Modelling the COVID-19 Data using an Integrated Integer-Valued Time Series Model

Yuvraj Sunecher [1], Naushad Manode Khan [2]
[1] University of Technology Mauritius, Port Louis, Mauritius
[2] University of Mauritius, Reduit, Mauritius
Email: [1] ysunecher@utm.ac.mu, [2] n.mamodekhan@uom.ac.mu

Abstract
A novel stable bivariate integer-valued autoregressive moving average (Bivariate IN-ARMA(1,1)) model that takes into account the joint Poisson distribution of the two innovations, is proposed in this study. The conditional maximum likelihood (CML) estimation approach, which is based on the principles of thinning and convolution, is used to estimate the model parameters. Based on certain combinations of the model parameters, a simulation study is carried out to evaluate the effectiveness of the CML technique. An application on COVID-19 data series is also provided.
Index Terms—Integer-valued, autoregressive, moving average, conditional maximum likelihood, simulation, COVID-19.

1. Introduction

Since the introduction of the family of integer-valued time series processes based on the binomial thinning operation [1] by Mckenzie [2, 3], lots of research has been conducted on the different features of such time series models. Two of the main contributions have been to alter the innovation distributions from the classical Poisson, Negative-Binomial (Poisson-Gamma mixture) [2, 3], Geometrics to Com- Poisson [3, 4], mixed Geometric (NGINAR(1)), Poisson-Lindley [5] and to shift the binomial thinning operator to other generalized thinning procedures (See the paper by Weib et al. [6] for a comprehensive review). Both changes are in view of improving the quality of fits under different scenarios of over, equi- and under-dispersion and lots of development are under process with new mixed Poisson models.

In these above references, the authors have mainly worked with the subclass of IN-ARMA process: the auto-regressive process of order 1 (IN-AR(1)) as referred above and the moving-average process of order 1 (IN-MA(1)) [7]. Thus, the general class of IN-ARMA model of order 1 (IN-ARMA(1,1)), has not been much explored until recently Weib et al. [6] proposed an IN-ARMA(1,1) model with Poisson (IN- ARMA(1,1)P) and Negative Binomial (NB) (IN-ARMA(1,1)NB) innovations un- der strict stationary conditions. This IN-ARMA(1,1) has shown important merits as compared to some popular existing IN-AR(1) models. Firstly, the moments of the simple IN-ARMA(1,1) with Poisson innovations (IN-ARMA(1,1)P) show that the model is appropriate for modelling over-dispersed time series of counts. Furthermore, in the data generating process in Weib et al. [6] where IN-AR(1)NB time series of counts was simulated, the IN-ARMA(1,1) with Poisson yields almost same mean and over-dispersion parameter estimates with practically similar standard errors as compared with the true model. While, also, the proposed IN-ARMA(1,1)P and IN-ARMA(1,1) NB models provides equally good fits and forecasting values as the IN-AR(1) counterparts. Another important finding in the paper by Weib et al. [6] was that there needs to explore other model selection criteria other than the most used Akaike information criterion (AIC) since in spite of the real-life data examples in the paper exhibit an ARMA structure, the AIC of the IN-AR(1)NB was slightly better than the IN-ARMA(1,1)P and IN-ARMA(1,1)NB respectively.

At this stage of the literature, the IN-ARMA(1,1) model has only been explored for univariate time series that exhibit over-dispersion. Given some of its interesting features, especially in modelling over-dispersion, it is important to assess the be- havior of this process in bivariate time series set-up. It is well-known that bivariate IN-AR(1) processes [8–10] have proved useful and yielded satisfactory results in applications related to day and night accidents, crime activities and stock trading and hence, the development of an IN-ARMA(1,1) model becomes an important competitor and provides a motivation to assess the quality of fits and forecasts in comparison with the popular IN-AR(1) models.

Thus, to initiate this novel process, this paper proposes a bivariate IN-ARMA(1,1) model with Poisson innovations (BINARMA(1,1)P) in correspondence with the models in Weib et al. [6] but where the cross correlation between the two IN-ARMA(1,1) series is induced by the correlated pair of Poisson innovations. In the case of the Bivariate IN-ARMA(1,1)P, the bivariate Poisson model proposed by Kocherlokota and Kocherlokota [11] is referred.

