Modelling the Number of Daily Stock Transactions using a Novel Time Series Model.

Y. Sunecher

June 27-30, 2022

Outline:

Principal Investigator

'Modelling the Number of Daily Stock Transactions using a Novel Time Series Model.'

- Dr Yuvraj Sunecher, Lecturer, University of Technology Mauritius

Motivation

- Bivariate count time series data recorded at consecutive time points:$\{[Y_t^{[1]}, Y_t^{[2]}]', \ t \in \mathbb{Z}\}$, influenced by some common time-dependent and/or time-independent explanatory variables $\boldsymbol{x}_t$.
- Some real-life examples can be found in:
- Road accidents (slight injury, injury and fatal)
- Financial data, among others.

Stock Data

$$
\begin{array}{cc}
Y_t^{[1]} & Y_t^{[2]} \\
\hline
Y_1^{[1]} & Y_1^{[2]} \\
Y_2^{[1]} & Y_2^{[2]} \\
\vdots & \vdots \\
Y_T^{[1]} & Y_T^{[2]}
\end{array}
$$

- where $Y_t^{[1]}$ indicates the number of intra-day transactions of Company 1 and $Y_t^{[2]}$ indicates the number of intra-day transactions of Company 2 for $t = 1, \ldots, T$.

Explanatory Variables

- Some common time-dependent and/or time-independent explanatory variables $\boldsymbol{x}_t$: news effect, time of the day effect, day effect among others.
- Collaboration with stock brokers.

Challenges

- Modelling such bivariate count time series recorded over consecutive time points which are
- correlated (serial- and cross-correlation),
- under the influence of time dependent/independent covariates
- and different dispersion parameters makes it challenging.

Outline:

Classical INARMA(1,1)

- The first INARMA model was developed by McKenzie (1988) under stationarity condition and was specified as:

$$Y_t = \rho_1 \circ Y_{t-1} + \rho_2 \circ R_{t-1} + R_t \quad \text{for} \quad t = 1, \ldots, T, \quad Y_t \geq 0 \tag{1}$$

  where Y_t are non-negative count values, ρ_k indicates the serial-correlation, $\circ$ is the binomial thinning operator such that $\rho \circ R_{t-1}$ accounts for the count nature of the series and R_t is the innovation term.
- This model can be extended as follows:

BINARMA(1,1) Model:

- $$Y_t^{[1]} = \gamma_{11} * Y_{t-1}^{[1]} + \gamma_{12} * R_{t-1}^{[2]} + R_t^{[1]} \quad (2)$$
$$Y_t^{[2]} = \gamma_{21} * Y_{t-1}^{[2]} + \gamma_{22} * R_{t-1}^{[2]} + R_t^{[2]} \quad (3)$$
- The pair of innovations $\{[R_t^{[1]}, R_t^{[2]}]\}$ is allowed to follow the bivariate CMP model.
- $R_t^{[k]} \sim COM - Poisson(\lambda_t^{[k]}, \tilde{\nu}_k)$. Note that $\lambda_t^{[k]} = (\theta_t^{[k]})^{1/\tilde{\nu}_k} - (\frac{\tilde{\nu}_k - 1}{2\tilde{\nu}_k})$, where $\theta_t^{[k]} = \exp(\boldsymbol{x}_t' \boldsymbol{\beta}^{[k]})$ with $\boldsymbol{\beta}^{[k]} = [\beta_1^{[k]}, \beta_2^{[k]}, \ldots, \beta_j^{[k]}, \ldots, \beta_p^{[k]}]'$ for $k \in \{1, 2\}$.

BINARMA(1,1) Model

- Hence, $E(R_t^{[k]}) = \lambda_t^{[k]}$, $\text{Var}(R_t^{[k]}) = \left[\frac{\lambda_t^{[k]}}{\tilde{\nu}_k} + \frac{\tilde{\nu}_k - 1}{2\tilde{\nu}_k^2}\right]$, while $\text{Corr}(R_t^{[1]}, R_t^{[2]}) = \alpha$.
- Finally, the relationship between $Y_t^{[k]}$ and $R_t^{[k]}$ are as follows: $\text{Cov}(Y_t^{[k]}, R_t^{[k]}) = \text{Var}(R_t^{[k]})$.

