Economics Bulletin

Volume 38, Issue 3

An INAR(1) model with Poisson-Lindley innovations

Tito Lívio
Universidade Federal do Piauí

Naushad Mamode Khan
University of Mauritius

Marcelo Bourguignon
Universidade Federal do Rio Grande do Norte

Hassan S. Bakouch
Tanta University

Abstract

Real count data time series often show the phenomenon of the overdispersion. In this paper, we introduce a first order non-negative integer valued autoregressive process with Poisson-Lindley innovations based on the binomial thinning operator. The new model is particularly suitable for time series of counts exhibiting overdispersion and therefore competes against others recently established. The main properties of the model are derived. The methods of conditional least squares, Yule-Walker and conditional maximum likelihood are used for estimating the parameters. The proposed model is also applied to a weekly sales of soap product data series.

Citation: Tito Lívio and Naushad Mamode Khan and Marcelo Bourguignon and Hassan S. Bakouch, (2018) "An INAR(1) model with Poisson-Lindley innovations", *Economics Bulletin*, Volume 38, Issue 3, pages 1505-1513
Contact: Tito Lívio - titolivio@ufpi.edu.br, Naushad Mamode Khan - n.mamodekhan@uom.ac.mu, Marcelo Bourguignon - m.p.bourguignon@gmail.com, Hassan S. Bakouch - hnbakouch@yahoo.com.
Submitted: December 21, 2017. **Published:** August 05, 2018.

1 Introduction

Time series of counts arise in different areas of applications related to finance, transport, socioeconomic activities, medical science and among others. Usually, it is perceived that these series of count observations are dynamic in nature and are thus subject to significant overdispersion relative to the means. In this regard, several observation-driven integer valued time series autoregressive models have been proposed in literature for the modelling of such data. However, much theoretical work has been concentrated on the use of the Poisson distribution as an integral feature of the model. In this context, McKenzie (1985) and Al-Osh and Alzaid (1987), independently, introduced the first order non-negative integer-valued autoregressive (INAR(1)) process with Poisson innovations, the Poisson distribution is not always suitable for modelling, since its mean and variance are the same and this property may be unacceptable for real data.

Several alternatives to models with a Poisson innovations have been proposed in the literature. Ristić et al. (2017) defined the negative binomial thinning operator and based on the such operator introduced the overdispersed integer-valued time series model with geometric marginal (NGINAR(1)). Jazi et al. (2012) proposed the geometric INAR(1) model with geometric innovations. Mohammadpour et al. (2018) proposed a first-order integer-valued autoregressive process with Poisson-Lindley marginals based on the binomial thinning. But, the innovation structure form is complex, consequently, the conditional probabilities of this model don't have a simple form. However, the PL distribution has many properties (see below), then it is worth study the INAR(1) model with PL innovations.

The main objective of this paper is to propose an new INAR(1) model with Poisson-Lindley (PL) innovations based on the binomial thinning operator (Steutel and van Harn, 1979), denoted by INARPL(1) model, for modeling nonnegative integer valued time series with overdispersion, with the hope that the new process may have a better fit in certain practical situations (see Section 5). The use of innovations that come from the PL distribution has many advantages, that the PL distribution belongs to compound Poisson family and has other common properties like unimodality, overdispersion, and infinite divisibility (Ghitany and Al-Mutairi, 2009), such as the negative binomial distribution. Furthermore, considering a PL distribution with parameter θ, then the PL distribution can be viewed as mixture of geometric with parameter $1/(1+\theta)$ and negative binomial with parameters 2 and $1/(1+\theta)$ with mixing proportions $\theta/(1+\theta)$ and $1/(1+\theta)$, respectively. Moreover, the skewness and kurtosis of the PL distribution are smaller than the negative binomial distribution (Ghitany and Al-Mutairi, 2009).

The rest of the paper unfolds as follows. In Section 2, the INAR(1) process with Poisson-Lindley innovation is presented. In the same section, the main properties of the model are derived. In Section 3, we propose estimation methods for the model parameters. In Section 4, we present some simulation results for the estimation methods. In Section 5, we illustrate the application of the model to the weeky sales series of soap product.

