, perform well. The second two scenarios are representative of overdispersion. Methods designed specifically for the Poisson context now perform uniformly badly, but simple, bias-corrected, Yule-Walker and least squares estimators perform well in all cases.

Key words Bias correction – Estimation – INAR models – Overdispersion – Small sample properties – Time series of counts.

1 Introduction

The analysis of time series of counts has received considerable attention lately. Recent contributions to this area include e.g. Brännäs and Hall (2001) and Brännäs and Hellström (2001). Introductory tracts can be found in the econometrics textbook of Greene (2000) and the monographs of Cameron and Trivedi (1998) and Winkelmann (2000). An attractive class of models from the perspective of time series analysis is the observation driven integer-valued autoregressive (INAR) model of Al-Osh and Alzaid (1987) and McKenzie (1988). Drawing a close parallel with the well-known class of Gaussian autoregressive-moving average processes, it not only offers a rich choice of dependence structure but also allows a variety of distributional assumptions for the different components of the process. There exist close relationships between the INAR class of models and the well established field of branching processes with immigration. This is exploited in Jung and Tremayne (2003) for testing purposes and will be utilized further in the estimation context of the present paper.

We consider the estimation of certain types of INAR models with a conditional linear first order autoregressive (CLAR(1)) dependence structure. The CLAR(1) structure is useful and often sufficient to capture the dependence in the data in a variety of practical applications. Moreover, it serves as a starting point for the extension to higher order INAR models allowing for richer dependence structures. Given the highly nonlinear character of these models, the required developments are by no means straightforward. These are subject of ongoing research to be reported elsewhere. In the analysis of count data a natural first choice for the distributional assumption is Poisson. Consequently, we focus here on description and evaluation of estimators for the parameters of CLAR(1) models with Poisson marginal distributions. In empirical applications, however, extra binomial variation, or overdispersion, is regularly found in count data, rendering the restrictive Poisson assumptions inadequate for valid inference. It, therefore, seems natural to investigate the robustness of these estimators to deviations from the Poisson distribution of forms that will generally be unknown to applied workers. It turns out that this question can be addressed quite elegantly within the framework of the CLAR(1) class of models described in Grunwald et al. (2000). The negative binomial (NB) model, will be employed in our analysis, since it is the parametric model most often used in empirical work to capture the overdispersion phenomenon. As part of our investigations we also employ a random coefficient formulation for the INAR(1) model as another means of generating overdispersion.

Novel features of the paper include: a different proposal for an exact maximum likelihood estimator; use of new moment conditions in conjunction with conditional and unconditional generalized method of moments estimators; and the use of bias correction techniques in this context for what we believe is the first time. The simulation evidence is more comprehensive than any hitherto available in the literature. Our evidence strongly indicates that erroneously maintaining the Poisson marginal structure of the data when it does not hold extracts a heavy price in estimator performance with procedures derived under this assumption. We conclude that some of the simplest estimators available are preferable across the range of scenarios considered, as long as bias correction is allowed for.

2 Some introductory theory

2.1 The class of conditional linear AR(1) models

The general framework considered here is the class of CLAR(1) models introduced by Grunwald et al. (2000). For the process $\{X_t; t = 0, \pm 1, \pm 2, \ldots\}$ the CLAR(1) structure is defined as

$$m(X_{t-1}) \equiv \mathrm{E}(X_t \mid X_{t-1}) = a\, X_{t-1} + \lambda \,. \tag{1}$$

Here a and λ are (possibly restricted) real numbers chosen in such a way that the value of $m(X_{t-1})$ remains in the admissible parameter space of the conditional distribution $f(x_t|x_{t-1})$. Appropriate restrictions will be given below for the specific models used here.

Alternatively, a stationary first order autoregressive structure can be defined by means of an exponentially decaying autocorrelation function (ACF)

$$\rho_k \equiv \mathrm{corr}(X_t, X_{t-k}) = a^k \qquad (k = 1, 2, \ldots)\,. \tag{2}$$

Grunwald et al. (2000) show that under very mild conditions (2) is implied by the CLAR(1) structure, but the converse is not true in general.

The usefulness of the structure (1) is based upon the fact that it includes many non-Gaussian AR(1) models that have been proposed in the literature. In particular it includes the class of CLAR(1) models with discrete support routinely employed for the modelling of times series of counts. In this paper we evaluate the properties of various estimators for the parameters of interest under three different scenarios by Monte Carlo methods. The first scenario is the fixed coefficient INAR(1) model with a Poisson marginal distribution. This is the standard framework usually chosen as a starting point for the analysis of time series of counts. It will be discussed in some detail in the next subsection. In empirical applications involving count data, overdispersion is frequently observed. To show that, under such circumstances, different estimators can give quite different results, we give an example using real data at the beginning of the penultimate section of the paper. This motivates the second and third scenarios employed which use negative binomial innovations in the simulations, in the contexts of fixed and random coefficient autoregression. While the first set of experiments presupposes equidispersion in the counts, the second and third are representative of the phenomenon of overdispersion. We can, therefore, analyze the behaviour of those estimators explicitly derived in the context of Poisson random variables under all three scenarios and give recommendations as to which estimators may be preferable when overdispersion of an unknown form is suspected in the data.

2.2 The INAR(1) benchmark model

The fixed coefficient INAR(1) model introduced by Al-Osh and Alzaid (1987) and McKenzie (1988) is defined on the discrete support $\mathbb{N}_0$ by means of the difference equation

$$X_t = a \circ X_{t-1} + W_t \qquad (t = 0, \pm 1, \pm 2, \ldots) . \tag{3}$$

It is assumed that the fixed parameter $a \in [0, 1)$ and that W_t is an independently and identically distributed (*iid*) discrete random variable sequence with finite first moment ($\mu_W > 0$) and second (central) moment ($\sigma_W^2 > 0$). W_t and X_{t-1} are presumed to be stochastically independent for all points in time. The process generated by (3) is stationary. The discreteness of the process $\{X_t\}$ is ensured by the binomial thinning operation (Steutel and van Harn, 1979)

$$a \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} Y_{i,t-1} , \tag{4}$$

where the $Y_{i,t-1}$ are assumed to be *iid* Bernoulli random variables with $P(Y_{i,t-1} = 1) = a$ and $P(Y_{i,t-1} = 0) = 1 - a$. Note that subsequent thinning operations are performed independently of each other with a constant probability a and that thinning is a random operation with an associated probability distribution.

In the following we state those properties of the INAR(1) process that are necessary to derive the estimators to be considered. Without employing any distributional assumption, the unconditional moments of X_t are given as follows: $\mathrm{E}(X_t) = \mu_W/(1-a)$; and $\mathrm{Var}(X_t) = (a\,\mu_W + \sigma_W^2)/(1-a^2)$. The ACF of the process is identical to (2) with the qualification that only positive autocorrelation is allowed. Both the regression function $\mathrm{E}(X_t|X_{t-1}) = aX_{t-1} + \mu_W$ and the conditional variance function $\mathrm{Var}(X_t|X_{t-1}) = a(1-a)X_{t-1} + \sigma_W^2$ are linear in X_{t-1}.

It can easily be shown that the INAR(1) process is structurally equivalent to the subcritical Bienaymé-Galton-Watson branching process with immigration (BGWI) process; for the definition of a BGWI process see e.g. Venkataraman (1982). This fact is exploited in the next section.

The properties of the marginal distribution of the INAR(1) process can be conveniently summarized by its probability generating function (pgf), which is given by (B.2) in Appendix B. By inspecting the pgf of the INAR(1) process it turns out that it satisfies the definition of a discrete self-decomposable distribution; see Steutel and van Harn (1979). This class contains the Poisson and negative binomial distributions as special cases and discrete stable distributions as a sub-class; the former special case is a member of this sub-class but the latter is not. Consequently, one can choose as the marginal distribution of an INAR(1) process any member of the class of discrete self-decomposable distributions. It also turns out that the marginal distribution of the INAR(1) process is completely specified through the choice of the distribution of the innovation W_t. If the distribution of W_t and the marginal distribution of X_t are to be from the same family of distributions (as is the case in the linear Gaussian AR(1) model), the possible choices are narrowed down to the class of discrete stable distributions; see Alzaid and Al-Osh (1988, p. 55) for related discussion. This follows directly from a comparison of the definition of the class of discrete stable distributions given in Steutel and van Harn (1979) and the pgf (B.2) of the INAR(1) process. Since the only discrete stable distribution with a finite first moment is the Poisson distribution, its choice for the innovation and the marginal distribution of the INAR(1) process can be justified on the basis of these theoretical considerations.

Assuming $W_t \sim \text{Po}(\lambda)$ it is then straightforward to show that $X_t \sim \text{Po}(\lambda/(1-a))$; see (B.4) and (B.5) for the relevant derivation. The resulting INAR(1) process will henceforth be denoted PoINAR(1).

For estimation and forecasting purposes the conditional distribution $f(x_t|x_{t-1})$ plays a key role. Two insightful ways to derive this distribution are provided in Appendix B. The resulting form is given by

$$\begin{aligned} f(x_t|x_{t-1}) &= \sum_{k=0}^{m} \binom{x_{t-1}}{k} a^k (1-a)^{x_{t-1}-k} \exp(-\lambda) \frac{\lambda^{x_t-k}}{(x_t-k)!} \\ &= x_{t-1}! \exp(-\lambda)\, C(x_{t-1}, x_t) \,, \end{aligned} \tag{5}$$

compare (B.7′).

3 Estimators considered

This section is concerned with the estimation of the two parameters of interest: the thinning parameter; and the mean of the immigration rate. The estimators discussed here have almost exclusively been proposed in the context of the INAR(1) model with a Poisson marginal distribution. In those cases where no (or only minimal) distributional assumptions are used in their derivation, this is indicated in Appendix A. The estimators are grouped according to three broad categories: moment-based; regression-based; and likelihood-based estimators. Certain feasible estimators drawn from the first category that have been proposed in the BGWI literature are not included in this analysis because of their rather unfavourable small sample properties. For an extensive discussion of this point see Jung (1999, Chapter 4). Outline details of the estimators used are provided in the next three paragraphs with a fuller treatment given in Appendix A.

