Stat Papers (2012) 53:951–970
DOI 10.1007/s00362-011-0399-9

REGULAR ARTICLE

Diagnostic checks for integer-valued autoregressive models using expected residuals

Yousung Park · Hee-Young Kim

Received: 23 January 2011 / Revised: 14 April 2011 / Published online: 27 July 2011

Abstract Integer-valued time series models make use of thinning operators for coherency in the nature of count data. However, the thinning operators make residuals unobservable and are the main difficulty in developing diagnostic tools for autocorrelated count data. In this regard, we introduce a new residual, which takes the form of predictive distribution functions, to assess probabilistic forecasts, and this new residual is supplemented by a modified usual residuals. Under integer-valued autoregressive (INAR) models, the properties of these two residuals are investigated and used to evaluate the predictive performance and model adequacy of the INAR models. We compare our residuals with the existing residuals through simulation studies and apply our method to select an appropriate INAR model for an over-dispersed real data.

Keywords Integer-valued AR(p) · Residuals · Probability integral transformation · Over-dispersion · Thinning parameter

Mathematics Subject Classification (2000) 62M10 · 62M20

1 Introduction

When the counts are of low-frequency and depend on their past observations, it is not appropriate to approximate their dynamic structure by continuous time series

Y. Park
Department of Statistics, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-701, Korea
e-mail: yspark@korea.ac.kr

H.-Y. Kim (✉)
Institute of Statistics, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-701, Korea
e-mail: starlike.kim@gmail.com

models such as the autoregressive-moving average (ARMA) processes. To obtain such an ARMA-like autoregressive structure, many integer-valued time series models make use of thinning operations in place of the multiplication of ARMA model for coherency in the nature of count data. Integer-valued time series models were surveyed by Kedem and Fokianos (2002) and McKenzie (2003). In particular, Weiß (2008a) summarized a broad variety of thinning operations and showed how they are applied to integer-valued ARMA (INARMA) models. Recently many new models on integer valued time series have been proposed; generalized-INAR models (Latour 1998; Brännäs and Hellström 2001), threshold-INAR models (Thyregod et al. 1999; Brännäs and Hellström 2001), geometric INAR(1) model (Ristić et al. 2009), and random coefficient INAR models (Zheng et al. 2006, 2007, 2008). Most researches have generally been limited to various integer-valued autoregressive processes because of the complicated form of the likelihood of the full INARMA models.

Although there have been many studies of integer-valued time series models, they have largely been focused on model building and parameter estimation. In addition, the main focus has been on evaluating point forecasts, not probabilistic forecasts. Recent contribution to the theory of probabilistic forecasts for integer-valued time series models can be found in Freeland and McCabe (2004b), Jung and Tremayne (2006), and McCabe et al. (2009). Jung and Tremayne (2003) presented test statistics for serial correlations using i.i.d. Poisson or negative binomial counts under the null against INAR(1) or INAR(2) under the alternative. Their method thus has a limitation if the null is a correlated sequence of counts or if the alternative is a general INAR(p) with unknown p. To circumvent this limitation, Bu and McCabe (2008) proposed $p+1$ sets of residuals in a Poisson INAR(p) model. The existence of any dependence structure in these $p+1$ sets of residuals would suggest the need for a more general specification. The idea behind our approach is basically the same as Bu and McCabe (2008) in that both approaches create new sets of residuals: two sets of residuals for ours and $p+1$ sets of residuals for Bu and McCabe.

Another important issue is the evaluation of predictive performance by comparing a probabilistic forecast, which is in the form of a predictive distribution, with the true data-generating distribution. For the case of continuous time series data, this comparison has been achieved by the uniformity of probability integral transform (PIT) values, which are supplemented by tests for independence because PIT values are i.i.d. samples from $U(0, 1)$ when the predictive distribution of one-step-ahead forecasts agrees with the true data generating distribution (Frühwirth-Schnatter 1996; Diebold et al. 1998). For the case of discrete data, randomized PIT values can be used for predictive performance (Frühwirth-Schnatter 1996; Liesenfeld et al. 2006) for i.i.d. underlying PIT values. The randomized PIT values are defined by adding some random noises to the discrete-valued PIT values so that they are i.i.d. continuous $U(0, 1)$ values.

In the case of serially correlated counts modelled by thinning operators, however, the predictive distribution is a convolution of discrete random variables because of thinning operators. For example, a one-step-ahead forecast is expressed by the sum of $p+1$ integer-valued random variables in the Poisson INAR(p) model suggested by Du and Li (1991): the first p variables are independent binomial random variables depending on the past counts from lag one to lag p, and the remaining one is a

Poisson error term. This makes residuals unobservable, and the predictive distribution of the one-step-ahead forecast varies with time, implying that the PIT values calculated from one-step-ahead forecasts can be neither identically distributed nor converted into randomized PIT values.

To overcome such non-homogenous and non-uniformly distributed PIT values, we introduce, so called, *expected residuals* to evaluate predictive performance and to examine serial correlations. The expected residuals are regarded as the estimates of the error term. Thus, by investigating their empirical distribution and correlation, one can check whether a hypothesized predictive distribution is the true data generating process (DGP).

One type of expected residuals is obtained from a conditional predictive distribution given in previous observations, and the PIT values calculated from the expected residuals are used as diagnostic tools: a S-shaped empirical cumulative distribution function (ECDF) of the PIT values indicates an overdispersed or underdispersed predictive distribution, and a U-shaped or humped-shaped ECDF points at underestimated or overestimated thinning parameter, or an underestimated or overestimated error mean. These informal diagnostics are supplemented by independent test of PIT values to specify the order p in the INAR(p) model of Du and Li (1991). However, the expected residuals turn out not to work well for the independent test. Thus, we consider modified usual residuals as a second type of expected residuals for better independent test and investigate their properties useful for parameter estimation and model specification.

Using both types of expected residuals, we first determine an appropriate order p in INAR(p), and then investigate whether the time series of counts under consideration is over-dispersed or underdispersed by examining the shape of ECDF of the first type of residuals. Finally, the correlations of both expected residuals will be used to identify if an estimated parameter is overestimated or underestimated.

In this article, we only consider the Poisson INAR(p) of Du and Li (1991) because it has not only the same autocorrelation structure as the continuous AR(p) models but also our method developed for the INAR(p) model can be easily extended to other integer-valued time series models such as Negative Binomial INAR(1) (McKenzie 1987), Random Coefficient INAR(p) (Zheng et al. 2006), General Poisson INAR(p) (Alzaid and Al-Osh 1993), and a hybrid model like the CINAR(p) model of Weiß (2008b) whenever their conditional distributions given past counts are available. Furthermore, since most applications in the practice of integer-valued time series modelling are focused on INAR(1) models, and most INAR(1) models appeared in literature have the same autocorrelation structure as that of the Poisson INAR(1), such INAR(1) models share the same properties with the Poisson INAR(l) model.

The remainder of this article proceeds as follows. Section 2 provides a theoretical framework for the distribution of discrete PIT values. After a brief review of the Poisson INAR(p) model, two expected residuals are defined and their properties are discussed. Section 3 assesses the predictive performance of the residuals in a graphical sense by plotting the ECDF of their PIT values. Simulation studies are performed to compare the two expected residuals with the residuals of Freeland and McCabe (2004a) and Bu and McCabe (2008). Section 4 includes an over-dispersed real data analysis for which we apply one i.i.d. Poisson model and four INAR(l) models. Section 5 concludes.

2 Probability integral transforms and expected residuals

2.1 The distribution of the PIT

Let $y_1, y_2, \ldots, y_T$ be a time series of count data, which are the realizations from the conditional probability mass functions (PMF) $\{f_t(y_t|\mathcal{F}_{t-1})\}_{t=1}^{T}$, where $\mathcal{F}_t$ is a sigma-field generated by $y_1, \ldots, y_t$. We denote by $p_t(y_t|\mathcal{F}_{t-1})$ the predictive model chosen by the data analyst corresponding to the true model $f_t(y_t|\mathcal{F}_{t-1})$ and define a random variable Z_t by $Z_t = \sum_{y=0}^{Y_t} p_t(y|\mathcal{F}_{t-1})$. Then we have the following results.