2 Poisson-Lindley INAR(1) model

A random variable X is said to have a Poisson-Lindley (PL) distribution (Sankaran, 1970) if its probability mass function (pmf) is given by

$$\Pr(X=k)=\frac{\theta^2(k+\theta+2)}{(\theta+1)^{k+3}}, \quad \theta>0, \quad k\in\mathbb{N}. \tag{1}$$

The mean and variance of the PL distribution are given, respectively, by

$$\mathrm{E}(X) = \frac{\theta+2}{\theta(\theta+1)} \text{ and } \mathrm{Var}(X) = \frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2} = \mathrm{E}(X)\left[1+\frac{\theta^2+4\theta+2}{\theta(\theta+1)(\theta+2)}\right].$$

Note that $\mathrm{Var}(X)/\mathrm{E}(X) > 1$, i.e, the PL distribution is overdispersed, for more details see Ghitany and Al-Mutairi (2009).

The skewness and kurtosis of the Poisson-Lindley distribution are given by

$$\gamma_1 = \frac{2(\theta+1)^4(\theta+2) - \theta^3(\theta+2)(\theta+3)}{[2(\theta+1)^3 - \theta^2(\theta+2)]^{3/2}}$$

and

$$\gamma_2 = 3 + \frac{2(\theta+1)^5[(\theta+3)^2-3] - \theta^4(\theta+2)[(\theta+4)^2-3]}{[2(\theta+1)^3 - \theta^2(\theta+2)]^2}.$$

In short, we name this distribution as the PL(θ) distribution. The probability generating function (pgf) of X, denoted by $\varphi_X(s) := \mathrm{E}[s^X]$, is given by

$$\varphi_X(s) = \frac{\theta^2}{\theta+1}\cdot\frac{2+\theta-s}{(\theta+1-s)^2}.$$

A discrete-time stationary non-negative integer-valued stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$ is said to be a first-order integer-valued autoregressive [INAR(1)] process if it satisfies the following equation

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{2}$$

where $\alpha \circ Y_{t-1} = \sum_{j=1}^{Y_{t-1}} B_j$ is the binomial thinning operator, $\{Y_j\}_{j\geq 1}$ is a sequence of independent and identically distributed Bernoulli random variables with $\Pr(B_j = 1) = 1 - \Pr(B_j = 0) = \alpha \in [0, 1)$, $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is an innovation sequence of independent and identically distributed non-negative integer-valued random variables not depending on past values of $\{Y_t\}_{t\in\mathbb{Z}}$ of mean μ_ϵ and variance σ_ϵ^2. It is also assumed that the B_j variables that define $\alpha \circ Y_{t-1}$ are independent of the variables from which other values of the series are calculated. Moreover, we assume that all B_j variables defining the thinning operations are independent of the innovation sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$.

Let $\{\epsilon_t\}_{t\in\mathbb{Z}}$ be a sequence of discrete i.i.d. random variables following a PL distribution with pmf given in (1). Thus, the mean and variance of $\{Y_t\}_{t\in\mathbb{Z}}$ are given, respectively, by

$$\mathrm{E}(Y_t) = \frac{\theta+2}{(1-\alpha)(1+\theta)\theta} \quad \text{and} \quad \mathrm{Var}(Y_t) = \frac{\theta+2}{\theta(\theta+1)(1-\alpha^2)}\left[1+\alpha+\frac{\theta^2+4\theta+2}{\theta(\theta+1)(\theta+2)}\right].$$

The process defined in Equation (2) with $\{\epsilon_t\}_{t\in\mathbb{Z}} \sim$ PL(θ) is Markovian, stationary, and ergodic. From Al-Osh and Alzaid (1987), we have that $\alpha \in [0, 1)$ and $\alpha = 1$ are the conditions of stationarity and non-stationarity of the process $\{Y_t\}_{t\in\mathbb{Z}}$, respectively. Also, $\alpha = 0$ and $\alpha > 0$ respectively imply the independence and dependence of the observations of $\{Y_t\}_{t\in\mathbb{Z}}$. In this paper, we restrict our study to the stationary case. In short, we name this model as the INARPL(1) process.

