BINAR(1) MODEL WITH GEOMETRIC MARGINALS: APPLICATION TO INTRA-DAY STOCK TRANSACTIONS IN MAURITIUS

[1]NAUSHAD MAMODE KHAN, [2]YUVRAJ SUNECHER, [3]VANDNA JOWAHEER, [4]ISVEN VEERASAWMY, [5]MUSLUN AZMI

University of Mauritius (UOM), University of Mauritius.
E-mail: [1]n.mamodekhan@uom.ac.mu, [2]ysunecher@umail.utm.ac.mu, [3]vandnaj@uom.ac.mu, [4]isvenveerasawmy@gmail.com, [5]azmi0314@yahoo.com

Abstract - This paper proposes a novel bivariate integer- valued autoregressive process of order 1 (BINAR(1)) where the counting series are Geometric while their corresponding innovations follow a mixed Geometric distribution with non-stationary moments due to the effect of time-dependent covariates. In contrast to the existing stationary BINAR(1) with marginal Geometric model, this paper assumes the cross-correlation between the series is induced by the correlated innovations only. As for the inferential part, the regression parameters are estimated by the generalized quasi-likelihood (GQL) method while the dependence parameters by the method of moments based on a robust autocorrelation specification. A simulation study is designed to assess the estimation procedures. The new model is also applied to analyse the trading intensity or volume of intra-day transactions of the two major banking institutions in Mauritius, namely Mauritius Commercial Bank Group Limited (MCB) and State Bank of Mauritius Holdings Ltd (SBMH), that are listed in the Stock Exchange of Mauritius (SEM). The root mean square errors (RMSEs) are computed and compared with the BINAR(1) with Poisson innovations.

Index Terms - BINAR(1), Geometric, GQL, Non-Stationary ,Over- Dispersion

I. INTRODUCTION

The study of BINAR(1) time series models [1] under different innovation distributions has preoccupied the interest of many researchers in the recent times. Originally BINAR(1) model with the bivariate Poisson innovations and bivariate negative binomial (NB) innovations [2] were developed assuming that the inter-relation between the two counting series were induced by the correlated innovations. In the same trend, other BINAR(1) with Geometric marginals [3] was proposed but under a more complicated structure where the cross-correlation was induced by assuming that the observation of one of the counting series at a specific time point were related with the previous-lagged observations of the other counting series with the innovation terms held independent. A similar complex cross-correlation structure was considered [4] but assuming the innovations of the counting series are additionally inter-related. Overall, these multivariate time series processes were mainly developed under the stationary moment conditions. Until, the BINAR(1) in [2] was extended to a non-stationary BINAR(1) process with Poisson innovations where the series were subject to the influence of some common time-varying covariates to induce the non-stationarity. On the same non-stationary condition and cross-correlated structure specification, the non-stationary BINAR(1) process with NB innovations [5] was developed, wherein the current observations were related with the previous-lagged observation of the same series via the binomial thinning operator. Notably, the non-stationary BINAR(1) with Geometric counting series has not yet been developed. In fact, the research outcomes in [6] motivates such a development. It was remarked that the stationary bivariate INAR(1) process under the Geometric marginals yield superior AICs and BICs than the unconstrained BINAR(1) process with Poisson innovations [4]. Besides, such a BINAR(1) process caters for over-dispersion and it is therefore important to compare the standard errors (s.e) of the regression estimates and other dependence parameters along with the root mean square errors (RMSEs) under the BINAR(1) with Poisson and Geometric marginals. Thus, this paper proposes to construct the non-stationary BINAR(1) with Geometric marginals under the simple assumption that the inter-relation between the series is induced by the paired correlated innovations only. It is noteworthy that the thinning distribution under the Geometric series follows the NB model with constant thinning coefficient. This is in contrast with the BINAR(1)s with Poisson or NB marginals where the thinning mechanism was binomial-based (See the references [7], [8], [9], [10], [11], [12], [13], [14]).

