A robust algorithm for estimating regression and dispersion parameters in non-stationary longitudinally correlated Com–Poisson data

Naushad Mamode Khan

Abstract

In recent years, Com–Poisson has emerged as one of the most popular discrete models in the analysis of count data owing to its flexibility in handling different types of dispersion. However, in a stationary longitudinal Com–Poisson count data set-up where the covariates are time independent, estimation of regression and dispersion parameters based on a generalized quasi-likelihood (GQL) approach involves some major computational difficulties particularly in the inversion of the joint covariance matrix. On the other hand, in practical real-life longitudinal studies, time-dependent covariates leading to non-stationary responses are more frequently encountered. This implies that further computational problems will now arise when estimating parameters under non-stationary set-ups. This paper overcomes this problem by approximating the inverse of the ill-conditioned covariance matrix in the GQL approach through a multidimensional conjugate gradient method. The performance of this novel version of the GQL approach is then assessed on simulations of AR(1) stationary and AR(1) non-stationary longitudinal Com–Poisson counts and on real-life epileptic seizure counts. However, there is not yet an algorithm to generate non-stationary longitudinal Com–Poisson counts nor a GQL algorithm to estimate the parameters under non-stationary set-ups. Thus, the paper also provides a framework to generate non-stationary AR(1) Com–Poisson counts along with the construction of a GQL equation under non-stationary set-ups.

1. Introduction

In practice, correlated count responses are either over- or underdispersed, thus limiting the application of the Poisson model. In the literature, such data can be modelled through some weighted or generalized form of the Poisson distributions [1], but these models do not belong to the exponential family of distributions. Recently, Shmueli *et al.* [7] revived the two-parameter Com–Poisson distribution (CMP), which satisfies the properties of the generalized linear models and can model all forms of dispersion. Owing to its elegant properties, Mamode Khan and Jowaheer [4] extended the cross-sectional CMP to longitudinal CMP by first developing its AR(1) Gaussian autocorrelation structure based on time-independent covariates followed by the construction of the generalized quasi-likelihood (GQL) that was based on a general autocorrelation structure developed by Sutradhar and Das [8]. Their GQL approach yields reliable and efficient estimates of the regression and dispersion parameters under both simulated AR(1) longitudinal stationary Com–Poisson data and real-life data applications. However, a huge number of non-convergent simulations were reported. Upon investigation, it was revealed that the covariance structure in the GQL technique was ill conditioned. Unfortunately, this computational deficiency limits the application of the GQL approach. Besides, these authors have developed the AR(1) generating process and the GQL equation based only on time-independent covariates, whilst, in many real-life studies, especially in medicine, time-varying covariates such as age, treatment, number of medical check-ups and attacks are commonly encountered. In fact, Mamode Khan and Jowaheer [4] did consider

time-dependent covariates in the analysis of breakdown data but the values of these covariates were averaged over a single time point to make their GQL work.

Non-stationarity through the inclusion of time-dependent covariates based on the Com–Poisson model has not yet been dealt with. This paper proposes a reconstruction of the GQL estimating equation based on non-stationary longitudinal Com–Poisson counts. The performance of this GQL equation is then assessed on simulation studies and on real-life data applications. Note that there is not yet an algorithm to generate non-stationary AR(1) longitudinal Com–Poisson counts. Thus, we also develop an algorithm to generate such types of counts based on the binomial thinning operation. On the other hand, the computational challenges in the estimation of parameters under the assumption of time-dependent covariates can already be foreseen given the performance of the GQL approach under stationary set-ups. To address this issue, this paper comes forward with the multiconjugate gradient algorithm to approximate the inverse of the covariance structure in the GQL approach and thereafter uses the Newton–Raphson technique to estimate jointly the regression and dispersion parameters.

The organization of the paper is as follows: in the next section, an overview of the longitudinal CMP model is presented. In §3, the derivation of AR(1) non-stationary Gaussian autocorrelation structure is performed along with its generating process. This is followed by a review of the GQL estimating equation under stationary set-up and the construction of GQL under non-stationarity. In the same section, the multidimensional conjugate gradient method (MCGM) approach is implemented. In §6, a simulation study of AR(1) stationary and non-stationary cases of the longitudinal CMP counts is carried out under various covariate designs and values of dispersion that encompass equi-, over- or underdispersion. The GQL approach based on Newton–Raphson and MCGM is then used to estimate the regression and dispersion parameters using these simulated data. This section also comprises a comparison between Newton–Raphson based on the ill-conditioned covariance matrix with Newton–Raphson based on the MCGM approach. In §7, both versions of GQL are applied to epileptic seizure counts. The conclusion is presented in the last section.