The conditional mean and the conditional variance are given by

$$\mathrm{E}(Y_{t+1}|Y_t) = \alpha Y_t + \frac{\theta+2}{\theta(\theta+1)}$$

and

$$\text{Var}(Y_{t+1}|Y_t) \;\; = \;\; \alpha(1-\alpha)Y_t + \frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2}.$$

The transition probabilities of this process are given by

$$\Pr(Y_t = k|Y_{t-1} = l) = \sum_{i=0}^{\min(k,l)} \binom{l}{i} \alpha^i (1-\alpha)^{l-i} \cdot \frac{\theta^2(k-i+\theta+2)}{(\theta+1)^{k-i+3}}, \quad k,l \geq 0, \qquad (3)$$

where $(\cdot)$ is the standard combinatorial symbol.

The autocorrelation function (ACF) at lag h is given by

$$\text{Corr}(Y_t, Y_{t-h}) = \rho_X(h) = \alpha^h, \quad h \geq 0.$$

3 Estimation and inference of the unknown parameters

This section is concerned with the estimation of the two parameters of interest. We consider three estimation methods, namely, conditional least squares, Yule-Walker and conditional maximum likelihood.

3.1 Conditional least squares estimation

The conditional least squares estimator $\widehat{\eta} = (\widehat{\alpha}_{cls}, \widehat{\theta}_{cls})^\top$ of $\eta = (\alpha,\theta)^\top$ is given by

$$\widehat{\eta} = \arg\min_{\eta}(S_n(\eta)),$$

where $S_n(\eta) = \sum\limits_{t=2}^{n}[Y_t - g(\eta, Y_{t-1})]^2$ and $g(\eta, Y_{t-1}) = \text{E}(Y_t|Y_{t-1})$. Thus, the conditional least squares (CLS) estimators of α and θ can be written in closed form as

$$\begin{aligned}
\widehat{\alpha}_{cls} &= \frac{(n-1)\sum_{t=2}^{n} Y_t Y_{t-1} - \sum_{t=2}^{n} Y_t \sum_{t=2}^{n} Y_{t-1}}{(n-1)\sum_{t=2}^{n} Y_{t-1}^2 - \left(\sum_{t=2}^{n} Y_{t-1}\right)^2}, \\
\widehat{\theta}_{cls} &= \frac{Y_{\widehat{\alpha}_{cls}} - 1 + \sqrt{Y_{\widehat{\alpha}_{cls}}^2 + 6Y_{\widehat{\alpha}_{cls}} + 1}}{2},
\end{aligned}$$

where $Y_{\widehat{\alpha}_{cls}} = (n-1)\left(\sum_{t=2}^{n} Y_t - \widehat{\alpha}_{cls}\sum_{t=2}^{n} Y_{t-1}\right)^{-1}$.

Theorem 3.1 *The estimators $\widehat{\alpha}_{cls}$ and $\widehat{\theta}_{cls}$ are strongly consistent for estimating α and θ, respectively, and satisfy the asymptotic normality*

$$\sqrt{n}\begin{pmatrix} \widehat{\alpha}_{cml} - \alpha \\ \widehat{\theta}_{cml} - \theta \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, c^2 \begin{bmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{bmatrix}\right),$$

where $c = \theta^3(\alpha-1)(\theta+1)^3[\mu_2\theta^4(1-\alpha)^2(\theta+1) - (\theta+2)^2(\theta^2+4\theta+2)]^{-1}$, r_{11}, r_{12}, r_{21} and r_{22} are given in the Appendix.

3.2 Yule-Walker estimation

The Yule-Walker (YW) estimator of α and θ, based upon the fact that $\rho_X(1) = \alpha$ and $\mathrm{E}(Y_t) = (\theta+2)[(1-\alpha)(1+\theta)\theta]^{-1}$. Thus, from a sample $Y_1, \ldots, Y_n$ of a stationary process $\{Y_t\}_{t\in\mathbb{Z}}$, the estimators of α and θ are defined as

$$\begin{aligned}
\widehat{\alpha}_{yw} &= \frac{\sum_{t=2}^{n}(Y_t-\overline{Y})(Y_{t-1}-\overline{Y})}{\sum_{t=1}^{n}(Y_t-\overline{Y})^2}, \\
\widehat{\theta}_{yw} &= \frac{Y_{\widehat{\alpha}_{yw}} - 1 + \sqrt{Y_{\widehat{\alpha}_{yw}}^2 + 6Y_{\widehat{\alpha}_{yw}} + 1}}{2},
\end{aligned}$$

where $Y_{\widehat{\alpha}_{yw}} = [\overline{Y}(1-\widehat{\alpha}_{yw})]^{-1}$ and $\overline{Y} = (1/n)\sum_{t=1}^{n} Y_t$ is the sample mean.