A standard Yule-Walker estimator for the parameters a and λ is the initial estimator employed. The precise variant used is given in (A.1) and (A.2). As an alternative we include an estimator for λ proposed in Greene (2000, Chapter 19.9.7) in the group of moment-based estimators to be analyzed; this variant substitutes (A.3) for (A.2) and will be denoted (GR). We also look at Generalized Method of Moments (GMM) estimators employing various combinations of conditional as well as unconditional moment restrictions. It turns out that GMM estimation based on the latter type of moment restrictions is sometimes numerically unstable and we, therefore, give it limited consideration in the subsequent analysis; see Sections A.1 and 5 for further details. The estimator to be used throughout carries the acronym GMM in the following discussion and the conditional moment restrictions employed with it are given in (A.5), (A.6) and (A.7).

The usual regression-based estimation method in the context of the INAR(1) model is the conditional least squares (CLS) method of Klimko and Nelson (1978). The estimators are given as (A.9) and (A.10) in Section A.2. Due to the fact that the conditional variance, $\mathrm{Var}(X_t|X_{t-1})$, of the INAR(1) process is not constant over time, weighted conditional least squares estimators seem attractive alternatives to consider. This type of estimator has been proposed in the branching process literature in the context of estimating the parameters of BGWI processes. Taking the structural equivalence between INAR(1) and subcritical BGWI processes into account we can, therefore, use the weighted conditional least squares estimation method given in Wei and Winnicki (1989) along with the weighting scheme proposed there. These parameter estimators will be denoted WCLS and are given in (A.12) and (A.13). A further variant has been advanced by Heyde and Lin (1992) in the context of estimating equations principles. This leads to a different set of weights that are data determined and provides an an asymptotic quasi-likelihood (AQL) estimator. See (A.15) and (A.16) for details of the estimators and (A.17) and (A.18) for how the weights are determined.

Two likelihood-based estimation methods are considered that differ in their treatment of the (unobservable) starting value, X_0, of the INAR(1) process. The conditional maximum likelihood (CML) method treats X_0 as given and the relevant likelihood to be maximized is given at (A.21). The exact maximum likelihood (EML) method is based on the full likelihood function including an estimate of X_0. Utilizing the time reversibility property of PoINAR(1) (McKenzie, 1988), an estimate of the starting value can be obtained by means of the backcasting method see e.g. Box, Jenkins and Reinsel (1994). Further details are provided in Section A.3 at (A.24), (A.25) and the surrounding discussion.

4 Simulation design

We initially conducted a series of Monte Carlo experiments to assess the small sample properties of the various estimators outlined in Section 3 under the scenario of the PoINAR(1) model. These simulations are designed to resemble situations typically encountered in empirical research. Series of short and moderate length consisting of low counts were generated using the PoINAR(1) model described in Section 2.2. The biases and mean squared errors (MSE) of the various estimators discussed in Section 3 are computed using realistic parameter combinations of the dependence parameter a and the mean of the innovation process λ. Details pertaining to the simulation design are presented here and a discussion of the results constitutes the next section.

The design parameters for the main body of simulations were chosen as follows. The level of the process $\mathrm{E}(X_t)$ was fixed at 5 in order to ensure the generation of low level count series. We experimented with other (low) values for the mean of the process and found qualitatively similar results. The dependence structure of the INAR(1) process is governed by the parameter a and this was allowed to vary in the range $[0.2, [0.1], 0.9]$. Due to the functional relationship between the mean of the process and the value of the parameters a and λ, the value for the latter parameter necessarily varies in the range from 0.5 to 4.

The sample sizes T used were $50, 100$ and 500, with the main discussion below focussing on the intermediate value, since many salient points arise in this context. For $T = 500$ all estimators generally show minimal biases when estimating either a or λ; full details are available from the authors on request. Higher sample sizes were, therefore, not employed. The simulations were carried out using programs written in GAUSS Version 3.2.

In order to prevent the inclusion of simulated data sets that are likely to lead to inadmissible estimates for a we used a filtering device in the data generating process. One possible filter would be to employ a suitable test of randomness (see Jung and Tremayne, 2003) and filter out those series that do not exhibit a significant dependence structure. An even simpler type of filtering device is to look at the first order sample autocovariance of the data set. Should this not be positive, the generated series is not used and is replaced with a new realization. This second type of filter is used here. The results stemming from the use of both filters are broadly similar. Of course these filters are able to prevent results with negative estimated values of a only for those estimators that directly rely on the first order autocorrelation, viz. YW, GR and CLS. These and most other estimators may nevertheless produce inadmissible results. For the calculation of summary statistics of interest, we simply discarded these invalid estimation results for the estimators concerned. (Alternatively, trimmed estimators could have been used but this would not change the overall picture.) Some details of the number of invalid results that have been discarded are reported below.

The class of likelihood-based estimators is the only group of estimators explicitly able to prevent inadmissible estimates. In the process of optimization of the likelihood function, either a constrained optimization routine can be used, or a suitable parameter transformation can be undertaken. Due to the availability of the well known MAXLIK routine in Gauss, we chose the latter approach in the simulations and transformed the parameters a and λ as follows: $\eta = \ln[a/(1-a)]$; and $\delta = \ln(\lambda)$. The resulting estimates for η and δ were then transformed back via: $a = \exp(\eta)/[1+\exp(\eta)]$; and $\lambda = \exp(\delta)$.

The results of 5 000 simulation experiments are exposited by means of graphs and tables whose entries summarize the outcomes through the bias, percentage bias and mean squared error (MSE) for the 8 estimators.

5 Results for the PoINAR(1) model

The YW, GR, CLS, WCLS, AQL, CML and EML estimators are compared for the PoINAR(1) model (using Poisson innovations) in the first Monte Carlo experiment. We also discuss GMM based on both unconditional and conditional moment restrictions. The results for sample size 100 were found to be representative for the various simulation designs and are therefore described in detail. The results for smaller (larger) sample sizes lead to higher (smaller) biases and MSEs but not to different conclusions.

The tendency of the different estimators to produce negative estimates for a differs markedly. While YW, CLS and GR are not affected by this problem, WCLS, AQL and GMM are, but to differing extents. For a weak dependence structure ($a = 0.2$) WCLS produces invalid estimates for the parameter a in about 4% of the cases. As the dependence increases this number rapidly decreases to zero.

In contrast to this behaviour, the AQL estimator does produce (rarely) negative estimates for the parameter a, but these can occur no matter how large within the chosen range is the true value. This behaviour is undesirable. With regard to the two GMM estimators that based on the unconditional moment restrictions, invalid estimates of a occur about 5% of the time; GMM based upon conditional moment restrictions also suffers from this defect but very rarely. For moderate values of the dependence parameter (say $0.3 \leq a \leq 0.8$) the bias properties of GMM based on unconditional moment restrictions are very good and parallel those based on conditional moment restrictions. This, however, is not the case for either $a = 0.2$ or $a = 0.9$ and, therefore, in view of greater numerical instability difficulties encountered with the former we do not report it further. The acronym GMM will henceforth be used to denote the estimator based on conditional moment restrictions.

		YW	GR	CLS	WCLS	AQL	GMM	CML	EML
	a								
Bias	0.3	−0.0213	−0.0213	−0.0213	−0.0188	−0.0200	−0.0001	−0.0128	−0.0105
	0.5	−0.0280	−0.0280	−0.0278	−0.0263	−0.0265	−0.0030	−0.0081	−0.0051
	0.7	−0.0322	−0.0322	−0.0319	−0.0320	−0.0296	−0.0019	−0.0049	−0.0023
	0.9	−0.0412	−0.0412	−0.0387	−0.0398	−0.0376	−0.0008	−0.0027	−0.0014
%-Bias	0.3	−7.08	−7.08	−7.09	−6.27	−6.65	−0.02	−4.26	−3.50
	0.5	−5.60	−5.60	−5.57	−5.26	−5.30	−0.61	−1.61	−1.03
	0.7	−4.60	−4.60	−4.56	−4.58	−4.23	−0.28	−0.69	−0.33
	0.9	−4.58	−4.58	−4.30	−4.42	−4.18	−0.09	−0.30	−0.16
MSE	0.3	0.0097	0.0097	0.0097	0.0117	0.0100	0.0085	0.0086	0.0086
	0.5	0.0092	0.0092	0.0092	0.0106	0.0095	0.0059	0.0054	0.0054
	0.7	0.0071	0.0071	0.0071	0.0075	0.0073	0.0027	0.0022	0.0022
	0.9	0.0049	0.0049	0.0046	0.0048	0.0049	0.0004	0.0004	0.0003
	λ								
Bias	3.5	0.1034	−0.0421	0.1040	0.0917	0.0975	−0.0622	0.0612	0.0451
	2.5	0.1364	0.0048	0.1359	0.1281	0.1292	−0.0472	0.0362	0.0180
	1.5	0.1564	0.0025	0.1551	0.1556	0.1435	−0.0238	0.0184	0.0030
	0.5	0.2020	0.0091	0.1915	0.1968	0.1863	−0.0028	0.0084	0.0008
%-Bias	3.5	2.96	0.71	2.97	2.62	2.79	−1.78	1.75	1.29
	2.5	5.45	0.19	5.44	5.12	5.17	−1.89	1.45	0.72
	1.5	10.43	0.17	10.34	10.37	9.57	−1.59	1.23	0.20
	0.5	40.41	1.81	38.30	39.37	37.26	−0.57	1.67	0.17
MSE	3.5	0.2747	0.3558	0.2750	0.3220	0.2819	0.3045	0.2439	0.2432
	2.5	0.2517	0.1780	0.2507	0.2857	0.2579	0.1762	0.1505	0.1487
	1.5	0.1878	0.0655	0.1872	0.1973	0.1944	0.0700	0.0586	0.0573
	0.5	0.1307	0.0120	0.1244	0.1299	0.1311	0.0104	0.0093	0.0086

Table 1 *Bias, percentage bias and MSE for the estimators* $\hat{a}_{\bullet}$ *(top panel) and* $\hat{\lambda}_{\bullet}$ *(bottom panel) of the PoINAR(1) model at* $T = 100$.