2. The Bivariate In-Arma (1,1) Model

Consider

$$Y_t^{[1]} = \rho_1 * Y_{t-1}^{[1]} + \rho_2 * R_{t-1}^{[1]} + R_t^{[1]} \quad (1)$$

$$Y_t^{[2]} = \rho_3 * Y_{t-1}^{[2]} + \rho_4 * R_{t-1}^{[2]} + R_t^{[2]} \quad (2)$$

where $Y_t^{[k]}$ is the counting random variable for the k^{th} series at the t^{th} time point with $R^{[k]}$ is the corresponding error term. $*$ represents the binomial thinning operator

$$\rho * Y = \begin{cases} \sum_{s=0}^{Y} b_s(\rho), & Y > 0, \\ 0 & Y \leq 0 \end{cases}$$

and hence $\rho * Y \,|Y \sim$ Binomial(Y, ρ) and E($\rho * Y\,|Y$) = $\rho E(Y)$ and Var($\rho * Y$)=$\rho(1-\rho)Y$, with $\rho \in [0, 1)$.

Let assume that the pair $\{R_t^{[1]}, R_t^{[2]}\}$ are correlated and follow the bivariate Poisson [11] and

$$Corr(R_t^{[1]}, R_{t+h}^{[2]}) = \begin{cases} \rho_{12} & h = 0 \\ 0 & h \neq 0 \end{cases} \quad (3)$$

Thus, $\left\{R_t^{[1]}\right\}$ is an i.i.d sequence of Poisson counts with mean $\lambda_t^{[k]}$. We further assume that Cov($Y_t^{[k]}, R_{t+h}^{[k]}$) = 0, h≠0 and for h=0, it is simply shown that Cov($Y_t^{[k]}, R_t^{[k]}$) = Var($R_t^{[k]}$).

Under strict stationary conditions,

$$E(Y_t^{[1]}) = \rho_1 E(Y_t^{[1]}) + \rho_1 E(R_t^{[1]}) + E(R_t^{[1]})$$

$$(1-\rho_1)E(Y_t^{[1]}) = (1 + \rho_2)E(R_t^{[1]})$$

$$E(Y_t^{[1]}) = \frac{(1 + \rho_2)}{(1 - \rho_1)} E(R_t^{[1]})$$

$$E\left(Y_t^{[1]}\right) = \frac{(1+\rho_2)}{(1-\rho_1)}\lambda^{[1]} = \mu^{[1]}$$

$$E\left(Y_t^{[1]}\right) = \mu^{[1]} \tag{4}$$

$$E(Y_t^{[2]}) = \frac{(1+\rho_4)}{(1-\rho_3)}E(R_t^{[2]})$$

$$E\left(Y_t^{[2]}\right) = \frac{(1+\rho_4)}{(1-\rho_3)}\lambda^{[2]} = \mu^{[2]}$$

$$E\left(Y_t^{[2]}\right) = \mu^{[2]} \tag{5}$$

$Var\left(Y_t^{[1]}\right) = Var\left(\rho_1 * Y_{t-1}^{[1]}\right) + Var\left(\rho_2 * R_{t-1}^{[1]}\right) + Var(R_t^{[1]})$ + 2Cov($\rho_1 * Y_{t-1}^{[1]}, \rho_2 * R_{t-1}^{[1]}$)

$Var\left(Y_t^{[1]}\right) = \rho_1(1-\rho_1)\mu^{[1]} + \rho_1^2\, Var(Y_t^{[1]}) + \rho_2(1-\rho_2)\lambda^{[1]} + \rho_2^2\, \lambda^{[1]} + 2\rho_1\rho_2\lambda^{[1]} + \lambda^{[1]}$