Lemma 2.1 *Let $Q_t(z_t) = P(Z_t \le z_t|\mathcal{F}_{t-1})$ be the conditional distribution of Z_t with the support defined by $J_t \equiv \{g_t(u)|u \in [0, 1]\}$, where $g_t(u) = inf\{z_t \big| u \le Q_t(z_t)\}$.*

(a) *If $z_t \in J_t$ and $p_t(y_t|\mathcal{F}_{t-1}) = f_t(y_t|\mathcal{F}_{t-1})$, then $Q_t(z_t) = z_t$.*
(b) *Additionally, if J_t is invariant over time $t = 1, 2, \ldots, T$, then $Z_1, Z_2, \ldots, Z_T$ are i.i.d. with $Q_t(z) = z$ for all $t = 1, 2, \ldots, T$ where $z \in J \equiv J_t$.*

Proofs are given in Appendix A. Lemma 2.1-(a) implicitly implies that $Q_{t_1}(z) \neq Q_{t_2}(z)$ if $t_1 \neq t_2$, contrary to the continuous case for which $Q_{t_1}(z) = Q_{t_2}(z)$ regardless of t_1 and t_2 because J_t is invariant over time t and takes real values from [0, 1]. Thus, for continuous case, Lemma 2.1-(b) is automatically met, and Lemma 2.1 is reduced to Rosenblatt (1952), Diebold et al. (1998), and Clements (2006). As each PIT at time t (i.e., Z_t) has jumps only at z_t included in J_t by 2.1-(a), the PIT's, $Z_1, \ldots, Z_T$, are i.i.d., by 2.1-(b), only when J_t's are all the same. This implies that the ECDF of the observed PIT $z_1, z_2, \ldots, z_T$ should be a straight line with the slope 1. We note that a histogram of the PIT values used in continuous random variables is useless in evaluating the predictive distribution because our PIT values are discrete and no longer uniformly distributed.

2.2 INAR(p) process and expected residuals

For a time series of count data, integer-valued analogues of the usual ARIMA models have been suggested (McKenzie 1985; Al-Osh and Alzaid 1987; Alzaid and Al-Osh 1990; Du and Li 1991; Kim and Park 2008). These integer-valued time series models possess many features in common with the ARMA models. Both can express their dynamics in the form of difference equations and share a common behavior in correlation structures. A primary difference between them is that the integer-valued time series models use thinning operators in place of the multiplication in the ARMA models. Weiß (2008a) summarized variety of thinning operators and their roles in defining integer-valued time series models.

We only consider the integer-valued autoregressive process with order p (INAR(p)) provided by Du and Li (1991) to avoid the complexity arising in model description because our approach remains the same for other INARMA models whenever their conditional distributions are given.

Du and Li (1991) defined their INAR(p) model:

$$y_t = \theta_1 \circ y_{t-1} + \theta_2 \circ y_{t-2} + \cdots + \theta_p \circ y_{t-p} + \varepsilon_t, \quad t = 1, 2, \ldots, T \tag{1}$$

where $\theta_k \circ y_{t-k} \sim Binomial\,(y_{t-k}, \theta_k)$ and $\{\theta_k \circ y_{t-k},\ k = 1, 2, \ldots, p\}$ are independent when $y_{t-1}, \ldots, y_{t-p}$ are conditioned, and $\{\varepsilon_t\}$ are i.i.d. sequence of counts. The conditional probability $P\left(y_t | y_{t-1}, \ldots, y_{t-p}\right)$ is easily obtained as given by

$$\begin{aligned} &P\left(y_t \left| y_{t-1}, \ldots, y_{t-p}\right.\right) \\ &= \sum_{i_1=0}^{min(y_t, y_{t-1})} \binom{y_{t-1}}{i_1} \theta_1^{i_1} (1-\theta_1)^{y_{t-1}-i_1} \\ &\quad \times \sum_{i_2=0}^{min(y_{t-2}, y_t - i_1)} \binom{y_{t-2}}{i_2} \theta_2^{i_2} (1-\theta_2)^{y_{t-2}-i_2} \\ &\quad \cdots \sum_{i_p=0}^{min\left(y_{t-p}, y_t - \sum_{k=1}^{p-1} i_k\right)} \binom{y_{t-p}}{i_p} \theta_p^{i_p} \left(1-\theta_p\right)^{y_{t-p}-i_p} \\ &\quad \times P\left(\varepsilon_t = y_t - \sum_{k=1}^{p} i_k\right) \end{aligned} \tag{2}$$

(Also see Bu et al. 2008).

This shows that the conditional distribution is a convolution of p binomial random variables and one error term, which are independent. Thus, the probability integral transform (PIT) $Z_t = \sum_{y=0}^{Y_t} P\left(y \left| y_{t-1}, \ldots, y_{t-p}\right.\right)$ has the support depending on $y_{t-1}, \ldots, y_{t-p}$ as binomial parameters, implying that each Z_t has a different support J_t, and thus $Z_1, \ldots, Z_T$ are not identically distributed.

This requires another measurement that produces an i.i.d. sequence of PIT values and a natural candidate is $\hat{\varepsilon}_t = y_t - \sum_{k=1}^{p} \hat{\theta}_k \circ y_{t-k}$, where $\hat{\theta}_k$ is an estimate of θ_k. If p is correctly specified and $\hat{\theta}_k$ is a good estimate in some sense, then the residual $\hat{\varepsilon}_t$ behaves like ε_t, and the PIT calculated from $\hat{\varepsilon}_t$ can in turn be treated as the realization of an i.i.d. sequence. However, $\hat{\varepsilon}_t$ is random and unobservable because of the binomial thinning operator '$\circ$' (even when $y_{t-1}, \ldots, y_{t-p}$ are known). This leads us to make estimates of $\hat{\varepsilon}_t$.

We propose the conditional expectation, $E\left(\varepsilon_t \left| y_t, y_{t-1}, \ldots, y_{t-p}\right.\right)$, as an estimate $\hat{\varepsilon}_t$ and denote it by $\tilde{\varepsilon}_{1t}$. Then, using the conditional distribution given in (2), we have the following explicit form of $\tilde{\varepsilon}_{1t}$.

Proposition 2.2 *For the stationary INAR(p) process given in (1),*

$$\widetilde{\varepsilon}_{1t} := E\left(\varepsilon_t | y_t, \ldots, y_{t-p}\right)$$
$$= \sum_{i_1=0}^{min(y_t, y_{t-1})} \sum_{i_2=0}^{min(y_{t-2}, y_t - i_1)} \cdots \sum_{i_p=0}^{min(y_{t-p}, y_t - \sum_{k=1}^{p-1} i_k)} \left\{ \left(y_t - \sum_{k=1}^{p} i_k \right) \right.$$
$$\left. \times \frac{\binom{y_{t-1}}{i_1} \theta_1^{i_1} (1-\theta_1)^{y_{t-1}-i_1} \cdots \binom{y_{t-p}}{i_p} \theta_p^{i_p} (1-\theta_p)^{y_{t-p}-i_p} P\left(\varepsilon_t = y_t - \sum_{k=1}^{p} i_k\right)}{P\left(y_t | y_{t-1}, \ldots, y_{t-p}\right)} \right\} \tag{3}$$

where $P(y_t \,|y_{t-1}, \ldots, y_{t-p})$ *is given in* (2).

The proof is given in Appendix B. When $p = 1$, in particular,

$$\widetilde{\varepsilon}_{1t} = \sum_{i=0}^{min(y_t, y_{t-1})} \left\{ (y_t - i) \times \frac{\binom{y_{t-1}}{i} \theta_1^i (1-\theta_1)^{y_{t-1}-i} P(\varepsilon_t = y_t - i)}{P(y_t \,|y_{t-1})} \right\}$$

where

$$P(y_t \,|y_{t-1}) = \sum_{i=0}^{min(y_t, y_{t-1})} \binom{y_{t-1}}{i} \theta_1^i (1-\theta_1)^{y_{t-1}-i} P(\varepsilon_t = y_t - i).$$

As shown in the proof, $\widetilde{\varepsilon}_{1t}$ is obtained whenever the two conditional probabilities, $P\left(y_t \,\middle|y_{t-1}, \ldots, y_{t-p}\right)$ and $P\left(\varepsilon_t, y_t \,\middle|y_{t-1}, \ldots, y_{t-p}\right)$, are specified. This is also true for other integer-valued time series models. For example, $\widetilde{\varepsilon}_{1t}$ for the combined INAR(p) (CINAR(p)) independent thinning model provided by Weiß (2008b) is

$$\widetilde{\varepsilon}_{1t} = E(\varepsilon_t \,|y_t, \ldots, y_{t-p})$$
$$= \sum_{k=0}^{y_t} \left\{ (y_t - k) \frac{P(\varepsilon_t = y_t - k) \sum_{i=1}^{p} \phi_i \binom{y_{t-i}}{k} \alpha^k (1-\alpha)^{y_{t-i}-k}}{P(y_t \,|y_{t-1}, \ldots y_{t-p})} \right\}$$

and

$$P(y_t \,|y_{t-1}, \ldots, y_{t-p}) = \sum_{k=0}^{y_t} P(\varepsilon_t = y_t - k) \sum_{i=1}^{p} \phi_i \binom{y_{t-i}}{k} \alpha^k (1-\alpha)^{y_{t-i}-k}$$

where ϕ_i is a probability with $\sum_{i=1}^{p} \phi_i = 1$ and α is a thinning parameter.