3.3 Conditional maximum likelihood estimation

The conditional log-likelihood function for the PLINAR(1) model is given by

$$\ell(\alpha,\theta) = \log\left[\prod_{t=2}^{n} \Pr(Y_t = y_t | Y_{t-1} = y_{t-1})\right] = \sum_{t=2}^{n} \log\left[\Pr(Y_t = y_t|Y_{t-1} = y_{t-1})\right], \quad (4)$$

with $\Pr(Y_t|Y_{t-1})$ as in (3). The conditional maximum likelihood (CML) estimators $\widehat{\alpha}_{cml}$ and $\widehat{\theta}_{cml}$ of α and θ are defined as the values of α and θ that maximize the conditional log-likelihood function in (4). There will be, in general, no closed form for the CML estimators and their obtention will need, in practice, numerical methods.

4 Simulation

In this section, a small Monte Carlo simulation experiment is conducted to evaluate the estimation of the INARPL(1) process parameters, i.e., the performances of the CLS, YW and CML estimators for a sample size of n observed values of Y_t is the motivation of this section. The simulation was performed using the R programming language (http://www.r-project.org). The number of Monte Carlo replications was 1000. The sample sizes considered are $n = 100, 200, 300,$ and 500. For the values of parameters, we considered $\alpha = 0.2, 0.4, 0.6$ and 0.8, and $\theta = 0.5, 1.0, 2.0$ and 4.0.

Tables 1 presents the empirical mean and mean squared error (in parentheses) of the estimates of the parameters of the INARPL(1) process. Note that as the sample size increases, the bias tends to zero in all three cases, confirming that the estimators are asymptotically unbiased. Furthermore, for CLS and YW methods, increasing α, the bias and MSE also increase. This indicates that these two estimation methods are sensitive to a process that is closer to the non-stationary boundary ($\alpha = 1$). The empirical investigation presented here suggests that, generally speaking, the CML is, in fact, much better than the CLS and YW. Thus, we recommend the use of the CML method to estimate the model parameters of an INARPL(1) model.

5 Application

In this section, consider the series of weekly sales (in integer units) of particular soap product in a supermarket. The data are taken from a database provided by the Kilts Center for

Table I: Empirical means and mean squared errors (in parentheses) of the estimates of the parameters for some values of α and θ.

T	$\hat{\alpha}_{yw}$	$\hat{\theta}_{yw}$	$\hat{\alpha}_{cml}$	$\hat{\theta}_{cml}$	$\hat{\alpha}_{cls}$	$\hat{\theta}_{cls}$
			(a) $\alpha = 0.2$ e $\theta = 0.5$			
100	0.1971 (0.0083)	0.5341 (0.0049)	0.2117 (0.0032)	0.5396 (0.0035)	0.1994 (0.0085)	0.5361 (0.0052)
200	0.2033 (0.0053)	0.4830 (0.0039)	0.2107 (0.0021)	0.4847 (0.0029)	0.2043 (0.0054)	0.4834 (0.0039)
300	0.1997 (0.0040)	0.4853 (0.0022)	0.2054 (0.0015)	0.4864 (0.0013)	0.2005 (0.0040)	0.4855 (0.0022)
500	0.2087 (0.0022)	0.5313 (0.0017)	0.2130 (0.0009)	0.5329 (0.0013)	0.2092 (0.0022)	0.5316 (0.0017)
			(b) $\alpha = 0.4$ e $\theta = 1$			
100	0.3846 (0.0093)	1.0214 (0.0303)	0.4050 (0.0038)	1.0395 (0.0192)	0.3885 (0.0094)	1.0282 (0.0321)
200	0.3891 (0.0047)	1.0090 (0.0139)	0.4013 (0.0018)	1.0206 (0.0088)	0.3910 (0.0047)	1.0119 (0.0143)
300	0.3924 (0.0035)	1.0033 (0.0109)	0.4024 (0.0014)	1.0128 (0.0066)	0.3939 (0.0035)	1.0055 (0.0111)
500	0.3961 (0.0019)	1.0063 (0.0061)	0.4005 (0.0008)	1.0100 (0.0035)	0.3969 (0.0019)	1.0075 (0.0062)
			(c) $\alpha = 0.6$ e $\theta = 2$			
100	0.5660 (0.0090)	1.9681 (0.1871)	0.5940 (0.0032)	2.0410 (0.1048)	0.5718 (0.0088)	1.9944 (0.1999)
200	0.5813 (0.0045)	1.9902 (0.0875)	0.5969 (0.0017)	2.0349 (0.0550)	0.5846 (0.0044)	2.0047 (0.0897)
300	0.5863 (0.0030)	1.9816 (0.0603)	0.5970 (0.0011)	2.0118 (0.0326)	0.5880 (0.0030)	1.9903 (0.0615)
500	0.5908 (0.0017)	1.9850 (0.0383)	0.5983 (0.0006)	2.0069 (0.0211)	0.5919 (0.0017)	1.9897 (0.0387)
			(d) $\alpha = 0.8$ e $\theta = 4$			
100	0.7492 (0.0077)	3.6750 (1.4570)	0.7907 (0.0019)	4.1480 (1.0560)	0.7582 (0.0070)	3.8335 (1.8526)
200	0.7764 (0.0030)	3.8851 (0.7749)	0.7959 (0.0008)	4.0987 (0.4213)	0.7810 (0.0028)	3.9627 (0.8222)
300	0.7815 (0.0020)	3.8372 (0.5165)	0.7972 (0.0005)	4.0095 (0.2100)	0.7844 (0.0019)	3.8829 (0.5299)
500	0.7912 (0.0011)	3.9724 (0.3599)	0.7986 (0.0003)	4.0420 (0.1389)	0.7928 (0.0010)	4.0013 (0.3724)

