Communications for Statistical Applications and Methods
2023, Vol. 30, No. 3, 273–289

https://doi.org/10.29220/CSAM.2023.30.3.273
Print ISSN 2287-7843 / Online ISSN 2383-4757

Integer-Valued HAR(p) model with Poisson distribution for forecasting IPO volumes

SeongMin Yu[a], Eunju Hwang[1,a]

[a]Department of Applied Statistics, Gachon University, Korea

Abstract

In this paper, we develop a new time series model for predicting IPO (initial public offering) data with non-negative integer value. The proposed model is based on integer-valued autoregressive (INAR) model with a Poisson thinning operator. Just as the heterogeneous autoregressive (HAR) model with daily, weekly and monthly averages in a form of cascade, the integer-valued heterogeneous autoregressive (INHAR) model is considered to reflect efficiently the long memory. The parameters of the INHAR model are estimated using the conditional least squares estimate and Yule-Walker estimate. Through simulations, bias and standard error are calculated to compare the performance of the estimates. Effects of model fitting to the Korea's IPO are evaluated using performance measures such as mean square error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) etc. The results show that INHAR model provides better performance than traditional INAR model. The empirical analysis of the Korea's IPO indicates that our proposed model is efficient in forecasting monthly IPO volumes.

Keywords: initial public offering, integer-valued heterogeneous autoregressive model, conditional least squares estimate, Yule-Walker estimate

1. Introduction

The integer-valued time series model has recently attracted increasing interest due to its potential applicability and big data in various fields such as financial markets (Catania and Di Mari, 2021; Freeland and McCabe, 2004), epidemiology (Appiah and Maposa, 2021; Kowal, 2019) and social science (Chang *et al.*, 2022; McCabe and Martin, 2005). A time series of count data that fluctuates over time is observed in industrial and economic fields. In particular, monthly IPO (initial public offering) volume is in this type of data with cycles. Over the past 20 years, the number of international IPOs and market capitalization have increased. The IPO wave is closely related to economic factors such as market regulation. For example, IPO activity decreased during the dot-com bubble (Boeh and Dunbar, 2014), and IPO volume was found to be sensitive to changes in market conditions in the Hong Kong stock market (Güçbilmez, 2015). Predicting a count time series containing heterogeneous effects is an important challenge. Therefore, we develop a new model for the count time series and conduct an empirical study of monthly IPOs.

As a prior study of the count time series, the first order integer-valued autoregressive (INAR(1)) model with Poisson distribution was introduced by McKenzie (1985) and Al-Osh and Alzaid (1988)

This research was supported by Research Fund of Gachon University (GCU-202104590001).
[1] Corresponding author: Department of Applied Statistics, Gachon University, 1342 Seongnamdaero, Sujeong-Gu, Seongnam-Si, Gyeonggi-Do 13120, South Korea. E-mail: ehwang@gachon.ac.kr

Published 31 May 2023 / journal homepage: http://csam.or.kr

introduced the q^{th}-order integer-valued moving average (INMA(q)) model. An integer-valued ARMA (INARMA) model for the dependent sequence of Poisson counts was investigated by McKenzie (1988). Al-Osh and Alzaid (1990), Du and Li (1991) expanded it to the p^{th}-order (INAR(p)) model. An integer-valued model with a conditional Poisson distribution was studied by Cardinal *et al.* (1999). The INGARCH model is studied by Ferland *et al.* (2006) as an integer-valued GARCH model, following the Poisson distribution for a discrete data. Orozco *et al.* (2021) studied a new mixed first-order integer-valued autoregressive process with Poisson innovations.

As for the study of count time series model for the IPO, Ivanov and Lewis (2008) provided a study to identify determinants of the IPO issue cycle using an autoregressive conditional count model and Wang and Ning (2022) studied a Markov regime-switching model for predicting IPO volumes and detecting issue cycles.

In this paper, we propose the integer-valued heterogeneous autoregressive (INHAR) model with a Poisson distribution to implement a non-negative integer-valued time series model. The parameters are estimated using the conditional least squares (CLS) method and the Yule-Walker method, and their asymptotic normalities are established. A simulation study is conducted for the INHAR models with several orders, which are compared by evaluating bias and standard error. As an empirical example, the Korea's IPO data are adopted. Out-of-Sample forecasts are performed in the IPO dataset to calculate one-step ahead predicted value. To evaluate the prediction performance of the proposed INHAR model, mean square error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE) and root relative square error (RRSE) are calculated and compared with those of the existing INAR model, as well as their efficiencies are evaluated.

The remainder of this paper is organized as follows. In Section 2, the INAR model on which the INHAR model is based are described and the characteristics of the INHAR model. Section 3 describes the CLSE method for estimating parameters and compares it with the Yule Walker method. Section 4 describes one-step ahead forecasting and evaluation of forecasting performance by applying the proposed model to the number of monthly IPO in the South Korea from Jan.2000 to Jul.2022. The conclusions are given in Section 5.

2. Integer-Valued heterogeneous autoregressive (INHAR) model

In this section, we introduce an integer-valued heterogeneous autoregressive (INHAR) model with the Poisson thinning operator. Most of integer-valued times series models use the binomial thinning operator which has been proposed by Steutel and van Harn (1979), and implemented by McKenzie (1985), Al-Osh and Alzaid (1987), Park and Oh (1997), Freeland and McCabe (2005) and Lu and Wang (2022). This work adopts the Poisson thinning operator, which has been first introduced by Ferland *et al.* (2006, p. 926), and applied by Orozco *et al.* (2020). Also, Yang *et al.* (2019) proposed a generalized Poisson INAR model with the generalized Poisson thinning operator. Mahmoudi and Rostami (2020) introduced an integer-valued moving average model with power series innovations based on a Poisson thinning operator. A main characteristic of the Poisson distribution is to count the number of random events. For this reason, Park and Oh (1997), Freeland and McCabe (2005) and Lu and Wang (2022) who adopted the binomial thinning operator considered the Poisson distribution as innovations. A difference between the binomial distribution and Poisson distribution is that the former has a bounded finite support while the latter has infinite support. In our proposed INHAR(p) model, though having small probability, all possible nonnegative integer values can be taken, which is an advantage of the Poisson distribution and is possibly reasonable in the practical counting process

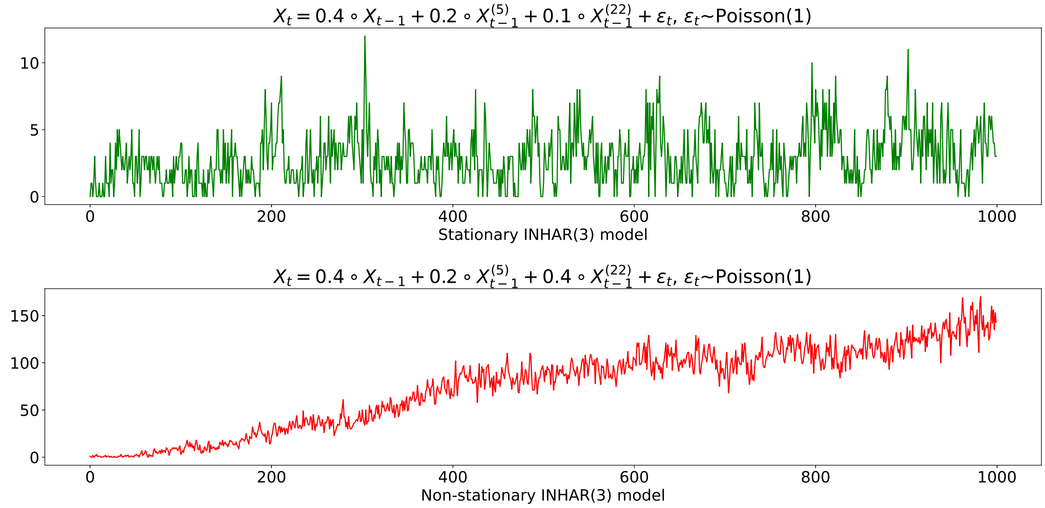

Figure 1: *Stationary and non-stationary of INHAR model.*

without any bounded assumption. In the following the definition of the Poisson thinning operator is stated as in Ferland *et al.* (2006, p. 926) and Orozco *et al.* (2020).

Definition 1. *Let X be a non-negative integer-valued random variable and α be a positive real number. The Poisson thinning operator $\circ$ is given by*

$$\alpha \circ X = \sum_{j=1}^{X} N_j, \tag{2.1}$$

where $\{N_j\}$ is a sequence of independently and identically distributed (i.i.d.) random variables of Poisson distribution with mean α and N_j is independent of X.

For a given non-negative discrete random variables X, the random variable $\alpha \circ X$ has the Poisson distribution with mean αX.