As for parameter estimation, the specification of the full or conditional likelihood equation is rather tedious since under the BINAR(1) with Geometric series, the corresponding marginal innovation series is mixed Geometric as illustrated in [3] and [15]. Thus, in this paper, we propose, similar to [16], [17], [18] and [19], to formulate a generalized quasi-likelihood (GQL) equation that relies only on the correct specification of the scores and corresponding mean function. The auto-covariance structure in the GQL may be approximated using the robust autocorrelation coefficients from the method of moments described in [16], [13], [14], [20] and [21]. This method is further encouraged since it is known to provide asymptotically equally efficient estimates as likelihood based methods satisfying the normality

assumptions and the consistency property [22]. The remainder of the paper is described as follows: In the next section, the non-stationary BINAR(1) model with Geometric marginal counting series is constructed and its statistical properties are derived. In section III, the GQL approach is presented to estimate the mean or regression effects while the serial and dependence parameters is handled by a robust moment estimating function. A real-life application on stock data is provided in section IV and the conclusion is provided in the last section.

II. MODEL DEVELOPMENT

The two INAR(1) series is specified as

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

$$Y_t^{[2]} = \rho_2 * Y_{t-1}^{[2]} + R_t^{[2]} \quad (2)$$

where $Y_t^{[k]}$ is Geometric with parameters $\left(\frac{\mu_t^{[k]}}{1+\mu_t^{[k]}}\right)$, that is Pr $(Y_t^{[k]} = y) = \left(\frac{(\mu_t^{[k]})^y}{(1+\mu_t^{[k]})^{y+1}}\right)$, $\mu_t^{[k]} \geq 0$. $\rho_k*$is the NB thinning operator where $\rho_k * Y_{t-1}^{[k]}$ indicates $\rho_k * Y_{t-1}^{[k]} = \sum_{i-1}^{Y_{t-1}^{[k]}} Z_i$ where $\{Z_i\}$ are i.i.d random variables with Geom $\left(\frac{\rho_k}{1+\rho_k}\right)$. That is, $\rho_k * Y_{t-1}^{[k]} | Y_{t-1}^{[k]}$ follows NB with parameters $Y_{t-1}^{[k]}$ and probability $\left(\frac{\rho_k}{1+\rho_k}\right)$. In the above, $<R_t^{[k]}>$ are independent series and the pair $<Y_t^{[k]}, R_{t'}^{[k]}>$ are mutually independent for $t \neq t'$. Note

$$\text{Corr}\,(R_t^{[1]}, R_{t'}^{[2]}) = \begin{cases} \rho_{12,t} & t = t', \\ 0 & t \neq t'. \end{cases} \quad (3)$$

Under these assumptions, it can can easily be verified that

Cov $(Y_t^{[k]}, R_t^{[k]})$ = Var $(R_t^{[k]})$ and

Cov $(Y_t^{[k]}, R_t^{[k']})$ = Cov $(R_t^{[k]}, R_t^{[k']})$, $k, k' \in \{1,2\}$ and $k \neq k'$.

Using the thinning properties from [15]:

1. E $(\rho_k * Y_{t-1}^{[k]}) = \rho_k E(Y_{t-1}^{[k]})$
2. Var$(\rho_k * Y_{t-1}^{[k]}) = \rho_k(1+\rho_k)E(Y_{t-1}^{[k]}) + \rho_k^2 \text{Var}(Y_{t-1}^{[k]})$

and with $E(Y_t^{[k]}) = \mu_t^{[k]}$ and Var$(Y_t^{[k]}) = \mu_t^{[k]}(1+\mu_t^{[k]})$, we derive the distribution of the innovation process $R_t^{[k]}$.

Theorem 1: Let $\varphi_{Y_t^{[k]}}, \varphi_Z$ and $\varphi_{R_t^{[k]}}$, be the probability generating functions of the random variables $Y_t^{[k]}$, Z and $R_t^{[k]}$) respectively. Following (1) and (2),

$$\varphi_{R_t^{[k]}}(s) = \frac{\varphi_{Y_t^{[k]}}(s)}{\varphi_{Y_t^{[k]}}(\varphi_Z(s))} = \frac{1}{1+\mu_t^{[k]}(1-s)}\left[1-\frac{\rho_k\mu_t^{[k]}}{(\mu_t^{[k]}-\rho_k)}\right] + \frac{1}{1+\rho_k-\rho_k s}\left[\frac{\rho_k\mu_t^{[k]}}{(\mu_t^{[k]}-\rho_k)}\right] \quad (4)$$