As the expected residual $\widetilde{\varepsilon}_{1t}$ is nonnegative, we can define the PIT of $\widetilde{\varepsilon}_{1t}$:

$$\widetilde{Z}_t = \sum_{\varepsilon=0}^{[\tilde{\varepsilon}_{1t}]} p_t(\varepsilon) \tag{4}$$

where $[x]$ is the nearest integer to x and $p_t(\varepsilon)$ is a predictive probability mass function (PMF) of ε_t in Model (1). Thus, when this $p_t(\varepsilon)$ equals the true data generating PMF of ε_t, denoted by $f_t(\varepsilon)$, and $\tilde{\varepsilon}_{1t}$ accurately represents $\hat{\varepsilon}_t$, Lemma 2.1 implies that the CDF of $\tilde{Z}_t$ is an identity function of its observation.

To check the correlation arising from the incorrect specification of the order p or underestimation (or overestimation) of the parameters involved in the INAR(p), we propose another expected residual $\widetilde{\varepsilon}_{2t}$ defined by

$$\begin{aligned}\tilde{\varepsilon}_{2t} &:= y_t - E(\theta_1 \circ y_{t-1} - \cdots - \theta_p \circ y_{t-p} \left| y_{t-1}, \ldots, y_{t-p} \right.) \\ &= y_t - \theta_1 y_{t-1} - \cdots - \theta_p y_{t-p}.\end{aligned} \tag{5}$$

Note that this residual $\tilde{\varepsilon}_{2t}$ is not the same as the usual residual $y_t - E(y_t \,|y_{t-1}, \ldots, y_{t-p})$, and hence calculation of $E\,(\varepsilon_t)$ is not required. However, as this modified residual can be negative, it can not be used to define a PIT. Nevertheless, as we show in the next section, $\tilde{\varepsilon}_{2t}$ is better than $\tilde{Z}_t$ (or equivalently, $\tilde{\varepsilon}_{1t}$) in detecting misspecified models and incorrectly estimated parameters through its time correlations with the following property.

Lemma 2.3 *Suppose that the INAR(p) process, $y_t = \theta_1 \circ y_{t-1} + \cdots + \theta_p \circ y_{t-p} + \varepsilon_t$, is stationary. Then $\{\tilde{\varepsilon}_{2t}, t = 1, \ldots, T\}$ are uncorrelated if $\theta_1, \ldots, \theta_p$ are correctly estimated, and Corr $(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k})$ has a negative (positive) value if θ_k is overestimated (underestimated) for $k = 1, \ldots, p$.*

The proof is given in Appendix C. Lemma 2.3 states that a positive (negative) lag correlation of $\tilde{\varepsilon}_{2t}$ with sufficiently large (small) value (e.g., greater (smaller) than $1.645/\sqrt{(T)}$ $(-1.645/\sqrt{(T)})$ implies a good indication of underestimation (overestimation) of the corresponding thinning parameter. In this regard, although the residual $\tilde{\varepsilon}_{2t}$ is exactly the same as that in the continuous counterpart of AR(p) with a zero mean error term, the signs of correlations of $\tilde{\varepsilon}_{2t}$ provide much more information for determining the order p and estimating thinning parameters, unlike correlations from the continuous AR(p) model. It is easy to show that using Weiß (2008b), $\tilde{\varepsilon}_{2t}$ in CINAR(p) is

$$\tilde{\varepsilon}_{2t} = y_t - \alpha \sum_{i=1}^{p} \phi_i y_{t-i}.$$

At this point of time, it is necessary to describe the relationship between our expected residuals and other residuals appeared in literature. Freeland and McCabe (2004a) and Bu and McCabe (2008) defined $p + 1$ residuals by $r_{kt} = E(\theta_k \circ y_{t-k} \left| y_t, y_{t-1}, \ldots, y_{t-p} \right.) - \theta_k y_{t-k}$ for $k = 1, \ldots, p$ and $r_{0t} = E\,(\varepsilon_t \,|y_t, \ldots, y_{t-p}\big) - E(\varepsilon_t)$ in a INAR(p) process and showed that the sum of $p + 1$ residuals is equal to $y_t - \sum_{i=1}^{p} \theta_i y_{t-i} - E(\varepsilon_t)$. As defined in (3), $\tilde{\varepsilon}_{1t}$ is $E(\varepsilon_t \,|y_t, \ldots, y_{t-p})$ and hence $\tilde{\varepsilon}_{1t} = r_{0t} + E(\varepsilon_t)$. As shown in Bu and McCabe (2008), the sum of $\sum_{k=0}^{p} r_{kt}$ is $y_t - \sum_{i=1}^{p} \theta_i y_{t-i} - E(\varepsilon_t)$ and thus we have the relationship of $\tilde{\varepsilon}_{2t} = \sum_{k=0}^{p} r_{kt} + E(\varepsilon_t)$. Therefore, $\tilde{\varepsilon}_{1t}$ and $\tilde{\varepsilon}_{2t}$ are directly obtained without an estimate of $E(\varepsilon_t)$, while $r_{0t}, \ldots, r_{pt}$ need $E(\varepsilon_t)$.

3 Simulation studies

The two expected residuals, $\tilde{\varepsilon}_{1t}$ and $\tilde{\varepsilon}_{2t}$, have a complementary relationship. The PIT values of $\tilde{\varepsilon}_{1t}$ are mainly used for evaluating a predictive distribution of the INAR(p) model, including overdispersion or underdispersion problem, whereas $\tilde{\varepsilon}_{2t}$ is used to test if any time lag correlation remains for a specified INAR(p) model and to check under or overestimation of model parameters.

3.1 Evaluation of predictive distribution

The $\tilde{\varepsilon}_{1t}$ is likely to have an abnormal value of ε_t when model parameters are not correctly estimated. The predictive distribution of ε_t in INAR(p), using $\tilde{\varepsilon}_{1t}$, can be evaluated by plotting an observed PIT value (i.e., $\tilde{z}_t$ in (4)) on its cumulative relative frequency (i.e., the number of observations less than or equal to $\tilde{z}_t$ divided by the total number of observations). As the CDF of PIT is an identity function by Lemma 2.1 when the predictive distribution of $\tilde{\varepsilon}_{1t}$ coincides with the true data generating function of ε_t, the plot should be a straight line with intercept = 0 and slope = 1 only if all parameters including order p and the distribution of ε_t are correctly specified and estimated.

We consider $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$ as the true data generating process (DGP) to generate $y_1, \ldots, y_{100}$ and apply these 100 counts to the same model as the DGP, to $y_t = 0.3 \circ y_{t-1} + \varepsilon_t$ with $p_t(\varepsilon)$ equal to Poisson(5), and to $y_t = 0.5 \circ y_{t-1} + \varepsilon_t$ with $p_t(\varepsilon)$ equal to Poisson(3) as three predictive processes. The ECDF plots for PIT values of the respective $\tilde{\varepsilon}_{1t}$ are displayed in Fig. 1a, b, and c. Figure 1a shows the ECDF that is a straight line with slope = 1 as expected because the DGP equals the predictive process, whereas Fig. 1b and c show a U-shaped ECDF. Thus, a U-shaped ECDF indicates underestimation of θ or λ in model $y_t = \theta \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(\lambda)$, from which we easily infer that a humped-shaped ECDF indicates an overestimation of θ and λ.