The biases of representative estimators are depicted in the upper panels of Figure 1 (and the subsequent figure in a different context). Table 1 provides a fuller account, including the MSE, for some of the combinations of a and λ, together with the percentage bias in each estimator of both parameters. To economize on the use of space, only a representative sample of simulation results is presented. Further results can be obtained from the authors on request. It clearly emerges that GMM, CML and EML perform well for both a and λ with respect to bias as well as MSE. Of the two maximum likelihood estimators EML has small but discernible advantages over CML. Moreover, the advantages are even evident at $T = 500$ and so we would, on balance, advocate the use of the more complicated EML in the

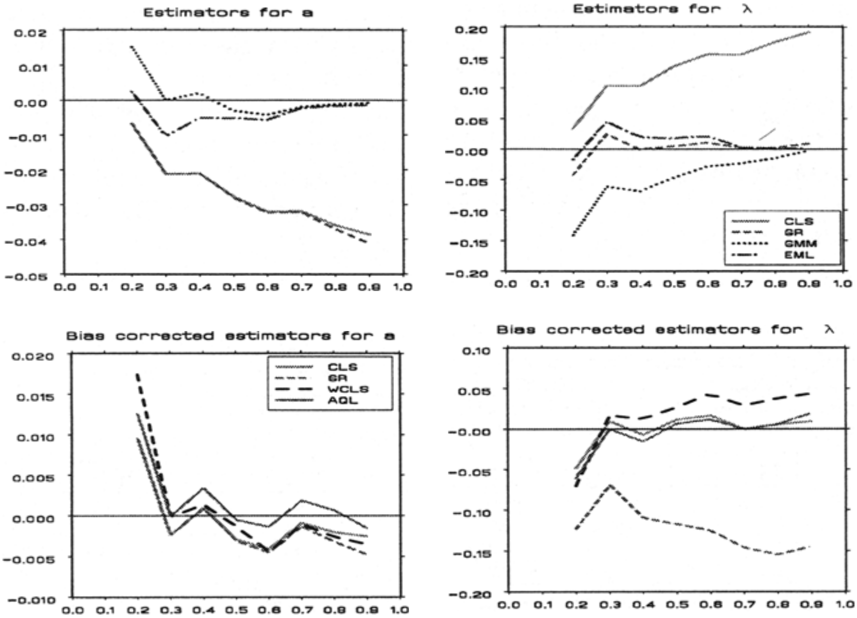

Figure 1 *Bias of representative estimators (upper panels) and bias corrected estimators (lower panels) for a and λ in the PoINAR(1) model for $T = 100$.*

context of a PoINAR(1) model.

The group of estimators containing YW, GR, ČLS, WCLS and AQL all exhibit a downward bias for a of approximately 5% at sample size 100. In view of the association between the estimators for a and λ, the estimates of the latter for these five estimators are generally upward biased, except for GR which has almost no bias. Observe that for large values of a the bias in λ can be quite high, e.g. around 40% for $a = 0.9$ and $\lambda = 0.5$. This is true for YW and the regression-based estimators, in which context there are rare instances of inadmissible estimates with WCLS and AQL.

Reducing T to 50 basically doubles all biases and increases the inadmissible results for WCLS to about 10% of the replications (almost all for $a = 0.2, 0.3$) and 7% for AQL, which are spread across all values of a. These arguments may mitigate against the use of these two regression-based estimators in samples of this size. For sample sizes less than this we would not recommend that an INAR(1) model be fitted to data because of the unacceptable finite sample properties that eventuate. It can be noted in passing from our Monte Carlo experiments that the sample size needed to obtain a reliable estimate of the parameters a and λ seems to be inversely related to the level of dependence found in the data. Freeland (1998, Chapter 4) reports the results of a small simulation experiment similar in spirit to aspects of our own. His findings are not at variance with ours and he also has a short discussion of asymptotic efficiency.

Summarizing our results in the case of the PoINAR(1) model, the likelihood-based estimation methods, GMM and possibly GR could be recommended. At this stage, certain of the moment-based and regression-based estimators might be regarded as inferior because of their finite sample biases. But these may be amenable to treatment by bias reduction methods and it is to this issue that we turn in the next section.

6 Bias correction in the PoINAR(1) model

The results of the last section indicate that the group of estimators comprising of YW, GR, CLS, WCLS and AQL may be unpromising because of unfavourable bias properties stemming from the estimation of the parameter a. But looking at the biases at different sample sizes it has already been mentioned that a doubling of T leads to a halving of the bias. Therefore, it may be the case that the bias function can be approximated by a linear function of T^{-1}. Moreover, the bias of estimators of a may be well approximated by a linear function of the parameter. This is suggested by some of the entries in Table 1; see also the upper panels of Figure 1.

Bias correction in Gaussian AR(1) models has attracted the attention of researchers for many years, beginning with Kendall (1954) and Marriott and Pope (1954). See also Andrews (1993) and MacKinnon and Smith (1998) for more recent contributions and Shaman and Stine (1988) for work extending the approach to higher order autoregressive models. An approximately unbiased estimator for the dependence parameter in the Gaussian AR(1) model has been derived by Orcutt and Winokur (1969) using Monte Carlo techniques. Given the links between the Gaussian AR(1) model and the PoINAR(1) model afforded by the CLAR(1) framework, it is of interest to see if bias correction methods used in the Gaussian context also improve performance in the PoINAR(1) case. We, therefore, consider the following bias corrected estimator

$$\check{a}_{\bullet} = \frac{1}{T-3}\left(T\,\hat{a}_{\bullet} + 1\right) \tag{6}$$

for the parameter a, where $\hat{a}_{\bullet}$ indicates the following estimators of a for which bias correction is available, viz. YW, GR, CLS, WCLS and AQL. Corresponding estimators $\check{\lambda}_{\bullet}$ arise naturally from using $\check{a}_{\bullet}$ rather than $\hat{a}_{\bullet}$ in their construction.

In order to evaluate the effects of the bias correction scheme (6) under the PoINAR(1) scenario, we conducted a Monte Carlo experiment. The simulations were carried out along the lines of Section 4, but using the bias corrected versions of the above mentioned estimators. The lower panels of Figure 1 depict the results for certain estimators and some results are presented in more detail in Table 2 for $T = 100$. At this sample size the bias in all five estimators of a disappears for all levels of interest (except, in fact, for $a = 0.2$, the results for which are not explicitly given). With respect to the estimation of the parameter λ it can be observed that for YW, CLS, WCLS and AQL the bias in these estimators also disappears.

	a	YW	GR	CLS	WCLS	AQL
Bias	0.3	−0.0023	−0.0023	−0.0024	−0.0002	−0.0001
	0.5	−0.0031	−0.0031	−0.0029	−0.0013	−0.0005
	0.7	−0.0013	−0.0013	−0.0009	−0.0011	0.0019
	0.9	−0.0048	−0.0048	−0.0025	−0.0035	−0.0016
%-Bias	0.3	−0.77	−0.77	−0.78	−0.05	−0.01
	0.5	−0.62	−0.62	−0.58	−0.26	−0.11
	0.7	−0.18	−0.18	−0.13	−0.15	0.26
	0.9	−0.53	−0.53	−0.28	−0.39	−0.17
MSE	0.3	0.0099	0.0099	0.0099	0.0121	0.0104
	0.5	0.0090	0.0090	0.0090	0.0105	0.0093
	0.7	0.0065	0.0065	0.0064	0.0069	0.0068
	0.9	0.0034	0.0034	0.0032	0.0034	0.0038
	λ					
Bias	3.5	0.0087	−0.0694	0.0093	0.0165	−0.0002
	2.5	0.0117	−0.1174	0.0113	0.0266	0.0064
	1.5	0.0013	−0.1462	−0.0001	0.0288	−0.0001
	0.5	0.0191	−0.1448	0.0094	0.0434	0.0192
%-Bias	3.5	0.25	−1.98	0.27	0.47	−0.01
	2.5	0.47	−4.70	0.45	1.06	0.26
	1.5	0.09	−9.74	−0.00	1.92	−0.00
	0.5	3.82	−28.97	1.88	8.68	3.85
MSE	3.5	0.2762	0.3558	0.2764	0.3270	0.3003
	2.5	0.2431	0.1865	0.2421	0.2770	0.2487
	1.5	0.1692	0.0825	0.1689	0.1747	0.1914
	0.5	0.0888	0.0336	0.0858	0.0856	0.0970

Table 2 *Bias, percentage bias and MSE for the bias corrected estimators* $\hat{a}_\bullet$ *and* $\hat{\lambda}_\bullet$ *of the PoINAR(1) model at* $T = 100$.

This is not true for the GR estimator where a limited downward bias (though up to 30% at $a = 0.9$) is introduced by virtue of the bias correction scheme. The bias reduction proposed, therefore, does not aid the performance of GR; without it a is estimated with bias but λ not, whilst with it the reverse obtains.

At sample size 50 the same results broadly hold, though a serious upward bias in λ when $a = 0.9$ of between 20 and 30% can arise with YW and the regression-based estimators. At this smaller sample size WCLS and AQL give inadmissible values some 20% of the time spread across all true values of a. Even YW and CLS exhibit this tendency 8 − 12% of the time (with $\check{a}_\bullet > 1$). Of course at $a = 0.9$ the bias correction scheme introduces an approximate 8% upward correction at $T = 50$. Thus bias corrected estimates of a larger than one may not be surprising.

The conclusions from our simulation studies under the PoINAR(1) scenario can now be updated as follows. The likelihood-based methods perform very well, as does GMM. After applying the bias correction scheme, YW and CLS both work well. The other regression-based estimators behave less satisfactorily because of the evident instability in the estimation results reported.