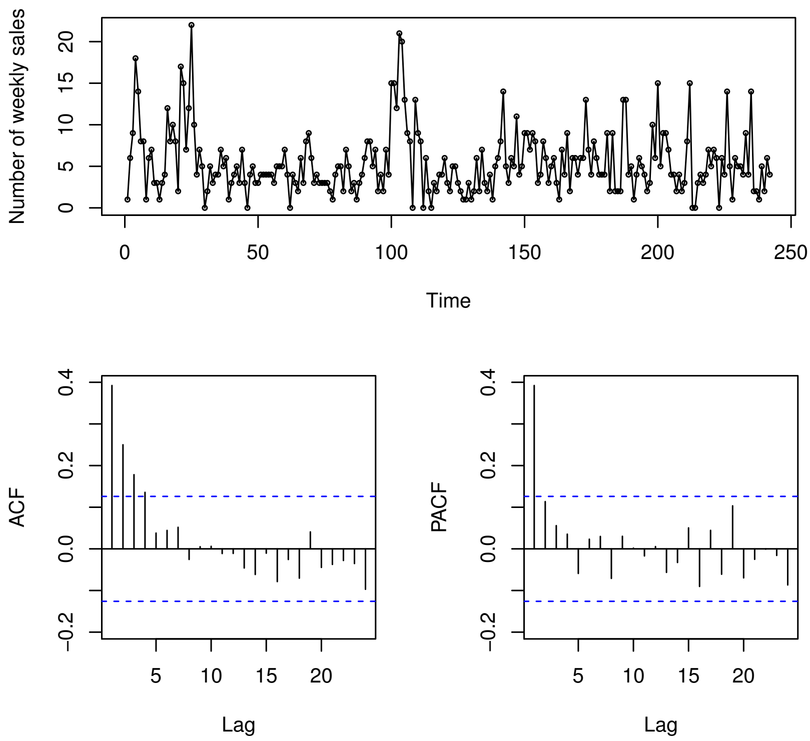
Marketing, Graduate School of Business of the University of Chicago, at: `http://gbswww.uchicago.edu/kilts/research/db/dominicks`. (The product is 'Zest White Water 15 oz.', with code 3700031165). The length, sample mean and variance are 242, 5.44, 15.40, respectively. Note that the sample variance is much larger than the sample mean; hence, the data seem to be overdispersed. We apply the test for overdispersion described by Schweer and Weiß (2014) with significance level at 5%. The p-value for the test being < 0.01 leads to a rejection of the null hypothesis of a Poisson INAR(1) process. Consequently, a Poisson marginal distribution would not be appropriate.

