A New Overdispersed Integer-Valued Moving Average Model with Dependent Counting Series

Kaizhi Yu and Huiqiao Wang *

School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China; yuksz@swufe.edu.cn
* Correspondence: 118020209004@smail.swufe.edu.cn

Abstract: A new integer-valued moving average model is introduced. The assumption of independent counting series in the model is relaxed to allow dependence between them, leading to the overdispersion in the model. Statistical properties were established for this new integer-valued moving average model with dependent counting series. The Yule–Walker method was applied to estimate the model parameters. The estimator’s performance was evaluated using simulations, and the overdispersion test of the INMA(1) process was applied to examine the dependence between counting series.

Keywords: integer-valued moving average model; counting series; dispersion test

1. Introduction

Integer-valued time series can be encountered in numerous fields, such as epidemiology, insurance, and intraday stock transitions. The most widely used model is the integer-valued autoregressive (INAR) model, a recursive model first introduced by Alzaid and Al-Osh [1] and is similar to the traditional autoregressive (AR) model. Du and Li [2] generalized the model to the $p$-th order (which was called INAR($p$) model) and proved the ergodic and Markov properties of the model. Similar to the continuous-valued moving average model, the $q$-th order integer-valued moving average model INMA($q$) was introduced by Al-Osh and Alzaid [3], which is a slightly different form proposed by McKenzie [4].

Many researchers generalize the INAR model to deal with different real-life situations. Weiß [5] presented a new INAR($p$) model showing possible marginal distributions of the DSD family. This model overcomes the difficulty of choosing the appropriate marginal distribution. Monteiro and Scotto [6] defined the periodic integer-valued autoregressive model, driven by a periodic sequence of independent Poisson-distributed random variables. Weiß [7] proposed the extended Poisson INAR(1) model, where the innovations are assumed to be serially dependent. Zhu [8] introduced a negative binomial INGARCH model to handle integer-valued time series with overdispersion and potential extreme observations. The study by Weiß [9] discussed threshold models for integer-valued time series with infinite range and briefly discussed new models for counting data time series with a finite range. Kang and Wang [10] generalized the mixture INAR(1) model based on mixing Pegram and binomial thinning operator. Li and Wang [11] proposed the first-order mixed integer-valued autoregressive process with zero-inflated generalized power series innovations, which contains the commonly used zero-inflated Poisson and geometric distributions. To handle the non-stationary integer-valued time series with a large dispersion, Kim and Park [12] introduced an integer-valued autoregressive process with a signed binomial thinning operator (INARS($p$)).

Various modified thinning operators have been proposed to capture the specificity of real data, and many new INAR-type models have been defined. For example, Zheng and Basawa [13] introduced the random coefficient thinning operator while Ristić and
Bakouch [14] proposed the negative binomial thinning operator. The $p$th-order integer-valued autoregressive process with signed generalized power series thinning operator was proposed by Zhang et al. [15].

The most significant generalization of the thinning operator was made by Ristić et al. [16] which they called the dependent Bernoulli thinning operator. They constructed the new sequence of the Bernoulli random variable allowing correlation between counting series. Based on this, Miletić Ilić et al. [17] proposed the new model, based on the mix of regular binomial thinning and dependent thinning operator. For more details of thinning operators, refer to Weiβ [18].

MA-type models are very important in time series analysis. The method of moving average is generally popular in statistical and mathematical analyses. Some researchers study forecasting using the moving average method. Winters [19] analyzed the exponentially weighted moving average method for forecasting sales. Cox [20] proposed the weighted moving average method to predict the Markov series. Landaukas et al. [21] introduced an algebraic approach to select the appropriate weight coefficients for weight moving averages performing better than classical moving average predictors. Wind speed prediction using combined time series model and neural network prediction was studied by Nan et al. [22]. Other applied weight moving average methods in control charts. Alevizakos et al. [23] proposed the triple exponentially weighted moving average control chart (TEWMA), which improves the detection ability of the classical control chart. Capizzi and Masarotto [24] proposed the adaptive exponentially weighted moving average control chart (AEWMA), which weights past observations of the monitored process using a suitable function of the current error. Adegoke et al. [25] studied the multivariate homogeneously weighted moving average (MHWMA) control chart for monitoring a process mean vector.

Some researchers constructed MA-type models from the perspective of count time series. Brännäs and Quoreshi [26] used the INMA process to model the number of transactions for intraday stocks and extended the model to include explanatory variables. Brännäs and Hall [27] mainly focused on the estimation in the INMA model. The construction of the thinning operation from these studies is based on independent counting series, which is a strong assumption. Thus, to make the model more flexible in capturing the specificity of different data types, this assumption should be relaxed.

For small counts of intraday transactions in stocks per minute, the decision of buyer or seller could be affected by the public news, which means decisions from different individuals may not be uncorrelated. The change in inventories can sometimes be described as an INMA process. However, during a certain period, the change can be influenced by the same external factor. For the INMA model applying a discrete risk model, the number of claims for certain insurance will be affected by the same factor, such as natural disasters. Thus, these claims are no longer independent. The independence between counting series should be relaxed. Allowing the correlation between counting series is natural, and therefore the INMA model based on dependent counting series can be derived to handle different real data situations.

The rest paper is organized as follows. Section 2 presents the model construction and discussions on some relevant statistical properties, and Section 3 discussed the estimation of unknown parameters. Section 4 shows the numerical simulation results and give dispersion test for dependence between counting series, while the conclusions are given in Section 5.