2. Overview of the longitudinal CMP model

For our discussion, let y_{it} be a count response and x_{it} be a p-dimensional vector of covariates for subject i $(i = 1, \ldots, I)$ observed at time t $(t = 1, \ldots, T)$. Let β be the $p \times 1$ vector of regression parameters. For the ith subject, let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})^T$ be the $T \times 1$ response vector and $X_i = (x_{i1}, \ldots, x_{iT})^T$ be the $T \times p$ matrix of covariates. The Com–Poisson density function of y_{it} is given by

$$f(y_{it}) = \frac{\lambda_{it}^{y_{it}}}{(y_{it}!)^\nu} \frac{1}{Z(\lambda_{it}, \nu)}, \tag{1}$$

where

$$Z(\lambda_{it}, \nu) = \sum_{j=0}^{\infty} \frac{\lambda_{it}^j}{(j!)^\nu}, \tag{2}$$

where $\lambda_{it} = \exp(x_{it}^T \beta)$ in (1); the parameter ν corresponds to the dispersion index. More specifically, the values $\nu = 1$, $\nu < 1$ and $\nu > 1$ correspond to equi-, over- and underdispersion. The moments of the CMP model derived by Shmueli et al. [7] are given by

$$E(Y_{it}) = \theta_{it} = \lambda_{it}^{1/\nu} - \frac{\nu - 1}{2\nu}, \quad \text{Var}(Y_{it}) = \frac{\theta_{it}}{\nu} + \frac{\nu - 1}{2\nu^2}. \tag{3}$$

Thus, $Y_{it} \sim \text{CMP}(\theta_{it}/\nu, \nu)$. Under the assumption of stationarity, that is, time-independent covariates, for $t = 1, 2, \ldots, T$,

$$\theta_{i1} = \theta_{i2} = \ldots = \theta_{iT}, \tag{4}$$

whereas, for time-dependent covariates, θ_{it} is different at each time point. Since the y_{it} are repeatedly recorded over time $t = 1, \ldots, T$, the T repeated responses under the ith cluster are more likely to be correlated. Note that under a stationary set-up, the correlation structure is constant for all individuals, whereas, for a non-stationary structure,

$$
C_i(\rho) = \begin{bmatrix}
1 & \rho_{i,12} & \rho_{i,13} & \cdots & \rho_{i,1T} \\
\rho_{i,21} & 1 & \rho_{i,23} & \cdots & \rho_{i,2T} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\rho_{i,T1} & \rho_{i,T2} & \rho_{i,T3} & \cdots & 1
\end{bmatrix},
\tag{5}
$$

where $\rho_{i,lm}$ is the correlation coefficient for the ith individual at lag $k = |l - m|$.

3. Derivation of the AR(1) Gaussian non-stationary Com–Poisson longitudinal counts

In this section, we provide the framework to generate sequences of dependent count responses following the CMP distribution based on the assumption of time-dependent covariates analogous to the dependent Com–Poisson counts generated under stationary set-ups. Here, the correlation matrix is not unique for all the I clusters, that is, $C_i(\rho) \neq C(\rho)$.

We assume, for $t = 1, \ldots, T$, $y_{i,t-1} \sim \text{CMP}(\theta_{i,t-1}/\nu, \nu)$ and $d_{it} \sim \text{CMP}((\theta_{it} - \rho\theta_{i,t-1})/\tilde{\nu}, \tilde{\nu})$, where

$$
\tilde{\nu} = \frac{q_0 + \sqrt{q_0^2 - 8q_1}}{4q_1}
\tag{6}
$$

with $q_0 = (2\theta_{it} - 2\rho\theta_{i,t-1} + 1)$, $q_1 = (1 - \rho^2)((\nu - 1)/2\nu^2) + \theta_{it}/\nu - (\rho\theta_{i,t-1}/\nu)(\rho + \nu(1 - \rho))$ and $0 < \rho < 1$. Then, following McKenzie [5] and Mamode Khan and Jowaheer [4], the non-stationary AR(1) model for the sequence $\{y_{it}\}$ can be formulated as

$$
y_{it} = \rho * y_{i,t-1} + d_{it},
\tag{7}
$$

where $0 < \rho < 1$. The symbol $*$ indicates the binomial convolution thinning operation such that

$$
\rho * y_{i,t-1} \mid y_{i,t-1} \sim \text{Binomial}(y_{i,t-1}, \rho).
\tag{8}
$$