Model Development

$$\mu_t^{[1]} \equiv E(Y_t^{[1]}) = \gamma_{11}\mu_{t-1} + \gamma_{12}\lambda_{t-1}^{[1]} + \lambda_t^{[1]}, \tag{4}$$

$$\mu_t^{[2]} \equiv E(Y_t^{[2]}) = \gamma_{21}\mu_{t-1} + \gamma_{22}\lambda_{t-1}^{[1]} + \lambda_t^{[1]}, \tag{5}$$

$$\begin{aligned}\text{Var}(Y_t^{[1]}) = {} & \gamma_{11}(1-\gamma_{11})\mu_{t-1}^{[1]} + \gamma_{11}^2\,\text{Var}(Y_{t-1}^{[1]}) + \gamma_{12}(1-\gamma_{12})\lambda_{t-1}^{[1]} \\ & + (\gamma_{12}^2 + 2\gamma_{11}\gamma_{12})\left[\frac{\lambda_{t-1}^{[1]}}{\tilde{\nu}_1} + \frac{\tilde{\nu}_1 - 1}{2\tilde{\nu}_1^2}\right] + + \left[\frac{\lambda_t^{[1]}}{\tilde{\nu}_1} + \frac{\tilde{\nu}_1 - 1}{2\tilde{\nu}_1^2}\right],\end{aligned} \tag{6}$$

Model Development

$$\text{Var}(Y_t^{[2]}) = \gamma_{21}(1-\gamma_{21})\mu_{t-1}^{[2]} + \gamma_{21}^2 \text{Var}(Y_{t-1}^{[2]}) + \gamma_{22}(1-\gamma_{22})\lambda_{t-1}^{[2]} + (\gamma_{22}^2 + 2\gamma_{21}\gamma_{22})\left[\frac{\lambda_{t-1}^{[2]}}{\tilde{\nu}_2} + \frac{\tilde{\nu}_2 - 1}{2\tilde{\nu}_2^2}\right] + \left[\frac{\lambda_t^{[2]}}{\tilde{\nu}_2} + \frac{\tilde{\nu}_2 - 1}{2\tilde{\nu}_2^2}\right]. \quad (7)$$

The cross-covariances between the two series are given by

$$\text{Cov}(Y_t^{[1]}, Y_t^{[2]}) = \gamma_{11}\gamma_{21}\text{Cov}(Y_{t-1}^{[1]}, Y_{t-1}^{[2]}) + (\gamma_{11}\gamma_{22} + \gamma_{12}\gamma_{21} + \gamma_{12}\gamma_{22})\alpha + \alpha\sqrt{\lambda_t^{[1]}}\sqrt{\lambda_t^{[2]}}. \quad (8)$$

Outline:

Estimation Approach (CMLE)

- This section describes the CML estimation method for estimating the regression and correlation parameters based on thinning and convolution properties following Pedeli and Karlis (2011).
- Thus, the conditional density of the BINARMA(1,1) model is given by

$$f_1(k) = \sum_{j_1=0}^{k} \binom{y_{t-1}^{[1]}}{j_1} \binom{r_{t-1}^{[1]} = y_{t-1}^{[1]} - k}{k - j_1} \gamma_{11}^{j_1}(1-\gamma_{11})^{y_{t-1}^{[1]}-j_1} \gamma_{12}^{k-j_1}(1-\gamma_{12})^{y_{t-1}^{[1]}-2k+j_1}, \quad (9)$$