2. The Model and Basic Properties

2.1. The Model Construction

The counting series $\{U_i\}_{i \in \mathbb{N}}$ of the integer-valued model is defined as:

$$U_i = (1 - V_i)W_i + V_iZ$$

$\{W_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d random variable with $W_i \sim B(1, \beta)$, $\beta \in [0, 1]$. $\{V_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d random variable with $V_i \sim B(1, \theta)$, $\theta \in [0, 1]$. $Z$ is a random variable with $Z \sim B(1, \beta)$. The operator $\circ_\theta$ is defined by $\circ_\theta \varepsilon_t = \sum_{i=1}^{\varepsilon_t} U_i$, it is the depen-
The DCINMA(q) model is defined as:

\( \text{Theorem 1.} \) The numerical characteristics of \( \{X_t\} \) in Definition 1 are as follows:

(i) \( E(X_t) = \mu_t(1 + \sum_{i=1}^{q} \beta_i) \)

(ii) \( \text{Var}(X_t) = \sigma_t^2 + \sum_{i=1}^{q} [\mu_i \beta_i + \mu_i^2 \theta^2 \beta_i (1 - \beta_i) - \mu_i (\theta^2 \beta_i - \theta^2 \beta_i^2 + \beta_i^2) + \beta_i^2 \sigma_t^2] \)

(iii) \( \text{cov}(X_t, X_{t-k}) = \begin{cases} \sigma_t^2 \sum_{i=0}^{d-k} \beta_i \beta_{i+k} & k = 1, \ldots, q \\ 0 & k \geq q + 1 \end{cases} \)

**Proof.** See Appendix A. \( \square \)

**Theorem 2.** \( \{X_t\} \) is the process defined in Definition 1, then \( \{X_t\} \) is a covariance stationary process.

**Proof.** It can be seen from the Theorem 1 that the unconditional mean and unconditional variance of \( X_t \) is a finite constant given the distribution of \( \epsilon_t \). Thus, \( X_t \) is a stationary process. \( \square \)

**Theorem 3.** \( \{X_t\} \) is the process defined in Definition 1, then \( \{X_t\} \) is ergodic in mean and autocovariance function.

**Proof.** See Appendix A. \( \square \)

2.3. The Probability Generating Functions for DCINMA Model

Ristić et al. [16] derived the probability generating function of the \( \sum_{i=1}^{n} U_i \) as follows:

\[ \Phi_{U_i} = E[s^{(U_1+U_2+\ldots+U_n)}] = (1 - \beta)(1 - \beta(1 - \theta)(1 - s))^n + \beta(1 - \beta + \theta(1 - \beta))(1 - s)^n \]

The above equation implies that the term \( \sum_{i=1}^{n} U_i \) has a distribution of:

\[ U_1 + U_2 + \ldots + U_n = \begin{cases} \text{Bin}(n, \beta(1 - \theta)) & w.p \ 1 - \beta \\ \text{Bin}(n, \beta + \theta - \beta \theta) & w.p \ \beta \end{cases} \]
Then the probability generating function (PGF) of \{X_t\} is:

\[
\phi_{X_t}(s) = [(1 - \beta) \cdot \phi_\varepsilon(1 - \beta(1 - \theta)(1 - s)) + \beta \cdot \phi_\varepsilon(1 - (\beta + \theta - \beta\theta)(1 - s))] \cdot \phi_\varepsilon(s)
\]

\(\phi_\varepsilon(s)\) is the probability generating function of the \(\varepsilon_t\). Given the distribution of \(\varepsilon_t\), the explicit expression of the probability generating function can be derived. Suppose Poisson distribution of \(\varepsilon_t\), then

\[
\phi_{X_0}(s) = [(1 - \beta) \cdot e^{-\lambda\beta(1-\theta)(1-s)} + \beta e^{-\lambda\beta(\beta+\theta-\beta\theta)(1-s)}] \cdot e^{\lambda(s-1)}
\]

The probability generating function is defined by the probabilities. The uniqueness of a power series expansion implies that the probability generating function in turn defines probabilities.

Therefore, we can derive the probability of \(X_t\). For \(X_t = j\), the probability mass function of \(X_t\) is:

\[
P(X_t = j) = \left[\frac{1}{j!} \frac{d^j \phi(s)}{ds^j}\right]_{s=0}
\]

The bivariate probability generating function of \(X_t\) is \(\Phi_{X_t,X_{t-1}}(s_1,s_2)\). Thus, deriving the explicit expression of bivariate probability generating function with Poisson innovation is easy for DCINMA(1) process.

\[
E(s_1^{X_t} s_2^{X_{t-1}}) = E(s_1^{\beta\varepsilon_{t-1}+\varepsilon_t} s_2^{\beta\varepsilon_{t-2}+\varepsilon_{t-1}}) = E(s_1^{\beta\varepsilon_{t-1}} \cdot s_1^{\beta\varepsilon_{t-2}} \cdot s_2^{\varepsilon_{t-1}}) = E(s_1^{\varepsilon_t}) \cdot E(s_1^{\beta\varepsilon_{t-2}}) \cdot E(s_2^{\beta\varepsilon_{t-1}})
\]