Under these conditions, it can be shown that

$$
E(Y_{it}) = \theta_{it}, \quad \text{Var}(Y_{it}) = \frac{\theta_{it}}{\nu} + \frac{\nu - 1}{2\nu^2}.
\tag{9}
$$

Similarly,

$$
\text{Cov}(Y_{it}, Y_{i,t-k}) = \rho\left(\frac{\theta_{i,t-k}}{\nu} + \frac{\nu - 1}{2\nu^2}\right)
$$

and thus the lag-k correlation is

$$
\text{Corr}(Y_{it}, Y_{i,t-k}) = \rho^k \frac{\sqrt{(\theta_{i,t-k}/\nu + (\nu - 1)/2\nu^2)}}{\theta_{i,t}/\nu + (\nu - 1)/2\nu^2} = \rho^k \frac{\sqrt{\lambda_{i,t-k}^{1/\nu}/\nu}}{\sqrt{\lambda_{it}^{1/\nu}/\nu}}
\tag{10}
$$

for $k = 1, \ldots, T - 1$. Note that since the mean parameter $(\theta_{it} - \rho\theta_{i,t-1}) > 0$ and ρ is a probability parameter, $0 < \rho < \min(\theta_{i2}/\theta_{i1}, \ldots, \theta_{it}/\theta_{i,t-1}, 1)$.

4. *GQL under non-stationarity*

In this section, we review the joint generalized quasi-likelihood estimating equation (GQL-I) technique proposed by Mamode Khan and Jowaheer [4] to estimate the regression and dispersion parameters in the model for stationary Com–Poisson longitudinal count data and use the true general autocorrelation structure suggested by Sutradhar and Das [8] to estimate the regression parameters and the working multivariate normality assumption of Prentice and Zhao [6] to estimate the dispersion parameters. This GQL-I estimating equation is constructed using observations y_{it} and their squares y_{it}^2 and is given by

$$f(\hat{\beta}, \hat{\nu}) = \sum_{i=1}^{I} D_i^T \widetilde{\Sigma_i}^{-1} (f_i - \mu_i) = 0, \tag{11}$$

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 $f_{it} = (y_{it}, y_{it}^2)$, $\mu_{it} = (\theta_{it}, m_{it})^T$, where $m_{it} = \lambda_{it}^{1/\nu}/\nu + \theta_{it}^2$. The gradient matrix is given by

$$D_i = [\partial \mu_i / \partial \beta^T, \partial \mu_i / \partial \nu] = [D_{i1}^T, \ldots, D_{it}^T, \ldots, D_{iT}^T]^T \tag{12}$$

with

$$D_{it} = \begin{bmatrix} \partial \theta_{it}/\partial \beta^T & \partial \theta_{it}/\partial \nu \\ \partial m_{it}/\partial \beta^T & \partial m_{it}/\partial \nu \end{bmatrix}$$

for $t = 1, \ldots, T$.

The covariance matrix of f_i is expressed as

$$\widetilde{\Sigma_i} = \begin{bmatrix} \widetilde{\Sigma_{i1}} & \widetilde{\Omega_{i12}} & \widetilde{\Omega_{i13}} & \ldots & \widetilde{\Omega_{i1T}} \\ & \widetilde{\Sigma_{i2}} & \widetilde{\Omega_{i23}} & \ldots & \widetilde{\Omega_{i2T}} \\ & & \widetilde{\Sigma_{i3}} & \ldots & \widetilde{\Omega_{i3T}} \\ & & & \ddots & \\ & & & & \widetilde{\Sigma_{iT}} \end{bmatrix}, \tag{13}$$

where the diagonal submatrix

$$\widetilde{\Sigma_{it}} = \begin{bmatrix} \mathrm{Var}(Y_{it}) & \mathrm{Cov}(Y_{it}, Y_{it}^2) \\ & \mathrm{Var}(Y_{it}^2) \end{bmatrix}$$

and, for $t \neq w$, the off-diagonal submatrix

$$\widetilde{\Omega_{itw}} = \begin{bmatrix} \mathrm{Cov}(Y_{it}, Y_{iw}) & \mathrm{Cov}(Y_{it}, Y_{iw}^2) \\ \mathrm{Cov}(Y_{it}^2, Y_{iw}) & \mathrm{Cov}(Y_{it}^2, Y_{iw}^2) \end{bmatrix}$$

for $t = 1, \ldots, T$ and $w = 1, \ldots, T$. Note that for time-independent covariates, these components are computed at a single time point only and the same expressions are used for $t = 1, 2, \ldots, T$. For more details on the stationary case, refer to Mamode Khan and Jowaheer [4].