Some basic properties of the Poisson thinning operator defined in (2.1):

$$\begin{aligned} E\left[\alpha \circ X \mid X\right] &= \alpha X, \\ Var\left[\alpha \circ X \mid X\right] &= \alpha X, \\ E\left[\alpha \circ X\right] &= \alpha E\left[X\right], \\ Var\left[\alpha \circ X\right] &= \alpha E\left[X\right] + \alpha^2 Var\left[X\right]. \end{aligned}$$

The INHAR(p) process $\{X_t\}$, based on the Poisson thinning operator, is defined as follows.

Definition 2. *We define a p^{th} order INHAR model $\{X_t : t \in \mathbb{Z}\}$ with Poisson distribution by*

$$X_t = \alpha_1 \circ X^{(1)}_{t-1} + \alpha_2 \circ X^{(2)}_{t-1} + \cdots + \alpha_p \circ X^{(p)}_{t-1} + Z_t, \tag{2.2}$$

where Z_t is a sequence of i.i.d. random variables of Poisson distribution with mean λ and for $i = 1, 2, \ldots, p$, $X_{t-1}^{(i)}$ is the nearest integer of $(X_{t-1} + \cdots + X_{t-h_i})/h_i$ with positive integers $1 = h_1 < h_2 < \cdots < h_p$, i.e., $|X_{t-1}^{(i)} - (X_{t-1} + \cdots + X_{t-h_i})/h_i| \leq 1/2$.

Also, note that $\alpha_i \circ X_{t-1}^{(i)} = \sum_{j=1}^{X_{t-1}^{(i)}} N_j^{(i)}$ where $N_j^{(i)}$ are i.i.d. Poisson random variables, independent of Z_t, with parameter α_i for each $i = 1, 2, \ldots, p$. Let $\mathcal{F}_{t-1}$ be all information set up to time $t-1$, i.e. σ-field generated by $\{X_{t-1}, X_{t-2}, \ldots\}$. The conditional mean and the conditional variance of the INHAR(p) model are, respectively, given as

$$E[X_t \mid \mathcal{F}_{t-1}] = \alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda,$$
$$Var[X_t \mid \mathcal{F}_{t-1}] = \alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda.$$

The following proposition provides a stationary condition for the model. The proof is given similarly to that of conventional HAR model. It needs to be formulated rigorously because of the difference of an integer $X_{t-1}^{(i)}$ and a real number $(X_{t-1} + \cdots + X_{t-h_i})/h_i$.

Proposition 1. *For the INHAR(p) model to be stationary, it is necessary that $0 \leq \sum_{i=1}^{p} \alpha_i < 1$.*

Its proof is given in Appendix, Figure 1 shows the simulation results of the stationary and non-stationary INHAR(p) models.

Under this stationary condition, unconditional mean and unconditional variance are given as

$$\mu = E[X_t] = \frac{\lambda}{1 - \sum_{i=1}^{p} \alpha_i},$$
$$\sigma^2 = Var[X_t] = \mu \sum_{i=1}^{p} \alpha_i + \lambda + \sum_{i=1}^{p} \alpha_i^2 \tilde{\sigma}_i^2,$$

where $\tilde{\sigma}_i^2$ is given in the proof of **Proposition 1**.

3. Estimations

Let $\theta = (\alpha_1, \ldots, \alpha_p, \lambda)^\top$, a vector of parameters. Suppose that we observed $\{X_{-h_p+1}, \ldots, X_{-1}, X_0, \ldots, X_n\}$ with the *true* parameter value denoted by $\theta^o = (\alpha_1^o, \ldots, \alpha_p^o, \lambda^o)^\top$.

First the conditional least squares (CLS) method is adopted for the estimation. Since the conditional expectation is given by $E[X_t|\mathcal{F}_{t-1}] = \alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda$, we consider

$$Q_n(\theta) = \sum_{t=1}^{n} \left[X_t - \alpha_1 X_{t-1}^{(1)} - \cdots - \alpha_p X_{t-1}^{(p)} - \lambda\right]^2. \tag{3.1}$$

The CLS estimate of θ is obtained by minimizing $Q_n(\theta)$ and given by

$$\hat{\theta}_{\text{CLS}} = \left(\hat{\alpha}_{1,\text{CLS}}, \ldots, \hat{\alpha}_{p,\text{CLS}}, \hat{\lambda}_{\text{CLS}}\right)^\top = \mathbb{X}^{-1}\mathbb{Y}, \tag{3.2}$$

where

$$\mathbb{Y} = \left(\sum_{t=1}^{n} X_t X_{t-1}^{(1)}, \ldots, \sum_{t=1}^{n} X_t X_{t-1}^{(p)}, \sum_{t=1}^{n} X_t\right)^\top.$$

$$\mathbb{X} = \begin{bmatrix} \sum_{t=1}^{n}\left(X_{t-1}^{(1)}\right)^2 & \sum_{t=1}^{n} X_{t-1}^{(2)}X_{t-1}^{(1)} & \cdots & \sum_{t=1}^{n} X_{t-1}^{(p)}X_{t-1}^{(1)} & \sum_{t=1}^{n} X_{t-1}^{(1)} \\ \sum_{t=1}^{n} X_{t-1}^{(1)}X_{t-1}^{(2)} & \sum_{t=1}^{n}\left(X_{t-1}^{(2)}\right)^2 & \cdots & \sum_{t=1}^{n} X_{t-1}^{(p)}X_{t-1}^{(2)} & \sum_{t=1}^{n} X_{t-1}^{(2)} \\ \vdots & & \ddots & & \vdots \\ \sum_{t=1}^{n} X_{t-1}^{(1)}X_{t-1}^{(p)} & \sum_{t=1}^{n} X_{t-1}^{(2)}X_{t-1}^{(p)} & \cdots & \sum_{t=1}^{n}\left(X_{t-1}^{(p)}\right)^2 & \sum_{t=1}^{n} X_{t-1}^{(p)} \\ \sum_{t=1}^{n} X_{t-1}^{(1)} & \sum_{t=1}^{n} X_{t-1}^{(2)} & \cdots & \sum_{t=1}^{n} X_{t-1}^{(p)} & n \end{bmatrix}.$$

Theorem 1. *In the model (2.2) with condition $\sum_{i=1}^{p}\alpha_i < 1$, the CLS estimate $\hat{\theta}_{CLS}$ in (3.2) has the following asymptotic property: As $n \to \infty$*

$$\sqrt{n}\left(\hat{\theta}_{CLS} - \theta^o\right) \xrightarrow{\text{d}} \mathcal{N}\left(0, V^{-1}WV^{-1}\right),$$

where $V = \lim_{n\to\infty}\mathbb{X}/n$, a.s. and $W = ((w_{ij}))_{(p+1)\times(p+1)}$, with the following w_{ij}:

$$w_{ij} = E\left[X_{t-1}^{(i)}X_{t-1}^{(j)}Var(\varepsilon_t \mid \mathcal{F}_{t-1})\right], \quad 1 \le i, j \le p,$$
$$w_{i(p+1)} = E\left[X_{t-1}^{(i)}Var(\varepsilon_t \mid \mathcal{F}_{t-1})\right], \; 1 \le i \le p, \quad \text{and} \quad w_{(p+1)(p+1)} = E\left[Var(\varepsilon_t \mid \mathcal{F}_{t-1})\right],$$

where $\varepsilon_t = X_t - \alpha_1 X_{t-1}^{(1)} - \cdots - \alpha_p X_{t-1}^{(p)} - \lambda$ for each t. The expectations are with respect to the stationary distribution.

Second, we consider the Yule-Walker estimate using the fact that the HAR model of order p is a form of an AR model of order h_p. Recall $E[X_t|\mathcal{F}_{t-1}] = \alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda$ and $\mu = (\alpha_1 + \cdots + \alpha_p)\mu + \lambda$. Let $Y_t = E[X_t|\mathcal{F}_{t-1}] - \mu$ for each t, then we have

$$Y_t = \alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda - \left(\alpha_1\mu + \cdots + \alpha_p\mu + \lambda\right) = \alpha_1 Y_{t-1}^{(1)} + \cdots + \alpha_p Y_{t-1}^{(p)}, \tag{3.3}$$

where $Y_{t-1}^{(i)} = X_{t-1}^{(i)} - \mu$ for $i = 1, \ldots, p$. Note that Y_t takes real values (rather than integers), and can be expressed, (see the proof of **Theorem 2**), as a linear AR(h_p) form of

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{h_p} Y_{t-h_p} + \delta_t \tag{3.4}$$

for some noise δ_t with mean zero, where $\beta_1 = \sum_{l=1}^{p}\alpha_l/h_l$; $\beta_2 = \cdots = \beta_{h_2} = \sum_{l=2}^{p}\alpha_l/h_l$; $\cdots$, in general, (with $h_0 = 0$)

$$\beta_{h_{i-1}+1} = \cdots = \beta_{h_i} = \sum_{l=i}^{p}\frac{\alpha_l}{h_l}, \;\text{ for }\; i = 1, 2, \ldots, p. \tag{3.5}$$