Given the Poisson distribution of the innovation term, we can obtain:

\[
E(s_1^{\varepsilon_t}) = e^{\lambda(s_1-1)} \quad E(s_2^{\beta\varepsilon_{t-1}}) = (1 - \beta) \cdot e^{-\lambda\beta(1-\theta)(1-s_2)} + \beta \cdot e^{-\lambda\beta(\beta+\theta-\beta\theta)(1-s_2)}
\]

\[
E(s_1^{\beta\varepsilon_{t-1}} \cdot s_2^{\varepsilon_{t-1}}) = (1 - \beta) \cdot e^{\lambda s_2[1-\beta(1-\theta)(1-s_2)]} + \beta \cdot e^{\lambda s_2[(1-\beta-\beta\theta)(1-s_1)]}
\]

2.4. Compare with the INMA(q) Model

The mean and the covariance of the \(q\)-th order integer-valued moving average model has the same expression, and the variance of the INMA\((q)\) process is:

\[
V_{\text{inma}} = \sigma^2_\varepsilon + \sum_{i=1}^q \left[\sigma^2_\varepsilon \beta_i + \mu_i \beta_i (1 - \beta_i)\right]
\]

From Theorem 1, the variance of the DCINMA process presents a more complicated expression than the INMA\((q)\) process due to the correlation between the counting series of parameter \(\theta\). For the Poisson innovation, the overdispersion index of the INMA and DCINMA model is as follows:

\[
I_{\text{inma}} = 1, I_{\text{dinma}} = \frac{1 + \beta + \beta \lambda \theta^2 (1 - \beta)}{1 + \beta}
\]

Since the value of \(\lambda\), \(\beta\) and \(\theta\) are all non-negative, the term \(1 + \beta \lambda \theta^2 (1 - \beta) > 1\). When two models (INMA(1) and DCINMA(1)) share the same \(\lambda\) and \(\beta\), the \(\theta\) will determine whether there is dependence between counting series \((\theta \neq 0)\). Thus, if we want to test \(\theta = 0\), it is equivalent to evaluating whether the model is overdispersed. If the value of \(\theta\) is 0, the model degenerates to INMA model.
2.5. Compare the Entropy with INMA(1) Model

Entropy is an important concept in physics, but it can also be applied to other disciplines, including cosmology and economics. Entropy is a measure of the randomness or disorder of a system. In our case, entropy can be seen as a dispersion measure for the INMA model. Thus, we evaluate the model from the perspective of entropy. The definition of Shannon entropy as follows:

\[ H(Y) = -\sum_{i=1}^{n} p(y_i) \cdot \ln p(y_i) \]

where \( Y \) is a discrete random variable with probability mass function taking values on \( y_1, \ldots, y_n \). We denote the \( \Phi_{\text{dinma}}(s) \) and \( \Phi_{\text{inma}}(s) \) as probability generating functions of the DCINMA(1) and INMA(1) process, respectively. Suppose the same innovation term for the two models follows the Poisson distribution. We can rewrite the probability generating function of them as:

\[ \Phi_{\text{dinma}}(s) = (1 - \beta) \cdot e^{(s - 1) \cdot [\lambda\beta(1 - \theta) + \lambda]} + \beta \cdot e^{(s - 1) \cdot [\lambda\beta + \theta - \beta \theta + \lambda]} \]
\[ \Phi_{\text{inma}}(s) = e^{(s - 1) \cdot (\lambda\beta + \lambda)} \]

Thus, we can conclude the distribution of DCINMA(1) and INMA(1) process based on the definition of the probability generating function. \( X_t^{\text{inma}(1)} \) and \( X_t^{\text{dinma}(1)} \) denote the sample from INMA(1) model and DCINMA(1) model.

\[ X_t^{\text{inma}(1)} \sim \text{Poi}!(\lambda\beta + \lambda) \]
\[ X_t^{\text{dinma}(1)} \sim \begin{cases} 
\text{Poi}!(\lambda\beta(1 - \theta) + \lambda) & \text{w.p} \ 1 - \beta \\
\text{Poi}!(\lambda\beta + \theta - \beta \theta + \lambda) & \text{w.p} \ \beta 
\end{cases} \]

The Shannon entropy for both models can be derived as follows:

\[ H(X_t^{\text{inma}(1)}) = (\lambda\beta + \lambda)[1 - \log((\lambda\beta + \lambda))] + e^{(\lambda\beta + \lambda)} \cdot \sum_{k=0}^{\infty} \frac{(\lambda\beta + \lambda)^{k} \log(k!)}{k!} \]
\[ H(X_t^{\text{dinma}(1)}) = (1 - \beta) \cdot \{(\lambda\beta(1 - \theta) + \lambda)[1 - \log((\lambda\beta(1 - \theta) + \lambda))] + e^{(\lambda\beta(1 - \theta) + \lambda)} \cdot \sum_{k=0}^{\infty} \frac{(\lambda\beta(1 - \theta) + \lambda)^{k} \log(k!)}{k!} \}
+ \beta \cdot \{(\lambda(\beta + \theta - \beta \theta) + \lambda)[1 - \log((\lambda(\beta + \theta - \beta \theta) + \lambda))] + e^{(\lambda(\beta + \theta - \beta \theta) + \lambda)} \cdot \sum_{k=0}^{\infty} \frac{(\lambda(\beta + \theta - \beta \theta) + \lambda)^{k} \log(k!)}{k!} \} \]

The expression of entropy for the DCINMA(1) model is more complicated than the INMA(1) model due to the additional parameter \( \theta \).

