Journal of Statistical Research *https://doi.org/10.3329/jsr.v58i1.75427*
2024, Vol. 58, No. 1, pp. 243-257 *ISSN 0256 - 422 X*

NEGATIVE BINOMIAL INTEGER-VALUED AUTO-REGRESSIVE PROCESS FOR LONGITUDINAL COUNT DATA

NAUSHAD MAMODE KHAN
Department of Economics and Statistics, University of Mauritius, Mauritius
Email: n.mamodekhan@uom.ac.mu

VANDNA JOWAHEER
Department of Mathematics, University of Mauritius, Mauritius
Email: vandnaj@uom.ac.mu

YUVRAJ SUNECHER*
Department of Accounting, Finance and Economics, University of Technology Mauritius, Mauritius
Email: ysunecher@utm.ac.mu

SUMMARY

This paper proposes a longitudinal integer-valued auto-regressive model of order one with Negative-Binomial marginals. The proposed model is suitable for analyzing repeated count data that exhibits significant over-dispersion at each time point and that is exposed to several time-dependent covariates. The estimation of the model parameters is handled by two non-likelihood approaches: The Generalized Quasi-likelihood (GQL) and the adaptive Quadratic Inference function (AQIF). The consistency of the model estimators is assessed via Monte Carlo simulation experiments and application to the epileptic seizures is made. The results demonstrate that both approaches GQL and AQIF yield reliable estimates but GQL provides better standard errors.

Keywords and phrases: Longitudinal, Counts, Negative Binomial, GQL, AQIF, Non-Stationarity

AMS Classification: 65C60, 62J12, 62H12, 62J20, 62J10

1 Introduction

Longitudinal count data is made up of repeated count observations for different independent subjects over a specified period of time. The repeated counts for each subject are most likely to be serially correlated. In addition, information on the time-dependent or time-independent covariates that are most likely to influence the repeated counts may be provided. Such types of data are commonly seen in medical studies where different patients are subjected to various types of treatment over a period of time and how these treatments influence their health outcomes. In the literature, Thall and Vail (1990) analyzed the epileptic seizures data where the number of seizures over four eight

* Corresponding author

week periods for fifty nine patients were examined against two different treatments: Placebo and Progabide. Such data arise in various fields of research (Sutradhar, 2011; Oyet and Sutradhar, 2013; Adachi and Willoughby, 2015; Sutradhar and Rao, 2016; Zhao et al., 2020; Tang et al., 2023).

The main interest in such longitudinal studies is to estimate the effects of the different influential factors on the outcomes. Thus, we require to establish the longitudinal model and identify the model parameters. This paper assumes that the inter-valued auto-regressive process of order 1 (INAR(1)) is the time series structure to represent the serial auto-correlation among the repeated observed responses. In this context, we refer to the papers by McKenzie (1986), Al-Zaid and Al-Osh (1990), Al-Osh and Aly (1992), Alosh (2010) and Oyet and Sutradhar (2021). These authors assumed that the current observation is related with previous lagged observation by a binomial thinning operator with fixed or random coefficient and a random innovation term. In this work, we assume that the count response in the INAR process is Negative Binomial (NB) marginal with NB innovations. We then propose to estimate the model parameters including the regression effects, over-dispersion coefficient and auto-correlation parameters using the GQL (Sutradhar and Das (1999)) and AQIF (Qu and Lindsay (2003)) approaches. In a nutshell, Liang and Zeger (1986) proposed a generalized estimating equation (GEE) approach to estimate model parameters in longitudinal count data analysis since the full likelihood was difficult to formulate. The GEE assumes some common 'working' correlation structures like the auto-regressive, equi-correlation and moving average to mimick the true serial structure (Lipsitz and Fitzmaurice (2008); Zhang et al. (2012). However, Crowder (1995) pointed out that the GEE may result in inconsistent estimates of the auto-correlation parameters while Sutradhar and Das (1999) noticed that the 'working' independence structure may provide more efficient estimates than the GEE based under other common working structures. Thereon, Sutradhar and Das (1999) proposed a general auto correlation structure and called the estimation approach as GQL. In the same lines of thought, Qu and Lindsay (2003) proposed an alternative estimation approach to GEE based on using a quadratic inference function (AQIF). Their approach yields more efficient estimates than GEE. However, as at now, there are very few studies comparing GQL and AQIF.

This paper explores the longitudinal INAR with NB marginals and compares the efficiency of the estimators between these two approaches. The organization of this paper is as follows: In section 2, we develop the Gaussian non-stationary AR(1) correlation model for longitudinal NB data. In the next section, we develop the joint generalized quasi-likelihood estimation (JGQL) approach based on these models. We will also develop the AQIF approach. Next, we compare the performance of these approaches for analyzing non-stationary and over-dispersed longitudinal count data through simulation studies and section 5 provides a real-life application to epileptic seizures data. The conclusion is presented in the last section.

2 The Longitudinal INAR(1) Model with NB marginal

We assume that the response y_{it} conditional on random effect γ_{it} has a Poisson distribution of the form

$$f(y_{it} \mid \gamma_{it}) = \frac{1}{y_{it}!} \exp\{y_{it}\eta_{it} - \exp(\eta_{it})\}, \tag{2.1}$$

with $E(Y_{it} \mid \gamma_{it}) = Var(Y_{it} \mid \gamma_{it}) = \exp(\eta_{it})$, where $\eta_{it} = x_{it}^T\beta + \log(\gamma_{it})$ and γ_{it} follows a gamma distribution denoted by $Gamma(\frac{1}{c}, \frac{1}{c})$, with the density function given by

$$g(\gamma_{it}) = \frac{(\gamma_{it})^{\frac{1}{c}-1}\exp(-\frac{\gamma_{it}}{c})}{\Gamma(\frac{1}{c})c^{\frac{1}{c}}}. \tag{2.2}$$

Marginally, y_{it} has the negative binomial distribution with the probability mass function

$$f(y_{it}) = \frac{\Gamma(c^{-1}+y_{it})}{\Gamma(c^{-1})y_{it}!}(\frac{1}{1+c\theta_{it}})^{c^{-1}}(\frac{c\theta_{it}}{1+c\theta_{it}})^{y_{it}}, \tag{2.3}$$

i.e.,

$$Y_{it} \sim \text{NeBin}(1/c, c\theta_{it}), \tag{2.4}$$

where c is the over-dispersion parameter. The expectation and variance of y_{it} are given by

$$E(Y_{it}) = \theta_{it} = \exp(x_{it}^T\beta), Var(Y_{it}) = \theta_{it} + c\theta_{it}^2, \tag{2.5}$$

where $c > 0$.