For the non-stationary case,

$$\partial \theta_{it}/\partial \beta^T = \frac{\lambda_{it}^{1/\nu}}{\nu} x_{it}^T,$$

$$\partial \theta_{it}/\partial \nu = \frac{1}{2}\frac{\nu - 1}{\nu^2} - \frac{1}{2\nu} - \frac{\lambda_{it}^{1/\nu} x_{it}^T \beta}{\nu^2},$$

$$\partial m_{it}/\partial \beta^T = x_{it}^T \left(\frac{2\lambda_{it}^{1/\nu} + 2\nu\lambda_{it}^{2/\nu} - \nu\lambda_{it}^{1/\nu}}{\nu^2} \right),$$

$$\partial m_{it}/\partial \nu = \frac{1}{2\nu^3}[2\lambda_{it}^{1/\nu}\nu\ln(\lambda_{it}) + \nu - 1 - 4\lambda_{it}^{2/\nu}\ln(\lambda_{it})\nu - 4\lambda_{it}^{1/\nu}\nu - 4\lambda_{it}^{1/\nu}\ln(\lambda_{it})].$$

Next, the construction of the covariance matrix $\widetilde{\Sigma_i}$ is performed in two phases.

4.1. Computing the block diagonal elements of $\widetilde{\Sigma_{it}}$

Using the moment generating formula of the CMP function [**4**, **7**],

$$\mathrm{Cov}(Y_{it}, Y_{it}^2) = \frac{2\lambda_{it}^{1/\nu} + 2\nu\lambda_{it}^{2/\nu} - \nu\lambda_{it}^{1/\nu}}{\nu^2},$$

$$\mathrm{Var}(Y_{it}^2) = \frac{\lambda_{it}^{1/\nu}\nu^2 + 4\lambda_{it}^{3/\nu}\nu^2 + 10\lambda_{it}^{2/\nu}\nu - 4\lambda_{it}^{1/\nu}\nu + 4\lambda_{it}^{1/\nu} - 4\lambda_{it}^{2/\nu}\nu^2}{\nu^3}.$$

4.2. Computing the block off-diagonal elements of $\widetilde{\Sigma_{it}}$

Following Jowaheer and Sutradhar [**2**] and Mamode Khan and Jowaheer [**4**], the off-diagonal covariance entries are derived by approximating the higher-order moments based on the assumption that $(y_{i1}, y_{i2}, \ldots, y_{iT}))$ follow a multivariate normal distribution. By letting $\sigma_{itw} = \rho_{itw}\sigma_{itt}\sigma_{iww}$ with $\sigma_{itt} = \sqrt{\lambda_{it}^{1/\nu}/\nu}$, we obtain

$$\mathrm{Cov}(Y_{it}, Y_{iw}) = \rho_{itw}\sqrt{\left(\frac{\lambda_{it}^{1/\nu}}{\nu}\right)\left(\frac{\lambda_{iw}^{1/\nu}}{\nu}\right)}.$$

As for $\mathrm{Cov}(Y_{it}, Y_{iw}^2)$ and $\mathrm{Cov}(Y_{it}^2, Y_{iw}^2)$, we require the two higher-order moments $E(Y_{it}Y_{iw}Y_{iw'})$ and $E(Y_{it}Y_{it'}Y_{iw}Y_{iw'})$ that were derived by Mamode Khan and Jowaheer [**4**]. In this paper, we extend their formulations to obtain

$$\mathrm{Cov}(Y_{it}, Y_{iw}^2) = 2\left(\rho_{itw}\sqrt{\left(\frac{\lambda_{it}^{1/\nu}}{\nu}\right)\left(\frac{\lambda_{iw}^{1/\nu}}{\nu}\right)} + \theta_{it}\theta_{iw}\right)\theta_{iw} - 2\theta_{it}\theta_{iw}^2, \tag{14}$$