From the above, the probability mass function (pmf) of the random variable $R_t^{[k]}$ is given by

$$\Pr(R_t^{[k]} = m) = \left[1 - \frac{\rho_k\mu_t^{[k]}}{(\mu_t^{[k]}-\rho_k)}\right]\frac{\mu_t^{[k]m}}{(1+\mu_t^{[k]})^{m+1}} + \left[\frac{\rho_k\mu_t^{[k]}}{\mu_t^{[k]}-\rho_k}\right]\frac{\rho_k^m}{(1+\rho_k)^{m+1}} \quad (5)$$

for m = 0,1,2,… and using the above properties, the mean and variance of the random variable $R_t^{[k]}$ are given as follows:

$$E(R_t^{[k]}) = \lambda_t^{[k]} = \mu_t^{[k]} - \rho_k\mu_{t-1}^{[k]} \quad (6)$$

and

$$\text{Var}(R_t^{[k]}) = \mu_t^{[k]}(1+\mu_t^{[k]}) - \rho_k(1+\rho_k)\mu_{t-1}^{[k]} - \rho_k^2\mu_{t-1}^{[k]}(1+\mu_{t-1}^{[k]}) \quad (7)$$

Under stationary assumptions, the moments of these innovations reduce to the same forms as specified by [15] and confirm that the innovations are also marginally mixed Geometric with parameters $\left(\frac{\mu_t^{[k]}}{1+\mu_t^{[k]}}\right)$ and $\left(\frac{\rho_k}{1+\rho_k}\right)$.

The serial-correlation and cross variances at lag-h are derived to obtain:

$$\text{Corr}(Y_t^{[k]}, Y_{t+h}^{[k]}) = \rho_k^h \frac{\sqrt{\mu_t^{[k]} + \mu_t^{[k]^2}}}{\sqrt{\mu_{t+h}^{[k]} + \mu_{t+h}^{[k]^2}}} \quad (8)$$

$$\text{Cov}(Y_t^{[1]}, Y_{t+h}^{[2]}) = \rho_2^h \text{Cov}(Y_t^{[1]} Y_t^{[2]}) \quad (9)$$

$$\text{Cov}(Y_{t+h}^{[1]}, Y_t^{[2]}) = \rho_1^h \text{Cov}(Y_t^{[1]} Y_t^{[2]}) \quad (10)$$

while the cross-covariances at the same time point is given by:

$$\text{Cov}(Y_t^{[1]}, Y_t^{[2]}) = \text{Cov}(\rho_1 * Y_{t-1}^{[1]} + R_t^{[1]}, \rho_2 * Y_{t-1}^{[2]} + R_t^{[2]}) = \rho_1\rho_2 \text{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) + [\rho_{12,t}\sqrt{\text{Var}(R_t^{[1]})}\sqrt{\text{Var}(R_t^{[2]})}] \quad (11)$$

III. ESTIMATION METHODOLOGY

The GQL function to estimate the regression parameters in the bivariate set-up follow the same procedure as in [16] and [19] and is of the form

$$D'_\beta \Sigma_\beta^{-1}(Y - \mu) = 0, \quad (12)$$

where in terms of the R built in function, Y = rbind [$Y^{[1]}, Y^{[2]}$] with $Y^{[k]} = [Y_1^{[k]}, Y_2^{[k]}, \ldots, Y_t^{[k]}, \ldots, Y_T^{[k]}]'$ and μ =rbind [$E(Y^{[1]})$, E $(Y^{[2]})$], and D_β is (2Txp) with $D_\beta = [D_\beta^{[1]}, D_\beta^{[2]}]$ with the j^{th} column $D_\beta^{[k]}$ given by $c(\frac{\delta\mu_1^{[k]}}{\delta\beta_j^{[k]}}, \ldots, \frac{\delta\mu_t^{[k]}}{\delta\beta_j^{[k]}}, \ldots, \frac{\delta\mu_T^{[k]}}{\delta\beta_j^{[k]}})$ with j =1 : p. Note, $\frac{\delta\mu_t^{[k]}}{\delta\beta_j^{[k]}} = \mu_t^{[k]} x_t$.