Consider the relationship of y_{it} and $y_{i,t-1}$, i.e.,

$$y_{it} = \alpha_{it} * y_{i,t-1} + d_{it}, \tag{2.6}$$

where $0 < \alpha_{it} < 1$ and $\alpha_{it} \sim \text{Beta}(\frac{\rho}{c}, \frac{1-\rho}{c})$. We further assume that

$$\begin{aligned} &\alpha_{it} * y_{i,t-1} \mid y_{i,t-1}, \alpha_{it} \sim \text{Binomial}(y_{i,t-1}, \alpha_{it}), \qquad (2.7)\\ &y_{i,t-1} \sim \text{NeBin}(\frac{1}{c}, c\theta_{i,t-1}), \text{ and}\\ &d_{it} \sim \text{NeBin}(\frac{(\theta_{it}-\rho\theta_{i,t-1})^2}{c(\theta_{it}^2-\rho\theta_{i,t-1}^2)}, \frac{c(\theta_{it}^2-\rho\theta_{i,t-1}^2)}{(\theta_{it}-\rho\theta_{i,t-1})}), \text{ for } t = 1, \ldots, T, \end{aligned}$$

where $\theta_{it} = \exp(x_{it}^T\beta)$, d_{it} and $y_{i,t-1}$ are independent. Note $y_{i0} \sim \text{NeBin}(\frac{1}{c}, c\theta_{i1})$. Following these assumptions, $Y_{it} \sim \text{NeBin}(1/c, c\theta_{it})$. Hence we have

$$E(Y_{it}) = \theta_{it}, \tag{2.8}$$

$$Var(Y_{it}) = \theta_{it}(1 + c\theta_{it}), \tag{2.9}$$

$$Cov(Y_{it}, Y_{i,t-k}) = \rho^k(\theta_{i,t-k} + c\theta_{i,t-k}^2). \tag{2.10}$$

The lag-k correlation

$$\begin{aligned} Corr(Y_{it}, Y_{i,t-k}) &= \frac{Cov(Y_{it}, Y_{i,t-k})}{\sqrt{\theta_{it}+c\theta_{it}^2}\sqrt{\theta_{i,t-k}+c\theta_{i,t-k}^2}}\\ &= \rho^k\frac{\sqrt{(\theta_{i,t-k}+c\theta_{i,t-k}^2)}}{\sqrt{\theta_{it}+c\theta_{it}^2}}, \qquad (2.11) \end{aligned}$$

for $k = 1, \ldots, T-1$. Note that since the mean parameter of d_{it}, i.e., $(\theta_{it} - \rho\theta_{i,t-1}) > 0$ and since ρ is a probability parameter, we have

$$0 < \rho < \min(\frac{\theta_{i2}}{\theta_{i1}}, \ldots, \frac{\theta_{it}}{\theta_{i,t-1}}, 1). \tag{2.12}$$

Refer to Appendix for detailed derivations of the moments.

3 Estimation of Parameters

3.1 The joint generalized quasi-likelihood estimating equation (JGQL)

This estimation approach is based on observation-driven non-stationary correlation models developed in section 2. The JGQL estimating equation to estimate the regression and over-dispersion parameters is given by

$$\sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1} (f_i - \mu_i) = 0, \tag{3.1}$$

where $f_i = (f_{i1}^T, \ldots, f_{it}^T, \ldots, f_{iT}^T)$, $\mu_i = (\mu_{i1}^T, \ldots, \mu_{it}^T, \ldots, \mu_{iT}^T)$ are $2T \times 1$ vectors with $f_{it} = (y_{it}, y_{it}^2)$, $\mu_{it} = (\theta_{it}, m_{it})^T$. $\theta_{it} = E(Y_{it})$ and $m_{it} = E(Y_{it}^2) = \theta_{it} + (c+1)\theta_{it}^2$, where $\theta_{it} = \exp(x_{it}^T\beta)$. $\widetilde{\Sigma_i}$ is the covariance matrix of the score vector f_i and D_i is the $2T \times (p+1)$ derivative matrix consisting of

$$D_i = [\partial\mu_i/\partial\beta^T, \partial\mu_i/\partial c] = [D_{i1}^T, \ldots, D_{it}^T, \ldots, D_{iT}^T]^T,$$

with

$$D_{it} = \begin{pmatrix} \partial\theta_{it}/\partial\beta^T & 0 \\ \partial m_{it}/\partial\beta^T & \partial m_{it}/\partial c \end{pmatrix},$$

where $\partial\theta_{it}/\partial\beta^T = \theta_{it}x_{it}^T$, $\partial m_{it}/\partial\beta^T = \theta_{it}x_{it}^T + 2(c+1)\theta_{it}^2 x_{it}^T$ and $\partial m_{it}/\partial c = \theta_{it}^2$. The construction of the covariance matrix $\widetilde{\Sigma_i}$ with the moment generating function of y_{it} is given by