We compared the INARPL(1) process with the PLINAR(1) process (Mohammadpour et al., 2018), with the Poisson INAR(1) process (Al-Osh and Alzaid, 1987), and also with the NGINAR(1) model with geometric marginal distribution (Ristić et al., 2017). These models can be capture overdispersion inherent in the analysis of integer-valued time series data. Thus, the use of these models for fitting this data set seems justified. In order to estimate the parameters of these models, we adopt the CML method (as discussed in Subsection 3.3) and all the computations were done using the R software (R Core Team, 2016). Since the Fisher information matrix is not available, the standard errors are obtained as the square roots of the elements in the diagonal of the inverse of the negative of the Hessian of the conditional log-likelihood calculated at the CML estimates (Bourguignon and Vasconcellos, 2015).

The time series data and their sample autocorrelation and partial autocorrelation functions are displayed in the Figure 1. Analyzing Figure 1, we conclude that a first order autoregressive model may be appropriate for the given data series, because of the clear cut-off after lag 1 in the partial autocorrelations.

Table II provides the estimates (with standard errors in parentheses) of the model parameters and three goodness-of-fit statistics: Akaike information criterion (AIC), Bayesian information criterion (BIC) and root mean square (RMS) (differences between observations and predicted

Figure 1: Plots of the time series, autocorrelation and partial autocorrelation functions for the number of weekly sales.



values). In general, it is expected that the better model to fit the data presents the smaller values for these quantities. From this table, we observe that the proposed model being better.

Table II: Estimates of the parameters (standard errors in parentheses), AIC, BIB, and RMS for the number of weekly sales.

Model	Par. 1	Par. 2	AIC	BIC	RMS
INARPL(1)	0.3202	0.4533	1249.07	1256.05	3.6094
(α, θ)	(0.0368)	(0.0324)			
PLINAR(1)	0.3152	0.3516	1287.12	1294.10	3.6282
(α, θ)	(0.0518)	(0.0241)			
Poisson INAR(1)	0.2340	4.1855	1363.81	1370.79	3.6515
(α, λ)	(0.0324)	(0.2110)			
NGINAR(1)	0.5770	4.6146	1296.43	1303.41	3.6883
(α, μ)	(0.0515)	(0.5558)			

References

Al-Osh, M.A. and Alzaid, A.A. (1987). "First-order integer valued autoregressive (INAR(1)) process" *Journal of Time Series Analysis* **8**, 261–275.

Bourguignon, M. and Vasconcellos, K.L.P. (2015). "Improved estimation for Poisson INAR(1) models" *Journal of Statistical Computation and Simulation* **85**, 2425–2441.

Ghitany, M.E. and Al-Mutairi, D.K. (2009). "Estimation Methods for the discrete Poisson-Lindley distribution" *Journal of Statistical Computation and Simulation* **79**, 1–9.

Jazi, M.A, Jones, G. and Lai, C.D. (2012). "Integer valued AR(1) with geometric innovations" *Journal of the Iranian Statistical Society* **11**, 173–190.

McKenzie, E. (1985). "Some simple models for discrete variate time series" *Journal of the American Water Resources Association* **21**, 645–650.

Mohammadpour, M., Bakouch, H.S. and Shirozhan, M. (2018). "Poisson-Lindley INAR(1) Model with Applications" *Brazilian Journal of Probability and Statistics* **32**, 262–280.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org.

Ristić, M. M., Bakouch, H.S. and Nastić, A.S. (2009). "A new geometric first-order integer-valued autoregressive (NGINAR(1)) process" *Journal of Statistical Planning and Inference* **139**, 2218–2226.

Sankaran, M. (1970). "The discrete Poisson-Lindley distribution" *Biometrics* **26**, 145–149.

Schweer, S. and Weiß, C. (2014). "Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion" *Computational Statistics and Data Analysis* **77**, 267–284.

Steutel, F.W. and van Harn, K. (1979). "Discrete analogues of self-decomposability and stability" *The Annals of Probability* **7**, 893–899.

Tjostheim, D. (1986). "Estimation in nonlinear time series models" *Stochastic Processes and Their Applications* **21**, 251–273.

Appendix

Proof of Theorem 1.