$(1-\rho_1^2)Var\left(Y_t^{[1]}\right) = \rho_1(1-\rho_1)\mu^{[1]} + \rho_2(1-\rho_2)\lambda^{[1]} + \rho_2^2\,\lambda^{[1]} + 2\rho_1\rho_2\lambda^{[1]} + \lambda^{[1]}$

$$(1-\rho_1^2)Var\left(Y_t^{[1]}\right) = \rho_1(1-\rho_1)\mu^{[1]} + [1+2\rho_1\rho_2+\rho_2]\lambda^{[1]}$$

$$Var\left(Y_t^{[1]}\right) = \frac{\rho_1(1-\rho_1)\mu^{[1]}+[1+2\rho_1\rho_2+\rho_2]\lambda^{[1]}}{(1-\rho_1^2)} \tag{6}$$

$$Var\left(Y_t^{[2]}\right) = \frac{\rho_3(1-\rho_3)\mu^{[2]}+[1+2\rho_3\rho_4+\rho_4]\lambda^{[2]}}{(1-\rho_3^2)} \tag{7}$$

$$\mathrm{Cov}(Y_t^{[1]}, Y_{t+h}^{[1]}) = \rho_1^h Var(Y_t^{[1]}) + \rho_1^{h-1}\rho_2\lambda^{[1]} \tag{8}$$

$$\mathrm{Cov}(Y_t^{[2]}, Y_{t+h}^{[2]}) = \rho_3^h Var(Y_t^{[2]}) + \rho_3^{h-1}\rho_4\lambda^{[2]} \tag{9}$$

$$\mathrm{Cov}(Y_t^{[1]}, Y_t^{[2]}) = \frac{(1+\rho_1\rho_4+\rho_2\rho_3+\rho_2\rho_4)\rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}}}{(1-\rho_1\rho_3)} \tag{10}$$

$$\mathrm{Cov}(Y_t^{[1]}, Y_{t+h}^{[2]}) = \rho_3^h\, \mathrm{Cov}(Y_t^{[1]}, Y_t^{[2]}) + \rho_3^{h-1}\rho_4\rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}} \tag{11}$$

$$\mathrm{Cov}(Y_{t+h}^{[1]}, Y_t^{[2]}) = \rho_1^h\, \mathrm{Cov}(Y_t^{[1]}, Y_t^{[2]}) + \rho_1^{h-1}\rho_2\rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}} \tag{12}$$

3. Estimation of Parameters

$$f_1(k) = \sum_{j_1=0}^{k}\binom{y_{t-1}^{[1]}}{j_1}\binom{r_{t-1}^{[1]} = y_{t-1}^{[1]} - k}{k-j_1}\rho_1^{j_1}(1-\rho_1)^{y_{t-1}^{[1]}-j_1}\rho_2^{k-j_1}(1-\rho_2)^{y_{t-1}^{[1]}-2k+j_1} \tag{13}$$

$$f_2(s) = \sum_{j_2=0}^{s}\binom{y_{t-1}^{[2]}}{j_2}\binom{r_{t-1}^{[2]} = y_{t-1}^{[2]} - s}{s-j_2}\rho_3^{j_2}(1-\rho_3)^{y_{t-1}^{[2]}-j_2}\rho_4^{s-j_2}(1-\rho_4)^{y_{t-1}^{[2]}-2s+j_2} \tag{14}$$

and considering the innovation terms' bivariate distribution

$f_3\left(r_t^{[1]} = y_{t-1}^{[1]} - k, r_{t-1}^{[2]} = y_{t-1}^{[2]} - s\right)$ = P($R_t^{[1]} = r_t^{[1]}, R_t^{[2]} = r_t^{[2]}$), where

$$f_3\left(r_t^{[1]} = y_{t-1}^{[1]} - k, r_{t-1}^{[2]} = y_{t-1}^{[2]} - s\right) = e^{-(\lambda^{[1]}+\lambda^{[2]}-\rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}})}\sum_{m=0}^{\min(k,s)}\{\left[\lambda^{[1]} - \rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}}\right]^{y_{t-1}^{[1]}-k-m}[\lambda^{[2]} - \rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}}]^{y_{t-1}^{[2]}-s-m}[\rho_{12}\sqrt{\lambda^{[1]}}\sqrt{\lambda^{[2]}}]^m\}/[\left(y_{t-1}^{[1]} - k - m\right)!\left(y_{t-1}^{[2]} - s - m\right)!] \tag{15}$$