The structure of $\sum\beta$ is as follows:

$$\sum\beta = \begin{pmatrix} \text{Var}(Y_t^{[1]}) & \text{Cov}(Y_t^{[1]}, Y_t^{[2]}) \\ \text{Cov}\left(Y_t^{[2]}, Y_t^{[1]}\right) & \text{Var}(Y_t^{[2]}) \end{pmatrix} \quad (13)$$

β = rbind [$\beta^{[1]}, \beta^{[2]}$] is estimated iteratively:

$$(\hat{\beta}_{r+1}) = (\hat{\beta}_r) + [D'_\beta \Sigma_\beta^{-1} D_\beta]^{-1}_r \quad [D'_\beta \Sigma_\beta^{-1}(Y - \mu)]_r \quad (14)$$

Note in the above $\sum\beta, \rho_k$ in (8) is estimated by using the moment estimator proposed by [16] based on the robust autocorrelation estimate by [20] and [23].

$$\hat{\rho}_k = \frac{T\sum_{t=1}^{T-1}\tilde{Y}_t^{[k]}\tilde{Y}_{t+1}^{[k]}}{\sum_{t=1}^{T}\tilde{Y}_t^{2\ [k]}\sum_{t=1}^{T-1}\sqrt{\frac{\mu_t^{[k]}+\mu_t^{[k]\ 2}}{\mu_{t+1}^{[k]}+\mu_{t+1}^{[k]\ 2}}}} \quad (15)$$

with $\tilde{Y}_t^{[k]} = \frac{\tilde{Y}_t^{[k]}-\mu_t^{[k]}}{\sqrt{\mu_t^{[k]}+\mu_t^{[k]\ 2}}}$ and the cross-correlation $\rho_{12,t}$ in (11) is estimated empirically using

$$\hat{\rho}_{12,t} = \frac{\tilde{\sigma}_{t,12}-\hat{\rho}_1\hat{\rho}_2\tilde{\sigma}_{t-1,1,12}}{\sqrt{\mathrm{Var}(R_t^{[1]})}\sqrt{\mathrm{Var}(R_t^{[1]})}} \quad (16)$$

where $\tilde{\sigma}_{t,12} = \frac{1}{T}\sum_{t=1}^{T}(Y_t^{[1]} - \hat{\mu}_t^{[1]})(Y_t^{[2]} - \hat{\mu}_t^{[2]})$, $\tilde{\sigma}_{t-1,12} = \frac{1}{T-1}\sum_{t=2}^{T}(Y_{t-1}^{[1]} - \hat{\mu}_{t-1}^{[1]})(Y_{t-1}^{[2]} - \hat{\mu}_{t-1}^{[2]})$ and $\hat{\mu}_t^{[k]}) = \exp(x_t'\hat{\beta}^{[k]})$.

$(\hat{\beta} - \beta)$ is shown to be asymptotically normal with mean 0 and covariance matrix specified by

$$[D'_\beta\Sigma_\beta^{-1}D_\beta]^{-1} \quad [D'_\beta\Sigma_\beta^{-1}(Y-\mu)(Y-\mu)'\Sigma_\beta^{-1}D_\beta] \quad X\,[D'_\beta\Sigma_\beta^{-1}D_\beta]^{-1} \quad (17)$$

The derivation of the above asymptotic covariance follows from [16] and [22].

The conditional expectation and variance of the one step ahead forecast $Y_{t+1}^{[k]}$ given $Y_t^{[k]}$ are:

$$E(Y_{t+1}^{[k]}|Y_t^{[k]}) = \hat{\rho}_k(Y_t^{[k]} - \hat{\mu}_t^{[k]}) + \hat{\mu}_{t+1}^{[k]} \quad (18)$$

$$\mathrm{Var}(Y_{t+1}^{[k]}|Y_t^{[k]}) = \hat{\rho}_k(1+\hat{\rho}_k)Y_t^{[k]} + \hat{\mu}_{t+1}^{[k]}(1+\hat{\mu}_{t+1}^{[k]}) - \hat{\rho}_k(1+\hat{\rho}_k)\hat{\mu}_t^{[k]} - \hat{\rho}_k^{\ 2}\hat{\mu}_t^{[k]}(1+\hat{\mu}_t^{[k]}) \quad (19)$$