To derive asymptotic properties of the CLS estimators we use Theorems 3.1 and 3.2 given in Tjostheim (1986). It can be verified that the conditions stated in the Theorems 3.1 and 3.2 by Tjostheim (1986) are satisfied by our model. Let $\boldsymbol{\eta} = (\alpha, \theta)$ and $g(\boldsymbol{\eta}, Y_{t-1}) = E(Y_t|Y_{t-1}) = \alpha Y_{t-1} + (\theta+2)[\theta(\theta+1)^{-1}]$, the first condition C1 of Theorem 3.1 is satisfied, because

$$E\left[\left(\frac{\partial g}{\partial \alpha}\right)^2\right] = E(Y_t^2) = \sigma^2 + \mu^2 < \infty, \quad E\left[\left(\frac{\partial g}{\partial \theta}\right)^2\right] = \left[\frac{\theta^2+4\theta+2}{\theta^2(\theta+1)^2}\right]^2 < \infty,$$

where $Var(Y_t) = \sigma^2$ and $E(Y_t) = \mu$.

To verify the second condition of Theorem 3.1, we need to find the real numbers a_1 and a_2 that makes

$$E\left[\left(a_1\frac{\partial g}{\partial \alpha} + a_2\frac{\partial g}{\partial \theta}\right)^2\right] = E[(a_1 Y_{t-1} + a_2 k)^2] = E\left(a_1^2 Y_{t-1}^2 + 2a_1 a_2 k Y_{t-1} + a_2^2 k^2\right) = 0, \quad (5)$$

where $k = \dfrac{\theta^2 + 4\theta + 2}{\theta^2(\theta+1)^2}$. Solving the Equation (5) in relation to a_1, we obtain $a_1 = \dfrac{-a_2 k\mu \pm |a_2|k\sigma i}{\sigma^2 + \mu^2}$, where $i = \sqrt{-1}$. Then $a_1 \notin \mathbb{R}$, unless $a_1 = a_2 = 0$. Let

$$f_{t|t-1} = Var(Y_t|Y_{t-1}) = \alpha(1-\alpha)Y_{t-1} + \sigma_\epsilon^2,$$

where $\sigma_\epsilon^2 = (\theta^3 + 4\theta^2 + 6\theta + 2)[\theta^2(\theta+1)]^{-1}$, after some algebra, we obtain

$$\begin{aligned}
\mathbf{R} &= E\left(\frac{\partial g}{\partial \boldsymbol{\beta}}(\alpha,\theta) f_{t+1|t} \frac{\partial g^T}{\partial \boldsymbol{\beta}}(\alpha,\theta)\right) \\
&= \begin{bmatrix} \alpha(1-\alpha)\mu_3 + \sigma_\epsilon^2\mu_2 & \dfrac{-(\theta^2+4\theta+2)[\alpha(1-\alpha)\mu_2 + \sigma_\epsilon^2\mu_1]}{\theta^2(\theta+1)} \\ \dfrac{-(\theta^2+4\theta+2)[\alpha(1-\alpha)\mu_2 + \sigma_\epsilon^2\mu_1]}{\theta^2(\theta+1)} & \dfrac{(\theta^4 + 8\theta^3 + 20\theta^2 + 4)[\alpha(1-\alpha)\mu_1 + \sigma_\epsilon^2]}{\theta^4(\theta+1)^2} \end{bmatrix},
\end{aligned}$$

where $\mu_r = E(Y_t^r)$, $r = 1, 2, 3$. Then, $\mathbf{R} < \infty$. Thus,

$$\mathbf{U} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \sigma^2 + \mu^2 & \dfrac{(\theta+2)(\theta^2+4\theta+2)}{\theta^3(\theta+1)^3(1-\alpha)} \\ \dfrac{(\theta+2)(\theta^2+4\theta+2)}{\theta^3(\theta+1)^3(1-\alpha)} & \dfrac{\theta^2+4\theta+2}{\theta^2(\theta+1)^2} \end{bmatrix},$$

where $a_{ij} \equiv E\left(\dfrac{\partial g}{\partial \beta_i}\dfrac{\partial g}{\partial \beta_j}\right)$, $i, j = 1, 2$. The covariance matrix is given by $\mathbf{U^{-1}RU^{-1}}$. Using the above expressions for $\mathbf{U}$ and $\mathbf{R}$, after some algebra, it can be shown that the covariance matrix is given by