7 Effects of overdispersion

The framework of the class of CLAR(1) models with discrete support is employed now to model the phenomenon of overdispersion in time series of counts. Extra binomial variation is a common phenomenon in observed count data. We first present an example of a count data set exhibiting overdispersion and estimate the parameters of an INAR(1) model using bias corrected Yule-Walker and CLS estimators on the one hand and by maximum likelihood on the other. This will be seen to provide strong motivation for the remaining discussion in this section. We use monthly strike data published by the U.S. Bureau of Labor Statistics for the period January 1994 to December 2002, a total of 108 observations. The counts consist of the number of work stoppages leading to 1 000 workers or more being idle in effect in the period. The observations range from zero to 14 ongoing work stoppages in a particular month and have a sample mean and variance 4.94 and 7.92, respectively. The first 10 sample autocorrelations are provided in the top panel of Table 3. They decay exponentially suggesting a CLAR(1) model may be appropriate. This is corroborated by the sample partial autocorrelation function (not shown).

k	1	2	3	4	5	6	7	8	9	10
SAC(k)	0.573	0.346	0.153	0.117	0.029	0.009	−0.028	−0.084	−0.016	0.048

$\breve{a}_{YW} =$ 0.6049 (0.0818)	$\breve{\lambda}_{YW} =$ 1.9536 (0.4020)
$\breve{a}_{CLS} =$ 0.6103 (0.0813)	$\breve{\lambda}_{CLS} =$ 1.9410 (0.4027)
$\hat{a}_{CML} =$ 0.5061 (0.0561)	$\hat{\lambda}_{CML} =$ 2.4603 (0.2989)
$\hat{a}_{EML} =$ 0.5069 (0.0549)	$\hat{\lambda}_{CML} =$ 2.3937 (0.2917)

Table 3 *Sample autocorrelations (SAC) for various lags k (top panel) and estimation results (bottom panel) for the strikes data. Asymptotic standard errors are provided in parentheses.*

The lower part of Table 3 provides estimates for the Yule-Walker (YW) and the conditional least squares (CLS) estimators with bias correction and maximum likelihood estimates based on the conditional (CML) and the exact (EML) likelihood function. Asymptotic standard errors are given in parentheses. Note that the YW and CLS estimates, while being close to one another, differ by about 20% from the two maximum likelihood estimates. In the simulations reported in Section 6 (and also in Section 5) it is rare indeed to find such divergence of individual estimates from one to another based on the same realization of data. This real data set does, of course, exhibit overdispersion and one is led to question whether it is wise to rely on the assumption of equidispersion inherent in Poisson innovations in making recommendations on the desirability of estimators. The evidence of Sections 5 and 6 is predicated on this assumption; this section relaxes it and provides convincing evidence explaining what is seen in Table 3.

Following Grunwald et al. (2000) we choose two types of models from the class of CLAR(1) models as exemplars of the situation of overdispersion. These are: the fixed coefficient first order autoregression; and the random coefficient autoregression, both with a NB innovation distribution. These models have previously been discussed in some detail by Al-Osh and Aly (1992) and McKenzie (1986). In what follows we briefly present the two types of models and list some of their properties. Then the design parameters for further Monte Carlo experiments will be outlined and, finally, the behaviour of the various estimators under both overdispersion scenarios will be described. As a practical matter, a user of these models will not know the distribution of the (unobservable) innovations. But any application of GMM, or likelihood-based estimators, for example, needs to make distributional, or at least moment, assumptions. The computations summarized below assume a misspecified Poisson form for the innovations, rather than their true overdispersed form. As might be expected, we find evidence of deleterious effects on estimators based on assuming particular aspects of innovation behaviour erroneously.

The first type of model is of the form (3) but with $W_t \sim \mathrm{NB}(n,p)$. The resulting marginal distribution is not necessarily NB, but exhibits overdispersion which is relevant for our analysis. Note that the interpretation as the estimation of the mean of the innovation process by means of the estimators $\hat{\lambda}_\bullet$ (where the notation is as introduced in (6)) is retained by the fact that, under the NB specification, $\mathrm{E}(W_t) = \lambda = np/(1-p)$. The second type of model is the random coefficient autoregressive model

$$X_t = a_t \circ X_{t-1} + W_t \qquad (t = 0, \pm 1, \pm 2, \ldots)\,. \tag{7}$$

It is assumed that the random variable a_t is beta distributed in order to keep the range of its possible values between 0 and 1 and so the notion of thinning is retained (McKenzie, 1986). The innovation W_t is again $NB(n,p)$. The resulting marginal distribution of X_t is not of central interest, but the fact that the process (7) is able to introduce overdispersion into the simulated counts is. What should be noted in passing is the possibility that the random coefficient autoregression introduces an extra source of overdispersion compared with the fixed coefficient model with NB innovations.

We generated data sets under both models and investigated the behaviour of the estimators given in Section 3. The design parameters of these experiments were chosen in such a way as to parallel the simulations under the PoINAR(1) scenario, but with the presence of overdispersion in the data. Note that the level of overdispersion now depends on the parameters of the negative binomial as well as that of the beta distribution. Since these parameters are changed across the different experiments in order to generate various degrees of dependence, the level of overdispersion varies; it is between 2 and 3 in the fixed coefficient and 3 and 4 in the random coefficient cases, respectively.

In the fixed coefficient CLAR(1) model with a NB innovation distribution, the mean of process remains 5, as it does under the random coefficient scenario. The thinning parameter a was allowed to vary from 0.2 to 0.9 as previously. Fixing p to 2/3, the index parameter n of the NB innovation distribution was then chosen according to $n = 5/2 \cdot (1 - a)$. (This relationship is easily derived using the fact that the mean of the process was set to 5 and the mean of the NB innovation distribution used is $np/(1 - p)$.) Results are again based on 5 000 simulation runs for $T = 50, 100, 500$, but we discuss the $T = 100$ case only. The same set of estimators discussed under equidispersion continues unchanged under both overdispersion scenarios. Recall that those expressly derived using properties of the Poisson distribution (e.g. GR, GMM, CML and EML) are being employed in situations for which they were not expressly designed. It is of interest to see how this impinges upon their performance. Where applicable, bias reduction methods can continue to be employed, if desired.

In the random coefficient CLAR(1) model with a NB innovation distribution the following design parameters for the Monte Carlo were used. To mimic the dependence structure of the experiments above now requires us to use combinations of the two parameters of the beta distribution in such a way that the mean thinning rate goes from 0.2 to 0.9 in steps of 0.1. The scale parameter of the NB distribution was again set to 2/3 while the index parameter, n, was computed according to $n = 5/2 \cdot [1 - \mathrm{E}(a_t)]$, where $\mathrm{E}(a_t)$ is the mean of the beta distribution.

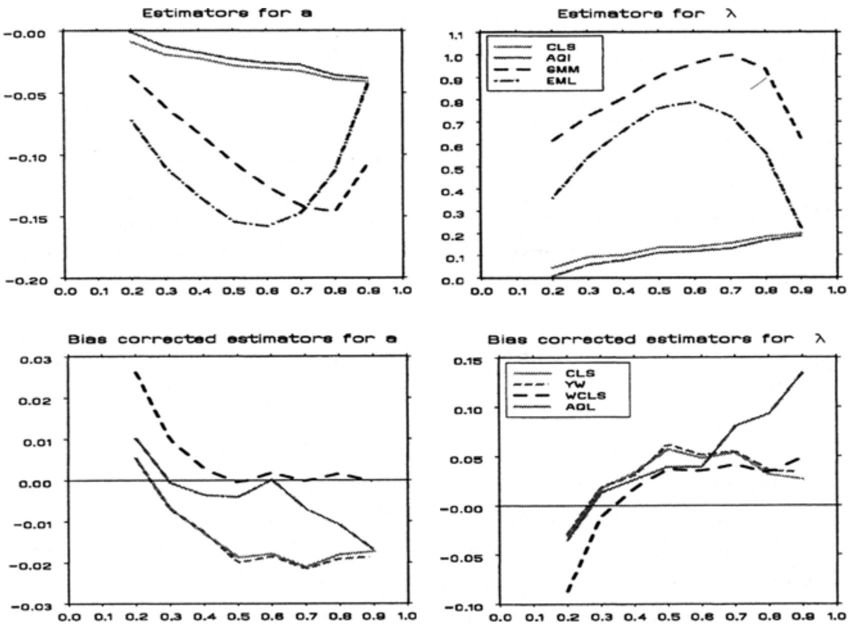

Figure 2 *Bias of representative estimators (upper panels) and bias corrected estimators (lower panels) for a and λ in the CLAR(1) model with NB innovations and random coefficient autoregression for* $T = 100$.

		YW	GR	CLS	WCLS	AQL	GMM	CML	EML
	a								
Bias	0.3	−0.0179	−0.0179	−0.0180	−0.0100	−0.0104	−0.0613	−0.1068	−0.1056
	0.5	−0.0281	−0.0281	−0.0276	−0.0306	−0.0222	−0.1076	−0.1464	−0.1444
	0.7	−0.0325	−0.0325	−0.0318	−0.0350	−0.0268	−0.1395	−0.1349	−0.1322
	0.9	−0.0411	−0.0411	−0.0381	−0.0440	−0.0383	−0.0897	−0.0351	−0.0330
%-Bias	0.3	−5.96	−5.96	−5.99	−3.33	−3.46	−20.43	−35.59	−35.19
	0.5	−5.63	−5.63	−5.51	−6.11	−4.43	−21.52	−29.28	−28.88
	0.7	−4.65	−4.65	−4.54	−5.00	−3.83	−19.93	−19.27	−18.89
	0.9	−4.57	−4.57	−4.24	−4.89	−4.25	−9.96	−3.90	−3.66
MSE	0.3	0.0096	0.0096	0.0095	0.0140	0.0105	0.0102	0.0157	0.0155
	0.5	0.0089	0.0089	0.0088	0.0144	0.0100	0.0174	0.0256	0.0251
	0.7	0.0074	0.0074	0.0072	0.0114	0.0083	0.0240	0.0219	0.0212
	0.9	0.0050	0.0050	0.0045	0.0068	0.0077	0.0102	0.0020	0.0019
	λ								
Bias	3.5	0.0839	5.4000	0.0852	0.0460	0.0461	0.7158	0.5340	0.5244
	2.5	0.1411	3.3951	0.1384	0.1549	0.1128	0.9043	0.7415	0.7274
	1.5	0.1515	1.7645	0.1475	0.1655	0.1229	0.9570	0.6676	0.6511
	0.5	0.1971	0.5408	0.1841	0.2150	0.1878	0.5147	0.1789	0.1701
%-Bias	3.5	2.40	154.29	2.43	1.31	1.32	20.45	15.26	14.98
	2.5	5.65	135.80	5.53	6.19	4.51	36.17	29.66	29.10
	1.5	10.10	117.63	9.83	11.03	8.19	63.80	44.51	43.41
	0.5	39.42	108.16	36.83	43.01	37.56	102.94	35.78	34.03
MSE	3.5	0.3368	33.6792	0.3387	0.4579	0.3621	0.8085	0.5430	0.5303
	2.5	0.2925	13.8195	0.2916	0.4421	0.3236	1.0781	0.7949	0.7704
	1.5	0.2219	4.0433	0.2185	0.3317	0.2474	1.1437	0.6597	0.6333
	0.5	0.1444	0.4998	0.1368	0.2011	0.2422	0.4319	0.0948	0.0863

Table 4 *Bias, percentage bias and MSE for the estimators* $\hat{a}_\bullet$ *(top panel) and* $\hat{\lambda}_\bullet$ *(bottom panel) of the fixed coefficient autoregressive model with NB innovations at* $T = 100$.