$$M_{y_{it}}(s) = \{1 + c\theta_{it} - c\theta_{it}\exp(s)\}^{-1/c}. \tag{3.2}$$

By deriving the moments for y_{it}^2, y_{it}^3 and y_{it}^4, we obtain

$$\begin{aligned} cov(Y_{it}, Y_{it}^2) &= E(Y_{it}^3) - E(Y_{it})E(Y_{it}^2) \\ &= \theta_{it} + 3\theta_{it}^2 + 3c\theta_{it}^2 + 3c\theta_{it}^3 + 2c^2\theta_{it}^3 + \theta_{it}^3 - \theta_{it}(\theta_{it} + (c+1)\theta_{it}^2) \\ &= \theta_{it}\{1 + (2+3c)\theta_{it} + 2c(1+c)\theta_{it}^2\}, \end{aligned} \tag{3.3}$$

$$\begin{aligned} Var(Y_{it}^2) &= E(Y_{it}^4) - E(Y_{it}^2)^2 \\ &= \theta_{it} + 7\theta_{it}^2 + 7c\theta_{it}^2 + 6\theta_{it}^3 + 18c\theta_{it}^3 + 12c^2\theta_{it}^3 + 6c\theta_{it}^4 \\ &\qquad + 11c^2\theta_{it}^4 + 6c^3\theta_{it}^4 + \theta_{it}^4 - (\theta_{it} + (c+1)\theta_{it}^2)^2 \\ &= \theta_{it} + (6+7c)\theta_{it}^2 + (4 + 16c + 12c^2)\theta_{it}^3 + (4c + 10c^2 + 6c^3)\theta_{i1}^4. \end{aligned} \tag{3.4}$$

The computation of elements of $\widetilde{\Omega_{itw}}$ is done using the relation

$$\begin{aligned}\sigma_{itw} &= \widetilde{\sigma_{itw}} - \theta_{it}\theta_{iw} \\ &= \rho_{itw}\sigma_{itt}\sigma_{iww},\end{aligned} \tag{3.5}$$

with $\sigma_{itt} = \sqrt{\theta_{it} + c\theta_{it}^2}$, where $\theta_{it} = E(Y_{it})$ and $\widetilde{\sigma_{itw}} = E(Y_{it}Y_{iw})$. The off-diagonal sub matrices are obtained from

$$E(Y_{it}Y_{iw}Y_{iw}) = 2\left(\rho_{itw}\sqrt{\left(\theta_{it} + c{\theta_{it}}^2\right)\left(\theta_{iw} + c{\theta_{iw}}^2\right)} + \theta_{it}\theta_{iw}\right)\theta_{iw} + (\theta_{iw} + (c+1)\theta_{iw}^2)\theta_{it} - 2\,\theta_{it}{\theta_{iw}}^2, \tag{3.6}$$

$$\begin{aligned}E(Y_{it}Y_{it}Y_{iw}Y_{iw}) &= \sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + 2\rho_{itw}^2(\theta_{it} + c\theta_{it}^2)(\theta_{iw} + c\theta_{iw}^2) \\ &+ 2((\theta_{it} + (c+1)\theta_{it}^2)\theta_{iw} \\ &+ 2(\rho_{itw}\sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2}\theta_{it}\theta_{iw})\theta_{it} - 2\theta_{it}^2\theta_{iw})\theta_{iw} \\ &+ 2(2(\rho_{itw}\sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + \theta_{it}\theta_{iw})\theta_{iw} + (\theta_{iw} + c\theta_{iw}^2 + \theta_{iw}^2)\theta_{it} \\ &- 2\theta_{it}\theta_{iw}^2)\theta_{it} - (\theta_{it} + c\theta_{it}^2 + \theta_{it}^2)\theta_{iw}^2 \\ &- 4(\rho_{itw}\sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + \theta_{it}\theta_{iw})\theta_{it}\theta_{iw} \\ &- (\theta_{iw} + (c+1)\theta_{iw}^2)\theta_{it}^2 + 3\theta_{it}^2\theta_{iw}^2.\end{aligned} \tag{3.7}$$

Thus we have

$$\begin{aligned}Cov(Y_{it}, Y_{iw}) &= E(Y_{it}Y_{iw}) - E(Y_{it})E(Y_{iw}) \\ &= \rho_{itw}\sqrt{(\theta_{it} + c{\theta_{it}}^2)(\theta_{iw} + c{\theta_{iw}}^2)},\end{aligned} \tag{3.8}$$

$$\begin{aligned}Cov(Y_{it}, Y_{iw}^2) &= E(Y_{it}Y_{iw}Y_{iw}) - E(Y_{it})E(Y_{iw}^2) \\ &= 2(\rho_{itw}\sqrt{(\theta_{it} + c{\theta_{it}}^2)(\theta_{iw} + c{\theta_{iw}}^2)} + \theta_{it}\theta_{iw})\theta_{iw} - 2\theta_{it}{\theta_{iw}}^2,\end{aligned} \tag{3.9}$$

$$\begin{aligned}Cov(Y_{it}^2, Y_{iw}^2) &= E(Y_{it}Y_{it}Y_{iw}Y_{iw}) - E(Y_{it}^2)E(Y_{iw}^2) \\ &= \sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + 2\rho_{itw}^2(\theta_{it} + c\theta_{it}^2)(\theta_{iw} + c\theta_{iw}^2) \\ &+ 2((\theta_{it} + (c+1)\theta_{it}^2)\theta_{iw} + 2(\rho_{itw}\sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + \theta_{it}\theta_{iw})\theta_{it} - 2\theta_{it}^2\theta_{iw})\theta_{iw} \\ &+ 2(2(\rho_{itw}\sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + \theta_{it}\theta_{iw})\theta_{iw} + (\theta_{iw} + c\theta_{iw}^2 + \theta_{iw}^2)\theta_{it} - 2\theta_{it}\theta_{iw}^2)\theta_{it} \\ &- (\theta_{it} + c\theta_{it}^2 + \theta_{it}^2)\theta_{iw}^2 - 4(\rho_{itw}\sqrt{\theta_{it} + c\theta_{it}^2}\sqrt{\theta_{iw} + c\theta_{iw}^2} + \theta_{it}\theta_{iw})\theta_{it}\theta_{iw} \\ &- (\theta_{iw} + (c+1)\theta_{iw}^2)\theta_{it}^2 + 3\theta_{it}^2\theta_{iw}^2 - (\theta_{it} + (c+1)\theta_{it}^2)(\theta_{iw} + (c+1)\theta_{iw}^2).\end{aligned} \tag{3.10}$$