$$\mathbf{U^{-1}RU^{-1}} = c^2\begin{bmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{bmatrix},$$

where $c = \theta^3(\alpha-1)(\theta+1)^3[\mu_2\theta^4(1-\alpha)^2(\theta+1) - (\theta+2)^2(\theta^2+4\theta+2)]^{-1}$ and

$$
\begin{aligned}
r_{11} &= \frac{1}{\theta}\left(\alpha-1\right)\left[-\theta^{3}\left(\alpha-1\right)\left(\theta+1\right)^{2}\left(\alpha\mu_{3}\left(\alpha-1\right)-\sigma_{\epsilon}^{2}\mu_{2}\right)+\left(\theta+2\right)\left(\alpha\mu_{2}\left(\alpha-1\right)-\mu_{1}\sigma_{\epsilon}^{2}\right)\left(\theta^{3}+4\theta+\right.\right.\\
&+ \frac{1}{\theta^{4}\left(\theta+1\right)^{2}}\left(\theta+2\right)\left[\theta^{3}\left(\alpha-1\right)\left(\theta+1\right)^{2}\left(\alpha\mu_{2}\left(\alpha-1\right)-\mu_{1}\sigma_{\epsilon}^{2}\right)\left(\theta^{3}+4\theta+2\right)\right.\\
&+ \left.\left(\theta+2\right)\left(\alpha\mu_{1}\left(\alpha-1\right)-\sigma_{\epsilon}^{2}\right)\left(\theta^{4}+8\theta^{3}+20\theta^{2}+4\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
r_{21} &= r_{12}=\frac{\mu_{2}}{\theta\left(\theta^{2}+4\theta+2\right)}\left(\alpha-1\right)\left(\theta+1\right)\left[\theta^{3}\left(\alpha-1\right)\left(\theta+1\right)^{2}\left(\alpha\mu_{2}\left(\alpha-1\right)-\mu_{1}\sigma_{\epsilon}^{2}\right)\left(\theta^{3}+4\theta+2\right)\right.\\
&+ \left.\left(\theta+2\right)\left(\alpha\mu_{1}\left(\alpha-1\right)-\sigma_{\epsilon}^{2}\right)\left(\theta^{4}+8\theta^{3}+20\theta^{2}+4\right)\right]\\
&+ \frac{1}{\theta^{2}\left(\theta+1\right)}\left(\theta+2\right)\left[-\theta^{3}\left(\alpha-1\right)\left(\theta+1\right)^{2}\left(\alpha\mu_{3}\left(\alpha-1\right)-\sigma_{\epsilon}^{2}\mu_{2}\right)\right.\\
&+ \left.\left(\theta+2\right)\left(\alpha\mu_{2}\left(\alpha-1\right)-\mu_{1}\sigma_{\epsilon}^{2}\right)\left(\theta^{3}+4\theta+2\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
r_{22} &= \frac{\mu_{2}\theta\left(\theta+1\right)^{2}}{\left(\theta^{2}+4\theta+2\right)^{2}}\left(\alpha-1\right)\left[\mu_{2}\theta\left(\alpha-1\right)\left(\theta+1\right)^{2}\left(\alpha\mu_{1}\left(\alpha-1\right)-\sigma_{\epsilon}^{2}\right)\left(\theta^{4}+8\theta^{3}+20\theta^{2}+4\right)\right.\\
&+ \left.\left(\theta+2\right)\left(\alpha\mu_{2}\left(a-1\right)-\mu_{1}\sigma_{\epsilon}^{2}\right)\left(\theta^{2}+4\theta+2\right)\left(\theta^{3}+4\theta+2\right)\right]\\
&+ \frac{1}{\theta^{2}+4\theta+2}\left(\theta+2\right)\left[\mu_{2}\theta\left(\alpha-1\right)\left(\theta+1\right)^{2}\left(\alpha\mu_{2}\left(\alpha-1\right)-\mu_{1}\sigma_{\epsilon}^{2}\right)\left(\theta^{3}+4\theta+2\right)\right.\\
&- \left.\left(\theta+2\right)\left(\alpha\mu_{3}\left(\alpha-1\right)-\sigma_{\epsilon}^{2}\mu_{2}\right)\left(\theta^{2}+4\theta+2\right)\right].
\end{aligned}
$$