In practice, however, the estimates of θ and λ are connected in such a way that $\hat{\lambda} = (1 - \hat{\theta})\bar{y}$, where $\bar{y}$ is the sample mean of y_t and $\hat{\theta}$ is the first-order sample autocorrelation as an estimate of θ. Accordingly, when we let $y_t = \hat{\theta} \circ y_{t-1} + \varepsilon_t$ and $\varepsilon_t \sim Poisson(\bar{y}(1 - \hat{\theta}))$ (i.e., $p_t(\varepsilon) = Poisson(\bar{y}(1 - \hat{\theta}))$) as a predictive process, we have a straight line as in Fig. 1a even for incorrect estimate of $\hat{\theta}$. For example, for DGP with $y_t = 0.5 \circ y_{t-1} + \varepsilon_t$ and $\varepsilon_t \sim Poisson(5)$, we still have a straight ECDF plot even for $\varepsilon_t \sim Poisson(\bar{y}(1 - \hat{\theta}))$ with incorrect $\hat{\theta} = 0.1$ as a predictive process. This implies that the PIT of $\tilde{\varepsilon}_{1t}$ may be useless unless a good

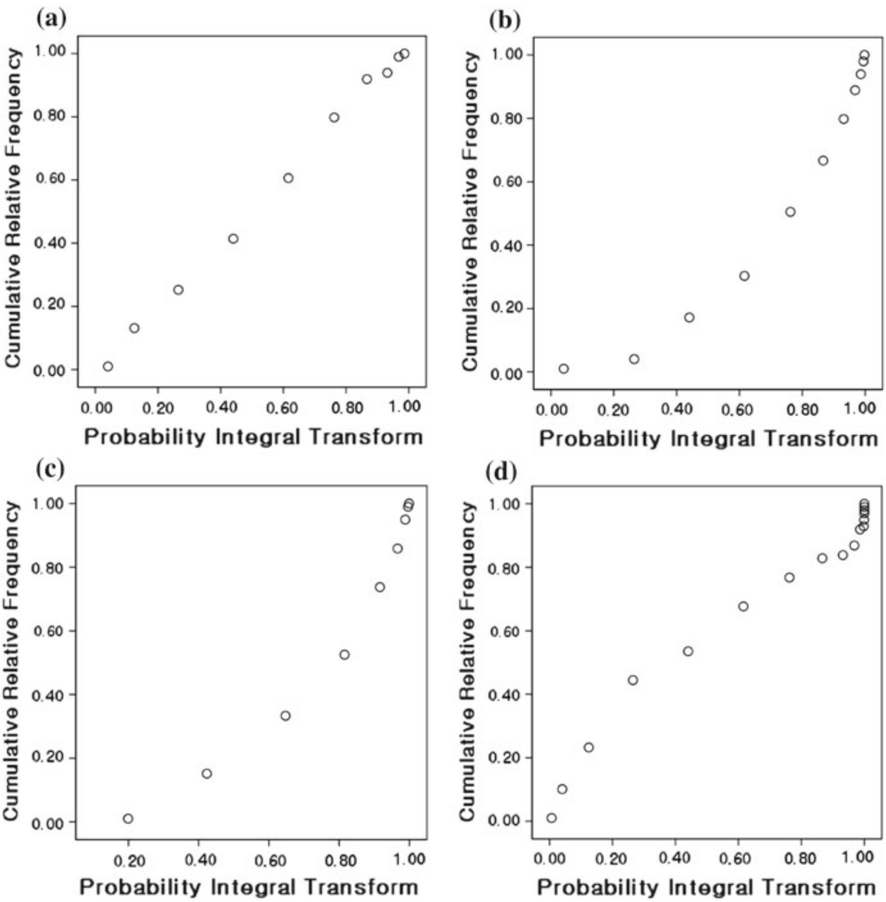


Fig. 1 ECDF plots of the PIT values for the expected residual $\tilde{\varepsilon}_{1t}$. The DGP is $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$ for (**a**)–(**c**), and $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim$ Negative Binomial random variable with mean $E(\varepsilon_t) = 5, Var(\varepsilon_t) \simeq 13.33$ for (**d**). The predicted processes are (**a**) $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$, (**b**) $y_t = 0.3 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$, (**c**) $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(3)$, and (**d**) $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$

estimate of θ is secured before evaluating the predictive process. This is done by $\tilde{\varepsilon}_{2t}$ in the following subsection.

The Poisson distribution is one of most important distributions in the analysis of count data but one frequently encounters the count data with a larger variance than its mean, called as the overdispersion problem. To examine such overdispersed data in a INAR(p) process, we apply the predictive process *of* $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$ to a sequence of counts generated from $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim$ Negative Binomial random variable with mean 5 and variance 13.33. That is, the DGP and the predictive process have the same thinning process and the same mean of the error term, but the variance in the DGP is 2.67 times larger than that in the predictive process. This case is illustrated by Fig. 1d, showing a reversed S-shaped plot for such overdispersed count data and leading to a S-shaped plot for underdispersed count data (not presented here to save space).

3.2 Correlations

We consider two correlations calculated from the PIT values of $\tilde{\varepsilon}_{1t}$ and from those of $\tilde{\varepsilon}_{2t}$ for the model specification of INAR(p) processes. To compare the correlations of $\tilde{\varepsilon}_{1t}, \tilde{\varepsilon}_{2t}$, and Bu and McCabe's $p+1$ residuals $r_{0t}, \ldots, r_{pt}$ defined in Sect. 2.2, we computed the probability rejecting the predictive process with a misspecified order p, an underestimated or overestimated thinning parameter, or an incorrect distribution of the error term, ε_t. Here, the mean of ε_t was set to be $\bar{y}(1-\sum_{i=1}^{p}\hat{\theta}_i)$ when the DGP does not coincide with a predictive process and the distribution of ε_t is not specified. In such a case, recall that ECDF of the PIT calculated by $\tilde{\varepsilon}_{1t}$ fails to detect an underestimated or overestimated thinning parameter as discussed in Sect. 3.1.

Simulation studies were performed on the INAR(l) and INAR(2) processes, and each scenario simulation was repeated 1,000 times to calculate rejection probabilities. We used the Ljung and Box statistic (Ljung and Box 1978), defined by $Q = T(T+2)\sum_{i=1}^{6}\rho_i^2/(T-i)$, where T is the number of observations, and ρ_i is the lag-i correlation calculated from $\tilde{\varepsilon}_{1t}, \tilde{\varepsilon}_{2t}, r_{it}$, or r_{0t}. This statistic follows a χ^2 distribution with the degree of freedom 6 when the DGP and predictive processes are the same, and with the degree of freedom 5 when they are not the same because of the estimate $\bar{y}(1-\sum_{i=1}^{p}\hat{\theta}_i)$ for $E(\varepsilon_t)$.

Table 1 shows the probabilities rejecting the predictive processes when the true DGP's are $y_t = \theta \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$ and the predictive processes are $y_t = \hat{\theta} \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$, where θ varies from 0.0 to 0.8 with increment 0.2 and $\hat{\theta}$ also varies from 0.0 to 0.8 for each θ. As $r_{1t} = 0$ when $\hat{\theta} = 0.0$, the rejection probabilities of r_{1t} are not available.

All four residuals satisfied the nominal significant level $\alpha = 0.05$, and $\tilde{\varepsilon}_{2t}$ had the highest rejection probabilities for most cases. We note that the rejection probabilities of r_{0t} and r_{1t} decreased as $\hat{\theta}$ moved from 0.6 to 0.8 when $\theta = 0.0$, suggesting that their power may not be a monotonic function of $|\hat{\theta}-\theta|$. The DD in Table 1 stands for the percentage of different decisions among the cases in which r_{0t} or r_{1t} rejected the null hypothesis of $\theta = \hat{\theta}$. The percentage ranged from 0.00 to 58.5%. Namely, the percentage that r_{0t} rejected a predictive process but r_{1t} did not reject it or vice versa ranged from 0.00 to 58.5%. Thus, we chose one of them, and our choice was r_{0t} as discussed below.

We examined the influence of overdispersion on the four residuals by letting the DGP: $y_t = 0.5 \circ y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ Negative Binomial vs. the predictive: $y_t = 0.5 \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$. Therefore, in contrast to the previous simulation study, this simulation had the same thinning parameter but different distributions of ε_t. In setting the DGP, we considered three distributions of ε_t, $\varepsilon_t \sim NEG(10, 2/3)$, $NEG(5, 1/2)$, and $NEG(3, 3/8)$ whose means are all 5, which are equal to the mean of ε_t in the predictive distribution, but variances are 1.5, 2, and 2.67 times larger than the variance of the predictive distribution (i.e., 5), respectively. The results are summarized in Table 2 and indicated that $\tilde{\varepsilon}_{2t}$ persisted with the nominal level $\alpha = 0.05$, whereas $\tilde{\varepsilon}_{1t}$, r_{0t}, and

Table 1 Probabilities rejecting the predictive processes given by $y_t = \hat{\theta} \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$ when the data generating process (DGP) is $y_t = \theta \circ y_{t-1} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$

θ	$\hat{\theta}$ = 0.0				0.2					0.4				
	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	DD (%)	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	DD (%)
0.0	.053	.058	.058	–	.290	.329	.288	.325	30.8	.723	.856	.668	.797	17.1
0.2	.307	.306	.306	–	.058	.053	.057	.058	53.0	.289	.338	.276	.312	33.3
0.4	.868	.889	.889	–	.379	.378	.361	.374	22.4	.054	.051	.059	.052	49.5
0.6	1	1	1	–	.917	.933	.916	.932	2.71	.473	.451	.450	.457	16.0
0.8	1	1	1	–	.999	.999	.999	.999	0.00	.970	.974	.973	.973	0.3