CMLE Approach

- $$f_2(s) = \sum_{j_2=0}^{s} \binom{y_{t-1}^{[2]}}{j_2} \binom{r_{t-1}^{[2]} = y_{t-1}^{[2]} - s}{s - j_2} \gamma_{21}^{j_2}(1-\gamma_{21})^{y_{t-1}^{[2]}-j_2} \gamma_{22}^{s-j_2}(1-\gamma_{22})^{y_{t-1}^{[2]}-2s+j_2}, \quad (10)$$

CMLE approach

- and a bivariate distribution of the innovation terms $f_3(r_t^{[1]} = y_{t-1}^{[1]} - k, r_t^{[2]} = y_{t-1}^{[2]} - s)$

$$= \left[\frac{(\lambda_t^{[1]})^{y_{t-1}^{[1]}-k}}{((y_{t-1}^{[1]}-k)!)^{\tilde{\nu}_1}}\right]\left[\frac{1}{Z(\lambda_t^{[1]},\tilde{\nu}_1)}\right]\left[\frac{(\lambda_t^{[2]})^{y_{t-1}^{[2]}-k}}{((y_{t-1}^{[2]}-k)!)^{\tilde{\nu}_2}}\right]\left[\frac{1}{Z(\lambda_t^{[2]},\tilde{\nu}_2)}\right] \tag{11}$$

CMLE

- The conditional density is written as
  $f((y_t^{[1]}, y_t^{[2]})|(y_{t-1}^{[1]}, y_{t-1}^{[2]}, r_{t-1}^{[1]}, r_{t-1}^{[2]}), \boldsymbol{\theta})$
  $= \sum_{k=0}^{g_1} \sum_{s=0}^{g_2} f_1(k) f_2(s) f_3(r_t^{[1]} = y_{t-1}^{[1]} - k, r_t^{[2]} = y_{t-1}^{[2]} - s)$,
  where $\boldsymbol{\theta} = [\gamma_{11}, \gamma_{12}, \gamma_{21}\gamma_{22}, \tilde{\nu}_1, \tilde{\nu}_2, \beta^{[k]}]$ is the vector of unknown parameters.
- The conditional likelihood function is given by

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{t=1}^{T} f((y_t^{[1]}, y_t^{[2]})|(y_{t-1}^{[1]}, y_{t-1}^{[2]}, r_{t-1}^{[1]}, r_{t-1}^{[2]}), \boldsymbol{\theta}) \quad (12)$$

- and the maximum likelihood estimators of $\boldsymbol{\theta}$ is obtained by maximizing

$$\log[L(\boldsymbol{\theta}|\boldsymbol{y})] = \log\left[\sum_{t=1}^{T} f((y_t^{[1]}, y_t^{[2]})|(y_{t-1}^{[1]}, y_{t-1}^{[2]}, r_{t-1}^{[1]}, r_{t-1}^{[2]}), \boldsymbol{\theta})\right] \quad (13)$$

Forecasting Equations

Based on the proposed model, the forecasting equations are derived as follows:

$$E(Y_{t+1}^{[1]}|y_t^{[1]}, r_t^{[1]}) = E(\gamma_{11} \circ Y_t^{[1]}|y_t^{[1]}) + E(\gamma_{12} \circ R_t^{[1]}|r_t^{[1]}) + E(R_{t+1}^{[1]})$$
$$= \gamma_{11} y_t^{[1]} + \gamma_{12} r_t^{[1]} + \lambda_{t+1}^{[1]} \quad (14)$$

$$E(Y_{t+1}^{[2]}|y_t^{[2]}, r_t^{[2]}) = E(\gamma_{21} \circ Y_t^{[2]}|y_t^{[2]}) + E(\gamma_{22} \circ R_t^{[2]}|r_t^{[2]}) + E(R_{t+1}^{[2]})$$
$$= \gamma_{21} y_t^{[1]} + \gamma_{22} r_t^{[2]} + \lambda_{t+1}^{[2]} \quad (15)$$

Application

- Two banking group of companies in Mauritius, namely Mauritius Commercial Bank Group Limited (MCB) and State Bank of Mauritius Holdings Ltd (SBMH).
- Data were collected on a 30 minutes interval from 4^{th} October to 10^{th} December 2021.
- Explanatory variables: Covid-19 news effect (x_{t1}), Friday effect (x_{t2})and time of the day effect (x_{t3}).
- SBMH has a mean (variance) of 1.0765 (1.4191) and 1.0393 (1.0154) for MCB.
- Both series are over-dispersed with a cross-correlation of 0.1617.