IV. APPLICATION OF BINAR(1) MODEL

We analyze the intra-day transactions of MCB and SBMH, collected from Stock Exchange of Mauritius (SEM) on a 30 minutes interval from 7th May to 22nd May 2018, that makes a total of 108 paired observations and three explanatory variables that influence these two series were considered: newseffect, Friday effect and time of the day effect. For the stock data, SBMH has a mean (variance) of 1.3611 (2.9445) and 1.2506 (2.5098) for MCB. The series exhibit a cross-correlation of 0.2671. The BINAR(1) model introduced in section II (BINAR(1)GEOM) is fittedand the GQL inference is applied to estimate the effects of these covariates and a comparison is conducted with the BINAR(1) model with Poisson innovations (BINAR(1)POIS) introduced by [16]. The results are given in Table II: Usingthe forecasting equation (18), we compute the one-step forecasts and root mean square errors (RMSEs) for the number intra-day transactions for SBMH and MCB for

TABLE II: INTRA-DAY TRANSACTIONS FOR SBMH AND MCB: GQL ESTIMATES OF THE REGRESSION AND SERIAL-CORRELATION PARAMETERS.

Time Series	Intercept	News	Friday	Time	$\hat{\rho}$	$\hat{\rho}_{12,1}$
			BINAR(1)	GEOM		
SBMH	0.2754 (0.2139)	0.2362 (0.0536)	0.1679 (0.0546)	0.1212 (0.0561)	0.3571	0.1734
MCB	0.2269 (0.1812)	0.1845 (0.0585)	0.1355 (0.0548)	0.1122 (0.0575)	0.2665	
			BINAR (1)	Pois		
SBMH	0.2438 (0.2217)	0.2088 (0.0642)	0.1463 (0.0607)	0.1335 (0.0631)	0.3543	0.1734
MCB	0.2051 (0.1971)	0.1557 (0.0625)	0.1446 (0.0613)	0.1286 (0.0618)	0.2551	

TABLE III: RMSES

Model	RMSE SBMH	RMSE MCB
BINAR (1) POIS	0.2422	0.2198
BINAR(1) GEOM	0.1918	0.1854

the BINAR(1)GEOM and also for the BINAR(1)POIS of [16], as shown in Table III: From the RMSE tables, we note that the BINAR(1)GEOM provides lower RMSE than BINAR(1)POIS.

CONCLUSION

This paper considers the modeling of the non-stationary bivariate time series using a BINAR(1) process with Geometric marginals where the means of the counting series were function of some common time-varying covariates. In section II, the model is developed and it is shown marginally thatthe moments of the corresponding innovation series follow the moment properties of a mixed Geometric distribution. The regression effects are estimated by the GQL approach while the serial and cross-correlation effects by the method of moments. The model was alsoapplied to analyze the intra-day transactions and providesreliable estimates and suitable RMSEs that will be surely helpful for policy makers.

REFERENCES

[1] M. Y. T. Z. A. Y. X. Zhang, Y. Liu and X. Li, "Comparative study of four time series methods in forecasting typhoid fever incidence in china." PLOS

[2] 1:http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0063116.

[3] X. Pedeli and D. Karlis, "A bivariate INAR(1) process with application," Statistical Modelling: An International Journal, vol. 11, pp. 325–349, 2011.