θ	$\hat{\theta}$ = 0.6					0.8				
	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	DD (%)	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	DD (%)
0.0	.942	.991	.839	.961	9.3	.945	1	.800	.917	11.9
0.2	.634	.818	.567	.684	27.1	.813	.987	.666	.759	23.6
0.4	.275	.324	.253	.276	41.4	.544	.760	.443	.466	42.4
0.6	.062	.051	.064	.048	57.1	.266	.308	.221	.232	58.5
0.8	.629	.629	.620	.627	6.7	.054	.046	.040	.049	50.6

Table 2 Significant levels for $y_t = 0.5 \circ y_{t-1} + \varepsilon_t$ where DGP: $\varepsilon_t \sim NEG(r, p)$ vs. predictive : $\varepsilon_t \sim Poisson(5)$ when the nominal level $\alpha = 0.05$

DGP: ε_t ~NEG (10, 2/3)				DGP: ε_t ~NEG (5, 1/2)				DGP: ε_t ~NEG (3, 3/8)			
$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}	$\tilde{\varepsilon}_{1t}$	$\tilde{\varepsilon}_{2t}$	r_{0t}	r_{1t}
.234	.069	.248	.850	.095	.057	.168	.655	.073	.062	.107	.473

r_{1t} did not satisfy the nominal level. In particular, r_{1t} was substantially influenced by the overdispersed data, making it difficult to identify whether the high correlation of r_{1t} was a result of overdispersed data or a bad estimation of thinning parameters. In this sense, together with the results in Table 1, r_{0t} was a better choice than $r_{1t.}$ Thus, we only consider r_{0t} among $p + 1$ residuals of Bu and McCabe (2008) from now on. One of main reasons resulted in Table 2 is that $\tilde{\varepsilon}_{2t}$ depends only on the autocorrelation structure of the INAR(p) model, whereas $\tilde{\varepsilon}_{1t}, r_{0t}$, and r_{1t} depend on the conditional distribution of the time series.

Finally, we investigated the lag-one and lag-two correlations and their signs when thinning parameters were overestimated or underestimated for the DGP: $y_t = 0.3 \circ y_{t-1} + 0.3 \circ y_{t-2} + \varepsilon_t, \varepsilon_t \sim Poisson(5)$. Thus, we fixed $\theta_1 = 0.3$ and $\theta_2 = 0.3$ in the DGP. Note that by Lemma 2.3, the sign of the lag-one or lag-two correlation of $\tilde{\varepsilon}_{2t}$ is positive (negative) if θ_1 or θ_2 is underestimated (overestimated). Denote the predictive processes by $y_t = \hat{\theta}_1 \circ y_{t-1} + \hat{\theta}_2 \circ y_{t-2} + \varepsilon_t, \varepsilon_t \sim Poissson(5)$, from which we

Table 3 Numbers of positive or negative correlations and probabilities rejecting the predictive processes for DGP: $y_t = 0.3 \circ y_{t-1} + 0.3 \circ y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim Poisson(5)$ (+(−): positive (negative) correlation)

	$\hat{\theta}_1 = 0.3$ $\hat{\theta}_2 = 0.0$			$\hat{\theta}_1 = 0.3$ $\hat{\theta}_2 = 0.1$			$\hat{\theta}_1 = 0.3$ $\hat{\theta}_2 = 0.2$			$\hat{\theta}_1 = 0.3$ $\hat{\theta}_2 = 0.4$			$\hat{\theta}_1 = 0.3$ $\hat{\theta}_2 = 0.5$		
	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power
$\tilde{\varepsilon}_{1t}$	+565	+997	.671	+537	+918	.424	+486	+822	.182	−466	−799	.121	−392	−934	.249
$\tilde{\varepsilon}_{2t}$	+533	+997	.706	+537	+965	.438	+467	+810	.167	−506	−860	.170	−404	−964	.344
r_{0t}	+535	+997	.677	+525	+965	.419	+463	+799	.159	−518	−851	.147	−446	−960	.306
	$\hat{\theta}_1 = 0.0$ $\hat{\theta}_2 = 0.3$			$\hat{\theta}_1 = 0.1$ $\hat{\theta}_2 = 0.3$			$\hat{\theta}_1 = 0.2$ $\hat{\theta}_2 = 0.3$			$\hat{\theta}_1 = 0.4$ $\hat{\theta}_2 = 0.3$			$\hat{\theta}_1 = 0.5$ $\hat{\theta}_2 = 0.3$		
	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power
$\tilde{\varepsilon}_{1t}$	+993	+795	.673	+953	+693	.370	+796	+632	.166	−793	−562	.130	−927	−598	.248
$\tilde{\varepsilon}_{2t}$	+995	+786	.682	+957	+682	.370	+795	+582	.145	−844	−611	.165	−958	−644	.324
r_{0t}	+994	+784	.670	+957	+677	.362	+787	+591	.141	−840	−631	.158	−952	−672	.276
	$\hat{\theta}_1 = 0.1$ $\hat{\theta}_2 = 0.1$			$\hat{\theta}_1 = 0.1$ $\hat{\theta}_2 = 0.2$			$\hat{\theta}_1 = 0.4$ $\hat{\theta}_2 = 0.4$			$\hat{\theta}_1 = 0.2$ $\hat{\theta}_2 = 0.4$			$\hat{\theta}_1 = 0.4$ $\hat{\theta}_2 = 0.2$		
	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power	lag-1	lag-2	power
$\tilde{\varepsilon}_{1t}$	+976	+991	.780	+963	+911	.588	−726	−796	.158	+821	−746	.155	−810	+747	.193
$\tilde{\varepsilon}_{2t}$	+974	+995	.804	+958	+908	.604	−766	−844	.208	+815	−774	.166	−853	+732	.225
r_{0t}	+972	+994	.796	+959	+906	.587	−760	−858	.167	+800	−773	.159	−858	+717	.211

considered three scenarios: $\hat{\theta}_1 = 0.3$ with varying $\hat{\theta}_2$ from 0.0 to 0.5, $\hat{\theta}_2 = 0.3$ with varying $\hat{\theta}_1$ from 0.0 to 0.5, and $\hat{\theta}_1$ and $\hat{\theta}_2$, both varying from 0.1 to 0.4.

Table 3 shows the numbers of positive or negative correlations of $\tilde{\varepsilon}_{1t}$, $\tilde{\varepsilon}_{2t}$, and r_{0t} and the probabilities rejecting hypothesized predictive processes. The '+' or '−' before each number in Table 3 indicates the sign of a correlation. For example, −799 means that 799 correlations were negative among 1,000 correlations. The number of positive (negative) lag-one or lag-two correlations of $\tilde{\varepsilon}_{2t}$ were much more than 500 when the first or second thinning parameter was underestimated (overestimated). The same results were found in the correlations of $\tilde{\varepsilon}_{1t}$ and r_{0t}.

For $\hat{\theta}_1 = 0.3$ and $\hat{\theta}_2 \neq \theta_2$, $\tilde{\varepsilon}_{2t}$ had the highest number of negative lag-two correlations when $\hat{\theta}_2 > \theta_2$ (i.e., overestimated) and of positive lag-two correlations when θ_2 is underestimated, except when $\hat{\theta}_2 = 0.2$, in which case $\tilde{\varepsilon}_{1t}$ had the highest number. Similar patterns were observed for lag-one correlations when $\hat{\theta}_1 \neq \theta_1$ and $\hat{\theta}_2 = 0.3$. However, no general pattern was observed for lag-one and lag-two correlations when $\hat{\theta}_1 \neq \theta_1$ and $\hat{\theta}_2 \neq \theta_2$.

Further, $\tilde{\varepsilon}_{2t}$ had the highest rejection probabilities, except when $\hat{\theta}_1 = 0.2$ and $\hat{\theta}_2 = 0.3$ and when $\hat{\theta}_1 = 0.3$ and $\hat{\theta}_2 = 0.2$ (i.e., slightly underestimated), in which cases $\tilde{\varepsilon}_{1t}$ rejected the predictive processes with the highest probabilities. These results are consistent with those in Table 1 for the INAR(1) processes. Therefore, according to Lemma 2.3, a positive (negative) lag correlation of $\tilde{\varepsilon}_{2t}$ with sufficiently large (small) value (e.g., greater (smaller) than $1.645/\sqrt{T}(-1.645/\sqrt{T})$) provided a good indication of underestimation (overestimation) of the corresponding thinning parameter.