Note that covariances will be different even for the responses which are same lags apart and for different individuals. We estimate the correlations for the different structures using a moment estimating equation approach following Jowaheer and Sutradhar (2005). We derive a moment estimate of ρ as follows:

AR(1) non-stationary model. From previous section, we obtain

$$E(Y_{it}) = \theta_{it}, Var(Y_{it}) = \theta_{it} + c\theta_{it}^2 \tag{3.11}$$

and

$$Corr(Y_{it}, Y_{i,t-k}) = \rho^k \frac{\sqrt{\theta_{i,t-k} + c\theta_{i,t-k}^2}}{\sqrt{\theta_{it} + c\theta_{it}^2}}. \tag{3.12}$$

Using the method of moments and equating

$$\frac{\sum_{i=1}^{I}\sum_{t=1}^{T-1} \tilde{Y_{it}} \tilde{Y_{i(t+1)}}/(T-1)}{\sum_{i=1}^{I}\sum_{t=1}^{T} \tilde{Y_{it}}^2/T} = \rho \frac{\sum_{i=1}^{I}\sum_{t=2}^{T}\sqrt{\frac{\theta_{i,t-1}+c\theta_{i,t-1}^2}{\theta_{it}+c\theta_{it}^2}}}{I(T-1)}. \tag{3.13}$$

Thus

$$\rho = \frac{IT\sum_{i=1}^{I}\sum_{t=1}^{T-1} \tilde{Y_{it}} \tilde{Y_{i,t+1}}}{[\sum_{i=1}^{I}\sum_{t=1}^{T} \tilde{Y_{it}}^2][\sum_{i=1}^{I}\sum_{t=2}^{T}\sqrt{\frac{\theta_{i,t-1}+c\theta_{i,t-1}^2}{\theta_{it}+c\theta_{it}^2}}]}, \tag{3.14}$$

where $\tilde{y_{it}} = \frac{y_{it}-\theta_{it}}{\sqrt{\theta_{it}+c\theta_{it}^2}}$. □

Thus using this value of ρ,

$$C_{i,t,t-k} = \rho^k \frac{\sqrt{\theta_{i,t-k} + c\theta_{i,t-k}^2}}{\sqrt{\theta_{it} + c\theta_{it}^2}}. \tag{3.15}$$

Note that ρ and the correlation matrix $C_{i,t,t-k}$ are obtained using known estimates for $\hat{\beta}$ and $\hat{c}$. The Newton-Raphson algorithm is applied to (3.1) to estimate the regression parameters, i.e.,

$$\begin{pmatrix} \hat{\beta}_{r+1} \\ \hat{c}_{r+1} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_r \\ \hat{c}_r \end{pmatrix} + [\sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1} D_i]_r^{-1} [\sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1} (f_i - \mu_i)]_r. \tag{3.16}$$

The estimators are consistent and under mild regularity conditions, for $I \to \infty$, it may be shown that $I^{\frac{1}{2}}((\hat{\beta}, \hat{c}) - (\beta, c))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I[\sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1} D_i]^{-1}[\sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1}(f_i-\mu_i)(f_i-\mu_i)^T \widetilde{\Sigma_i}^{-1} D_i][\sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1} D_i]^{-1}$ (Sutradhar and Das (1999)).

3.2 The AQIF approach

In this section, the AQIF approach under the non-stationary case is derived such that the basic score vector, the covariance structure and the derivative matrix vary with time.

1.
$$V = (1/I)\sum_{i=1}^{I}(f_i - \mu_i)(f_i - \mu_i)^T, \tag{3.17}$$
where $f_i = (f_{i1}^T, \dots, f_{it}^T, \dots, f_{iT}^T)$, $\mu_i = (\mu_{i1}^T, \dots, \mu_{it}^T, \dots, \mu_{iT}^T)$ are $2T \times 1$ vectors with $f_{it} = (y_{it}, y_{it}^2)$, $\mu_{it} = (\theta_{it}, m_{it})^T$. $\theta_{it} = E(Y_{it})$ and $m_{it} = E(Y_{it}^2) = \theta_{it} + (c+1)\theta_{it}^2$ where $\theta_{it} = \exp(x_{it}^T\beta)$. The steps to estimate the parameters can be summarized as follows:

2.
$$g_I = \begin{bmatrix} \sum_{i=1}^{I} D_i^T(f_i - \mu_i) \\ \sum_{i=1}^{I} \hat{\alpha}^T D_i^T V(f_i - \mu_i) \end{bmatrix}, \tag{3.18}$$
where $\hat{\alpha}$ is the orthogonal vector and
$$D_i = [\partial\mu_i/\partial\beta^T, \partial\mu_i/\partial c] = [D_{i1}^T, \dots, D_{it}^T, \dots, D_{iT}^T]^T,$$
with
$$D_{it} = \begin{pmatrix} \partial\theta_{it}/\partial\beta^T & 0 \\ \partial m_{it}/\partial\beta^T & \partial m_{it}/\partial c \end{pmatrix},$$
where $\partial\theta_{it}/\partial\beta^T = \theta_{it}x_{it}^T$, $\partial m_{it}/\partial\beta^T = \theta_{it}x_{it}^T + 2(c+1)\theta_{it}^2 x_{it}^T$ and $\partial m_{it}/\partial c = \theta_{it}^2$. Following Qu and Lindsay (2003), we rescale the D_i, V and the score vector $(f_i - \mu_i)$ to $A_i^{-\frac{1}{2}}D_i$,$A_i^{-\frac{1}{2}}VA_i^{-\frac{1}{2}}$ and $A_i^{-\frac{1}{2}}(f_i - \mu_i)$, where A_i is a $2T \times 2T$ block diagonal matrix of the form diag$[A_{i1}, \dots, A_{it}, \dots, A_{iT}]$, where
$$A_{it} = \begin{pmatrix} Var(Y_{it}) & \\ & Var(Y_{it}^2) \end{pmatrix},$$
with
$$Var(Y_{it}) = \theta_{it} + c\theta_{it}^2,$$
$$Var(Y_{it}^2) = \theta_{it} + (6+7c)\theta_{it}^2 + (4+16c+12c^2)\theta_{it}^3 + (4c+10c^2+6c^3)\theta_{it}^4.$$