Some results of the Monte Carlo experiments under both the fixed and random coefficient autoregression scenarios with no bias correction method applied at this stage are provided in Tables 4 and 5 for $T = 100$. For illustrative purposes Figure 2 depicts representative results for the random coefficient autoregression case only. The likelihood-based estimators CML and EML under the overdispersion scenarios are based on the wrong likelihood, although they might be considered quasi-likelihood (compare continuous models based on a Gaussian likelihood). But it clearly emerges from the simulation results that both CML and EML evidence wholly unsatisfactory performance. This includes high biases in estimating the dependence structure as well as the mean of the innovations. A similar problem is found with GMM. So, in essence, all 3 estimators that did well with no bias correction under the PoINAR(1) model fail completely when extra binomial variation of either type is present.

		YW	GR	CLS	WCLS	AQL	GMM	CML	EML
	a								
Bias	0.3	−0.0191	−0.0191	−0.0189	−0.0086	−0.0121	−0.0612	−0.1113	−0.1101
	0.5	−0.0282	−0.0282	0.0279	−0.0268	−0.0228	−0.1062	−0.1564	−0.1545
	0.7	−0.0332	−0.0332	−0.0324	−0.0328	−0.0272	0.1420	−0.1505	−0.1476
	0.9	−0.0430	−0.0430	−0.0408	0.0436	−0.0382	−0.1064	−0.0443	−0.0423
%-Bias	0.3	−6.37	−6.37	−6.32	−2.87	−4.05	−20.40	−37.07	−36.70
	0.5	−5.64	−5.64	−5.57	−5.36	−4.57	−21.25	−31.27	−30.89
	0.7	−4.75	−4.75	−4.62	−4.69	−3.88	−20.29	−21.49	−21.08
	0.9	−4.77	−4.77	−4.53	−4.84	−4.25	−11.82	−4.92	−4.70
MSE	0.3	0.0108	0.0108	0.0107	0.0142	0.0114	0.0111	0.0169	0.0167
	0.5	0.0102	0.0102	0.0102	0.0147	0.0112	0.0183	0.0293	0.0287
	0.7	0.0080	0.0080	0.0079	0.0108	0.0089	0.0255	0.0269	0.0260
	0.9	0.0054	0.0054	0.0050	0.0068	0.0069	0.0141	0.0032	0.0031
	λ								
Bias	3.5	0.0919	5.8257	0.0916	0.0409	0.0576	0.7226	0.5513	0.5416
	2.5	0.1384	3.7516	0.1367	0.1320	0.1117	0.9044	0.7781	0.7638
	1.5	0.1578	2.0925	0.1556	0.1598	0.1299	1.0017	0.7452	0.7306
	0.5	0.2054	0.6726	0.1970	0.2153	0.1864	0.6260	0.2295	0.2182
%-Bias	3.5	2.63	166.45	2.62	1.17	1.65	20.65	15.75	15.84
	2.5	5.54	150.07	5.47	5.28	4.47	36.18	31.12	30.55
	1.5	10.52	139.50	10.37	10.65	8.66	66.78	49.68	48.70
	0.5	41.08	134.51	39.40	43.06	37.28	125.20	45.89	43.63
MSE	3.5	0.3532	38.9958	0.3526	0.4469	0.3706	0.8237	0.5657	0.5542
	2.5	0.3059	16.6189	0.3080	0.4262	0.3305	1.0904	0.8688	0.8433
	1.5	0.2363	5.5844	0.2383	0.3227	0.2685	1.2549	0.7913	0.7644
	0.5	0.1493	0.7396	0.1454	0.2038	0.2121	0.5921	0.1434	0.1320

Table 5 *Bias, percentage bias and MSE for the estimators* $\hat{a}_\bullet$ *(top panel) and* $\hat{\lambda}_\bullet$ *(bottom panel) of the random coefficient autoregressive model with NB innovations at* $T = 100$.

	a	YW	GR	CLS	WCLS	AQL	λ	YW	GR	CLS	WCLS	AQL
Bias	0.3	0.0010	0.0010	0.0012	0.0080	0.0096	3.5	−0.0105	5.1603	−0.0097	−0.0054	−0.0524
	0.5	−0.0032	−0.0032	−0.0028	−0.0058	0.0035	2.5	0.0162	3.1077	0.0142	0.0761	−0.0033
	0.7	−0.0016	−0.0016	−0.0008	−0.0042	0.0043	1.5	−0.0031	1.4415	−0.0072	0.0633	−0.0281
	0.9	−0.0049	−0.0049	−0.0020	−0.0092	−0.0028	0.5	0.0149	0.2233	0.0019	0.0902	0.0152
%-Bias	0.3	0.35	0.35	0.39	2.66	3.20	3.5	−0.30	147.44	−0.28	−0.15	−1.50
	0.5	−0.65	−0.65	−0.56	−1.16	0.69	2.5	0.65	124.31	0.57	3.05	−0.13
	0.7	−0.23	−0.23	−0.11	−0.59	0.62	1.5	−0.21	96.10	−0.48	4.22	−1.87
	0.9	−0.54	−0.54	−0.22	−1.02	−0.32	0.5	2.99	44.66	0.38	18.04	3.04
MSE	0.3	0.0099	0.0099	0.0098	0.0152	0.0112	3.5	0.3396	31.0013	0.3413	0.4744	0.3759
	0.5	0.0086	0.0086	0.0085	0.0144	0.0100	2.5	0.2774	11.7923	0.2774	0.4292	0.5280
	0.7	0.0067	0.0067	0.0066	0.0108	0.0085	1.5	0.1993	2.8811	0.1970	0.3048	0.2395
	0.9	0.0035	0.0035	0.0032	0.0051	0.0063	0.5	0.0976	0.1831	0.0941	0.1459	0.2932

Table 6 *Bias, percentage bias and MSE for the bias corrected estimators* $\hat{a}_\bullet$ *and* $\hat{\lambda}_\bullet$ *of the fixed coefficient autoregressive model with NB innovations at* $T = 100$.

	a	YW	GR	CLS	WCLS	AQL	λ	YW	GR	CLS	WCLS	AQL
Bias	0.3	−0.0071	−0.0071	−0.0069	0.0099	−0.0006	3.5	0.0186	8.0351	0.0176	−0.0112	0.0129
	0.5	−0.0200	−0.0200	−0.0188	−0.0006	−0.0041	2.5	0.0622	6.0612	0.0570	0.0369	0.0391
	0.7	−0.0216	−0.0216	−0.0211	−0.0003	−0.0069	1.5	0.0550	3.6302	0.0536	0.0415	0.0803
	0.9	−0.0187	−0.0187	−0.0174	−0.0004	−0.0170	0.5	0.0341	0.9286	0.0267	0.0483	0.1345
%-Bias	0.3	−2.35	−2.35	−2.28	3.28	−0.19	3.5	0.53	229.57	0.50	−0.32	0.37
	0.5	−4.00	−4.00	−3.76	−0.13	−0.82	2.5	2.49	242.45	2.28	1.48	1.56
	0.7	−3.08	−3.08	−3.02	−0.04	−0.99	1.5	3.67	242.02	3.57	2.77	5.35
	0.9	−2.08	−2.08	−1.94	−0.05	−1.88	0.5	6.81	185.71	5.34	9.66	26.91
MSE	0.3	0.0147	0.0147	0.0146	0.0154	0.0152	3.5	0.3758	72.9509	0.3734	0.4261	0.4200
	0.5	0.0154	0.0154	0.0152	0.0142	0.0177	2.5	0.3399	42.3486	0.3377	0.3424	0.4623
	0.7	0.0121	0.0121	0.0121	0.0098	0.0161	1.5	0.2289	15.9774	0.2300	0.2082	0.5873
	0.9	0.0058	0.0058	0.0057	0.0045	0.0128	0.5	0.1011	1.5309	0.1008	0.1049	0.7272

Table 7 *Bias, percentage bias and MSE for the bias corrected estimators* $\hat{a}_\bullet$ *and* $\hat{\lambda}_\bullet$ *of the random coefficient autoregressive model with NB innovations at* $T = 100$.

The bias correction scheme (6) is now applied to all moment-based estimators except GMM, for which it is unavailable, and to the regression-based estimators. The low biases in a ($< 3\%$ downward in almost all cases) are noteworthy. Tables 6 and 7 and the lower panels of Figure 2 summarize the relevant Monte Carlo results. Additionally, there are only small upward biases for λ for all remaining estimators except GR, unless $\lambda = 0.5$; see Table 7. It is, perhaps, worth noting in passing that Tables 4 and 5 indicate that biases of the order of 20% occur with CML and EML for moderate values of a, whereas Tables 6 and 7 indicate that YW and CLS exhibit almost no bias for such values. This is congruent with the evidence from the empirical example at the beginning of this section. Finally, AQL and, to a lesser extent, WCLS give inadmissible results up to 14% of the time, spread across all true values of a. Even YW and CLS are affected by this phenomenon about 5% of the time for the lowest dependence structure analyzed. On balance though, the overall performance of the two last mentioned estimators is quite satisfactory.