$$\begin{aligned}
\mathrm{Cov}(Y_{it}^2, Y_{iw}^2) = {} & \sqrt{\frac{\lambda_{it}^{1/\nu}}{\nu}}\sqrt{\frac{\lambda_{iw}^{1/\nu}}{\nu}} + 2\rho_{itw}^2\left(\frac{\lambda_{it}^{1/\nu}}{\nu}\right)\left(\frac{\lambda_{iw}^{1/\nu}}{\nu}\right) \\
& + 2\left(\left(\theta_{it} + \frac{\lambda_{it}^{1/\nu}}{\nu}\right)\theta_{iw} + 2\left(\rho_{itw}\sqrt{\frac{\lambda_{it}^{1/\nu}}{\nu}}\sqrt{\frac{\lambda_{iw}^{1/\nu}}{\nu}} + \theta_{it}\theta_{iw}\right)\theta_{it} - 2\theta_{it}^2\theta_{iw}\right)\theta_{iw} \\
& + 2\left(2\left(\rho_{itw}\sqrt{\frac{\lambda_{it}^{1/\nu}}{\nu}}\sqrt{\frac{\lambda_{iw}^{1/\nu}}{\nu}} + \theta_{it}\theta_{iw}\right)\theta_{iw} + (\theta_{iw} + c\theta_{iw}^2 + \theta_{iw}^2)\theta_{it} - 2\theta_{it}\theta_{iw}^2\right)\theta_{it} \\
& - \left(\frac{\lambda_{it}^{1/\nu}}{\nu} + \theta_{it}^2\right)\theta_{iw}^2 - 4\left(\rho_{itw}\sqrt{\frac{\lambda_{iw}^{1/\nu}}{\nu}}\sqrt{\theta_{iw} + c\theta_{iw}^2} + \theta_{it}\theta_{iw}\right)\theta_{it}\theta_{iw} \\
& - \left(\theta_{iw} + \frac{\lambda_{iw}^{1/\nu}}{\nu}\right)\theta_{it}^2 + 3\theta_{it}^2\theta_{iw}^2 - \left(\theta_{it} + \frac{\lambda_{it}^{1/\nu}}{\nu}\right)\left(\theta_{iw} + \frac{\lambda_{iw}^{1/\nu}}{\nu}\right).
\end{aligned}$$

In comparison with the stationary model, here the covariances are different between the counts for the same lags apart and for different individuals. To estimate the AR(1) correlation parameters we use a moment estimating equation approach following Jowaheer

and Sutradhar [**3**]. Using the fact that

$$\text{Corr}(Y_{it}, Y_{i,t-k}) = \rho^k \frac{\sqrt{\lambda_{i,t-k}^{1/\nu}/\nu}}{\sqrt{\lambda_{it}^{1/\nu}/\nu}} \tag{15}$$

[refer to §1] and equating

$$\frac{\sum_{i=1}^I \sum_{t=1}^{T-1} \tilde{Y}_{it} \tilde{Y}_{i(t+1)}/(T-1)}{\sum_{i=1}^I \sum_{t=1}^T \tilde{Y}_{it}^2/T} \tag{16}$$

with

$$\rho \frac{\sum_{i=1}^I \sum_{t=2}^T \sqrt{(\lambda_{i,t-1}^{1/\nu}/\nu)/(\lambda_{it}^{1/\nu}/\nu)}}{I(T-1)}, \tag{17}$$

we obtain

$$\rho = \frac{IT \sum_{i=1}^I \sum_{t=1}^{T-1} \tilde{Y}_{it} \tilde{Y}_{i,t+1}}{[\sum_{i=1}^I \sum_{t=1}^T \tilde{Y}_{it}^2][\sum_{i=1}^I \sum_{t=2}^T \sqrt{(\lambda_{i,t-1}^{1/\nu}/\nu)/(\lambda_{it}^{1/\nu}/\nu)}\,]} \tag{18}$$

where $\tilde{y}_{it} = (y_{it} - \theta_{it})/\sqrt{\lambda_{it}^{1/\nu}/\nu}$. Thus, using this value of ρ, $C_{i,t,t-k} = \rho^k \sqrt{\lambda_{i,t-k}^{1/\nu}/\nu}/\sqrt{\lambda_{it}^{1/\nu}/\nu}$.

Note that ρ and the correlation matrix $C_{i,t,t-k}$ are obtained using known estimates for $\hat{\beta}$ and $\hat{\nu}$.