[4] M. Ristic, A. Nastic, K. Jayakumar, and H. Bakouch, "A bivariate INAR(1) time series model with geometric marginals," Applied Mathematical Letters, vol. 25(3), pp. 481–485, 2012.
[5] X. Pedeli and D. Karlis, "Some properties of multivariate INAR(1)processes."Computational Statistics and Data Analysis, vol. 67, pp 213–225, 2013a.
[6] Y. Sunecher, N. Mamodekhan, and V. Jowaheer, "Estimating the parameters of a BINMA Poisson model for a non-stationary bivariate time series," Communication in Statistics: Simulation and Computation, 2016.
[7] A. Nastic, M. Ristic, and P. Popovic, "Estimation in a bivariate integervalued autoregressive process," Communication in Statistics-Theory and Methods, vol. 45(19), pp. 5660–5678, 2016b.
[8] E. Aly and N. Bouzar, "Stationary solutions for integer-valued autoregressive
[9] processes," International Journal of Mathematics and Mathematical Sciences, vol. 2005(1), pp. 1–18, 2005.
[10] M. Bourguignon and K. Vasconcellos, "Improved estimation for Poisson INAR(1) models," Journal of Statistical Computation and Simulation, vol. 85(12), pp. 2425–2441, 2015.
[11] E. McKenzie, "Some ARMA models for dependent sequences of Poisson counts," Advances in Applied Probability, vol. 20, pp. 822– 835, 1988.
[12] M. Silva and V. Oliveira, "Difference equations for the higher order moments and cumulants of the INAR(1) model," Journal of Time Series Analysis, vol. 25, pp. 317–333, 2004.
[13] F. Steutel and K. Van Harn, "Discrete analogues of selfdecomposability and stability," The Annals of Probability, vol. 7, pp. 3893–3899, 1979.
[14] C. Wei_, "Thinning operations for modelling time series of counts-a survey." AStA Advances in Statistical Analysis, vol. 92(3), pp. 319– 341, 2008b.
[15] V. Jowaheer and B. Sutradhar, "On AR(1) versus MA(1) models for non-stationary time series of Poisson counts: part I (theory)," Wseas Transaction on Mathematics, vol. 4, pp. 1109–2769, 2005.
[16] N. Mamode Khan, Y. Sunecher, and V. Jowaheer, "Modelling AR(1) stationary time series of COM-Poisson counts." Proceedings of the 9th International Conference on Appplied Mathematics, Simulation, Modelling, 2015, pp. 104–110.
[17] P. Borges, M. Bourguignon, and F. Molinares, "A generalised NGINAR(1) process with inflated-parameter geometric counting series," Australian and New Zealand Journal of Statistics, vol. 59(1), pp. 137– 150, 2017.
[18] N. Mamode Khan, Y. Sunecher, and V. Jowaheer, "Modelling a nonstationary BINAR(1) Poisson process," Journal of Statistical Computation and Simulation, vol. 86, pp. 3106–3126, 2016b.
[19] J. Wilson, "A generalized quasi-likelihood model application to modelling poverty of asian american women." Journal of Modern Applied Statistical Methods, vol. 3(1), pp. 165–175, 2004.
[20] R. Alzghool, "Estimation for stochastic volatility model: Quasilikelihood and asymptotic quasi-likelihood approaches." Journal of King Saud University, vol. 29, pp. 114–118, 2017.
[21] Y. Sunecher, N. Mamodekhan, and V. Jowaheer, "A gql estimation approach for analysing non-stationary over-dispersed BINAR(1) time series," Journal of Statistical Computation and Simulation, 2017.
[22] B. Sutradhar and K. Das, "On the efficiency of regression estimators in generalised linear models for longitudinal data," Biometrika, vol. 86, pp. 459–65, 1999.
[23] V. Jowaheer and B. Sutradhar, "Analysing longitudinal count data with overdipsersion," Biometrika, vol. 89, pp. 389–399, 2002.
[24] B. Sutradhar, V. Jowaheer, and P. Rao, "Remarks on asymptotic efficient estimation for regression effects in stationary and non-stationary models for panel count data," Brazilian Journal of Probability and Statistics, vol. 28(2), pp. 241–254, 2014.
[25] T. Mallick and B. Sutradhar, "GQL versus conditional GQL inferences for non-stationary time series of counts with overdispersion," Journal of Time Series Analysis, vol. 29(2), pp. 402–420, 2008.
[26] M. Ristic, H. Bakouch, and A. Nastic, "A new geometric firstorder integer-valued autoregressive (NGINAR(1)) process," Journal of Statistical Planning and Inference, vol. 136, pp. 2218–2226, 2009.

★★★