8 Conclusions

We have examined a wide range of estimators for the two parameters of first order autoregressive models for count data. The performance of these estimators is assessed in three different frameworks: under equidispersion; under overdispersion resulting from negative binomial innovations; and, finally, overdispersion is additionally induced by random coefficient variation. In the first situation likelihood-based estimators based upon assuming Poisson innovations work extremely well at sample sizes from 100 upward, as does an estimator based upon a conditional Generalized Method of Moments approach. In this case other moment-based and regression-based estimators perform in quite predictable ways. But they frequently evidence downward bias in estimation of the parameter governing dependence structure and concomitant upward bias in the innovation mean. It turns out that many of these bias problems can be attenuated, and often eliminated almost entirely, by use of simple bias correction techniques.

Where overdispersion is present in data those estimators specifically designed for the Poisson context no longer perform well. This may be important because it indicates that the Poisson assumption is not innocuous and that it will not be appropriate to maintain the notion that it can continue to be used when it does not hold. We do not recommend this approach with low integer counts as a means of providing an acceptable basis for a pseudo-likelihood function, for example, a course of action sometimes adopted with the Gaussian specification in a continuous world. In turns out, however, that our evidence quite strongly indicates that bias corrected estimates based on suitable Yule-Walker and least squares approaches perform appealingly in all cases considered.

Acknowledgements The authors are grateful for the constructive comments and guidance of an Editor and two anonymous referees.

Appendix A

A.1 Moment-based estimators

For the PoINAR(1) model the Yule-Walker (YW) estimators for the parameter a is just the first order sample autocorrelation. Jung (1999, Chapter 4) discusses several variants of this estimator. It turns out that in small and medium size samples the following version is to be recommended

$$\hat{a}_{YW} = \frac{(T-1)^{-1}\sum_{t=2}^{T}(X_t - \bar{X})(X_{t-1} - \bar{X})}{T^{-1}\sum_{t=1}^{T}(X_t - \bar{X})^2}\,, \tag{A.1}$$

where $\bar{X} = T^{-1}\sum_{t=1}^{T} X_t$.

Estimation of λ is based on the moment condition arising from the marginal distribution and is $\mathrm{E}(X_t) = \lambda/(1-a)$

$$\hat{\lambda}_{YW} = (1 - \hat{a}_{YW})\, T^{-1}\sum_{t=1}^{T} X_t\,. \tag{A.2}$$

The YW estimators serve as a benchmark in many studies concerned with the estimation of the parameters of the INAR(1) model.

Greene (2000, Chapter 19.9.7) proposes a method of moments estimator based on the following two moments: $\mathrm{Var}(X_t) = \lambda/(1-a)$; and $\mathrm{Cov}(X_t, X_{t-1}) = a\lambda/(1-a)$. While the resulting estimator for a is identical to the the Yule-Walker estimator (A.1), a new estimator for λ results given by

$$\hat{\lambda}_{GR} = (1 - \hat{a}_{YW})\, T^{-1}\sum_{t=1}^{T}(X_t - \bar{X})^2\,. \tag{A.3}$$

Generalized Method of Moments (GMM) estimation in the context of the PoINAR(1) model has been introduced by Brännäs (1994); the close relationship to the Yule-Walker and method of moments estimators is discussed by the author and a range of moment conditions is provided. In principle both unconditional as well as conditional moments can serve as a basis for the GMM estimation. We experimented with both types and found that, in general, the conditional moments fulfil the moment restrictions more satisfactorily as compared to their unconditional counterparts.

That is to say, for moderate values of a the behaviour of the two variants is very similar but this is not the case when a takes extreme values of 0.2 and 0.9 in our experiments. In addition we found numerical instability problems in the course of the estimation process of GMM based on unconditional moment restrictions. Consequently, we exclude estimators based on the unconditional moments from our detailed analysis and concentrate on Generalized Methods of Moments estimators based on conditional moment restrictions in the latter part of the paper.

In general such estimation is based on minimization of the quadratic form

$$Q_{GMM} = \mathbf{m}(\boldsymbol{\theta})'\hat{\mathbf{W}}^{-1}\mathbf{m}(\boldsymbol{\theta}) , \tag{A.4}$$

where $\hat{\mathbf{W}}$ is a symmetric, positive definite weighting matrix and $\mathbf{m}(\boldsymbol{\theta})$ is the vector of moment conditions. Usually the estimation is carried out in two steps. In the first step $\hat{\mathbf{W}}$ is set to the identity matrix $\boldsymbol{I}$. The resulting GMM estimators $\widehat{\boldsymbol{\theta}}$ are consistent, but not necessarily efficient. In a second step $\widehat{\boldsymbol{\theta}}$ is used to obtain $\hat{\mathbf{W}}$ which, in turn, leads to second step efficient estimators for the parameters contained in $\boldsymbol{\theta}$. For further details see e.g. Greene (2000, Chapter 11). Preliminary work not reported here, and the findings of Brännäs (1994), suggest that the gain of using $\hat{\mathbf{W}}$ instead of $\boldsymbol{I}$ in (A.4) is negligible. We, therefore, report only GMM results based on the first-step estimation.

We experimented with various conditional moment restrictions and found the following three to work best:

(i) conditional first moment condition

$$\mathrm{E}(X_t - aX_{t-1} - \lambda | X_{t-1}) = 0 ; \tag{A.5}$$

(ii) conditional variance condition

$$\mathrm{E}\left[(X_t - aX_{t-1} - \lambda)^2 - a(1-a)X_{t-1} - \lambda \,|\, X_{t-1}\right] = 0 ; \tag{A.6}$$

(iii) and conditional covariance of X_t and X_{t-1} condition

$$\mathrm{E}\left[(X_t - aX_{t-1} - \lambda) \cdot (X_{t-1} - aX_{t-2} - \lambda) \,|\, X_{t-1}, X_{t-2}\right] = 0 . \tag{A.7}$$

For information we report that the unconditional moment restrictions that we found most satisfactory were based upon: the unconditional first moment; the unconditional uncentred second moment; and the uncentred unconditional sample analogue of
$\mathrm{E}(X_t \cdot X_{t-1})$.

A.2 Regression-based estimators

The conditional least squares (CLS) estimator of Klimko and Nelson (1978) was used by Winnicki (1988) and Wei and Winnicki (1989) to estimate the parameters of a simple BGWI process. In the course of the derivation of the CLS estimators for this kind of problem only minimal distributional assumptions are employed (see Wei and Winnicki, 1989). As a result we expect the CLS estimator to be quite robust with respect to deviations from a Poisson assumption in the INAR(1) model.

The CLS estimator is based on the criterion function

$$Q(\boldsymbol{\theta}) = \sum_t (X_t - E(X_t|\mathcal{F}_{t-1}))^2 , \tag{A.8}$$

where $\mathcal{F}_{t-1} = X_{t-1}$ in the PoINAR(1) model. $Q(\boldsymbol{\theta})$ has to be minimized with respect to the parameters a and λ in the PoINAR(1) model. The resulting estimators can be derived explicitly and are of the following form

$$\hat{a}_{CLS} = \frac{(T-1)\sum_{t=2}^T X_t X_{t-1} - \sum_{t=2}^T X_t \sum_{t=2}^T X_{t-1}}{(T-1)\sum_{t=2}^T X_{t-1}^2 - \left(\sum_{t=2}^T X_{t-1}\right)^2} \tag{A.9}$$

and

$$\hat{\lambda}_{CLS} = (T-1)^{-1}\left(\sum_{t=2}^T X_t - \hat{a}_{CLS}\sum_{t=2}^T X_{t-1}\right) . \tag{A.10}$$

The CLS estimator has been studied among others in Al-Osh and Alzaid (1987), Brännäs (1994) and Park and Oh (1997).

As provided in Section 2 the conditional variance of $X_t|X_{t-1}$ is not constant in the PoINAR(1) model. To explicitly account for this heteroskedasticity several weighted conditional least squares estimators have been proposed in the literature. The appropriate criterion function to be minimized takes the form

$$Q^*(\boldsymbol{\theta}) = \sum_t \left[\frac{X_t - E(X_t|\mathcal{F}_{t-1})}{\sqrt{\mathrm{Var}(X_t|\mathcal{F}_{t-1})}}\right]^2 \tag{A.11}$$

(see Winnicki, 1988, where a motivation for the estimators is provided). In the general setting of the BGWI process $\mathrm{Var}(X_t|\mathcal{F}_{t-1}) = c^2 X_{t-1} + d^2$, where c^2 is the variance of the offspring component (in our INAR(1) notation that of the Bernoulli distributed $Y_{i,t}$'s) and d^2 is the variance of the immigration rate. Since under the minimal assumptions employed in the BGWI literature both variances are unknown, Winnicki (1988) and Wei and Winnicki (1989) propose simplified weights given by $(1+X_{t-1})^{-1/2}$. The resulting estimators, hereafter WCLS, (with the appropriate modification of the multiplicative factor involving the sample size, to compensate for the loss of degrees of freedom) are as follows

$$\hat{a}_{WCLS} = \frac{\sum_{t=2}^{T} X_t \sum_{t=2}^{T}(1 + X_{t-1})^{-1} - (T-1)\sum_{t=2}^{T} X_t(1 + X_{t-1})^{-1}}{\sum_{t=2}^{T}(1 + X_{t-1})\sum_{t=2}^{T}(1 + X_{t-1})^{-1} - (T-1)^2} \tag{A.12}$$

$$\hat{\lambda}_{WCLS} = \frac{\sum_{t=2}^{T} X_{t-1}\sum_{t=2}^{T} X_t(1+X_{t-1})^{-1} - \sum_{t=2}^{T} X_t \sum_{t=2}^{T} X_{t-1}(1+X_{t-1})^{-1}}{\sum_{t=2}^{T}(1 + X_{t-1})\sum_{t=2}^{T}(1 + X_{t-1})^{-1} - (T-1)^2}\,. \tag{A.13}$$

An alternative, equivalent way of writing (A.13) is given by

$$\hat{\lambda}_{WCLS} = \left(\sum_{t=2}^{T}(1 + X_{t-1})^{-1}\right)^{-1} \times \tag{A.13'}$$
$$\left(\sum_{t=2}^{T} X_t(1 + X_{t-1})^{-1} - \hat{a}_{WCLS}\sum_{t=2}^{T} X_{t-1}(1 + X_{t-1})^{-1}\right).$$

Estimators of this type are considered in Brännäs (1994).