The conditional density is

$f((y_t^{[1]}, y_t^{[2]})|(y_{t-1}^{[1]}, y_{t-1}^{[2]}, r_{t-1}^{[1]}, r_{t-1}^{[2]}), \theta))=$
$\sum_{k=0}^{g_1}\sum_{s=0}^{g_2} f_1(k)\, f_2(s) f_3\left(r_t^{[1]} = y_{t-1}^{[1]} - k, r_{t-1}^{[2]} = y_{t-1}^{[2]} - s\right),$
where g_1=min($y_t^{[1]}, y_{t-1}^{[1]}$), g_2=min($y_t^{[2]}, y_{t-1}^{[2]}$) and
$\theta = (\rho_1, \rho_2, \rho_3, \rho_4, \lambda^{[1]}, \lambda^{[2]})$.

$$\mathbf{L}(\theta|y) = \prod_{t=1}^{T} f((y_t^{[1]}, y_t^{[2]})|\left(y_{t-1}^{[1]}, y_{t-1}^{[2]}, r_{t-1}^{[1]}, r_{t-1}^{[2]}\right), \theta) \quad (16)$$

$$\mathbf{Log}\ [\mathbf{L}(\theta|y)] = \log[\sum_{t=1}^{T} f((y_t^{[1]}, y_t^{[2]})|\left(y_{t-1}^{[1]}, y_{t-1}^{[2]}, r_{t-1}^{[1]}, r_{t-1}^{[2]}\right), \theta)] \quad (17)$$

4. Simulation

This section generates Bivariate IN-ARMA(1,1) time series data using Equations (1)-(2) for T = 100, 200, 600, with $\lambda^{[k]} = \exp(\beta^{[k]}x')$, where

$$\boldsymbol{x_{t1}} = \begin{cases} -\cos(5\pi t) + 0.05 & (t = 1, \dots, \frac{T}{6}) \\ \sin(2\pi t) + 0.05 & (t = \left(\frac{T}{6}\right) + 1, \dots, \frac{3T}{6}) \\ \cos(2\pi t) + 0.10 & t = \left(\frac{3T}{6}\right) + 1, \dots, T) \end{cases}$$

$$\boldsymbol{x_{t2}} = \begin{cases} t + 0.02 & (t = 1, \dots, \frac{T}{3}) \\ t - 0.02 & (t = \frac{T}{3}, \dots T) \end{cases}$$

where $(\beta_1^{[1]}, \beta_1^{[1]}) = (0.7, 1.2)$, $(\beta_1^{[2]}, \beta_1^{[2]}) = (0.5, 1.7)$. The other parameters are assumed to be: $(\rho_1, \rho_2) = (0.5, 0.7)$, $(\rho_3, \rho_4) = (0.4, 0.6)$.

After 3000 simulated runs, the outcomes are presented in the tables below:

Table 1: Parameter Estimates

	$\beta_1^{[1]}$	$\beta_2^{[1]}$	$\beta_1^{[2]}$	$\beta_2^{[2]}$
100	0.6833 (0.1322)	1.1871 (0.1455)	0.5311 (0.1156)	1.6856 (0.1257)
200	0.6911 (0.1058)	1.1931 (0.1027)	0.5134 (0.1002)	1.6966 (0.0987)
600	0.6961 (0.0589)	1.1987 (0.0587)	0.5018 (0.0447)	1.6988 (0.0547)
T	ρ_1	ρ_2	ρ_3	ρ_4
100	0.4858 (0.1047)	0.6859 (0.1056)	0.4327 (0.1165)	0.5817 (0.0999)
200	0.4917 (0.0815)	0.6920 (0.0889)	0.3962 (0.0933)	0.5911 (0.0844)
600	0.4978 (0.0391)	0.6979 (0.0206)	0.3970 (0.0311)	0.5990 (0.0255)

Table 2: Parameter Estimates

T	$\beta_1^{[1]}$	$\beta_2^{[1]}$	$\beta_1^{[2]}$	$\beta_2^{[2]}$
100	0.6878 (0.1218)	1.1864 (0.1528)	0.4983 (0.1228)	1.6828 (0.1286)
200	0.6924 (0.0987)	1.1935 (0.0924)	0.5124 (0.1014)	1.6915 (0.0964)