3. Then a GMM is used to construct an objective function given by
$$Q_I(\beta, c) = g_I^T C_I^{-1} g_I, \tag{3.19}$$
where C_I is the sample variance of g_I:
$$\begin{pmatrix} \sum_{i=1}^{I} D_i^T V D_i & (\sum_{i=1}^{I} D_i^T V^2 D_i)\hat{\alpha} \\ \hat{\alpha}^T(\sum_{i=1}^{I} D_i^T V^2 D_i) & \hat{\alpha}^T(\sum_{i=1}^{I} D_i^T V^3 D_i)\hat{\alpha} \end{pmatrix}. \tag{3.20}$$

4. By minimizing this function and applying the Newton-Raphson algorithm, we have

$$\begin{pmatrix} \hat{\beta}_{r+1} \\ \hat{c}_{r+1} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_r \\ \hat{c}_r \end{pmatrix} - [\ddot{Q}_I(\hat{\beta}_r, \hat{c}_r)]^{-1}[\dot{Q}_I(\hat{\beta}_r, \hat{c}_r)], \tag{3.21}$$

where asymptotically, $\dot{Q}_I(\beta, c) = 2\dot{g}_I^T C_I^{-1} g_I$ and $\ddot{Q}_I(\beta, c) = 2\dot{g}_I^T C_I^{-1}\dot{g}_I$, with $\dot{g}_I = [\frac{\partial g_I}{\partial \beta^T}, \frac{\partial g_I}{\partial c}]$ of dimension $(p+2) \times (p+1)$.

The estimators obtained in this way are consistent and asymptotically normal with $((\hat{\beta}, \hat{c}) - (\beta, c))^T \sim$ N$\{0, (E[\dot{g}_I^T]E[C_I^{-1}]E[\dot{g}_I])^{-1}\}$ (Qu and Lindsay, 2003).

4 Simulation Study

We first simulate non-stationary responses by constructing a time-dependent covariate design, where the first covariate is given by

$$x_{it1} = \begin{cases} \text{rbinom(0.5)} & (t = 1, \ldots, T/2), (i = 1, \ldots, I), \\ -\text{rbinom(0.5)} & (t = (T/2) + 1, \ldots, T), (i = 1, \ldots, I). \end{cases}$$

Here $rbinom(0.5)$ is the random binary variable with parameter 0.5, x_{it2} is generated from the standard normal distribution, for $t = 1, \ldots, T$. We generate $T = 4$ correlated negative binomial counts under AR(1) with autocorrelation coefficient $\rho = 0.9$. The true regression parameters are $\beta_0 = \beta_1 = 0.01$. We consider different values of c, and for each structure, we run 10,000 simulations. The following tables provide the simulated mean of the estimates where the entry in brackets represent their corresponding standard errors. The results are shown in Table 1-2.

Table 1: Estimates of parameters and standard errors under non-stationary AR(1) negative binomial model

		$c = 0.05$			$c = 0.2$		
I	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{c}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{c}$
60	JGQL	0.0112	0.0111	0.0502	0.0114	0.0113	0.1765
		(0.1856)	(0.1477)	(0.2805)	(0.2215)	(0.1872)	(0.1878)
	AQIF	0.0117	0.0115	0.04678	0.0118	0.0116	0.1666
		(0.1885)	(0.1490)	(0.3578)	(0.2277)	(0.1901)	(0.2115)
100	JGQL	0.0105	0.0101	0.0511	0.0102	0.0103	0.1976
		(0.1456)	(0.1045)	(0.1976)	(0.1132)	(0.1052)	(0.1256)
	AQIF	0.0105	0.0106	0.0479	0.0101	0.0107	0.1890
		(0.1494)	(0.1076)	(0.2987)	(0.1178)	(0.1055)	(0.1510)
500	JGQL	0.0101	0.0104	0.0498	0.0101	0.0102	0.1999
		(0.0989)	(0.0876)	(0.1678)	(0.0952)	(0.0835)	(0.1038)
	AQIF	0.0103	0.0106	0.0487	0.0102	0.0101	0.1956
		(0.1018)	(0.0894)	(0.2416)	(0.0978)	(0.0867)	(0.1207)
		$c = 0.5$			$c = 1$		
60	JGQL	0.0101	0.0101	0.4896	0.0102	0.0106	1.0162
		(0.1899)	(0.1546)	(0.2313)	(0.2167)	(0.1976)	(0.1267)
	AQIF	0.0105	0.0101	0.4698	0.0105	0.0104	0.9843
		(0.1901)	(0.1568)	(0.3967)	(0.2187)	(0.1982)	(0.1567)
100	JGQL	0.0101	0.0101	0.5015	0.0102	0.0106	1.0012
		(0.1150)	(0.1214)	(0.1556)	(0.1256)	(0.1265)	(0.1045)
	AQIF	0.0101	0.0106	0.4918	0.0106	0.0105	1.0175
		(0.1189)	(0.1281)	(0.2661)	(0.1292)	(0.1299)	(0.1442)
500	JGQL	0.0105	0.0101	0.4996	0.0101	0.0101	0.9989
		(0.0129)	(0.0137)	(0.0110)	(0.0578)	(0.0465)	(0.1015)
	AQIF	0.0104	0.0106	0.4984	0.0101	0.0101	0.9981
		(0.0131)	(0.0142)	(0.0221)	(0.0585)	(0.0474)	(0.1276)