The general principle of estimating equations (Godambe, 1960, and subsequent papers) has been applied to the estimation of branching processes with immigration by Heyde and Lin (1992). (The paper is reprinted in part in Heyde, 1997, Chapter 5 under the heading "Asymptotic Quasi-Likelihood".) Under minimal distributional assumptions the estimation is based on the vector-valued quasi-score estimation function

$$Q_T = \begin{bmatrix} \sum_{t=1}^{T} X_{t-1}(c^2 X_{t-1} + d^2)^{-1}(X_t - aX_{t-1} - \lambda) \\ \sum_{t=1}^{T}(c^2 X_{t-1} + d^2)^{-1}(X_t - aX_{t-1} - \lambda) \end{bmatrix} \tag{A.14}$$

using the martingale $\{\sum_{s=1}^{t} X_s - \mathrm{E}(X_s|\mathcal{F}_{s-1})\}$, where $\mathcal{F}_{s-1}$ is an appropriate filtration. Here c^2 denotes the (unknown) variance of the offspring distribution and d^2 the (unknown) variance of the immigration distribution.

Note the close relationship to the WCLS estimators discussed above. The difference lies in the fact that, instead of the simplified weights $(1 + X_{t-1})^{-1/2}$, here $(c^2 X_{t-1} + d^2)^{-1/2}$ are used and the unknown variances are replaced by consistent estimators. (Heyde and Lin, 1992 term these asymptotic quasi-likelihood (AQL) estimators).

Closed form solutions for the AQL are given by

$$\hat{a}_{AQL} = \frac{\sum_{t=2}^{T} w_t^{-1} \sum_{t=2}^{T} X_t X_{t-1} w_t^{-1} - \sum_{t=2}^{T} X_t w_t^{-1} \sum_{t=2}^{T} X_{t-1} w_t^{-1}}{\sum_{t=2}^{T} w_t^{-1} \sum_{t=2}^{T} X_{t-1}^2 w_t^{-1} - \left(\sum_{t=2}^{T} X_{t-1} w_t^{-1}\right)^2} \tag{A.15}$$

and

$$\hat{\lambda}_{AQL} = \frac{\sum_{t=2}^{T} X_t w_t^{-1} \sum_{t=2}^{T} X_{t-1}^2 w_t^{-1} - \sum_{t=2}^{T} X_t X_{t-1} w_t^{-1} \sum_{t=2}^{T} X_{t-1} w_t^{-1}}{\sum_{t=2}^{T} w_t^{-1} \sum_{t=2}^{T} X_{t-1}^2 w_t^{-1} - \left(\sum_{t=2}^{T} X_{t-1} w_t^{-1}\right)^2}, \tag{A.16}$$

with the weights $w_t = \hat{c}^2 X_{t-1} + \hat{d}^2$. Consistent estimators for the parameters c and d are provided by

$$\hat{c} = \frac{(T-1)\sum_{t=2}^{T} X_{t-1}\,\hat{u}_t^2 - \sum_{t=2}^{T} X_{t-1}\,\hat{u}_t^2}{(T-1)\sum_{t=2}^{T} X_{t-1}^2 - \left(\sum_{t=2}^{T} X_{t-1}\right)^2} \tag{A.17}$$

$$\hat{d} = \frac{\sum_{t=2}^{T} \hat{u}_t^2 \sum_{t=2}^{T} X_{t-1}^2 - \sum_{t=2}^{T} X_{t-1} \sum_{t=2}^{T} X_{t-1}\,\hat{u}_t^2}{(T-1)\sum_{t=2}^{T} X_{t-1}^2 - \left(\sum_{t=2}^{T} X_{t-1}\right)^2}, \tag{A.18}$$

where we take $\hat{u}_t = X_t - \hat{a}_{CLS} X_{t-1} - \hat{\lambda}_{CLS}$. Again an appropriate correction to the multiplicative factor involving the sample size is introduced, to compensate for loss of degrees of freedom. Note that in the case of the PoINAR(1) model the unknown variances of the offspring as well as the immigration distribution can be derived explicitly. But using these values in $\mathbf{Q}_T$ instead of c and d may destroy the robustness of the approach and we did not pursue this.

A.3 Likelihood-based estimation

Maximum likelihood estimation for the PoINAR(1) model is based upon the fact that the joint density $f_{X_0,\ldots,X_T}(x_0,\ldots,x_T|a,\lambda)$ can be factored, due to the Markov property, as

$$f_{X_0,\ldots,X_T}(x_0,\ldots,x_T|a,\lambda) = f_{X_0}(x_0) \cdot f_{X_1|X_0}(x_1|x_0) \prod_{t=2}^{T} f_{X_t|X_{t-1}}(x_t|x_{t-1}), \tag{A.19}$$

where $f_{X_0}(x_0)$ is the (marginal) density of the starting value X_0. The resulting unconditional, or exact, log-likelihood function $\ell(a,\lambda|X_0,\ldots,X_T)$ is then given, apart from irrelevant constants, by

$$\ell(a,\lambda|X_0,\ldots,X_T)=\ln[f_{X_0}(x_0)]+\ln[f_{X_1|X_0}(x_1|x_0)]+\sum_{t=2}^{T}\ln[f_{X_t|X_{t-1}}(x_t|x_{t-1})] \tag{A.20}$$

Since X_0 is not observable, the standard approach in the literature is to condition the likelihood function on this pre-sample value and maximize the conditional likelihood function. Employing the Poisson assumption for X_t the conditional likelihood function (apart from irrelevant constants) for the PoINAR(1) model can be shown to be (Al-Osh and Alzaid, 1987)

$$\ell_c(a,\lambda|X_1,\ldots,X_T) = -(T-1)\lambda + \sum_{t=2}^{T}\ln[C(X_{t-1},X_t)]\ , \tag{A.21}$$

where $C(x_{t-1},x_t)$ is given explicitly in (B.7′) below.

The score functions of the conditional likelihood function are given by

$$\frac{\partial \ell_c}{\partial \lambda} = \sum_{t=2}^{T}\frac{C(X_{t-1},X_t-1)}{C(X_{t-1},X_t)} - (T-1) \tag{A.22}$$

and

$$\frac{\partial \ell_c}{\partial a} = \sum_{t=1}^{T}\frac{(X_t - aX_{t-1}) - \lambda[C(X_{t-1},X_t-1)/C(X_{t-1},X_t)]}{a(1-a)}\ . \tag{A.23}$$

In small and medium sized samples the information about the unobservable starting value X_0 and its incorporation into the likelihood function should lead to improved estimation results. Al-Osh and Alzaid (1987) provide the exact likelihood function and set the starting value X_0 to the mean of the process. We use the spirit of their approach but a different means of accounting for the presence of X_0 by utilizing the time reversibility property (McKenzie, 1988) of the PoINAR(1) process. This can easily be demonstrated by means of the bivariate pgf given in the companion appendix as (B.8). It follows immediately that the inverse regression function $\mathrm{E}(X_{t-1}|X_t)$ is linear in X_t in the same way that $\mathrm{E}(X_t|X_{t-1})$ is. This property of the PoINAR(1) process will be used in in the context of maximum likelihood estimation using a backcasting algorithm (Box, Jenkins and Reinsel, 1994, p. 218) to obtain an estimate $\hat{X}_0$ for the unobservable starting value of the series.

Since the minimum mean squared error one-step ahead predictor for X_0 is of the form

$$\hat{X}_0 = \mathrm{E}(X_0|X_1,\ldots,X_T) = a\,X_1 + \lambda \tag{A.24}$$

a data coherent integer value will almost never arise. Our proposal is to use the mode of the conditional distribution of the back forecast $f(x_0|x_1;\hat{\alpha},\hat{\lambda})$. The mass points of this distribution can be computed directly via

$$f(x_0|x_1) = \sum_{k=0}^{m} \binom{x_0}{k} a^k (1-a)^{x_0-k} \exp(-\lambda) \frac{\lambda^{x_1-k}}{(x_1-k)!} \qquad \text{(A.25)}$$
$$= x_0! \exp(-\lambda)\, C^\star(x_0, x_1) \,,$$

where

$$C(x_0, x_1) = \sum_{k=0}^{m^\star} \frac{a^k\,(1-a)^{x_0-k}\,\lambda^{x_1-k}}{k!\,(x_1-k)!\,(x_0-k)!} \qquad \text{(A.25')}$$

and $m^\star = \min(x_0, x_1)$ for different values of x_0. The value with the highest probability mass is chosen as $\hat{X}_0$ (Freeland and McCabe, 2003a, call this procedure point mass forecasting.)

On the basis of a Poisson assumption for X_t the exact likelihood function (apart from irrelevant constants) is then of the form

$$\ell(a,\lambda|\hat{X}_0, X_1, \ldots, X_T) = \frac{\lambda(a-2)}{1-a} + \hat{X}_0 \ln\left(\frac{\lambda}{1-a}\right) + \ln[C(\hat{X}_0, X_1)]$$
$$-(T-1)\lambda + \sum_{t=2}^{T} \ln[C(X_{t-1}, X_t)] \,. \qquad \text{(A.26)}$$

The score functions for the exact likelihood can be compared to those for the conditional likelihood (A.22) and (A.23). They are given by

$$\frac{\partial \ell}{\partial \lambda} = \frac{a-2}{1-a} + \frac{\hat{X}_0}{\lambda} + \frac{C(\hat{X}_0, X_1 - 1)}{C(\hat{X}_0, X_1)} + \frac{\partial \ell_c}{\partial \lambda} \qquad \text{(A.27)}$$

and

$$\frac{\partial \ell}{\partial a} = \frac{(X_1 - a\hat{X}_0) - \lambda[C(\hat{X}_0, X_1 - 1)/C(\hat{X}_0, X_1)]}{a(1-a)}$$
$$- \frac{\lambda}{(1-a)^2} + \frac{\hat{X}_0}{1-a} + \frac{\partial \ell_c}{\partial a} \,. \qquad \text{(A.28)}$$

To obtain the exact maximum likelihood (EML) estimators, numerical procedures have to be employed. The EML-estimators are, of course, asymptotically equivalent to the CML-estimators, but may differ in finite samples. Likelihood-based estimators are also studied in Freeland and McCabe (2003b).