3. Parameter Estimation

The estimation of the INMA model is complicated. Brännäs and Hall [27] discussed the Yule–Walker estimator, generalized moment method (GMM) based on the probability generating function (PGF) function, and the conditional least square method. Here, we did not attempt to use maximum likelihood estimation, which requires density functions that are generally not easily obtained in the INMA model, especially for this dependent situation. The results of generalized moments method-based probability generating function (PGF) function estimator are highly correlated with the values of \( z_1 \) and \( z_2 \) in \( \Phi_{\lambda}(z_1, z_2) \), which are not stable. On the other hand, for the conditional least square method, the number of estimation equations are less than the number of parameters. This means that there is an additional parameter \( \theta \) to be estimated.
Therefore, we derive the Yule–Walker estimator to obtain the unknown parameters for the Poisson DCINMA(q) model. We denote the following symbols: $\mu_X$ is the sample mean of $\{X_t\}$, $\mu_\varepsilon$ is the mean of innovation term $\varepsilon_t$, $\gamma_0$ is the sample variance of $\{X_t\}$, $\gamma_1, \gamma_2, \ldots, \gamma_q$ are the sample covariance of 1-th, 2-th, $\ldots$, $q$-th order. The $\beta_0$ for the equations below is 1.

Then, the unknown parameters in the DCINMA(q) model can be solved by equations through the sample moments function, which is as follows:

\[
\begin{align*}
\mu_X &= \mu_\varepsilon (1 + \beta_1 + \beta_2 + \ldots + \beta_q) \\
\gamma_0 &= \sigma_\varepsilon^2 + \sum_{i=1}^{q} [\mu_i \beta_i + \mu_\varepsilon^2 \theta^2 \beta_i (1 - \beta_i) - \mu_\varepsilon (\theta^2 \beta_i \theta^2 \beta_i^2 + \beta_i^2) + \beta_i^2 \sigma_\varepsilon^2] \\
\gamma_1 &= \lambda (\beta_0 \beta_1 + \beta_1 \beta_2 + \ldots + \beta_{q-1} \beta_q) \\
\gamma_2 &= \lambda (\beta_0 \beta_2 + \beta_1 \beta_3 + \ldots + \beta_{q-2} \beta_q) \\
&\vdots \\
\gamma_{q-1} &= \lambda (\beta_0 \beta_{q-1} + \beta_1 \beta_q) \\
\gamma_q &= \lambda (\beta_0 \beta_q)
\end{align*}
\]

4. Simulation Study
4.1. Estimation of the Model Parameters

In this section, we present some simulation results to show the performance of the estimator using different sample sizes (n = 100, 300, 700, 1000). We focused on the Poisson DCINMA(1) model. The three group parameter values considered in this model are as follows:

- **Model A**: ($\lambda = 1$, $\theta = 0.6$, $\beta = 0.2$)
- **Model B**: ($\lambda = 4$, $\theta = 0.7$, $\beta = 0.1$)
- **Model C**: ($\lambda = 5$, $\theta = 0.5$, $\beta = 0.1$)

We used the above parameter groups to generate the data and applied Yule–Walker method, then computed the bias and standard error based on 10,000 replications for each parameter group. The estimation results and their performance are reported in Table 1.

As shown in Table 1, we obtained nearly convergent estimators in all cases. In all cases, as the sample size increased, the bias decreased.

### Table 1. Some numerical results of the estimates for true values of the parameters $\lambda$, $\theta$ and $\beta$.  

| Sample Size | $\hat{\lambda}$ | $\hat{\theta}$ | $\hat{\beta}$ |
|-------------|------------------|----------------|---------------|
| (a) True values: $\lambda = 1$, $\theta = 0.6$, $\beta = 0.2$ | | | |
| 100 | 0.9875 | 0.5999 | 0.2641 |
| Bias | 0.0124 | 0.0001 | -0.0641 |
| Standard Error | 0.1755 | 0.2065 | 0.1705 |
| 300 | 0.9858 | 0.6192 | 0.2627 |
| Bias | 0.0141 | -0.0192 | -0.0627 |
| Standard Error | 0.1082 | 0.2049 | 0.1085 |
| 700 | 0.9831 | 0.6489 | 0.2663 |
| Bias | 0.0168 | -0.0489 | -0.0663 |
| Standard Error | 0.0699 | 0.1922 | 0.0717 |
| 1000 | 0.9817 | 0.6718 | 0.2684 |
| Bias | 0.0182 | -0.0718 | -0.0684 |
| Standard Error | 0.0597 | 0.1829 | 0.0600 |
### Table 1. Cont.

| Sample Size | $\hat{\lambda}$ | $\hat{\theta}$ | $\hat{\beta}$ |
|-------------|-----------------|----------------|---------------|
| **(b) True values: $\lambda = 4$, $\theta = 0.7$, $\beta = 0.1$** | | | |
| 100         | 3.7816          | 0.6173         | 0.1690        |
| Bias        | 0.2183          | 0.0826         | -0.0690       |
| Standard Error | 0.5296        | 0.2031         | 0.1353        |
| 300         | 3.8872          | 0.6611         | 0.1394        |
| Bias        | 0.1127          | 0.0388         | -0.0394       |
| Standard Error | 0.3364        | 0.1767         | 0.0855        |
| 700         | 3.9123          | 0.7208         | 0.1324        |
| Bias        | 0.0876          | -0.0208        | -0.0324       |
| Standard Error | 0.2337        | 0.1353         | 0.0599        |
| 1000        | 3.9144          | 0.7430         | 0.1322        |
| Bias        | 0.0855          | -0.0430        | -0.0322       |
| Standard Error | 0.2012        | 0.1142         | 0.0514        |
| **(c) True values: $\lambda = 5$, $\theta = 0.5$, $\beta = 0.1$** | | | |
| 100         | 4.7799          | 0.5709         | 0.1556        |
| Bias        | 0.2200          | -0.0709        | -0.0556       |
| Standard Error | 0.5777        | 0.2059         | 0.1142        |
| 300         | 4.9129          | 0.5703         | 0.1279        |
| Bias        | 0.0870          | -0.0703        | -0.0279       |
| Standard Error | 0.3715        | 0.1852         | 0.0730        |
| 700         | 4.9343          | 0.5688         | 0.1243        |
| Bias        | 0.0656          | -0.0688        | -0.0243       |
| Standard Error | 0.2615        | 0.1329         | 0.0520        |
| 1000        | 4.9340          | 0.5738         | 0.1245        |
| Bias        | 0.0659          | -0.0738        | -0.0245       |
| Standard Error | 0.2211        | 0.1330         | 0.0442        |