Table 2: Estimates of parameters and standard errors under non-stationary AR(1) negative binomial model

		$c = 1.75$			$c = 3$		
I	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{c}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{c}$
60	JGQL	0.0105	0.0101	1.7386	0.0101	0.0101	2.8812
		(0.2143)	(0.1786)	(0.2678)	(0.2078)	(0.1988)	(0.2123)
	AQIF	0.0105	0.0105	1.7254	0.0101	0.0104	2.7892
		(0.2159)	(0.1790)	(0.3327)	(0.2089)	(0.2009)	(0.2879)
100	JGQL	0.0101	0.0101	1.7378	0.0102	0.0101	3.0132
		(0.0985)	(0.0995)	(0.1154)	(0.0983)	(0.0948)	(0.1034)
	AQIF	0.0106	0.0105	1.7276	0.0101	0.0101	2.8769
		(0.1001)	(0.1012)	(0.1365)	(0.1012)	(0.0954)	(0.1323)
500	JGQL	0.0105	0.0102	1.7492	0.0101	0.0101	2.8459
		(0.0367)	(0.0289)	(0.0899)	(0.0174)	(0.0189)	(0.0712)
	AQIF	0.0101	0.0101	1.7619	0.0101	0.0104	2.7861
		(0.0382)	(0.0311)	(0.0918)	(0.0181)	(0.0199)	(0.0852)

To estimate the regression and over-dispersion parameters, we assume small initial values for both parameters in both methods. We note the same pattern in the efficiency between the two methods. JGQL has shown slightly more efficient results than AQIF,i.e, for $I = 60$, we note large standard errors for the over-dispersion parameter in the AQIF approach while for JGQL, the standard errors for over-dispersion parameter are comparatively lower. In fact, as the number of clusters increases, the standard errors under both methods decrease but JGQL gains more efficiency irrespective of the values for c. For almost all number of clusters, we note that for $c = 0.2, 0.5, 1$, AQIF yields bigger standard errors for the over-dispersion parameter. As the number of clusters increases, the standard errors decrease but still, JGQL shows more efficiency. As for the regression estimates, JGQL yields slightly more efficient results than AQIF for almost all cluster sizes. For $I = 60$ and $c = 0.05$, JGQL has yielded approximately 2250 non-convergent simulations while AQIF fail in 1990 simulations respectively. As c increases to 1.75 and when $I = 60$, the number of non-convergent simulations increases to 3125 in the JGQL approach and for larger c, there is a further increase in the non-convergent simulations of the JGQL approach. However, for the AQIF approach, these number of non-convergent simulations are lesser. AQIF has failed in 1800 simulations. However, AQIF yields unreliable estimates of the over-dispersion parameter in many simulations. We note that, for large c, AQIF performs extremely slow. This is justified by the high number of flop counts. The average correlation estimates for each cluster under each case are provided in Table 3.

Table 3: Estimates of the non-stationary AR(1) correlations

c	I	$\hat{\rho_{12}}$	$\hat{\rho_{13}}$	$\hat{\rho_{14}}$	$\hat{\rho_{23}}$	$\hat{\rho_{24}}$	$\hat{\rho_{34}}$
0.05	60	0.9378	0.7867	0.6956	0.9276	0.7652	0.9256
	100	0.9245	0.7897	0.7267	0.9116	0.8256	0.9156
	500	0.9156	0.7992	0.6876	0.9124	0.8245	0.8997
0.2	60	0.9345	0.7892	0.6725	0.9242	0.7951	0.9166
	100	0.9232	0.7967	0.7275	0.9245	0.8243	0.9347
	500	0.9234	0.8124	0.7166	0.9256	0.7998	0.9431
0.5	60	0.8932	0.7867	0.7014	0.8876	0.7854	0.8917
	100	0.9152	0.8476	0.7286	0.9285	0.7999	0.9156
	500	0.8999	0.7967	0.7551	0.9098	0.8176	0.9101
1	60	0.8967	0.7816	0.7156	0.9256	0.8244	0.9267
	100	0.8956	0.8498	0.6987	0.9135	0.8246	0.9236
	500	0.8984	0.8136	0.7474	0.9187	0.7998	0.9076
1.75	60	0.9076	0.8267	0.6987	0.8998	0.7885	0.9156
	100	0.9132	0.8278	0.6891	0.9143	0.7982	0.8985
	500	0.9015	0.8187	0.7126	0.9126	0.7886	0.8842
3	60	0.8945	0.8278	0.6967	0.9257	0.8231	0.9243
	100	0.9032	0.8176	0.6997	0.8965	0.7864	0.8997
	500	0.9102	0.8287	0.6874	0.8997	0.7865	0.8897

5 Application

The epileptic seizures data from Thall and Vail (1990) consists of the number of seizures occurring at each of four successive two weekly clinic visits for 59 patients. The summary statistics for these responses are given in Table 4.

Table 4: Summary statistics of the epileptic seizure counts

	Visit 1	Visit 2	Visit 3	Visit 4
Sample mean	8.949	8.356	8.441	7.305
Sample variance	220.084	103.785	200.182	93.112

We notice that the variances are greater than their corresponding means for each time point indicating that the data are clearly over-dispersed. Information on their age, gender and mode of treatment are also available. We thus consider five covariates namely the intercept term denoted by x_{it1}, the treatment parameter x_{it2}, coded as 0 for placebo and 1 for progabide, the baseline-seizure rates denoted by x_{it3}, the age of the person denoted by x_{it4} and the interaction effect between treatment and baseline seizure rates x_{it5}. The mean parameter of the negative binomial distribution for the ith person is given by $\theta_i = \exp(x_i^T \beta)$ with $x_i = (x_{it1}, x_{it2}, \ldots, x_{it5})^T$ for $t = 1, \ldots, 4$. Here β is the 5×1 vector of regression parameters. Jowaheer and Sutradhar (2002) used the negative binomial model and estimated the model parameters using the true general autocorrelation structure based JGQL approach. We analyze the epileptic data assuming that the counts follow the unstructured covariance matrix and estimate the parameters using the AQIF approach.