Let $\beta = (\beta_1, \ldots, \beta_{h_p})^\top$ and $\hat{\beta}_{YW} = (\hat{\beta}_{1,YW}, \ldots, \hat{\beta}_{h_p,YW})^\top$ be the Yule-Walker estimate of β from the AR(h_p) model (3.4). Using backward substitution in (3.5), the Yule-Walker estimate $\hat{\alpha}_{YW} = (\hat{\alpha}_{1,YW}, \ldots, \hat{\alpha}_{p,YW})^\top$ of $\alpha = (\alpha_1, \ldots, \alpha_p)^\top$ are obtained:

$$\hat{\alpha}_{p,YW} = h_p\left(\frac{\sum_{j=h_{p-1}+1}^{h_p}\hat{\beta}_{j,YW}}{h_p - h_{p-1}}\right), \quad \hat{\alpha}_{p-1,YW} = h_{p-1}\left(\frac{\sum_{j=h_{p-2}+1}^{h_{p-1}}\hat{\beta}_{j,YW}}{h_{p-1} - h_{p-2}} - \frac{\hat{\alpha}_{p,YW}}{h_p}\right), \ldots,$$

in general, (with $h_0 = 0$)

$$\hat{\alpha}_{i,YW} = h_i\left(\frac{\sum_{j=h_{i-1}+1}^{h_i}\hat{\beta}_{j,YW}}{h_i - h_{i-1}} - \sum_{l=i+1}^{p}\frac{\hat{\alpha}_{l,YW}}{h_l}\right), \;\text{ for }\; i = 1, 2, \ldots, p. \tag{3.6}$$

Table 1: Estimation results of INHAR(2) models with $(h_1, h_2) = (1,7), (1,14), (1,21); \theta = (\alpha_1, \alpha_2, \lambda) = (0.4, 0.25, 1), n = 100, 500, 1000$

(h_1, h_2)		n	$\alpha_1 = 0.4$		$\alpha_2 = 0.25$		$\lambda = 1$	
			bias	(s.e.)	bias	(s.e.)	bias	(s.e.)
(1,7)	CLS	100	−0.0245	(0.0112)	−0.0924	(0.0177)	0.2406	(0.0367)
		500	−0.0025	(0.0023)	−0.0161	(0.0034)	0.0338	(0.0064)
		1000	−0.0007	(0.0011)	−0.0092	(0.0017)	0.0219	(0.0033)
	YW	100	0.0200	(0.0113)	−0.2195	(0.0110)	0.3954	(0.0335)
		500	0.0471	(0.0022)	−0.1811	(0.0023)	0.2568	(0.0065)
		1000	0.0498	(0.0011)	−0.1779	(0.0011)	0.2473	(0.0035)
(1,14)	CLS	100	−0.0226	(0.0117)	−0.1641	(0.0245)	0.3767	(0.0526)
		500	−0.0047	(0.0021)	−0.0329	(0.0039)	0.0772	(0.0083)
		1000	−0.0006	(0.0011)	−0.0169	(0.0019)	0.0334	(0.0040)
	YW	100	−0.0103	(0.0106)	−0.2417	(0.0106)	0.4781	(0.0330)
		500	0.0218	(0.0023)	−0.2120	(0.0022)	0.3745	(0.0066)
		1000	0.0276	(0.0011)	−0.2112	(0.0011)	0.3613	(0.0033)
(1,21)	CLS	100	−0.0334	(0.0113)	−0.2648	(0.0363)	0.5770	(0.0761)
		500	−0.0055	(0.0021)	−0.0426	(0.0047)	0.0946	(0.0094)
		1000	−0.0028	(0.0011)	−0.0206	(0.0022)	0.0485	(0.0045)
	YW	100	−0.0119	(0.0113)	−0.2550	(0.0102)	0.5125	(0.0332)
		500	0.0183	(0.0022)	−0.2274	(0.0021)	0.4113	(0.0068)
		1000	0.0193	(0.0011)	−0.2248	(0.0011)	0.4070	(0.0034)

Now its related estimate of λ is given by

$$\hat{\lambda} = \frac{1}{n}\sum_{t=1}^{n}\left(X_t - \hat{\alpha}_{1,YW}X_{t-1}^{(1)} - \cdots - \hat{\alpha}_{p,YW}X_{t-1}^{(p)}\right) \quad \text{or} \quad \hat{\mu}\left(1 - \hat{\alpha}_{1,YW} - \cdots - \hat{\alpha}_{p,YW}\right), \tag{3.7}$$

where $\hat{\mu} = \bar{X} = \sum_{t=1}^{n} X_t/n$.

The next theorem presents the asymptotic normality of $\hat{\mu}$ and $\hat{\alpha}_{YW}$ in the INHAR(p) model, rather than $\hat{\lambda}$. Let H be a $p \times h_p$ matrix given in the proof of **Theorem 2**.

Theorem 2. *In the model (2.2) with condition $\sum_{i=1}^{p}\alpha_i < 1$, we have following asymptotic property of the YW estimate $\hat{\alpha}_{YW}$: as $n \to \infty$*

$$\sqrt{n}\left(\hat{\alpha}_{YW} - \alpha\right) \xrightarrow{\text{d}} \mathcal{N}\left(0, \sigma^2 H\Gamma^{-1}H^{\top}\right),$$

where $\Gamma = ((\gamma(k-j)))_{k,j=1,\ldots,h_p}$, $h_p \times h_p$ matrix of autocovariances $\gamma(k-j) = Cov(X_{t+k}, X_{t+j})$, $\sigma^2 = \gamma(0) - \beta^{\top}\gamma$ with $\gamma = (\gamma(1), \ldots, \gamma(h_p))^{\top}$ and H is $p \times h_p$ matrix with components of functions of $(h_1, \ldots, h_p)$ such that $\alpha = H\beta$. Furthermore, we have $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma_\mu^2)$ where $\sigma_\mu^2 = \sum_{j=-\infty}^{\infty}\gamma(j)$.

4. Monte-Carlo simulation

In this section, a simulation study is conducted to see the performance of the estimates of the proposed model. Simulated data are generated by INHAR(p) models with order $p = 2$ and 3 with Poisson distribution. For order $p = 2$, lag $h_p = (1,7), (1,14)$, and $(1,21)$ are used with parameters $\theta = (\alpha_1, \alpha_2, \lambda) = (0.4, 0.25, 1)$, while for order $p = 3$, $h_p = (1,7,14)$ is used with parameters $\theta = (\alpha_1, \alpha_2, \alpha_3, \lambda) = (0.3, 0.45, 0.05, 2); (0.15, 0.12, 0.1, 1)$. The sample size $n = 100, 500$, and 1000 are adopted.

Table 2: Estimation results of INHAR(3) models with lags (1,7,14); $\theta = (\alpha_1, \alpha_2, \alpha_3, \lambda) = (0.3, 0.45, 0.05, 2)$; $(0.15, 0.12, 0.1, 1)$, $n = 100, 500, 1000$

	n	$\alpha_1 = 0.3$		$\alpha_2 = 0.45$		$\alpha_3 = 0.05$		$\lambda = 2$	
		bias	(s.e.)	bias	(s.e.)	bias	(s.e.)	bias	(s.e.)
CLS	100	−0.0334	(0.0140)	−0.0544	(0.0328)	−0.0683	(0.0300)	1.1731	(0.1439)
	500	−0.0085	(0.0026)	−0.0036	(0.0058)	−0.0328	(0.0060)	0.3579	(0.0237)
	1000	−0.0062	(0.0012)	0.0051	(0.0032)	−0.0087	(0.0029)	0.0808	(0.0100)
YW	100	0.1022	(0.0058)	0.0025	(0.0079)	−0.2567	(0.0059)	0.9407	(0.0201)
	500	0.0421	(0.0017)	0.0594	(0.0070)	−0.2042	(0.0051)	0.2922	(0.0253)
	1000	0.0161	(0.0008)	0.0297	(0.0018)	−0.0906	(0.0022)	−0.0333	(0.0188)
	n	$\alpha_1 = 0.15$		$\alpha_2 = 0.12$		$\alpha_3 = 0.1$		$\lambda = 1$	
		bias	(s.e.)	bias	(s.e.)	bias	(s.e.)	bias	(s.e.)
CLS	100	−0.0036	(0.0122)	−0.0882	(0.0287)	−0.1697	(0.0449)	0.2916	(0.0485)
	500	0.0051	(0.0021)	−0.0041	(0.0052)	−0.0539	(0.0057)	0.0499	(0.0063)
	1000	−0.0062	(0.0012)	−0.0007	(0.0026)	−0.0162	(0.0030)	0.0236	(0.0028)
YW	100	0.0343	(0.0036)	0.0567	(0.0074)	−0.0231	(0.0041)	0.3055	(0.0089)
	500	−0.0032	(0.0022)	−0.0015	(0.0037)	−0.0058	(0.0028)	0.3064	(0.0052)
	1000	−0.0019	(0.0013)	0.0047	(0.0023)	0.0035	(0.0024)	0.2880	(0.0032)

The CLS estimates and the Yule-Walker estimates are calculated and compared along with their bias and standard error. When the order $p = 2$ in Table 1, estimates of the CLS method decrease in bias and standard error as the sample size n increases. But, the Yule-Walker estimate yields somewhat inconsistent bias for α_1 as the sample size n increases. It is due to the effect of a big size of matrix H in the Yule-Walker estimates. As for the estimates of Poisson parameter λ, the CLS estimates have efficient results in bias, between 0.0219 and 0.0485 as $n = 1000$, whereas the Yule-Walker estimates have bias between 0.2473 and 0.4070 as $n = 1000$. The Yule-Walker method does not improve the bias of the estimates of λ as size increases due to affect of the big size of matrix H via the Equation (3.7).