Note that an incorrectly estimated thinning parameter affected not only the corresponding lag correlation but also other lag correlations. For $\hat{\theta}_1 = 0.3$ and $\hat{\theta}_2 \neq 0.3$, the number of negative lag-one correlations increased as θ_2 was more seriously overestimated, whereas the number of positive lag-one correlations increased as θ_2 was more seriously underestimated. For $\hat{\theta}_1 \neq 0.3$ and $\hat{\theta}_2 = 0.3$, the number of positive lag-two correlations increased as θ_1 was more seriously underestimated. For $\hat{\theta}_1 \neq \theta_1$ and $\hat{\theta}_2 \neq \theta_2$, the numbers of positive lag-one and lag-two correlations increased as θ_1 and θ_2 were both underestimated. This leads us to use an autocorrelation diagnostics such as the Ljung and Box statistic to measure lack of fit in integer-valued time series models.

4 Application

We apply our methods to select a INAR(p) model fitted well to 267 counts that are the numbers of downloads of a free TeX editor, called CWß TeXpert, during 267 days. The counts are between 0 and 14 and exhibit a serially correlated stationary series of counts as discussed in Weiß (2008a).

Figure 2 shows the autocorrelation function for the download counts and their ECDF plot drawn under an i.i.d. assumption of the download counts. It reveals a clear first-order autocorrelation as shown in Fig. 2a and the over-dispersion problem from the reversed S-shaped PIT plot as shown in Fig. 2b. Based on the first-order autocorrelation, we applied a Poisson INAR(l) model, $y_t = \theta \circ y_{t-1} + \varepsilon_t$ with moment estimates, $\hat{\lambda} = 1.813$ and $\hat{\theta} = 0.2448$ as provided in Table 4. Figure 3 shows no autocorrelation of $\tilde{\varepsilon}_{2t}$ but still a reversed S-shaped PIT plot of $\tilde{\varepsilon}_{1t}$ whose PIT values are calculated from $p_t(\varepsilon)$ equal to Poisson(1.813), implying that a INAR(l) model well reflects the autocorrelation structure of the download counts but a Poisson marginal does not fit well for variation of the counts. Thus, another INAR(l) with other than a Poisson marginal is needed.

We consider the following three alternative INAR(l) models.

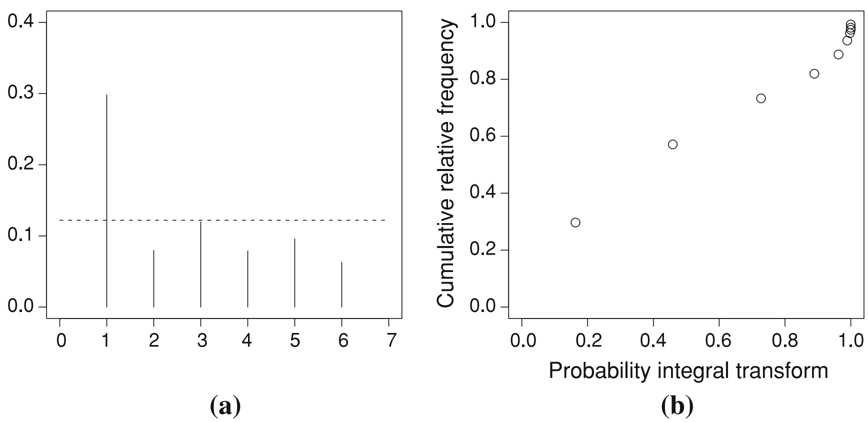
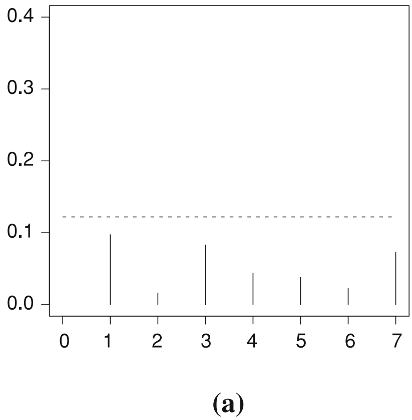


Fig. 2 Analysis of download counts. **a** autocorrelation of download counts and **b** ECDF plot of the PIT values for download counts

Table 4 Estimates and model diagnosis

Model	Estimates	Correlations		Q-statistic	
		$\tilde{\varepsilon}_{1_t}$	$\tilde{\varepsilon}_{2_t}$	$\tilde{\varepsilon}_{1_t}$	$\tilde{\varepsilon}_{2_t}$
i.i.d. Poisson	$\lambda = 2.40$	0.298*	0.245*	34.79*	24.85*
Poisson INAR (1)[†]	$(\lambda, \theta) = (1.813, 0.2448)$	0.097	0.008	5.57	3.71
NBINAR (1)	$(r, p, \theta) = (1.1227, 0.3186, 0.2448)$	0.030	0.008	3.66	3.71
NBRINAR (1)	$(r, p, \theta) = (1.1227, 0.3186, 0.2448)$	0.146*	0.008	11.05*	3.71
GPINAR (1)	$(\lambda, \rho, \theta) = (1.3552, 0.4355, 0.2448)$	0.150*	0.008	11.26*	3.71

[†] Bu and McCabe's correlations: $r_{0t} = .051$ and $r_{1t} = -0.174$*
* Significantly different from zero under $\alpha = 0.05$



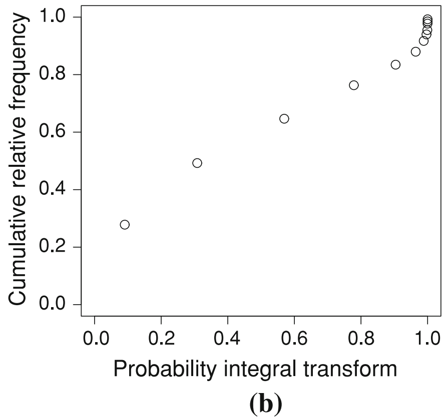


Fig. 3 Analysis of residuals from Poisson $y_t = 0.2448 \circ y_{t-1} + \varepsilon_t$. **a** autocorrelation of $\tilde{\varepsilon}_{2t}$ and **b** ECDF plot of the PIT values for $\tilde{\varepsilon}_{1t}$

NBINAR(1): $y_t = \theta \circ y_{t-1} + \varepsilon_t$ where the marginal of y_t follows a negative binomial distribution with parameters r and p (NB(r,p)) (see, McKenzie 1987 for details)
NBRINAR(1): $y_t = \beta_{r,t} \circ y_{t-1} + \varepsilon_t$ where $y_t \sim NB(r, p)$ and $\beta_{r,t} \sim$ i.i.d. $Beta$ $(r\theta, r(1-\theta))$ (see, Zheng et al. 2007)
GPINAR(1): $y_t = \theta \circ y_{t-1} + \varepsilon_t$ where y_t follows a general Poisson distribution with parameters λ and p (see Alzaid and Al-Osh 1993)

For all these models, we used moment (MM) estimates provided by Weiß (2008a) which are listed in the second column of Table 4. Since these three INAR(1) models have the same autocorrelation function (i.e., $corr(y_t, y_{t-k}) = \theta^k$) and the same residual $\tilde{\varepsilon}_{2t} = y_t - \theta y_{t-1}$ as those of Poisson INAR(1) model, the moment estimates of thinning parameter θ and the autocorrelations of $\tilde{\varepsilon}_{2t}$ for all four INAR(1) models including Poisson INAR(1) should be the same (i.e., $\hat{\theta} = 0.2448$ and $corr(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-1}) = 0.008$) as provided in Table 4. This table also provides the two correlations suggested by Bu and McCabe (2008), r_{0t} and r_{1t} where $r_{0t} = 0.051$ and $r_{1t} = -0.174$. The non-significant r_{0t} supported a INAR(1) model but the significant r_{1t} required a INAR(p) model with $p > 1$.