Application

Models	Series	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
BINARMA(1,1)CMP	$Y_t^{[1]}$	0.2268	0.2633	0.1482	0.1399
	s.e	(0.1975)	(0.0581)	(0.0288)	(0.0295)
	$Y_t^{[2]}$	0.1543	0.2456	0.1297	0.1193
	s.e	(0.2259)	(0.0447)	(0.0350)	(0.0587)
BINMA(1)CMP	$Y_t^{[1]}$	0.2122	0.2530	0.1533	0.1264
	s.e	(0.2089)	(0.0615)	(0.0314)	(0.0322)
	$Y_t^{[2]}$	0.1650	0.2627	0.1074	0.1317
	s.e	(0.2451)	(0.0566)	(0.0416)	(0.0661)
BINAR(1)CMP	$Y_t^{[1]}$	0.2412	0.2490	0.1231	0.1211
	s.e	(0.2159)	(0.0691)	(0.0375)	(0.0388)
	$Y_t^{[2]}$	0.1856	0.2659	0.0926	0.1262
	s.e	(0.2517)	(0.0612)	(0.0487)	(0.0701)

Table: Intra-day transactions for MCB and SBMH: Estimates of the regression parameters.

Application

- The record of any Covid-19 news on the stock market leads to an increase in the number of transactions.
- There is high trading intensity on Fridays rather than non-Fridays.
- It is remarked that more trading takes place in the afternoon than in the morning sessions.
- The Covid-19 news causes an expected increase in the trading frequency by 30.1 percent for SBMH and 27.8 percent for MCB.

Conclusion

- This paper introduces a non-stationary BINARMA(1,1) model with COM-Poisson innovations.
- The mean, variance and covariance expressions are derived under the assumption of non-stationarity.
- The model parameters are estimated using the CML method.
- These estimates prove to be efficient and reliable.
- The BINARMA(1,1) model is applied on a bivariate stock data.

Outline:

References

- Jowaheer, Mamode Khan and Sunecher (2018). A BINAR(1) Time series model with Cross-correlated COM-Poisson innovations. Communication in Statistics-Theory and Methods 47(5), 1133-1154.
- Karlis and Pedeli (2013). Flexible bivariate INAR(1) processes using copulas. Communications in Statistics Theory and Methods, 42, 723-740.
- Kocherlakota and Kocherlakota. (2001). Regression in the bivariate Poisson distribution. Communications in Statistics-Theory and Methods 30, 815-825.
- Marshall and Olkin. (1985). A family of bivariate distributions generated by the bivariate Bernoulli distribution. Journal of American Statistical Association 80, 332-338.
- McKenzie (1985). Some simple models for discrete variate time series. Water Resources Bulletin, 21, 645-650.

References

- Nastic, Ristic and Popovic (2016). Estimation in a bivariate integer valued autoregressive process. Communication in Statistics Theory and Methods 45, 5660-5678.
- Pedeli and Karlis (2011). A bivariate INAR(1) process with application. Statistical Modelling An International Journal 11, 325-349.
- Pedeli and Karlis (2013) Some properties of multivariate INAR(1) process. computational statistics and data analysis,67,213-225.
- Ristic, Nastic, Jayakumar and Bakouch (2012). A bivariate INAR(1) time series model with geometric marginals. Applied Mathematical Letters 25, 481-485.
- Sutradhar, Jowaheer, and Rao (2014). Remarks on asymptotic efficient estimation for regression effects in stationary and non-stationary models for panel count data. Brazilian Journal of Probability and Statistics 28, 241-254.