When the order of $p = 3$ in Table 2, not all biases of the CLS method decrease as n increases, but have small values of 0.0051 and −0.0062. For the estimates of λ, when $n = 1000$, the CLS estimates have effective bias between 0.0236 and 0.0808, while the Yule-Walker estimates have bias between −0.0333 and 0.2880.

From the simulation results, we conclude that the CLS estimates have the smaller values of bias and standard error than those of the Yule-Walker estimates. In the next section, we handle the empirical example with real data set and compare the INHAR model with the existing INAR model by illustrating the forecasting along with various performance measures.

5. Empirical example

5.1. Data description

This section examines the prediction of Korea's IPO data through the INHAR model. We get the monthly IPO volume (using KOSDAQ) of the Korean stock market from Jan.2000 to Jul.2022 on the Korea investor's network for disclosure system (https://kind.krx.co.kr) which includes 1851 IPOs over 271 months.

Figure 2 and Table 3 show Korea's monthly IPO volume and its descriptive statistics. The monthly IPO volumes are stationary based on augmented Dickey Fuller (ADF) in Table 3.

Financial-Macro time series data such as economic growth rate, stock return and IPO are known to have excess kurtosis, thick tail distribution characteristics, and positive or negative skewness, and

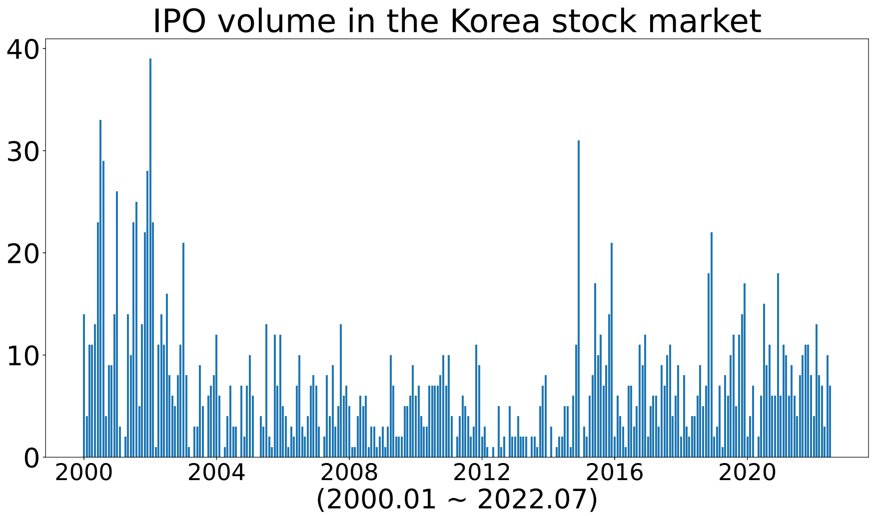

Figure 2: *Monthly IPO volume in the Korea stock market.*

Table 3: Descriptive statistics for the monthly IPO variable

n	Mean	Median	Maximum	Minimum	Std	Skewness	Kurtosis	ADF statistic	ADF p-value
271	6.830	6.000	39.000	0.000	6.104	1.985	5.408	−2.914	0.044

these characteristics have non-normality characteristics (Xiaochun, 2019).

HAR models are known to reflect the heterogeneity of time series. In other words, it means that it can reflect the long-term memory according to the moving average of the monthly IPO volume. Since IPO volume is an integer, it is modeled as an integer-valued HAR model with Poisson distribution.

5.2. Estimation results

To evaluate the performance of the INHAR model, we compare the coefficient estimates and standard errors according to the parameters. Considering that it is monthly data, lags h_p were selected by year (1,12), semi-annual (1,6,12).

The CLS estimate and Yule-Walker estimate for the fitted INHAR(p) model by order are shown in Table 4. The graph of the fitted model for CLS estimation is shown in Figure 3.

5.3. Out-of-Sample forecasting

From Jan.2000 to Dec.2020, 252 in-sample observations are used for model estimation, and from Jan.2021 to Jul.2022, 19 observations are used for out-of-sample forecasting comparisons, the number of which is denoted by m. We evaluate the five measurements of mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE) and root relative square error (RRSE) to investigate the one-step ahead forecasting performance.

Assume that $\{X_1, \ldots, X_n\}$ is the data observed during the in-sample period and X_{n+1} is the data used to calculate out-of-sample forecasts. $X_{n+1|n}$ is one-step ahead forecast. N represents the time epoch of out-of-sample to get one-step ahead forecast. $N - m + 1, \ldots, N$ are the time index used to

Table 4: Coefficients and standard errors (s.e.) of estimates for INHAR models with order p = 2, 3

		CLS		YW	
		coef.	(s.e.)	coef.	(s.e.)
INHAR(2) $h_p = (1,12)$	α_1	0.3225	(0.0617)	0.3205	(0.0618)
	α_2	0.4392	(0.1056)	0.5205	(0.1058)
	λ	1.6492	(0.6555)	1.1577	(0.6565)
INHAR(3) $h_p = (1,6,12)$	α_1	0.3069	(0.0667)	0.3025	(0.0668)
	α_2	0.1141	(0.1840)	0.1981	(0.1843)
	α_3	0.3442	(0.1862)	0.3404	(0.1865)
	λ	1.6411	(0.6564)	1.1782	(0.6575)

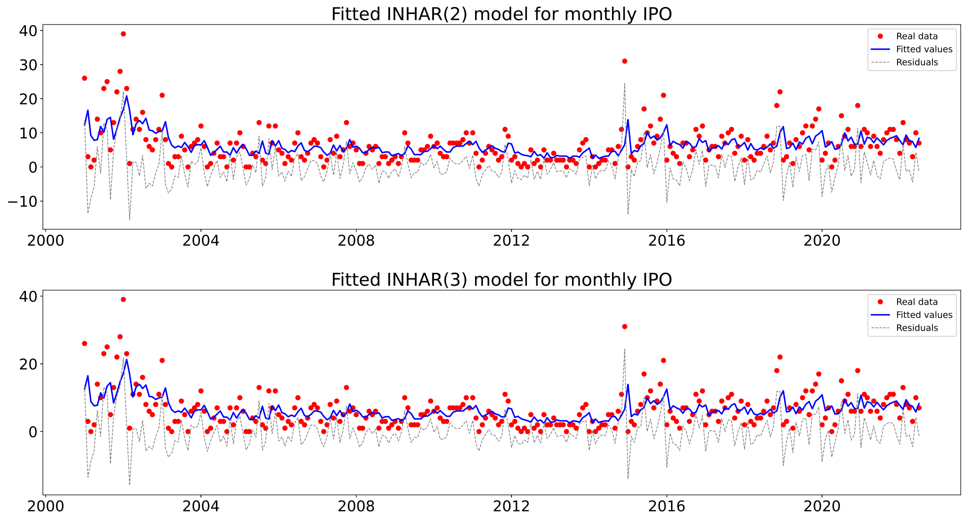

Figure 3: *Fitted INHAR(p) models for monthly IPO, p = 2, 3.*

compute the difference between the actual and predicted values, and X_N is the last observation in the total sample. The five performance measures are defined as follows.

$$\text{MAE} = \frac{1}{m}\sum_{n=N-m}^{N-1}\left|\hat{X}_{n+1|n} - X_{n+1}\right|,$$

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{n=N-m}^{N-1}\left(\hat{X}_{n+1|n} - X_{n+1}\right)^2},$$

$$\text{MAPE} = \frac{100}{m}\sum_{n=N-m}^{N-1}\left|\frac{\hat{X}_{n+1|n} - X_{n+1}}{X_{n+1}}\right|,$$

$$\text{SMAPE} = \frac{100}{m}\sum_{n=N-m}^{N-1}\frac{\left|X_{n+1} - \hat{X}_{n+1|n}\right|}{\left|X_{n+1}\right| + \left|\hat{X}_{n+1|n}\right|},$$

$$\text{RRSE} = \sqrt{\frac{\sum_{n=N-m}^{N-1}\left(\hat{X}_{n+1|n} - X_{n+1}\right)^2}{\sum_{n=N-m}^{N-1}\left(\bar{X}_1 - X_{n+1}\right)^2}},$$

Table 5: Comparison of forecasting performance of INHAR and INAR models; (2021.01–2022.07)