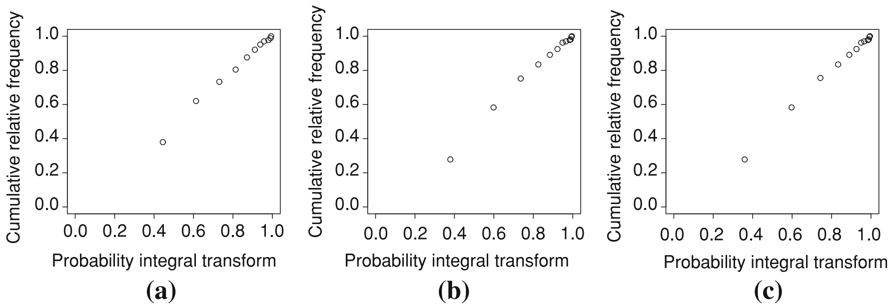


Fig. 4 ECDF plots of the PIT values for $\tilde{\varepsilon}_{1t}$ for three INAR (1) models. **a** NBINAR(1), **b** NBRINAR(1) and **c** GPINAR(1)

To select an appropriate model from the three models, two different aspects should be considered. First, we examine which model produces the ECDF plot closest to a straight line with slope 1. The ECDF plots in Fig. 4 look like almost identical straight lines and hence the three INAR(l) models seem to well reflect the overdispersed character of the download data. Secondly, we should check the autocorrelation for residuals $\tilde{\varepsilon}_{1t}$ and $\tilde{\varepsilon}_{2t}$. By using Ljung and Box statistic (Q-statistic), we can examine if there is any remaining autocorrelation. Here, the Q-statistic was calculated using from lag-one to lag-six correlations of the residuals. Table 4 clearly shows that a serial dependence in the download data should be considered from the i.i.d. Poisson with the large correlations of $\tilde{\varepsilon}_{1t}$ and $\tilde{\varepsilon}_{2t}$. Although the Poisson INAR(l) model revealed no correlation, it suffered from the overdispersion problem as observed in Fig. 3. Except NBINAR(l) model, all models appeared to have autocorrelated residuals of at least $\tilde{\varepsilon}_{1t}$ or $\tilde{\varepsilon}_{2t}$. Thus, the NBINAR(l) is most reasonable for CWß TeXpert data.

5 Conclusion

We presented two expected residuals for integer-valued time series models. These two residuals are easily obtained whenever the conditional distribution given in past observations is explicitly defined for integer-valued time series models. One residual is useful for examining the overdispersion (underdispersion) problem, and the other is useful for the independent test, parameter estimation, and model selection. As Du and Li's INAR(p) (Du and Li 1991) model has great appeal for modeling count data and our method can be easily extended to other integer-valued time series models when their conditional distributions are available, we conclude our work by presenting a general rule on how the two residuals diagnose the Du and Li's INAR(p) process. For a temporarily selected order p, we first examine lag-k correlations calculated from the PIT values of $\tilde{\varepsilon}_{1t}$ and from those of $\tilde{\varepsilon}_{2t}$ where $k \geq p+1$. If order p is not enough, some lag-k correlations for $k \geq p+1$ might be significantly positive. Second, after selecting the order p, the ECDF of PIT values of $\tilde{\varepsilon}_{1t}$ should be drawn. If it is S-shaped or reversed S-shaped, the data may be underdispersed or overdispersed, respectively, implying that the assumption of a Poisson error does not provide a good fit to the data. A U-shaped (humped-shaped) ECDF indicates underestimated (overestimated) thinning parameters or underestimated (overestimated) mean of the error term. The ECDF often appears to be a straight line with slope = 1 even when the parameters in INAR(p) are incorrectly estimated. For this case, we finally test lag-k correlations of $\tilde{\varepsilon}_{1t}$ and $\tilde{\varepsilon}_{2t}$ where $1 \leq k \leq p + m$ for some positive integer m. A significant positive (negative) lag-k correlation for $1 \leq k \leq p$ means that the k-th thinning parameter is underestimated (overestimated).

Appendix A

Proof of Lemma 2.1 (a). Denote $f_t\left(y_t|\mathcal{F}_{t-1}\right)$ and $p_t\left(y_t|\mathcal{F}_{t-1}\right)$ by $f_t\left(y_t\right)$ and $p_t\left(y_t\right)$, respectively, for simple notation. For real number z_t, note that

$$\begin{aligned} P\left(Z_t \leq z_t\,|Y_t = y_t\right) &= P\left(\sum_{y=0}^{Y_t} p_t(y) \leq z_t\,|Y_t = y_t\right) \\ &= P\left(\sum_{y=0}^{y_t} p_t\left(y\right) \leq z_t\,|Y_t = y_t\right) \\ &= 1_{\left[\sum_{y=0}^{y_t} p_t(y),\infty\right)}(z_t), \end{aligned}$$

where $1_A(z)$ is an indicator function with 1 if $z \in A$ and zero otherwise. From this expression, we have

$$\begin{aligned} Q_t(z_t) \equiv P\left(Z_t \leq z_t|\mathcal{F}_{t-1}\right) &= \sum_{y_t=0}^{\infty} P\left(Z_t \leq z_t|Y_t = y_t\right) P\left(Y_t = y_t|\mathcal{F}_{t-1}\right) \\ &= \sum_{y_t=0}^{\infty} 1_{\left[\sum_{y=0}^{y_t} p_t(y),\infty\right)}(z_t) P\left(Y_t = y_t|\mathcal{F}_{t-1}\right) = \sum_{y_t=0}^{\infty} 1_{\left[\sum_{y=0}^{y_t} p_t(y),\infty\right)}(z_t) f_t(y_t), \end{aligned}$$

or, equivalently,

$$Q_t(z_t) = \begin{cases} 0 & \text{if } z_t < p_t(0), \\ \sum_{i=0}^{l-1} f_t(i) & \text{if } \sum_{i=0}^{l-1} p_t(i) \leq z_t < \sum_{i=0}^{l} p_t(i), \quad l \geq 1 \end{cases}. \tag{6}$$

Since $g_t\left(u\right) = inf\left\{z_t|u \leq Q_t\left(z_t\right)\right\}$, using (6), one can observe that

$$g_t(u) = \begin{cases} p_t(0) & \text{if u} \leq f_t(0), \\ \sum_{i=0}^{l} p_t(i) & \text{if } \sum_{i=0}^{l-1} f_t(i) < u \leq \sum_{i=0}^{l} f_t(i), \quad l \geq 1, \end{cases}$$

implying that $J_t = \left\{\sum_{i=0}^{l} p_t(i), l = 0, 1, \ldots\right\}$. Since we only consider $z_t \in J_t$, the Eq. 6 is reduced to

$$Q_t(z_t) = \begin{cases} 0 & \text{if } z_t < p_t(0), \\ \sum_{i=0}^{l-1} f_t(i) & \text{if } \sum_{i=0}^{l-1} p_t(i) = z_t, \quad l \geq 1. \end{cases} \tag{7}$$

Finally, the assumption of $p_t(y_t) = f_t(y_t)$ and (7) give the result.

Proof of Lemma 2.1 (b). Note from Lemma 2.1-(a) that

$$\begin{aligned} & P\left(Z_1 \leq z_1, Z_2 \leq z_2, \ldots, Z_T \leq z_T\right) \\ & = \prod_{t=1}^{T} P\left(Z_t \leq z_t | \mathcal{F}_{t-1}\right) = \prod_{t=1}^{T} z_t I\left(z_t \in J_t\right) \end{aligned} \tag{8}$$

where $I(A) = 1$ if A holds and 0 if not. Since $J_t = \left\{\sum_{i=0}^{l} p_t(i), l = 0, 1, \ldots\right\}$ from Lemma 2.1-(a), the assumption of J_t being invariant over time t(i.e., $J_t = J$) implies that $p_1(i) = p_2(i) = \cdots = p_T(i)$ for each fixed $i = 0, 1, \ldots$. Thus, the PMF $p_t(i)$ does not depend on time t and (8) can be rewritten as

$$\prod_{t=1}^{T} z_t I\left(z_t \in J_t\right) = \prod_{t=1}^{T} z_t I\left(z_t \in J\right) = \prod_{t=1}^{T} P\left(Z_t \leq z_t\right).$$

Finally, since we assumed $p_t(i) = f_t(i)$ for $i = 0, 1, 2, \ldots$, the distribution of Z_t should be identical over time t.

Appendix B

Proof of Proposition 2.2 From the INAR(p) model defined in (1), it is easy to show that

$$\begin{aligned} P\left(\varepsilon_t, y_t | y_{t-1}, \ldots, y_{t-p}\right) = & \binom{y_{t-1}}{i_1} \theta_1^{i_1} \left(1 - \theta_1\right)^{y_{t-1} - i_1} \\ & \cdots \binom{y_{t-p}}{i_p} \theta_p^{i_p} \left(1 - \theta_p\right)^{y_{t-p} - i_p} P\left(\varepsilon_t = y_t - \sum_{k=1}^{p} i_k\right), \end{aligned}$$

where $0 \leq i_1 \leq min\left(y_t, y_{t-1}\right), 0 \leq i_2 \leq min\left(y_{t-2}, y_t - i_1\right), \ldots, 0 \leq i_p \leq min\left(y_{t-p}, y_t - \sum_{k=1}^{p-1} i_k\right)$.