4.2. Testing for Dependence between Counting Series

From Section 2.5, when the two processes have the same $\lambda$ and $\beta$, the DCINMA model always presents overdispersion due to $\theta$. We tested whether the value of $\theta$ is 0, which is equivalent to assessing whether the DCINMA process presents overdispersion. Aleksandrov and Weiß [28] proposed the diagnostic test for the INMA process. In this section, only the overdispersion test was applied. Under the null hypothesis $H_0: \theta = 0$ (the process does not present overdispersion), the distribution of index dispersion has the following form:

$$\hat{I}_{\text{disp}} \xrightarrow{d} N\left(1 - \frac{1}{T} \frac{1 + 3\beta}{1 + \beta}, \frac{1}{T} \left(2 + 4\left(\frac{-\beta}{1 + \beta}\right)^2\right)\right)$$

We then analyzed the simulation results to assess the performance of the overdispersion test. The nominal level $\alpha = 0.05$ was employed, for sample sizes, $n = 100, 150, 250$. We did 1000 replications to calculate the size and power of the test. The following parameter groups were considered:

- **Model D**: $(\lambda = 3, \theta=0.4, \beta = 0.5)$
- **Model E**: $(\lambda = 5, \theta=0.4, \beta = 0.3)$
- **Model F**: $(\lambda = 6, \theta=0.3, \beta = 0.4)$
- **Model G**: $(\lambda = 7, \theta=0.7, \beta = 0.8)$

From Table 2, given the same values for $\lambda$ and $\beta$, under the null hypothesis, $\theta = 0$, the size of $\hat{I}_{\text{disp}}$ are close to nominal level. Under the alternative situation, $\theta \neq 0$, for all cases, as $n$ increase the power quickly increases to 1.
| $\lambda$ | $\beta$ | $\theta$ | $n$ 100 | 250 | 500 |
|---------|--------|---------|--------|----|----|
| 3       | 0.4    | 0       | 0.054  | 0.057 | 0.058 |
|         |        | 0.5     | 0.431  | 0.877 | 0.937 |
| 5       | 0.4    | 0       | 0.053  | 0.057 | 0.047 |
|         |        | 0.3     | 0.447  | 0.824 | 0.901 |
| 6       | 0.3    | 0       | 0.052  | 0.055 | 0.047 |
|         |        | 0.4     | 0.63   | 0.8    | 0.99  |
| 7       | 0.7    | 0       | 0.052  | 0.046 | 0.053 |
|         |        | 0.8     | 0.998  | 1    | 1      |

5. Real Data Example

We then applied the proposed model to a real dataset. Unlike the integer-valued autoregressive model, since the density function of this DCINMA process is hard to obtain, the Akaike information criterion (AIC) is difficult to use in measuring the fitness of such an INMA-type model. To assess the performance of the model, we adapted the parametric resampling method by Jung et al. [29].

We used crime data available from the Forecasting Principles site. The data consist of 144 observations of monthly larceny counts for the City of Pittsburgh from January of 1990 to December of 2001. It would be highly improbable for criminals to remain in the same place for a long time. They probably would flee in various directions to commit offenses, so that the INMA-type model is appropriate in this case.

The calculated of sample mean (4.73) and variance (6.40) suggest that the dataset is overdispersed. In addition, the sample autocorrelation function (ACF) in Figure 1 shows that the order of the model should be set to one. We fitted the data set into the DCINMA(1) model and INMA(1) model to evaluate the model performance for overdispersed data. The estimation results for the two models are presented in Table 3.

| $\lambda$ | $\beta$ | $\theta$ |
|----------|--------|---------|
| DCINMA(1) | 3.21   | 0.47    | 0.80    |
| INMA(1)  | 3.82   | 0.24    | 0       |

The parametric resampling method can be performed in several steps:

**Step 1:** Generate samples with length equal to the original dataset from the fitted model for $R$ times.

**Step 2:** Compute the autocorrelation function (ACF) for each sample series. Then derive the empirical sample autocorrelation function (ACF) for the different lag orders.

**Step 3:** Using the results from **Step 2**, compute the $100(1 - \alpha/2)$ and $100(\alpha/2)$ quantiles.

For the given example, $R$ was set to 5000, and $\alpha$ was set to 0.05. We plotted the acceptance envelope for DCINMA(1) model and INMA(1) model, and the results are shown in Figures 2 and 3. The sample autocorrelations for the real dataset lie within the acceptance envelopes of the DCINMA(1) model except for only one point, while the important second order of the autocorrelation function (ACF) exceed the upper bound of the acceptance envelopes of the INMA(1) model. It is clear that the acceptance envelopes of the DCINMA(1) model performs better than the acceptance envelopes of the INMA(1) model. Thus, the proposed model is suitable for this dataset, adequately representing the autocorrelation for real dataset.
Figure 1. Time series plot (top figure) and autocorrelation function (below figure panel) for actual Larceny dataset.