600	0.6985 (0.0516)	1.1985 (0.0474)	0.5013 (0.0411)	1.6975 (0.0408)
T	ρ_1	ρ_2	ρ_3	ρ_4
100	0.4818 (0.1214)	0.6890 (0.1157)	0.3824 (0.1152)	0.5871 (0.1111)
200	0.4917 (0.0946)	0.6924 (0.0951)	0.3926 (0.0925)	0.5919 (0.0935)
600	0.4973 (0.0398)	0.6958 (0.0256)	0.3964 (0.0345)	0.5959 (0.0263)

We infer from Tables 1-2 that the estimates of the various model parameters are reliable, and that, as would be expected, the standard errors decrease with increasing time sizes.

5. Assessment of Covid-19 Data Sequences

The daily fresh COVID-19 infection and death series for Mauritius were taken from the official European data site (see https://data.europa.eu/data/ datasets/ covid19coronavirus-data?locale=en). The series covered the time period from 18 March 2020 to 25 April 2021 and totaled 404 observations. The following table includes some illustrative figures for Mauritius's daily new COVID-19 infection and death series:

Table 3: Descriptive data for the Mauritius COVID-19 active cases and deaths series

Descriptive Data	**Fresh COVID-19 Cases**	**Deaths**
Mean	3.00	0.04
Variance	57.9	0.07

The COVID-19 Stringency Index (SI), the vaccination event (Vaccine), the reproduction rate (ReR), and the Relative COVID-19 Risk due to Weather and Air Pollution (CRW) were taken into account as covariates. Therefore, the COVID-19 and deaths data series are subjected to the Bivariate IN-ARMA(1,1) model, with the following results:

Table 4: Estimates of Regression Coefficients

Series	SI	Vaccine	ReR	CRW
COVID-19 ($Y_t^{[1]}$)	-0.1110	-0.115	0.045	0.079
	(0.006)	(0.001)	(0.1286)	(0.0124)
Deaths ($Y_t^{[2]}$)	-0.1020	-0.091	0.051	0.049
	(0.009)	(0.002)	(0.0964)	(0.0117)

Based on the findings in Table 4, we can see that, in comparison to 'CRW' and 'ReR,' the variables 'ReR' and 'Vaccine' significantly decreased the number of infections and fatalities in Mauritius.

6. Conclusion

We present a novel stationary bivariate IN-ARMA(1,1) model with poisson innovations in this research. The formulae for joint and marginal covariance are generated. The CML technique is used to estimate the model parameters in terms of the inferential procedures. On some simulated data with trustworthy estimations, the proposed model is tested. The variables' contribution to the daily new COVID-19 infection and death series for Mauritius has also been demonstrated.

References

[1] F. Steutel, K. Van Harn, The Annals of Probability **7**, 3893 (1979)

[2] E. McKenzie, Water Resources Bulletin **21(4)**, 645 (1985)

[3] E. McKenzie, Advanced Applied Probability **18**, 679 (1986)

[4] A. Alzaid, M. Al Osh, Journal of Applied **27**, 314 (1990)

[5] M. Shirozhan, M. Mohammadpour, H. Bakouch, Brazilian Journal of Probabilty and Statistics **32(2)**, 262 (2018)

[6] C. Weib, M. Feld, N. Mamodekhan, Y. Sunecher, Stats **2**, 289 (2019)

[7] K. Brannas, A. Quoreshi, Applied Financial Economics **No.20(18)**, 1429 (2010)

[8] X. Pedeli, D. Karlis, Statistical Modelling: An International Journal **11**, 325 (2011)

[9] D. Karlis, X. Pedeli, Communications in Statistics-Theory and Methods **42**, 723 (2013)

[10] M. Ristic, A. Nastic, K. Jayakumar, H. Bakouch, Applied Mathematical Letters **25(3)**, 481 (2012)

[11] S. Kocherlakota, K. Kocherlakota, *Bivariate Discrete Distributions*, vol. 132, illustrated edn. (CRC Press, 1992).