	CLS					YW				
	MAE	RMSE	MAPE	SMAPE	RRSE	MAE	RMSE	MAPE	SMAPE	RRSE
INHAR(2)	2.4699	2.9643	39.1116	16.5429	1.108	2.4731	2.9821	40.2379	16.5121	1.1146
INAR(2)	2.7350	3.1520	40.7012	18.4818	1.1781	2.7306	3.1500	40.6475	18.4700	1.1774
INHAR(3)	2.4897	2.9743	39.1858	16.7153	1.1117	2.4891	3.0040	40.9191	16.5691	1.1228
INAR(3)	2.8006	3.1744	41.3265	18.8416	1.1865	2.7781	3.1682	41.2272	18.6899	1.1842

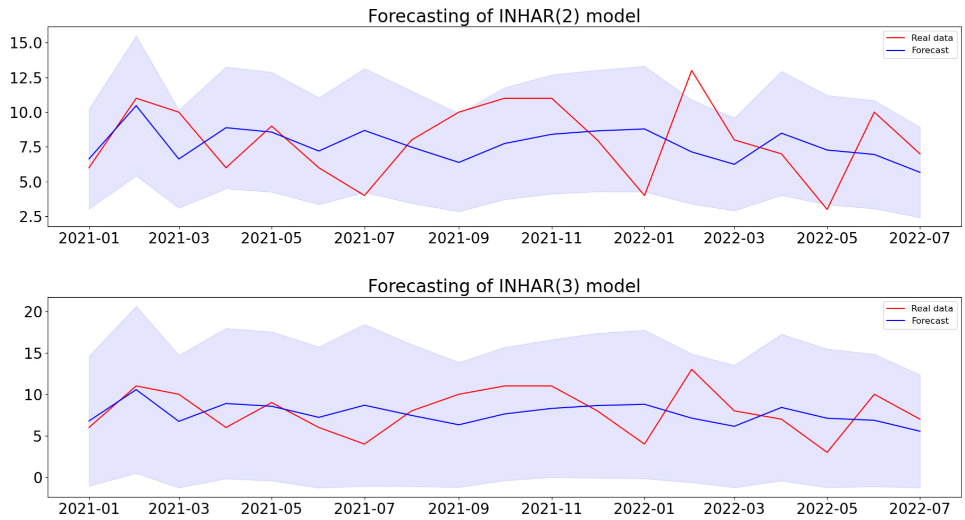

Figure 4: *Forecasting of INHAR(p) models, p = 2, 3.*

where $\bar{X}_1 = (\sum_{n=N-m}^{N-1} X_{n+1})/m$.

The one-step ahead forecasting performance of the INHAR and INAR models is presented in Table 5. The INHAR model showed better forecasting performance than the INAR model.

The forecasts plot for the CLS estimate can be seen in Figure 4, and the 95% prediction interval is displayed between Jan.2021 and Jul.2022.

Also, to elaborate on the comparison with the INHAR model, the efficiency of the proposed and existing model is calculated using two metrics of MAE, RMSE, MAPE and RRSE: The INHAR Model Efficiency, relative to the benchmark INAR model, is defined by

$$Effi_f = \left(\frac{f_0 - f_1}{f_1}\right) \times 100,$$

where f_0 is MAE, RMSE, MAPE, SMAPE and RRSE of the INAR model, respectively and f_1 is those of the INHAR model.

Table 6 displays the INHAR model efficiency.

Except for SMAPE, the CLS method is more efficient than the Yule-Walker method and the INHAR model can improve the performance error of the conventional INAR model.

Table 6: Comparison with INHAR model by computing efficiency

Efficiency	CLS					YW				
	MAE	RMSE	MAPE	SMAPE	RRSE	MAE	RMSE	MAPE	SMAPE	RRSE
p = 2	10.73	6.33	4.06	11.72	6.33	10.41	5.63	1.02	11.86	5.63
p = 3	12.49	6.73	5.46	12.72	6.73	11.61	5.47	0.75	12.8	5.47

6. Conclusion

In this study, estimation and forecasting of IPO time series data in Korea is studied by applying the INHAR model instead of the existing INAR model. The time series model for long-memory structure data is an HAR model that has recently attracted much attention in the field of financial time series analysis. The INHAR model is analyzed by combining the HAR model and the INAR model that considers integer values. For model selection, we evaluate the coefficient estimates by the conditional least squares (CLS) method and the Yule Walker method. It can be seen that the actual value of the IPO and the fit value by the INHAR model have a small error in the residuals along with the well-matched fit as shown in Figure 3. As a measure of forecasting performance, MAE, RMSE, MAPE, SMAPE and RRSE are evaluated in one-step head prediction. 95% prediction intervals and prediction values can be found in out-of-sample forecasting.

This study is the first attempt of the INHAR model for IPO data in the literature and verifies that the INHAR model provides a good prediction model for IPO data. Parameter estimation and forecasting are described with good fit and accuracy. Therefore, the proposed model can be an efficient tool for IPO analysis. We explored estimates and forecasts using the INHAR model for Korea's IPO time series data. However, the IPO data for a specific interval shows a heterogeneous variance. By applying the INHAR model as a regime-switching model, the discussion on this work can be extended.

For example, a recent Wang and Ning (2022) applied the NB-GARCH model to a regime-switching model to model the IPO data. The regime-switching INHAR model would be an interesting model for IPO data showing heterogeneity. This extension will be studied in future studies.

Appendix

Proof of Proposition 1. First we observe mean $\mu_t \equiv E[X_t] = E[E[X_t|\mathcal{F}_{t-1}]] = E[\alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda]$. Recall that $X_{t-1}^{(i)}$ is the nearest integer of $(X_{t-1} + \cdots + X_{t-h_i})/h_i$, i.e., $|X_{t-1}^{(i)} - (X_{t-1} + \cdots + X_{t-h_i})/h_i| \leq 1/2$. Let $\delta_t^{(i)} = X_{t-1}^{(i)} - (X_{t-1} + \cdots + X_{t-h_i})/h_i$. Without loss of a generality, we may assume that $E[\delta_t^{(i)}] = 0$, $\mathrm{Var}[\delta_t^{(i)}] = c \leq 1/4$ and $\delta_t^{(i)}$ is independent of $E[X_t|\mathcal{F}_{t-1}]$. Then we have

$$\mu_t = E\left[\frac{\alpha_1}{h_1}\sum_{j=1}^{h_1} X_{t-j} + \cdots + \frac{\alpha_p}{h_p}\sum_{j=1}^{h_p} X_{t-j} + \sum_{i=1}^{p}\alpha_i\delta_t^{(i)} + \lambda\right] = \beta_1\mu_{t-1} + \beta_2\mu_{t-2} + \cdots + \beta_{h_p}\mu_{t-h_p} + \lambda,$$

where β_j are given in Equation (3.5), since $E[\sum_{i=1}^{p}\alpha_i\delta_t^{(i)}] = 0$. Note that $\sum_{j=1}^{h_p}\beta_j = \sum_{i=1}^{p}\alpha_i$ (see Hwang and Shin, 2014). The recursion of μ_t above is of the form of AR model of order h_p and we write it using the backshift operator B as $(1 - \sum_{j=1}^{h_p}\beta_j B^j)\mu_t = \lambda$. Let $D(B) = 1 - \sum_{j=1}^{h_p}\beta_j B^j$. Then $\mu_t = D(B)^{-1}\lambda = \sum_{j=0}^{\infty}\psi_j B^j\lambda$ where ψ_j are coefficients of z^j in the Taylor expansion of $1/D(z)$. Hence $\mu_t = \sum_{j=0}^{\infty}\psi_j\lambda$ and noting $\sum_{j=0}^{\infty}\psi_j = 1/D(1) = 1/(1 - \sum_{j=1}^{h_p}\beta_j) = 1/(1 - \sum_{i=1}^{p}\alpha_i)$. Thus $\mu_t = \lambda/(1 - \sum_{i=1}^{p}\alpha_i)$, which is denoted by μ. The nonnegative INHAR model $\{X_t\}$ must satisfy necessarily the condition $1 - \sum_{i=1}^{p}\alpha_i > 0$.