Figure 2. Acceptance envelope of the DCINMA(1) model for the autocorrelation function for the larceny dataset.
6. Conclusions

In this paper, we constructed a new integer-valued moving average model with dependent counting series. The statistical properties of the proposed model were discussed and evaluated. The parameter estimation of the proposed model is based on the Yule–Walker method. The new model presents overdispersion due to the dependence parameter $\theta$, which means the dependence between counting series can be verified by an overdispersion test. Numerical simulation results were used to evaluate the performance of the estimation and overdispersion test.

Future extensions for this study are as follows. First, we only focused on the stationary DCINMA(1) model in this paper. However, cases with non-stationary series are more common. The switched system is an important model in studying hybrid systems, particularly from the perspective of control science and engineering. Refer to ([30–32]) for more detailed discussion. The mechanism conducts the transformation between different subsystems, providing an approach to generalize the DCINMA process into a non-stationary case. A function can be introduced to control the change for different parameter values of the stationary case, which characterize non-stationary in series. Second, the parameter $\theta$ in our model provides a probability for overdispersion. In a switched system, the switching signal is a piecewise constant function, depending on time, external signal, output, and its own past value. Thus, the weighted switching signal function with weights $\theta$ and $(1-\theta)$ can be considered, where the weight $\theta$ gives the probability for switching to different subsystems.

Author Contributions: Conceptualization, K.Y. and H.W.; methodology, K.Y.; software, H.W.; validation, K.Y. and H.W.; formal analysis, K.Y.; investigation, H.W.; resources, K.Y.; data curation, K.Y.; writing—original draft preparation, H.W.; writing—review and editing, K.Y.; visualization, H.W.; supervision, K.Y.; project administration, K.Y.; funding acquisition, K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used to support the findings of this study are included within the article.

Acknowledgments: This research was supported by the National Social Science Fund of China (No. 18BTJ039) and Fundamental Research Funds for the Central Universities (No. JBK2102021).
addition, the authors will be grateful for the editor and reviewers’ comments and suggestions, which led to improvements of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Appendix A**

**Proof of Theorem 1.** we give the detailed derivation process for 1-th order, the \( q \)-th order can be derived in the same manner. term (i) and (ii) of the Theorem 1 are easy to verify, so we focus on the term (iii).

\[
\text{Var}(X_t) = \text{Var}(\beta \circ \epsilon_{t-1}) + \text{Var}(\epsilon_t) = V[E[\beta \circ \epsilon_{t-1} | \epsilon_{t-1}]] + \text{Var}(\epsilon_t)
\]

Then, we focus on the term \( V[\beta \circ \epsilon_{t-1} | \epsilon_{t-1}] \).

\[
V[\beta \circ \epsilon_{t-1} | \epsilon_{t-1}] = [E[(\sum_{i=1}^{\epsilon_{t-1}} U_i)^2 \epsilon_{t-1}]] - [E[\sum_{i=1}^{\epsilon_{t-1}} U_i | \epsilon_{t-1}]]^2
\]

\[
= \epsilon_{t-1} \cdot E[U_1^2] + (\epsilon_{t-1} - 1) \cdot E[U_1 \cdot U_2] - [\epsilon_{t-1} \cdot \beta]^2
\]

\[
= \epsilon_{t-1} \cdot (\beta \cdot (1 - \beta) + \beta^2) + \epsilon_{t-1} \cdot (\epsilon_{t-1} - 1) \cdot (\theta^2 \cdot \beta \cdot (1 - \beta) + \beta^2)
\]

\[
- \beta^2 \cdot \epsilon_{t-1}^2
\]

\[
E[V[\beta \circ \epsilon_{t-1} | \epsilon_{t-1}]] = (\beta \cdot (1 - \beta) + \beta^2) \cdot E[\epsilon_{t-1}] - [\beta^2 \cdot \epsilon_{t-1}^2]
\]

\[
+ E[\epsilon_{t-1} \cdot (\epsilon_{t-1} - 1)] \cdot (\theta^2 \cdot \beta \cdot (1 - \beta) + \beta^2)
\]

\[
= \lambda \cdot (\beta \cdot (1 - \beta) - \beta^2) - \beta^2 \cdot (\lambda + \lambda^2)
\]

\[
+ (\lambda + \lambda^2 - \lambda) \cdot (\theta^2 \cdot \beta \cdot (1 - \beta) + \beta^2)
\]

\[
= \lambda^2 \cdot \beta \cdot (1 - \beta) \cdot \theta^2 + \beta \cdot \lambda - \beta^2 \cdot \lambda
\]

\[
\text{Var}(X_t) = V[\beta \cdot \epsilon_{t-1}] \lambda^2 \cdot \beta \cdot (1 - \beta) \cdot \theta^2 + \beta \cdot \lambda - \beta^2 \cdot \lambda + \lambda
\]

\[
= \beta \cdot \lambda + \lambda^2 \cdot \beta \cdot (1 - \beta) \cdot \theta^2 + \lambda
\]

\[
\square
\]

**Proof of Theorem 3.** let \( Y_t = X_t - \mu_X, \{X_t\} \) is the process defined in Section 2. To prove the \( Y_t \) is ergodic in autocovariance function, it is sufficient to show that:

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} \{E(Y_t Y_{t+v} Y_{t+v+l}) - R^2(v)\} = 0
\]

It is obvious \( E(Y_t) < \infty \). For \( i = 1, 2, 3, 4 \), let \( E(Y_t^2) = c_2, E(Y_t^4) = c_4 \).