Appendix B: Probability generating functions

The probability generating function (pgf) $\mathcal{P}$ of the thinning operation (4) can easily be obtained to be

$$\begin{aligned}\mathcal{P}_{a\circ X_{t-1}}(s) &= \mathcal{P}_{X_{t-1}}\left(\mathcal{P}_{Y_{i,t-1}}(s)\right) \\ &= \mathcal{P}_{X_{t-1}}(1-a+as)\ .\end{aligned} \tag{B.1}$$

The pgf of the INAR(1) process $\{X_t\}$ is given by

$$\mathcal{P}_{X_t}(s) = \mathcal{P}_{X_{t-1}}(1-a+a\,s)\cdot\mathcal{P}_{W_t}(s)\ , \tag{B.2}$$

where $\mathcal{P}_{W_t}(s) = \exp[-\lambda(1-s)]$ is the pgf of the innovation process W_t.

In contradistinction to Gaussian processes, a knowledge of the first and second order moments does not suffice to describe the dependence structure of the process entirely. Note that, due to the Markovian property of the INAR(1) model, the relevant tool for this purpose is the bivariate distribution function, or the bivariate pgf which is given by

$$\mathcal{P}_{X_t,X_{t-1}}(s_1,s_2) = \mathcal{P}_{X_{t-1}}(s_1(1-a-a\,s_2))\cdot\mathcal{P}_{W_t}(s_2)\ . \tag{B.3}$$

Assuming $W_t \sim \mathrm{Po}(\lambda)$ with $\lambda > 0$ the marginal distribution of the process $\{X_t\}$ can than be derived (see Al-Osh and Aly, 1992) by inserting the pgf of W_t into (B.2). The functional difference thus obtained

$$\begin{aligned}\mathcal{P}_{X_t}(s) &= \exp\left[-\lambda(1-s+a-as+a^2-a^2s+\ldots+a^T-a^Ts)\right] \\ &\quad\times\mathcal{P}_{X_0}(1-a^T+a^Ts) \\ &= \exp\left[-\frac{\lambda}{1-a}(1-s)(1-a^{T-1})\right]\cdot\mathcal{P}_{X_0}(1-a^T+a^Ts)\ ,\end{aligned} \tag{B.4}$$

then has to be solved iteratively. For $T\to\infty$ we obtain the result

$$\mathcal{P}_{X_t}(s) = \exp\left[-\frac{\lambda}{1-a}(1-s)\right]\ , \tag{B.5}$$

which shows that $X_t \sim \mathrm{Po}(\lambda/(1-a))$.

One way of obtaining the conditional distribution $f(x_t|\,x_{t-1})$ of the PoINAR(1) process is via the conditional pgf $\mathcal{P}_{X_t|X_{t-1}}(s)$. This is the product of the pgf of a binomial random variable and the pgf of the innovation process

$$\begin{aligned}\mathcal{P}_{X_t|X_{t-1}}(s) &= \mathcal{P}_{a\circ X_{t-1}|X_{t-1}}(s)\cdot\mathcal{P}_W(s) \\ &= [as+(1-a)]^{X_{t-1}}\cdot\exp\left[-\lambda(1-s)\right]\ .\end{aligned} \tag{B.6}$$

Jung (1999, p. 45-47) demonstrates how to obtain $f(x_t|X_{t-1} = x_{t-1})$ by means of an appropriate series expansion of the conditional pgf.

A second possibility that is computationally less cumbersome is as follows. Since (B.6) indicates that the conditional distribution $f(x_t|\, x_{t-1})$ is a convolution of a binomial distribution and a Poisson distribution, the appropriate formula for the probability distribution of a convolution of discrete random variates (see Kotz and Johnson, 1982 p. 187) can be applied. The resulting form is

$$\begin{aligned} f(x_t|x_{t-1}) &= \sum_{k=0}^{m} \binom{x_{t-1}}{k} a^k (1-a)^{x_{t-1}-k} \exp(-\lambda) \frac{\lambda^{x_t-k}}{(x_t-k)!} \\ &= x_{t-1}! \exp(-\lambda)\, C(x_{t-1}, x_t)\,, \end{aligned} \tag{B.7}$$

where

$$C(x_{t-1}, x_t) = \sum_{k=0}^{m} \frac{a^k\,(1-a)^{x_{t-1}-k}\,\lambda^{x_t-k}}{k!\,(x_t-k)!\,(x_{t-1}-k)!} \tag{B.7'}$$

and $m = \min(x_t, x_{t-1})$.

The bivariate pgf of the PoINAR(1) process is given by (for its derivation see Alzaid and Al-Osh, 1988)

$$\begin{aligned} \mathcal{P}_{X_t,X_{t-1}}(s_1,s_2) &= \mathcal{P}_{X_{t-1}}\Big(s_1(1-a-a\,s_2)\Big)\cdot \mathcal{P}_{W_t}(s_2) \\ &= \exp\left\{-\frac{\lambda}{(1-a)}\Big[1-s_1(1-a+as_2)\Big]\right\}\cdot \exp[-\lambda(1-s_2)] \\ &= \exp\left\{-\frac{\lambda}{(1-a)}\Big[2-a-(1-a)(s_1+s_2)-as_1s_2\Big]\right\} \end{aligned} \tag{B.8}$$

and is clearly symmetric in its arguments s_1 and s_2.

References

Al-Osh, M.A. and Aly, E.-E.A.A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Comm. Statist. Theory Methods* **21**, 2483-2492.

Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **8**, 261-275.

Alzaid, A.A. and Al-Osh, M.A. (1988). First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Statist. Neerlandica* **42**, 53-61.

Andrews, D.W.K. (1993). Exactly median-unbiased estimation of first-order autore gressive/unit-root models. *Econometrica* **61**, 139-165.

Box, G.E.P.; Jenkins, G.M. and Reinsel, G.C. (1994). *Time series analysis: forecasting and control* (3rd ed.). Prentice Hall, Englewood Cliffs.

Brännäs, K. (1994). Estimating and testing in integer-valued AR(1) models. *Working Paper No. 335, Department of Economics, University of Umeå, Sweden.*

Brännäs, K. and Hall, A. (2001). Estimation in integer-valued moving average models. *Appl. Stoch. Models Bus. Ind.* **17**, 277-291.

Brännäs, K. and Hellström, J. (2001). Generalized integer-valued autoregression. *Econometric Rev.* **20**, 425-433.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression analysis of count data.* Cambridge University Press, Cambridge.

Freeland, K. (1998). *Statistical analysis of discrete time series with application to the analysis of worker's compensation claims data.* Ph.D. thesis, University of British Columbia.

Freeland, K. and McCabe, B.P.M. (2003a). Forecasting discrete valued low count time series. *Internat. J. Forecasting*, forthcoming.

Freeland, K. and McCabe, B.P.M. (2003b). Analysis of count data by means of the Poisson autoregressive model. *J. Time Ser. Anal.*, forthcoming.

Godambe, V.P. (1960). An optimum property of regular maximum-likelihood estimation. *Ann. of Math. Stat.* **31**, 1208-1211.

Greene, W.H. (2000). *Econometric analysis* (4th ed.). Prentice Hall, Upper Saddle River.

Grunwald, G.K., Hyndman,R., Tedesco, L. and Tweedie, R.L. (2000). Non-Gaussian conditional linear AR(1) models. *Aust. N. Z. J. Stat.* **42**, 479-495.

Heyde, C.C. (1997). *Quasi-likelihood and its applications.* Springer, New York.

Heyde, C.C. and Lin, Y.-X. (1992). On quasi-likelihood methods and estimation for branching processes and heteroscedastic regression models. *Aust. J. Stat.* **34**, 199-206.

Jung, R. (1999). *Zeitreihenanalyse für Zähldaten. Eine Untersuchung ganzzahliger Autoregressiver-Moving-Average-Prozesse.* Josef Eul Verlag, Lohmar.

Jung, R.C. and Tremayne, A.R. (2003). Testing for serial dependence in time series models of counts. *J. Time Ser. Anal.* **24**, 65-84.

Kendall, M.G. (1954). Note on the bias in the estimation of autocorrelation. *Biometrika* **41**, 403-404.

Kotz, S. and Johnson, N.L. (1982). *Encyclopedia of statistical sciences. Volume 2.* Wiley, New York.

MacKinnon, J.G. and Smith, A.A., Jr. (1998). Approximate bias correction in econometrics. *J. Econometrics* **85**, 205-230.

Marriott, F.H.C. and Pope, J.A. (1954). Bias in the estimation of autocorrelations. *Biometrika* **41**, 390-402.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv. in Appl. Probab.* **18**, 679-705.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Adv. in Appl. Probab.* **20**, 822-835.

Orcutt, G.H. and Winokur, H.S., Jr. (1969). First order autoregression: inference, estimation, and prediction. *Econometrica* **37**, 1-14.

Park, Y. and Oh, C.W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statist. Papers* **38**, 287-302.

Shaman, P. and Stine, R.A. (1988). The bias of autoregressive coefficient estimators. *J. Amer. Statist. Assoc.* **83**, 842-848.

Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**, 893-899.

Venkataraman, K.N. (1982). A time series approach to the study of the simple subcritical Galton–Watson process with immigration. *Adv. Appl. Probab.* **14**, 1-20.

Wei, C.Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stochastic Process. Appl.* **31**, 261-282.

Winkelmann, R. (2000). *Econometric analysis of count data.*(3rd ed.) Springer, Berlin.

Winnicki, J. (1988). Estimation theory for the branching process with immigration. *Contemp. Math.* **80**, 301-322.