Second, we observe

$$\begin{aligned}\mathrm{Var}\left[X_t\right] &= E\left[\mathrm{Var}\left[X_t \mid \mathcal{F}_{t-1}\right]\right] + \mathrm{Var}\left[E\left[X_t \mid \mathcal{F}_{t-1}\right]\right] \\ &= E\left[\alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda\right] + \mathrm{Var}\left[\alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda\right] \\ &= \left(\alpha_1 + \cdots + \alpha_p\right)\mu + \lambda + \sum_{i=1}^{p} \alpha_i^2 \mathrm{Var}\left[X_{t-1}^{(i)}\right],\end{aligned}$$

and now $\mathrm{Var}[X_{t-1}^{(i)}] = \mathrm{Var}[Y_{t-1}^{(i)}]$, recalling $Y_{t-1}^{(i)} = X_{t-1}^{(i)} - \mu = (\sum_{j=1}^{h_i}(X_{t-j} - \mu) + \delta_t^{(i)})/h_i$. Also, we observe

$$E\left[Y_{t-1}^{(i)} - \frac{1}{h_i}\sum_{j=1}^{h_i} Y_{t-j}\right] = E\left[E\left[\frac{1}{h_i}\sum_{j=1}^{h_i}\left(X_{t-j} - \mu\right) + \delta_t^{(i)} - \frac{1}{h_i}\sum_{j=1}^{h_i} Y_{t-j} \mid \mathcal{F}_{t-1}\right]\right] = E\left[\delta_t^{(i)}\right] = 0,$$

where $Y_t = E[X_t|\mathcal{F}_{t-1}] - \mu$, (see Equation (3.3) on page 5). Thus we may write $Y_{t-1}^{(i)} = (\sum_{j=1}^{h_i} Y_{t-j} + \delta_t^{(i)})/h_i$ a.s. and

$$Y_t = \alpha_1\left(\frac{1}{h_1}\sum_{j=1}^{h_1} Y_{t-j}\right) + \cdots + \alpha_p\left(\frac{1}{h_p}\sum_{j=1}^{h_p} Y_{t-j}\right) + \delta_t,$$

where $\delta_t = \sum_{i=1}^{p} \alpha_i \delta_t^{(i)}$ with mean $E[\delta_t] = 0$. Therefore, Equation (3.4) is obtained along with (3.5). Under the condition $\sum_{j=1}^{h_p} \beta_j = \sum_{i=1}^{p} \alpha_i < 1$, $\{Y_t\}$ is stationary and $\mathrm{Cov}(Y_t, Y_{t-j}) \equiv \gamma(j)$ is independent of t. Thus $\mathrm{Var}[X_{t-1}^{(i)}] = \mathrm{Var}[Y_{t-1}^{(i)}] =$

$$\mathrm{Var}\left(\frac{1}{h_i}\sum_{j=1}^{h_i} Y_{t-j} + \delta_t^{(i)}\right) = \frac{1}{h_i}\left[\gamma(0) + 2\sum_{j=1}^{h_i-1}\left(1 - \frac{j}{h_i}\right)\gamma(j)\right] + c =: \tilde{\sigma}_i^2,$$

where $c = \mathrm{Var}(\delta_t^{(i)}) \le 1/4$. It is denoted by $\tilde{\sigma}_i^2$. Thus, $\mathrm{Var}[X_t] = \mu\sum_{i=1}^{p}\alpha_1 + \lambda + \sum_{i=1}^{p}\alpha_i^2\tilde{\sigma}_i^2 < \infty$.

Finally, we observe $\mathrm{Cov}(X_t, X_{t-k})$ for $k = 1, 2, \ldots$, which is equal to

$$E\left[\mathrm{Cov}\left(X_t, X_{t-k} \mid X_{t-k}\right)\right] + \mathrm{Cov}\left(E\left[X_t \mid X_{t-k}\right],\ E\left[X_{t-k} \mid X_{t-k}\right]\right).$$

In its first term, we have $\mathrm{Cov}(X_t, X_{t-k}|X_{t-k}) = E[(X_t - E[X_t|X_{t-k}])(X_{t-k} - E[X_{t-k}|X_{t-k}])|X_{t-k}] = 0$. Thus, $\mathrm{Cov}(X_t, X_{t-k}) = \mathrm{Cov}(E[X_t|X_{t-k}], E[X_{t-k}|X_{t-k}])$. If $k = 1$,

$$\begin{aligned}\mathrm{Cov}\left(X_t, X_{t-1}\right) &= \mathrm{Cov}\left(E\left[X_t \mid X_{t-1}\right], E\left[X_{t-1} \mid X_{t-1}\right]\right) = \mathrm{Cov}\left(\alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda, X_{t-1}\right) \\ &= \mathrm{Cov}\left(\sum_{i=1}^{p}\alpha_i\left[\frac{1}{h_i}\sum_{j=1}^{h_i} X_{t-j} + \delta_t^{(i)}\right] + \lambda, X_{t-1}\right) = \mathrm{Cov}\left(\sum_{j=1}^{h_p}\beta_j X_{t-j} + \delta_t + \lambda, X_{t-1}\right).\end{aligned}$$

Hence, $\mathrm{Cov}(X_t, X_{t-1}) = \sum_{j=1}^{h_p}\beta_j \mathrm{Cov}(X_{t-j}, X_{t-1})$. This equality is the same as the autocovariance function of lag one in the AR model of order h_p. If $k \ge 2$, then $E[X_t|X_{t-k}] = E[E[X_t|X_{t-1}, X_{t-k}]]$

$$= E\left[\alpha_1 X_{t-1}^{(1)} + \cdots + \alpha_p X_{t-1}^{(p)} + \lambda \mid X_{t-k}\right] = E\left[\sum_{j=1}^{h_p}\beta_j X_{t-j} + \delta_t + \lambda \mid X_{t-k}\right] = \sum_{j=1}^{h_p}\beta_j E\left[X_{t-j} \mid X_{t-k}\right] + \lambda,$$

and thus

$$\mathrm{Cov}\left(X_t, X_{t-k}\right) = \mathrm{Cov}\left(\sum_{j=1}^{h_p}\beta_j E\left[X_{t-j} \mid X_{t-k}\right] + \lambda,\ X_{t-k}\right) = \sum_{j=1}^{h_p}\beta_j \mathrm{Cov}\left(E\left[X_{t-j} \mid X_{t-k}\right],\ X_{t-k}\right).$$

We will show that $\mathrm{Cov}(E[X_{t-j}|X_{t-k}], X_{t-k}) = \mathrm{Cov}(X_{t-j}, X_{t-k})$. Just for notational simplicity, let $\mu_{t,j,k} = E[X_{t-j}|X_{t-k}]$. Since $E[X_{t-j} - \mu_{t,j,k}] = 0$, we have

$$\mathrm{Cov}\left(X_{t-j} - E\left[X_{t-j} \mid X_{t-k}\right],\ X_{t-k}\right) = \mathrm{Cov}\left(X_{t-j} - \mu_{t,j,k},\ X_{t-k}\right) = E\left[\left(X_{t-j} - \mu_{t,j,k}\right)\left(X_{t-k} - \mu\right)\right],$$

and it is equal to $E[E[(X_{t-j} - \mu_{t,j,k})(X_{t-k} - \mu)]|X_{t-k}] = E[(X_{t-k} - \mu)E[(X_{t-j} - \mu_{t,j,k})|X_{t-k}]] = E[(X_{t-k} - \mu)(E[X_{t-j}|X_{t-k}] - \mu_{t,j,k})] = 0$. Thus we have

$$\mathrm{Cov}\left(X_t, X_{t-k}\right) = \sum_{j=1}^{h_p}\beta_j \mathrm{Cov}\left(X_{t-j}, X_{t-k}\right),$$

which is the autocovariance function relation of the conventional stationary AR model of order h_p since $\sum_{j=1}^{h_p}\beta_j = \sum_{i=1}^{p}\alpha_i < 1$. Therefore, $\{X_t\}$ has covariance-stationary property along with constant mean and finite variance. □

Proof of Theorem 1: We follow the spirit of Klimko and Nelson (1978), who discussed the strong consistency and asymptotic joint normality of the CLS estimation for stochastic process including stationary ergodic process and Markov processes.

Recall $Q_n(\theta)$ in Equation (3.1) given by

$$Q_n(\theta) = \sum_{t=1}^{n}\left[X_t - \alpha_1 X_{t-1}^{(1)} - \cdots - \alpha_p X_{t-1}^{(p)} - \lambda\right]^2.$$

The CLS estimator $\hat{\theta}_n$ minimizes $Q_n(\theta)$ and thus satisfies

$$\frac{\partial Q_n(\theta)}{\partial \alpha_i} = 0,\ \ (i = 1, 2, \ldots, p) \quad \text{and} \quad \frac{\partial Q_n(\theta)}{\partial \lambda} = 0, \tag{6.1}$$

which implies the CLS estimator $\hat{\theta}_n \equiv \hat{\theta}_{\mathrm{CLS}} = \mathbb{X}^{-1}\mathbb{Y}$ in (3.2). The expression in (6.1) along with the CLS is written, in a vector form, as $\partial Q_n(\hat{\theta}_n)/\partial\theta = 0^{(p+1)\times 1}$.