5. Estimating GQL

The GQL-I estimating equation (11) is solved by the Newton–Raphson iterative method such that

$$\begin{bmatrix} \hat{\beta}_{r+1} \\ \hat{\nu}_{r+1} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_r \\ \hat{\nu}_r \end{bmatrix} + \left[\sum_{i=1}^I D_i^T \widetilde{\Sigma}_i^{-1} D_i \right]_r^{-1} \left[\sum_{i=1}^I D_i^T \widetilde{\Sigma}_i^{-1} (f_i - \mu_i) \right]_r, \tag{19}$$

where $\hat{\beta}_r$ is the value of $\hat{\beta}$ at the rth iteration. Here, $[.]_r$ is the value of the expression at the rth iteration. The algorithm works as follows: for initial values of $\hat{\beta}$ and $\hat{\nu}$, we calculate $\hat{\rho}$ using (18) to obtain the correlation structure (5) and then use these two sets of parameters to update the values of $\hat{\beta}$ and $\hat{\nu}$. Then the new set of parameters is used to calculate $\hat{\rho}_l$ and the iteration continues in this way until convergence. The estimators are consistent and, under mild regularity conditions, for $I \to \infty$, it may be shown that $I^{1/2}((\hat{\beta}, \hat{\nu}) - (\beta, \nu))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I[\sum_{i=1}^I D_i^T \widetilde{\Sigma}_i^{-1} D_i]^{-1}[\sum_{i=1}^I D_i^T \widetilde{\Sigma}_i^{-1} (f_i - \mu_i)(f_i - \mu_i)^T \widetilde{\Sigma}_i^{-1} D_i][\sum_{i=1}^I D_i^T \widetilde{\Sigma}_i^{-1} D_i]^{-1}$. From §§4.1 and 4.2, it is clear that the covariance matrix $\widetilde{\Sigma}_i$ has a complex structure. In fact, Mamode Khan and Jowaheer [**4**] showed that under simple cases of time-independent covariates, the covariance matrices become ill conditioned for each subject $i = 1, 2, \ldots, I$. To overcome this issue in the case of non-stationary time-dependent covariates, we propose using a multi-dimensional conjugate gradient method (MCGM) to approximate the inverse of the covariance matrix, thus yielding another version of the GQL-I algorithm which we redefine as GQL-II. The MCGM algorithm works as follows:

(1) Assume an initial C_0 and calculate $R_0 = \widetilde{\Sigma}_i - AC_0$, where $A = \widetilde{\Sigma}_i \widetilde{\Sigma}_i^T$ for the ith observation, and let $P_1 = R_0$.

(2) For $k = 1, 2, \ldots$ until convergence, the following iterations are repeated:
   (a) $\alpha_k = \|R_{k-1}\|^2/P_k.(AP_k)$;
   (b) $C_k = C_{k-1} + \alpha_k P_k$;

    (c) $R_k = \widetilde{\Sigma_i} - AC_k$;

    (d) $\eta_k = \|R_k^2\|/\|R_{k-1}^2\|$;

    (e) $P_{k+1} = R_k + \eta_k P_k$.

If C_k converges based on $\|R_k\| < 10^{-10}$, then $\widetilde{\Sigma_i}^{-1} = C_k^T$, else the iterations continue until convergence. Note that this procedure is executed for each $i = 1, 2, 3, \ldots, I$. In contrast to GQL-I, the convergence rate of GQL-II is slower since there is a sub-iteration at each iteration for each subject i. To assess the performance of GQL-II, we simulate both longitudinal AR(1) stationary and non-stationary Com–Poisson counts in the next section and estimate the regression and dispersion parameters under each set-up.

6. Simulation studies

This section is made up of the following parts: in the first subsection, we first generate AR(1) stationary and non-stationary longitudinal Com–Poisson counts following the binomial thinning operation derived in §1. Then, assuming $T = 4$ and for $I = 60, 100, 500$ with true values of $\beta_0 = \beta_1 = 1$ for values of $\nu = 0.5, 1, 1.5$, GQL-II is implemented. For each cluster size and values of ν, 10 000 simulations are carried out under both set-ups. Note that $\nu = 1$ indicates equidispersion or Poisson, $\nu = 0.5$ indicates overdispersion and $\nu = 1.5$ is underdispersion. To assess the performance of GQL-I and GQL-II, the numbers of non-convergent simulations under both set-ups are compared.

    Following the AR(1) stationary autocorrelation structures developed by Mamode Khan and Jowaheer [4], we extend their algorithm to simulate AR(1)-type longitudinal CMP count responses under non-stationary set-ups, as illustrated in the next section.