Since $P\left(\varepsilon_t | y_t, \ldots, y_{t-p}\right) = \frac{P\left(\varepsilon_t, y_t | y_{t-1} \ldots, y_{t-p}\right)}{P\left(y_t | y_{t-1}, \ldots, y_{t-p}\right)}$, the claim is immediate.

Appendix C

Proof of Lemma 2.3 Due to the stationary assumption, we can let $Cov\,(y_t, y_{t\pm i}) = \gamma_i$ for $i = 1, 2, \ldots$. Denote θ_k^* be an estimate of θ_k for $k = 1, \ldots, p$. Then

$$\begin{aligned}
&Cov\left(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}\right)\\
&\quad= Cov\left(y_t - \theta_1 y_{t-1} - \cdots - \theta_k^* y_{t-k} - \cdots - \theta_p y_{t-p}, y_{t-k}\right.\\
&\qquad\left. - \cdots - \theta_k^* y_{t-2k} - \cdots - \theta_p y_{t-p-k}\right)\\
&\quad= \gamma_k - \theta_1 \gamma_{k+1} - \cdots - \theta_k^* \gamma_{2k} - \cdots - \theta_p \gamma_{p+k}\\
&\qquad -\theta_1\left(\gamma_{k-1} - \theta_1 \gamma_k - \cdots - \theta_k^* \gamma_{2k-1} - \cdots - \theta_p \gamma_{p+k-1}\right)\\
&\qquad\quad \vdots\\
&\qquad -\theta_k^*\left(\gamma_0 - \theta_1 \gamma_1 - \cdots - \theta_k^* \gamma_k - \cdots - \theta_p \gamma_p\right)\\
&\qquad\quad \vdots\\
&\qquad -\theta_p\left(\gamma_{-p+k} - \theta_1 \gamma_{-p+k-1} - \cdots - \theta_k^* \gamma_{-p+2k} - \cdots - \theta_p \gamma_k\right).
\end{aligned}$$

After some algebraic calculation, $Cov\left(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}\right)$ can be rewritten by

$$\begin{aligned}
Cov(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}) = \gamma_k - \sum_{j=1}^{p} \theta_j \gamma_{|k-j|} - \theta_1 &\left(\gamma_{k+1} - \sum_{j=1}^{p} \theta_j \gamma_{|k+1-j|}\right)\\
&+ \cdots - \theta_k^* \left(\gamma_{2k} - \sum_{j=1}^{p} \theta_j \gamma_{|2k-j|}\right)\\
&- \cdots - \theta_p \left(\gamma_{p+k} - \sum_{j=1}^{p} \theta_j \gamma_{|p+k-j|}\right)\\
&+ \left(\theta_k - \theta_k^*\right)\left(\gamma_0 - \sum_{\substack{j=1\\ \neq k}}^{p} \theta_j \gamma_j - \theta_k^* \gamma_k\right).
\end{aligned}$$

The first $p+1$ terms of this equation are all zero by the stationarity of INAR(p) model as proved in Du and Li (1991). Thus, we have

$$Cov(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}) = (\theta_k - \theta_k^*)\left(\gamma_0 - \sum_{\substack{j=1\\ \neq k}}^{p} \theta_j \gamma_j - \theta_k^* \gamma_k\right) \tag{9}$$

$$= \frac{(\theta_k - \theta_k^*)}{\gamma_0}\left(1 - \sum_{\substack{j=1\\ \neq k}}^{p} \theta_j \rho_j - \theta_k^* \rho_k\right)$$

where $\rho_j = \gamma_j / \gamma_0$ is the correlation between y_t and y_{t-j} As shown in Du and Li (1991) and Kim and Park (2008), we have $0 \leq \rho_j \leq 1$, $j \geq 1$ due to the non-negative thinning parameters $\theta_1, \ldots, \theta_p$. Therefore

$$1 - \sum_{\substack{j=1\\ \neq k}}^{p} \theta_j \rho_j - \theta_k^* \rho_k \geq 1 - \sum_{\substack{j=1\\ \neq k}}^{p} \theta_j - \theta_k^* > 0$$

where the last inequality is from the stationary condition. Thus, $Cov(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}) = 0$ when $\theta_k = \theta_k^*$, $Cov(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}) > 0$ when $\theta_k > \theta_k^*$, and $Cov(\tilde{\varepsilon}_{2t}, \tilde{\varepsilon}_{2,t-k}) < 0$ when $\theta_k < \theta_k^*$.

References

Al-Osh MA, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(l)) process. J Time Ser Anal 8:261–275

Alzaid AA, Al-Osh MA (1990) An integer-valued pth-order autoregressive structure (INAR(p)) process. J Appl Probab 27:314–323

Alzaid AA, Al-Osh MA (1993) Some autoregressive moving average processes with generalized poisson marginal distributions. Ann Inst Stat Math 45:223–232

Brännäs K, Hellström J (2001) Generalized integer-valued autoregression. Econom Rev 20:425–443

Bu R, McCabe B (2008) Model selection, estimation and forecasting in INAR(p) models: a likelihood-based markov chain approach. Int J Forecast 24:151–162

Bu R, McCabe BPM, Hadri K (2008) Maximum likelihood estimation of higher-order integer-valued autoregressive processes. J Time Ser Anal 29:973–994

Clements MP (2006) Evaluating the survey of professional forecasters probability distributions of expected inflation based on derived event probabüity forecasts. Empir Econ 31:49–64

Diebold FX, Gunther TA, Tay AS (1998) Evaluating density forecasts with applications to financial risk management. Int Econ Rev 39:863–883

Du J-G, Li Y (1991) The integer-valued autoregressive (INAR(p)) model. J Time Ser Anal 12:129–142

Freeland RK, McCabe BPM (2004a) Analysis of low count time series data by Poisson autoregression. J Time Ser Anal 25:701–722

Freeland R, McCabe BPM (2004b) Forecasting discrete valued low count time series. Int J Forecast 20: 427–434

Frühwirth-Schnatter S (1996) Recursive residuals and model diagnostics for normal and non-normal state space models. Environ Ecol Stat 3:291–309

Jung R, Tremayne AR (2003) Testing for serial dependence in time series models of counts. J Time Ser Anal 24:65–84

Jung R, Tremayne AR (2006) Coherent forecasting in integer time series models. Int J Forecast 22:223–238

Kedem B, Fokianos K (2002) Regression models for time series analysis. Wiley, New York

Kim H-Y, Park Y (2008) A non-stationary integer-valued autoregressive model. Stat Pap 49:485–502

Latour A (1998) Existence and stochastic structure of a non-negative integer-valued autoregressive process. J Time Ser Anal 19:439–455

Liesenfeld R, Nolte I, Pohlmeier W (2006) Modeling financial transaction price movements: a dynamic integer count data model. Empir Econ 30:795–825

Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. Biometrika 65: 297–303

McCabe BPM, Martin GM, Harris D (2009) Optimal probabilistic forecasts for counts. Monash University, Working Paper 7/09, 2009

McKenzie E (1985) Some simple models for discrete variate time series. Water Resour Bull 21:645–650

McKenzie E (1987) Innovation distributions for gamma and negative binomial autoregressions. Scand J Stat 14:79–85

McKenzie E (2003) Discrete variate time series. In: Shanbhag DN, Rao CR (eds) Handbook of statistics. Elsevier, Amsterdam pp 573–606

Ristić MM, Bakouch HB, Nastić AS (2009) A new geometric first-order integer-valued autoregressive(NGINAR(1)) process. J Stat Plan Inference 139:2218–2226

Rosenblatt M (1952) Remarks on a multivariate transformation. Ann Math Stat 23:470–472

Thyregod P, Carstensen J, Madsen H, Arnbjerg-Nielsen K (1999) Integer-valued autoregressive models for tipping bucket rainfall measurements. Environmetrics 10: 395–411

Weiß CH (2008) Thinning operations for modeling time series of counts—a survey. Adv Stat Anal 92: 319–341

Weiß CH (2008) The combined INAR(p) models for time series of counts. Stat Probab Lett 78:1817–1822

Zheng H, Basawa IV, Datta S (2006) Inference for pth-order random coefficient integer-valued autoregressive processes. J Time Ser Anal 137:411–440

Zheng H, Basawa IV, Datta S (2007) First-order random coefficient integer-valued autoregressive processes. J Stat Plan Inference 173:212–229

Zheng H, Basawa IV, Datta S (2008) First-order observation-driven integer-valued autoregressive processes. Stat Probab Lett 78:1–9