Now to establish its asymptotic normality, consider the Taylor expansion of $Q_n(\theta)$ about θ^o:

$$Q_n(\theta) = Q_n(\theta^o) + (\theta - \theta^o)^\top \partial Q_n(\theta^o)/\partial\theta + \frac{1}{2}(\theta - \theta^o)^\top \partial^2 Q_n(\theta^*)/\partial\theta^2 (\theta - \theta^o),$$

where θ^* is an intermediate point between θ^0 and θ, in the neighborhood of θ^0 with $0 < \|\theta^0 - \theta^*\| < \|\theta^0 - \theta\| < \delta$ for some $\delta > 0$. Note that $\partial Q_n(\theta^o)/\partial\theta$ is a $(p + 1) \times 1$ column vector and $\partial^2 Q_n(\theta^*)/\partial\theta^2$ is a $(p + 1) \times (p + 1)$ matrix. Let V_n and T_n be the $(p + 1) \times (p + 1)$ matrices, respectively, such that

$$V_n = \partial^2 Q_n(\theta^o)/\partial\theta^2 \quad \text{and} \quad T_n(\theta^*) = \partial^2 Q_n(\theta^*)/\partial\theta^2 - V_n.$$

Then we have

$$0^{(p+1)\times 1} = \frac{1}{\sqrt{n}}\partial Q_n\left(\hat{\theta}_n\right)/\partial\theta = \frac{1}{\sqrt{n}}\partial Q_n(\theta^o)/\partial\theta + \frac{1}{2n}\left(V_n + T_n(\theta^*)\right)\sqrt{n}\left(\hat{\theta}_n - \theta^o\right). \tag{6.2}$$

First we may show that

$$\lim_{n\to\infty}\frac{1}{2n}\left(V_n + T_n\left(\theta^*\right)\right) = \lim_{n\to\infty}\frac{1}{n}\mathbb{X} =: V \ \ a.s., \tag{6.3}$$

and second

$$\frac{1}{2\sqrt{n}}\partial Q_n\left(\theta^o\right)/\partial\theta \ \stackrel{\mathrm{d}}{\longrightarrow}\ \mathcal{N}\left(0, W\right). \tag{6.4}$$

Equation (6.3) is obtained straightforwardly by the second partial derivatives of $Q_n(\theta)$. In order to show (6.4), write $(\partial Q_n(\theta^o)/2)/\partial\theta = -(A_1, \ldots, A_p, A_{p+1})^\top$, where $A_i = \sum_{t=1}^n \varepsilon_t X_{t-1}^{(i)}$, $(i = 1, \ldots, p)$, and $A_{p+1} = \sum_{t=1}^n \varepsilon_t$, with ε_t given in **Theorem 1**. Straightforwardly using the conditional expectation and conditional variance along with the fact $E[\varepsilon_t|\mathcal{F}_{t-1}] = 0$ a.s., we have $E[A_i/\sqrt{n}] = \sqrt{n}E[X_{t-1}^{(i)}E[\varepsilon_t|\mathcal{F}_{t-1}]] = 0$ for each i. Also, the $(p+1)\times(p+1)$ variance-covariance matrix is given by $E[(A_1, \ldots, A_p, A_{p+1})^\top(A_1, \ldots, A_p, A_{p+1})]/n = W = ((w_{ij}))$ as in **Theorem 1**. Indeed, $\mathrm{Cov}(A_i, A_j)/n = E[E[A_iA_j|F_{t-1}]]/n$ and $E[A_iA_j|F_{t-1}]/n =$

$$\frac{1}{n}E\left[\sum_{t_1=1}^{n}\varepsilon_{t_1}X_{t_1-1}^{(i)}\sum_{t_2=1}^{n}\varepsilon_{t_2}X_{t_2-1}^{(j)} \mid F_{t-1}\right] = \frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}X_{t_1-1}^{(i)}X_{t_2-1}^{(j)}E\left[\varepsilon_{t_1}\varepsilon_{t_2} \mid F_{t-1}\right]. \tag{6.5}$$

If $t_1 \neq t_2$, then $E[\varepsilon_{t_1}\varepsilon_{t_2}|F_{t-1}] =$

$$\begin{aligned}
&E\left[\left(X_{t_1} - \alpha_1X_{t_1-1}^{(1)} - \cdots - \alpha_pX_{t_1-1}^{(p)} - \lambda\right)\left(X_{t_2} - \alpha_1X_{t_2-1}^{(1)} - \cdots - \alpha_pX_{t_2-1}^{(p)} - \lambda\right) \mid F_{t-1}\right]\\
&= E\left[\left(\sum_{j=1}^{X_{t_1-1}^{(1)}}\left(N_j^{(1)} - \alpha_1\right) + \cdots + \sum_{j=1}^{X_{t_1-1}^{(p)}}\left(N_j^{(p)} - \alpha_p\right) + Z_{t_1} - \lambda\right)\right.\\
&\quad\left.\times\left(\sum_{j=1}^{X_{t_2-1}^{(1)}}(N_j^{(1)} - \alpha_1) + \cdots + \sum_{j=1}^{X_{t_2-1}^{(p)}}\left(N_j^{(p)} - \alpha_p\right) + Z_{t_2} - \lambda\right)\middle| F_{t-1}\right].
\end{aligned}$$

Noting that $E[Z_t - \lambda] = 0$, the last is of the form of a linear combination of $E[(N_{j_1}^{(i_1)} - \alpha_{i_1})(N_{j_2}^{(i_2)} - \alpha_i)|F_{t-1}]$, which is zero by the independence of $\{N_j^{(i)}\}$ with mean α_i, for $i = i_1, i_2$; $j = j_1, j_2$. Therefore, (6.5) is equal to $\sum_{t=1}^n X_{t-1}^{(i)}X_{t-1}^{(j)}E[\varepsilon_t^2|F_{t-1}]/n = X_{t-1}^{(i)}X_{t-1}^{(j)}\mathrm{Var}(\varepsilon_t|F_{t-1})$ by the stationarity. Thus $\mathrm{Cov}(A_i, A_j)/n = E[X_{t-1}^{(i)}X_{t-1}^{(j)}\mathrm{Var}(\varepsilon_t|F_{t-1})] = w_{ij}$. By a central limit theorem of a stationary process, the joint asymptotic normality in (6.4) is obtained. Hence, by (6.2), (6.3) and (6.4), we have $\sqrt{n}(\hat{\theta}_N - \theta^o) = \mathcal{N}(0, V^{-1}WV^{-1})$. □

Proof of Theorem 2: We recall the discussion about $\{Y_t\}$ on (3.3) and (3.4) along with the argument in the proof of Proposition 2.1. The Yule-Walker equation of (3.4) can be derived as follows:

$$\begin{aligned}
\gamma(j) &= \beta_1\gamma\left(j-1\right) + \cdots + \beta_{h_p}\gamma(j-h_p), \quad j = 1, 2, \ldots, h_p,\\
\mathrm{Var}\left(\delta_t\right) \equiv \sigma^2 &= \gamma\left(0\right) - \beta_1\gamma\left(1\right) - \cdots - \beta_{h_p}\gamma\left(h_p\right),
\end{aligned}$$

where $\gamma(j) = \text{Cov}(Y_t, Y_{t+j})$. From the Yule-Walker equation, we obtain the Yule-Walker estimator $\hat{\beta}_{j,YW}$ of β_j:

$$\begin{bmatrix} \hat{\beta}_{1,YW} \\ \hat{\beta}_{2,YW} \\ \vdots \\ \hat{\beta}_{h_p,YW} \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}(1) & \dots & \hat{\rho}(h_p-1) \\ \hat{\rho}(1) & 1 & \dots & \hat{\rho}(h_p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}(h_p-1) & \hat{\rho}(h_p-2) & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(h_p) \end{bmatrix},$$

where $\hat{\rho}(j)$ is sample autocorrelation function of lag j of $\{Y_t\}$. However, since $Y_t = E[X_t|F_{t-1}] - \mu$, we may compute $\hat{\rho}(j)$ from the observation $\{X_1, \dots, X_n\}$ as follows:

$$\hat{\rho}(j) = \frac{\sum_{t=j+1}^{n} \left(X_t - \bar{X}\right)\left(X_{t-j} - \bar{X}\right) / (n-j)}{\sum_{t=1}^{n} \left(X_t - \bar{X}\right)^2 / n}.$$

$j = 1, 2, \dots, h_p$, where $\bar{X} = \sum_{t=1}^{n} X_t / n$. And then straightforwardly the desired YW estimates $\hat{\alpha}_{i,YW}$ and $\hat{\lambda}_{YW}$ are obtained as in (3.6) and (3.7).