Table 5: Epileptic data:estimates of the regression and over-dispersion parameters using JGQL and AQIF approaches

Method	INTC	TR	BR	Age	INTA	$\hat{c}$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
$JGQL$	0.4582	-0.2471	0.0027	0.0210	0.0011	0.5142	0.5222	0.3371	0.2030
	(0.4321)	(0.1521)	(0.0040)	(0.0109)	(0.0048)	(0.3121)			
$AQIF$	0.4221	-0.2671	0.0056	0.0222	0.0012	0.5431			
	(0.4372)	(0.1603)	(0.0055)	(0.0111)	(0.0056)	(0.3914)			

We choose small starting values for the longitudinal correlations and small positive values for the regression and over-dispersion parameters. To obtain JGQL estimates, we apply the iterative equation (3.16) and calculate $\hat{\rho}$ using equation (3.14). To obtain AQIF estimates, we use the iterative equation (3.21). However, we note that in AQIF some starting values yield non-convergent estimates as shown in simulations. Some starting values also lead to computational difficulties in calculating the inverse of the covariance matrix in the JGQL approach and the hessian matrix in the AQIF approach. Between these two techniques, JGQL works slightly faster and converges in less iterates compared to AQIF.

The autocorrelation values under JGQL are large, indicating high longitudinal correlations. The values of c under both methods justify that the data are over-dispersed. The treatment parameter in both methods is negative which indicate that the predicted seizure counts will be less in the treatment group than in the placebo group. The age factor is positive indicating that as age increases, the patients are more likely to obtain more epileptic attacks. The interaction between the treatment and the baseline seizure rate does not appear to be significant in both methods. Based on the standard errors, we note that the estimates of the parameters under the JGQL approach are more efficient than AQIF especially for the over-dispersion parameter.

6 Conclusion

This paper proposes the longitudinal INAR(1) model with NB marginal, based on a random thinning procedure. The estimation of the model parameters is conducted using the JGQL and AQIF approaches. Simulation results show that the JGQL yield lower standard errors and estimates with lower bias under non-stationary set-up. The proposed model is applied to analyze the epileptic seizures data and yields reliable estimates of the different model parameters. It is also noticeable that the JGQL provides more efficient estimates than the AQIF approach. Thus, the longitudinal INAR(1) with NB marginal is a commendable model and we conclude that for such setting the JGQL is a reliable inferential estimation approach.

Appendix

A Detailed Derivation

$$\begin{aligned} E(Y_{it}) &= E_{Y_{i,t-1}} E_{\alpha_{it}} E(\alpha_{it} * y_{i,t-1} | Y_{i,t-1}, \alpha_{it}) + E(d_{it}) \\ &= E(\rho Y_{i,t-1}) + E(d_{it}) = \theta_{it}. \end{aligned} \tag{A.1}$$

$$\begin{aligned} Var(Y_{it}) &= Var(\alpha_{it} * Y_{i,t-1}) + Var(d_{it}) \\ &= Var_{Y_{i,t-1}}[E_{\alpha_{it}} E(\alpha_{it} * Y_{i,t-1} | Y_{i,t-1}, \alpha_{it})] + E_{Y_{i,t-1}}[Var_{\alpha_{it}}(E(\alpha_{it} * Y_{i,t-1}) | Y_{i,t-1}, \alpha_{it}) \\ &+ E(Var(Y_{it} | Y_{i,t-1}, \alpha_{it}))] + Var(d_{it}) \\ &= Var[E_{\alpha_{it}}(\alpha_{it} Y_{i,t-1})] + E[Var_{\alpha_{it}}(\alpha_{it} Y_{i,t-1}) + E_{\alpha_{it}}(Y_{it} \alpha_{it}(1 - \alpha_{it}))] + Var(d_{it}) \\ &= Var(\rho Y_{i,t-1}) + E[\frac{\rho(1-\rho)c}{1+c} Y_{i,t-1}^2 + \rho Y_{it} - Y_{it}(\frac{\rho(1-\rho)c}{1+c} + \rho^2)] + Var(d_{it}) \\ &= \rho^2(\theta_{i,t-1} + [c\theta_{i,t-1}^2 + \frac{\rho(1-\rho)c}{1+c}(\theta_{i,t-1} + (c+1)\theta_{i,t-1}^2) + \rho\theta_{it} - \theta_{it}(\frac{\rho(1-\rho)c}{1+c} + \rho^2)] \\ &+ (\theta_{it} - \rho\theta_{i,t-1}) + c(\theta_{it}^2 - \rho\theta_{i,t-1}^2) \\ &= \theta_{it}(1 + c\theta_{it}). \end{aligned} \tag{A.2}$$

$$\begin{aligned} E(Y_{it} Y_{i,t-1}) &= E_{Y_{i,t-1}}[Y_{i,t-1} E(Y_{it} | Y_{i,t-1})] \\ &= E_{Y_{i,t-1}}[Y_{i,t-1} E_{\alpha_{it}} E(Y_{it} | Y_{i,t-1}, \alpha_{it})] \\ &= E_{Y_{i,t-1}}[Y_{i,t-1}(\rho Y_{i,t-1} + \theta_{it} - \rho\theta_{i,t-1})] \\ &= \rho(\theta_{i,t-1} + c\theta_{i,t-1}^2 + \theta_{i,t-1}^2) + \theta_{it}\theta_{i,t-1} - \rho\theta_{i,t-1}^2 \\ &= \rho(\theta_{i,t-1} + c\theta_{i,t-1}^2) + \theta_{it}\theta_{i,t-1}. \end{aligned} \tag{A.3}$$