6.1. A review of the AR(1) stationary Gaussian autocorrelation structure for simulating correlated Com–Poisson counts

We generate a CMP variable y_{i0} with mean θ_{i1} and covariance index ν using the recurrence relation $P(Y_{i0} = y_{i0} + 1) = P(Y_{i0} = y_{i0})\lambda_{i1}(y_{i0} + 1)^{-\nu}$, where $P(Y_{i0} = 0) = \{\exp(\nu\lambda_{i1}^{1/\nu})/\lambda_{i1}^{(\nu-1)/2\nu}(2\pi)^{(\nu-1)/2}\sqrt{\nu}\}^{-1}$. Using the discrete analog of the inverse transformation method, we calculate the sum of these probabilities starting from $P(Y_{i0} = 0)$ until the sum exceeds the value of a simulated uniform $(0, 1)$ variable and at this cut-off point the value obtained becomes a Com–Poisson variable. A Com–Poisson variable d_{i1} with mean parameter $(1-\rho)\theta_{i1}$ and covariance index $\tilde{\nu}$ is generated in the same way, where $\tilde{\nu}$ is calculated using (10) with $\theta_{it} = \theta_{i1}$ for all $t = 1, 2, \ldots, T$. Then y_{i1} is obtained by using the Gaussian AR(1) relationship

$$y_{i1} = \rho * y_{i0} + d_{i1},$$

where $\rho * y_{i0}$ is a simulated binomial random variable with mean ρy_{i0} following the binomial thinning. In this way, we generate values of d_{it} with mean $(1 - \rho)\theta_{i1}$ and dispersion index $\tilde{\nu}$ given in (6) and a binomial random variable with mean ρy_{it-1} for $t = 2, \ldots, T$ to obtain successive values of y_{it}.

6.2. Generating AR(1) non-stationary Com–Poisson counts

Here, as well, we generate a CMP variable y_{i0} with mean θ_{i1} and covariance index ν in the same way as in §6.1. Then we generate a Com–Poisson variable d_{i1} with mean $(1-\rho)\theta_{i1}$ and covariance index $\tilde{\nu}$, where $\tilde{\nu}$ is given by (10). By using the relation (11), we generate successive values of y_{it} for $t = 2, \ldots, T$ using simulated values of d_{it}, which is generated with mean $(\theta_{it} - \rho\theta_{i,t-1})$ and covariance index $\tilde{\nu}$ given in (10) and a binomial random variable with mean

$\rho y_{i,t-1}$. As for the covariate design, we consider two covariates x_{it1} and x_{it2} under designs k, where $k = 1, 2$ as follows:

– under the stationary set-up,

$$x_{it1} = \begin{cases} k-1 & (i = 1, \ldots, I/5), \\ \text{rbinom}(k, 0.5) & (i = (I/5) + 1, \ldots, 2I/5), \\ \text{rpois}(1, k) & (i = (2I/5) + 1, \ldots, 3I/5), \\ 1 & (i = (3I/5) + 1, \ldots, 4I/5), \\ 2^k & (i = (4I/5) + 1, \ldots, I) \end{cases}$$

and the second covariate x_{it2} is generated from a standard normal distribution;

– under the non-stationary set-up,

$$x_{it1} = \begin{cases} (k-1) * t & (i = 1, \ldots, I/5), \\ \text{rbinom}(k, 0.5) + t & (i = (I/5) + 1, \ldots, 2I/5), \\ \text{rpois}(1, k) & (i = (2I/5) + 1, \ldots, 3I/5), \\ 1 + t & (i = (3I/5) + 1, \ldots, 4I/5), \\ 2^k & (i = (4I/5) + 1, \ldots, I) \end{cases}$$

and the second covariate x_{it2} is generated from a Poisson distribution with mean parameter k and for various values of t. By applying GQL-I and GQL-II on these simulated responses with the respective covariate designs, the following results were obtained.

The above estimates of the regression and dispersion parameters were obtained iteratively using GQL-I and GQL-II estimating equations based on small initial values of these parameters. We note that, under both methods and set-ups, the values of these estimates converged close to the true values of β and ν. As the cluster size increases from $I = 60$ to $I = 500$, the standard errors of the various estimated parameters decrease under both methods, designs and set-ups, thus conforming to the central limit theorem. More specifically, the standard errors under GQL-I are slightly superior than GQL-II under both stationary and non-stationary covariate designs. However, in terms of computations, both methods yield some non-convergent simulations, with GQL-II yielding the least number of failures, as illustrated in Table 3.