**The case** \( v = 0 \):

\[
E(Y_t Y_{t+v} Y_{t+v+l}) = E(Y_t^2 Y_{t+l})
\]

If \( l = 1 \),

\[
E(Y_t Y_{t+v} Y_{t+v+l}) = E(Y_t^2 Y_{t+l}) \leq (E(|Y_t^2 Y_{t+l}|^2))^{1/2} = c_4
\]
If \( I > 1 \), \( Y_t \) and \( Y_{t+1} \) are irrelevant, then
\[
E(Y_tY_{t+v}Y_{t+1}Y_{t+v+1}) = E(Y_t^2Y_{t+1}^2) = c_2^2 = \gamma_T^2(0)
\]
\[
\frac{1}{T} \sum_{t=0}^{T} \{ E(Y_tY_{t+v}Y_{t+1}Y_{t+v+1}) - R^2(v) \} \leq \frac{1}{T} \{ c_4 - \gamma_T^2(0) \} \to 0
\]

The case \( v \geq 1 \):
If \( I \leq v + 1 \):
\[
E(Y_tY_{t+v}Y_{t+1}Y_{t+v+1}) \leq [E(|Y_tY_{t+v}|^2|Y_{t+1}Y_{t+v+1}|^2)]^{1/2} \leq [E(|Y_t|^4|Y_{t+v}|^4|Y_{t+1}|^4|Y_{t+v+1}|^4)]^{1/2} = c_4
\]
If \( I > v + 1 \), \( Y_tY_{t+1} \) and \( Y_{t+1}Y_{t+v+1} \) are irrelevant, then
\[
E(Y_tY_{t+v}Y_{t+1}Y_{t+v+1}) = E(Y_tY_{t+v})E(Y_{t+1}Y_{t+v+1}) = R^2(v)
\]
\[
\frac{1}{T} \sum_{t=0}^{T} \{ E(Y_tY_{t+v}Y_{t+1}Y_{t+v+1}) - R^2(v) \} \leq \frac{1}{T} \{ (c_4 - R^2(v))(v + 1) \} \to 0
\]
We can obtain:
\[
l_{\text{lim}_{T \to \infty}}E[|\hat{R}^2_T(v) - R^2(v)|^2] = 0
\]

Which implies:
\[
\hat{R}^2(v) \overset{p}{\to} R^2(v)
\]
The above equation is holding for variable \( Y_t \):
\[
Y_t = X_t - \mu_X, \hat{R}^2_T(v) \overset{p}{\to} R^2_T(v)
\]
We need to prove that
\[
P(|\hat{R}^2_T(v) - R^2_X(v)| \geq \epsilon) \to 0
\]
Because of \( R^2_X(v) = R^2_T(v) \), rewriting the above equation as
\[
P(|\hat{R}^2_X(v) - R^2_X(v)| \geq \epsilon) = P(|\hat{R}^2_X(v) - R^2_T(v) + R^2_T(v) - R^2_X(v)| \geq \epsilon)
\]
\[
\leq P(|\hat{R}^2_T(v) - R^2_T(v)| \geq \epsilon/2) + P(|R^2_T(v) - R^2_X(v)| \geq \epsilon/2)
\]
\( R^2_T(v) \) and \( R^2_X(v) \) expressing as
\[
R^2_X(v) = \frac{1}{T} \sum_{t=1}^{T-k} (X_{t+k} - X)(X_t - X), R^2_T(v) = \frac{1}{T} \sum_{t=1}^{T-k} Y_{t+k}Y_t = \frac{1}{T} \sum_{t=1}^{T-k} (X_{t+k} - \mu_X)(X_t - \mu_X)
\]

Then the following expression can be derived
\[
P(|\hat{R}^2_X(v) - R^2_X(v)| \geq \epsilon/2) = P((X^2 - \mu_X^2) + (\mu_X - X)(X_{t+k} + X_T) \geq \epsilon/2)
\]
\[
\leq P(X^2 - \mu_X^2 \geq \epsilon/4) + P((\mu_X - X)(X_{t+k} + X_T) \geq \epsilon/4)
\]
Since \( X \overset{p}{\to} \mu_X \), by Slutsky’s theorem
\[
X^2 - \mu_X^2 \overset{p}{\to} 0, (\mu_X - X)(X_{t+k} + X_T) \overset{p}{\to} 0
\]
Consequently,
\[ P(\left| \hat{R}_{X}(v) - R_{X}(v) \right| \geq \epsilon) \to 0, \ T \to \infty \]