$$\begin{aligned} E(Y_{it} Y_{i,t-2}) &= E_{Y_{i,t-2}} E_{Y_{i,t-1}}[Y_{i,t-2} E(Y_{it} | Y_{i,t-1}, Y_{it-2})] \\ &= E_{Y_{i,t-2}} E_{Y_{i,t-1}}[Y_{i,t-2} E_{\alpha_{it}} E(Y_{it} | Y_{i,t-1}, Y_{i,t-2}, \alpha_{it})] \\ &= E_{Y_{i,t-2}}[Y_{i,t-2}(\rho E_{Y_{i,t-1}}(Y_{i,t-1} | Y_{i,t-2}) + (\theta_{it} - \rho\theta_{i,t-1}))] \\ &= \rho^2(\theta_{i,t-2} + c\theta_{i,t-2}^2 + \theta_{i,t-2}^2) + \theta_{i,t-1}\theta_{i,t-2} - \rho^2\theta_{i,t-2}^2 \\ &= \rho^2(\theta_{i,t-2} + c\theta_{i,t-2}^2) + \theta_{i,t-1}\theta_{i,t-2}. \end{aligned} \tag{A.4}$$

$$Cov(Y_{it}, Y_{i,t-1}) = E[Y_{it}Y_{i,t-1}] - E[Y_{it}]E[Y_{i,t-1}] = \rho(\theta_{i,t-1} + c\theta^2_{i,t-1}). \tag{A.5}$$

$$Cov(Y_{it}, Y_{i,t-2}) = E[Y_{it}Y_{i,t-2}] - E[Y_{it}]E[Y_{i,t-2}] = \rho^2(\theta_{i,t-2} + c\theta^2_{i,t-2}). \tag{A.6}$$

Thus

$$Cov(Y_{it}, Y_{i,t-k}) = \rho^k(\theta_{i,t-k} + c\theta^2_{i,t-k}). \tag{A.7}$$

The lag-k correlation

$$Corr(Y_{it}, Y_{i,t-k}) = \frac{Cov(Y_{it}, Y_{i,t-k})}{\sqrt{\theta_{it} + c\theta^2_{it}}\sqrt{\theta_{i,t-k} + c\theta^2_{i,t-k}}} = \rho^k \frac{\sqrt{(}\theta_{i,t-k} + c\theta^2_{i,t-k})}{\sqrt{\theta_{it} + c\theta^2_{it}}}, \tag{A.8}$$

for $k = 1, \ldots, T-1$.

References

Adachi, P. and Willoughby, T. (2015), “Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science,” *European Journal of Developmental Psychology*, 12, 116–128.

Al-Osh, M. and Aly, E. (1992), “First order autoregressive time series with negative binomial and geometric marginals,” *Commun. Statist-Theory Meth*, 21, 2483–2492.

Al-Zaid, A. and Al-Osh, M. (1990), “An integer-valued pth order autoregressive structure (INAR(P)) process,” *Journal of Applied Probability*, 27, 314–324.

Alosh, M. (2010), “Modeling longitudinal count data with dropouts,” *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, 9, 35–45.

Crowder, M. (1995), “On the use of a working correlation matrix in using generalized linear models for repeated measures,” *Biometrika*, 82, 407–10.

Jowaheer, V. and Sutradhar, B. (2002), “Analysing longitudinal count data with overdipsersion,” *Biometrika*, 89, 389–399.

Jowaheer, V. and Sutradhar, B. (2005), “Fitting lower order nonstationary autocorrelation models to the time series of Poisson counts,” *Transactions on Mathematics*, 4, 427–434.

Liang, K. and Zeger, S. (1986), “Longitudinal data analysis using generalized linear models,” *Biometrika*, 73, 13–22.

Lipsitz, S. and Fitzmaurice, G. (2008), “Generalized estimation equations for longitudinal data analysis,” *Longitudinal data analysis. New York: Chapman & Hall/CRC*, 43–78.

McKenzie, E. (1986), “Autoregressive moving-average processes with negative binomial and geometric marginal distrbutions,” *Advanced Applied Probability*, 18, 679–705.

Oyet, A. J. and Sutradhar, B. C. (2013), "Longitudinal modeling of infectious disease," *Sankhya B*, 75, 319–342.

— (2021), "Analyzing Unevenly Spaced Longitudinal Count Data," *Sankhya B*, 83, 342–373.

Qu, A. and Lindsay, B. (2003), "Building adaptive estimating equations when inverse of covariance estimation is difficult," *Journal of Royal Statistical Society*, 65, 127–142.

Sutradhar, B. and Das, K. (1999), "On the efficiency of regression estimators in generalised linear models for longitudinal data," *Biometrika*, 86, 459–65.

Sutradhar, B. C. (2011), "Familial Longitudinal Models for Count Data," *Dynamic Mixed Models for Familial Longitudinal Data*, 423–454.

Sutradhar, B. C. and Rao, R. P. (2016), "Inferences in Longitudinal Count Data Models with Measurement Errors in Time Dependent Covariates," *Sankhya B*, 78, 39–65.

Tang, W., He, H., and Tu, X. M. (2023), *Applied categorical and count data analysis*, CRC Press.

Thall, P. and Vail, S. (1990), "Some covariance models for longitudinal count data with overdispersion," *Biometrika*, 46, 657–71.

Zhang, H., Yu, Q., Feng, C., Gunzler, D., Wu, P., and Tu, X. (2012), "A new look at the difference between the GEE and the GLMM when modeling longitudinal count responses," *Journal of Applied Statistics*, 39, 2067–2079.

Zhao, Y., Qin, L., Zhang, P., Li, K., Liang, L., Sun, J., Xu, B., Dai, Y., Li, X., Zhang, C., et al. (2020), "Longitudinal COVID-19 profiling associates IL-1RA and IL-10 with disease severity and RANTES with mild disease," *JCI insight*, 5.

Received: February 4, 2024

Accepted: April 9, 2024