As for the asymptotic normality of the YW estimator $\hat{\beta}_{YW}$ of the AR model, it is well known that $\sqrt{n}(\hat{\beta}_{YW} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2\Gamma^{-1})$. Notice that $\alpha = H\beta$ and $\hat{\alpha}_{YW} = H\hat{\beta}_{YW}$ (that is, $\alpha, \hat{\alpha}_{YW}$ are linear transformation of $\beta, \hat{\beta}_{YW}$, respectively), by (3.6), where H is $p \times h_p$ matrix given by

$$\begin{bmatrix} \frac{h_1}{h_1-h_0} \cdots \frac{h_1}{h_1-h_0} & \frac{-h_1}{h_2-h_1} \cdots \frac{-h_1}{h_2-h_1} & \frac{-h_1}{h_3-h_2} \cdots \frac{-h_1}{h_3-h_2} & \cdots & \cdots & \frac{-h_1}{h_p-h_{p-1}} \cdots \frac{-h_1}{h_p-h_{p-1}} \\ 0 \cdots 0 & \frac{h_2}{h_2-h_1} \cdots \frac{h_2}{h_2-h_1} & \frac{-h_2}{h_3-h_2} \cdots \frac{-h_2}{h_3-h_2} & \cdots & \cdots & \frac{-h_2}{h_p-h_{p-1}} \cdots \frac{-h_2}{h_p-h_{p-1}} \\ \vdots & 0 \cdots 0 & \frac{h_3}{h_3-h_2} \cdots \frac{h_3}{h_3-h_2} & & & \vdots \\ \vdots & \vdots & 0 \cdots 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \frac{h_{p-1}}{h_{p-1}-h_{p-2}} \cdots \frac{h_{p-1}}{h_{p-1}-h_{p-2}} & \frac{-h_{p-1}}{h_p-h_{p-1}} \cdots \frac{-h_{p-1}}{h_p-h_{p-1}} \\ 0 \cdots 0 & 0 \cdots 0 & 0 \cdots 0 & \cdots & 0 \cdots 0 & \frac{h_p}{h_p-h_{p-1}} \cdots \frac{h_p}{h_p-h_{p-1}} \end{bmatrix}.$$

Since $nE[(\hat{\alpha}_{YW} - \alpha)(\hat{\alpha}_{YW} - \alpha)^\top] = nE[H(\hat{\beta}_{YW} - \beta)(\hat{\beta}_{YW} - \beta)^\top H^\top] = H(nE[(\hat{\beta}_{YW} - \beta)(\hat{\beta}_{YW} - \beta)^\top])H^\top \to H\sigma^2\Gamma^{-1}H^\top$ as $n \to \infty$, the desired normality of $\hat{\alpha}_{YW}$ with the asymptotic variance is completed.

Finally, we derive the asymptotic normality of $\sqrt{n}(\hat{\mu} - \mu)$. Noting that

$$\sum_{t=1}^{n} X_t = \sum_{t=1}^{n} \left(\sum_{j=1}^{X_{t-1}^{(1)}} N_j^{(1)} + \dots + \sum_{j=1}^{X_{t-1}^{(p)}} N_j^{(p)} + Z_t \right),$$

which is the sum of independent Poisson random variables, by the central limit theorem for the Poisson distributions, it suffices to find the asymptotic variance of $\sqrt{n}(\hat{\mu} - \mu) = \sum_{t=1}^{n} (X_t - \mu)/\sqrt{n}$.

Asymptotic normality of the sample mean in INAR(p) model was given in Jentsch and Weiß (2017), where the central limit theorem (CLT) from Ibragimov (1962) was applied. Our case has the same result for the asymptotic variance. To be precise, we have

$$\text{Var}\left(\sqrt{n}(\hat{\mu} - \mu)\right) = \frac{1}{n}\text{Cov}\left(\sum_{t=1}^{n} X_t, \sum_{s=1}^{n} X_s\right) = \frac{1}{n}\left[n\gamma(0) + \sum_{j=1}^{n-1} (n-j)(\gamma(j) + \gamma(-j))\right] =: V_n.$$

We will show that $|\sigma_\mu^2 - V_n| \to 0$ as $n \to \infty$. Using $\gamma(-j) = \gamma(j)$ we have

$$
\begin{aligned}
\left|\sigma_\mu^2 - V_n\right| &= \left|\sum_{j=-\infty}^{\infty} \gamma(j) - \gamma(0) - \sum_{j=1}^{n-1} \gamma(j) + \frac{1}{n}\sum_{j=1}^{n-1} j\gamma(j) - \sum_{j=-1}^{-n+1} \gamma(j) + \frac{1}{n}\sum_{j=-1}^{-n+1} |j|\,\gamma(j)\right| \\
&\le \left|\sum_{j=1}^{\infty} \gamma(j) - \sum_{j=1}^{n-1} \gamma(j)\right| + \left|\sum_{j=-1}^{-\infty} \gamma(j) - \sum_{j=-1}^{-n+1} \gamma(j)\right| + \left|\frac{1}{n}\sum_{j=1}^{n-1} j\gamma(j)\right| + \left|\frac{1}{n}\sum_{j=-1}^{-n+1} |j|\gamma(j)\right| \\
&\le 2\left|\sum_{j=n}^{\infty} \gamma(j)\right| + \frac{2}{n}\sum_{j=1}^{n-1} j\,|\gamma(j)|,
\end{aligned}
$$

where the first term tends to zero as $n \to \infty$. Note that the autocovariance function $\gamma(j)$ has the same structure of the conventional HAR model, which is an AR model, and thus $|\gamma(j)|$ is exponentially decreasing. Hence the second term $2(\sum_{j=1}^{n-1} j|\gamma(j)|)/n \to 0$ as $n \to \infty$. Therefore we complete the proof. □

References

Al-Osh MA and Alzaid AA (1987). First-Order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, **8**, 261–275.

Al-Osh MA and Alzaid AA (1988). Integer-Valued moving average (INMA) process, *Statistical Papers*, **29**, 281–300.

Al-Osh MA and Alzaid AA (1990). An integer-valued p^{th}-order autoregressive structure INAR(p) process, *Journal of Applied Probability*, **27**, 314–324.

Appiah S and Maposa D (2021). An integer-valued modelling and forecast of COVID-19 cases in Ghana: A review and comparative study, *Transylvanian Review*, **29**, 15712–15722.

Boeh K and Dunbar C (2014). IPO waves and the issuance process, *Journal of Corporate Finance*, **25**, 455–473.

Cardinal M, Roy R, and Lambert J (1999). On the application of integer-valued time series models for the analysis of disease incidence, *Statistics in Medicine*, **18**, 2025–2039.

Catania L and Di Mari R (2021). Hierarchical Markov-switching models for multivariate integer-valued time-series, *Journal of Econometrics*, **221**, 118–137.

Chang L, Liu X, Wang D, Jing Y, and Li C (2022). First-Order random coefficient mixed-thinning integer-valued autoregressive model, *Journal of Computational and Applied Mathematics*, **410**, 114222.

Jentsch C and Weiß CH (2017). Bootstrapping INAR Models, University of Mannheim, Department of Economics.

Du JG and Li Y (1991). The integer-valued autoregressive (INAR(p)) model, *Journal of Time Series Analysis*, **12**, 129–142.

Ferland R, Latour A, and Oraichi D (2006). Integer-Valued GARCH process, *Journal of Time Series Analysis*, **27**, 923–942.

Freeland R and McCabe B (2004). Analysis of low count time series data by Poisson autoregression, *Journal of Time Series Analysis*, **25**, 701–722.

Freeland R and McCabe B (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model, *Statistics and Probability Letters*, **73**, 147–153.

Güçbilmez U (2015). IPO waves in China and Hong Kong, *International Review of Financial Analy-*

sis, **40**, 14–26.
Hwang E and Shin D (2014). Infinite-Order, long-memory heterogeneous autoregressive models, *Computational Statistics and Data Analysis*, **76**, 339–358.
Ibragimov IA (1962). Some limit theorems for stationary processes, *Theory of Probability and Its Applications*, **7**, 349–382.
Ivanov V and Lewis CM (2008). The determinants of market-wide issue cycles for initial public offerings, *Journal of Corporate Finance*, **14**, 567–583.
Klimko LA and Nelson PI (1978). On conditional least squares estimation for stochastic processes, *The Annals of Statistics*, **6**, 629–642.
Kowal D (2019). Integer-Valued functional data analysis for measles forecasting, *Biometrics*, **75**, 1321–1333.
Lu F and Wang D (2022). A new estimation for INAR(1) process with Poisson distribution, *Computational Statistics*, **37**, 1185–1201.
Mahmoudi E and Rostami A (2020). First-Order integer-valued moving average process with power series innovations, *Journal of Statistical Theory and Applications*, **19**, 415–431.
McCabe BPM and Martin GM (2005). Bayesian prediction of low count time series, *International Journal of Forecasting*, **21**, 315–330.
McKenzie E (1985). Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 635–644.
McKenzie E (1988). Some ARMA models for dependent sequences of Poisson counts, *Advances in Applied Probability*, **20**, 822–835.
Orozco DLR, Sales LOF, Fernández LMZ, and Pinho ALS (2021). A new mixed first-order integer-valued autoregressive process with Poisson innovations, *AStA Advances in Statistical Analysis*, **105**, 559–580.
Park YS and Oh CH (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals, *Statistical Papers*, **38**, 287–302.
Steutel FW and van Harn K (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability*, **7**, 893–899.
Wang X and Ning C (2022). A new Markov regime-switching count time series approach for forecasting initial public offering volumes and detecting issue cycles, *Journal of Forecasting*, **41**, 118–133.
Xiaochun L (2019). On tail fatness of macroeconomic dynamics, *Journal of Macroeconomics*, **62**, 103154.
Yang K, Kang Y, Wang D, Li H, and Diao Y (2019). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued autoregressive processes, *Metrika*, **82**, 863–889.

Received August 25, 2022; Revised March 10, 2023; Accepted March 29, 2023