References
1. Alzaid, A.A.; Al-Osh, M.A. An integer-valued pth-order autoregressive structure (INAR(p)) process. *J. Appl. Probab.* 1990, 27, 314–324. [CrossRef]
2. Du, J.; Li, Y. The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.* 1991, 12, 129–142.
3. Al-Osh, M.A.; Alzaid, A.A. Integer-valued moving average (INMA) process. *Stat. Pap.* 1988, 29, 281–300. [CrossRef]
4. McKenzie, E. Some ARMA models for dependent sequences of Poisson count. *Adv. Appl. Probab.* 1988, 20, 822–835. [CrossRef]
5. Weiβ, C.H. The combined INAR(p) models for time series of counts. *Stat. Probab. Lett.* 2008, 78, 1817–1822. [CrossRef]
6. Monteiro, M.; Scotto, M.G.; Pereira, I. Integer-valued autoregressive processes with periodic structure. *J. Stat. Plan. Infer.* 2010, 140, 1529–1541. [CrossRef]
7. Weiβ, C.H. A Poisson INAR(1) model with serially dependent innovations. *Metrika* 2015, 78, 829–851. [CrossRef]
8. Zhu, F. Modeling time series of counts with compoisson INGARCH models. *Math. Comput. Model.* 2012, 56, 191–203. [CrossRef]
9. Möller, T.; Weiβ, C.H. Threshold models for integer-valued time series with infinite or finite range. In *Stochastic Models, Statistics and Their Applications*; Steland, A., Rafałowićz, E., Szajowski, K., Eds.; Springer: Geneva, Switzerland, 2015; pp. 327–334.
10. Yao, K.; Wang, D. A new INAR(p) process with bounded support for counts showing equidispersion, underdispersion and overdispersion. *Stat. Pap.* 2019, 62, 745–767.
11. Li, C.; Wang, D. First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. *J. Korean Stat. Soc.* 2015, 44, 232–246. [CrossRef]
12. Kim, H.Y.; Park, Y. A non-stationary integer-valued autoregressive model. *Stat. Pap.* 2008, 49, 485. [CrossRef]
13. Zheng, H.; Basawa, I.V. Inference for pth-order random coefficient integer-valued autoregressive processes. *J. Time Ser. Anal.* 2006, 27, 411–440. [CrossRef]
14. Ristić, M.M.; Bakouch, H.S. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *J. Stat. Plan. Infer.* 2009, 139, 2218–2226. [CrossRef]
15. Zhang, H.; Wang, D.; Zhu, F. Inference for INAR(p) processes with signed generalized power series thinning operator. *J. Stat. Plan. Infer.* 2010, 140, 667–683. [CrossRef]
16. Ristić, M.M.; Nastić, A.S.; Miletic Ilić, A.V. A geometric time series model with dependent Bernoulli counting series. *J. Time Ser. Anal.* 2013, 34, 466–476. [CrossRef]
17. Miletic Ilic, A.V.; Ristic, M.M.; Nastic, A.S.; Bakouch, H.S. An INAR(1) model based on a mixed dependent and independent counting series. *J. Stat. Comput. Simul.* 2018, 88, 290–304. [CrossRef]
18. Weiβ, C.H. Thinning operations for modeling time series of counts—A survey. *Adv. Appl. Probab.* 2008, 92, 319. [CrossRef]
19. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Manag. Sci.* 1960, 6, 231–362. [CrossRef]
20. Cox, D.R. Prediction by exponentially weighted moving averages and related methods. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* 1961, 23, 414–422. [CrossRef]
21. Landaukas, M.; Navickas, Z.; Vainoras, A.; Ragulskis, M. Weighted moving averaging revisited—An algebraic approach. *Comput. Appl. Math.* 2017, 36, 1545–1558. [CrossRef]
22. Nan, X.; Li, Q.; Qiu, D.; Zhao, Y.; Guo, X. Short-term wind speed syntheses correcting forecasting model and its application. *Int. J. Electr. Power Energy Syst.* 2013, 49, 264–268. [CrossRef]
23. Alevizakos, V.; Chatterjee, K.; Koukouvinos, C. The triple exponentially weighted moving average control chart. *Qual. Technol. Quant. Manag.* 2021, 18, 326–354. [CrossRef]
24. Capizzi, G.; Masarotto, G. An adaptive exponentially weighted moving average control chart. *Technometrics* 2003, 45, 199–207. [CrossRef]
25. Adegoke, N.A.; Abbasi, S.A.; Smith, A.N.H.; Anderson, M.J.; Pawley, M.D.M. A multivariate homogeneously weighted moving average control chart. *IEEE Access* 2019, 7, 9586–9597. [CrossRef]
26. Brännäs, K.; Quoreshi, A.M.M.S. Integer-valued moving average modelling of the number of transactions in stocks. *Appl. Financ. Econ.* 2010, 20, 1429–1440. [CrossRef]
27. Brännäs, K.; Hall, A. Estimation in integer-valued moving average models. *Appl. Stoch. Models. Bus. Ind.* 2001, 17, 277–291. [CrossRef]
28. Aleksandrov, B.; Weiβ, C.H. Parameter estimation and diagnostic tests for INMA(1) processes. *Test* 2019, 29, 196–232. [CrossRef]
29. Jung, R.C.; McCabe, P.M.; Tremayne, A.R. Model validation and diagnostics. In *Handbook of Discrete-Valued Time*; Davis, R.A., Holan, S.H., Lund, R., Ravishanker, N., Eds.; Chapman & Hall/CRC Press: Boca Raton, FL, USA, 2015; pp. 189–218.
30. Liberzon, D.; Morse, A.S. Basic problems in stability and design of switched systems. *IEEE Control Syst. Mag.* 1999, 19, 59–70.
31. Blanchini, F.; Miani, S. *Set-Theoretic Methods in Control*, 2nd ed.; Birkhäuser: Basel, Switzerland, 2015.
32. Jin, C.; Wang, R.; Wang, Q. Stabilization of switched systems with time-dependent switching signal. *J. Frankl. Inst.-Eng. Appl. Math.* 2020, 357, 13552–13568. [CrossRef]