In more detail, in the stationary and non-stationary cases, the intermediate computation in GQL-I fails in a very high number of simulations, particularly for $I = 60$ and $I = 100$ (see the above), because the covariance matrices $\widetilde{\Sigma_i}$ become singular. On the other hand, GQL-II fails because, in some simulations, the converged estimates were very far from the true values of the parameters, which may be due to the choice of initial values. Note that the number of failures is higher in the non-stationary case than the stationary case. Overall, GQL-II is an extremely robust approach as it has overcome in totality the ill-conditioned problem of the covariance matrix in simulations where GQL-I failed. Note that GQL-II took more iterations than GQL-I to converge, which was as expected, particularly in the non-stationary case.

7. *Application to epileptic seizure counts*

We revisit the epileptic dataset that was analyzed by Jowaheer and Sutradhar [2] (see also Thall and Vail [9]). The data consists of 59 epileptic patients for whom the numbers of seizures

occurring at each of four successive two-weekly clinic visits were reported. The summary statistics for these responses are given in the following table.

	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 note that the variances are greater than their corresponding means for each time point, indicating that the data are clearly overdispersed. We consider six covariates, namely the intercept term denoted by x_{it1}, the treatment parameter x_{it2}, coded as 0 for placebo and 1

TABLE 1. *Simulated mean estimates of the parameters and standard errors under designs 1 and 2 for GQL-I and GQL-II under stationary set-up.*

I	Design 1	$\hat{\beta}_{0\,\text{GQL-I}}$	$\hat{\beta}_{1\,\text{GQL-I}}$	$\hat{\nu}_{\text{GQL-I}}$	Design 1	$\hat{\beta}_{0\,\text{GQL-II}}$	$\hat{\beta}_{1\,\text{GQL-II}}$	$\hat{\nu}_{\text{GQL-II}}$
60	$\nu = 0.5$	1.1110	0.9994	0.4998	$\nu = 0.5$	1.1110	1.0012	0.4985
		(0.0622)	(0.0890)	(0.1835)		(0.0632)	(0.0894)	(0.1915)
100		1.0121	1.0111	0.5011		1.0124	1.0110	0.4998
		(0.0589)	(0.0782)	(0.1663)		(0.0601)	(0.0787)	(0.1680)
500		1.0101	1.0101	0.4999		1.001	1.0101	0.5010
		(0.0392)	(0.0409)	(0.1330)		(0.0401)	(0.0410)	(0.1331)
60	$\nu = 1$	1.0001	1.0109	0.9985	$\nu = 1$	1.0001	1.0105	0.9991
		(0.0781)	(0.0718)	(0.1515)		(0.0790)	(0.0720)	(0.1523)
100		1.0008	0.9991	1.0010		1.0010	1.0002	1.0003
		(0.0615)	(0.0531)	(0.1101)		(0.0620)	(0.0534)	(0.1105)
500		0.9991	1.0101	0.9999		0.9995	1.0101	1.0005
		(0.0380)	(0.0421)	(0.0995)		(0.0383)	(0.0430)	(0.1005)
60	$\nu = 1.5$	1.0111	1.0110	1.485	$\nu = 1.5$	1.0110	1.0109	1.4888
		(0.0723)	(0.0820)	(0.1010)		(0.0725)	(0.0823)	(0.1018)
100		0.9998	1.1001	1.5001		1.0101	1.1002	1.4990
		(0.0433)	(0.0555)	(0.0536)		(0.0435)	(0.0560)	(0.0538)
500		0.9990	1.0103	1.4950		1.0011	1.0101	1.4948
		(0.0366)	(0.0325)	(0.0316)		(0.0371)	(0.0330)	(0.0319)
I	**Design 2**	$\hat{\beta}_{0\,\text{GQL-I}}$	$\hat{\beta}_{1\,\text{GQL-I}}$	$\hat{\nu}_{\text{GQL-I}}$	**Design 2**	$\hat{\beta}_{0\,\text{GQL-II}}$	$\hat{\beta}_{1\,\text{GQL-II}}$	$\hat{\nu}_{\text{GQL-II}}$
60	$\nu = 0.5$	1.1110	0.9994	0.4998	$\nu = 0.5$	1.1111	1.0010	0.4978
		(0.0622)	(0.0890)	(0.1835)		(0.0631)	(0.0892)	(0.1878)
100		1.0121	1.0111	0.5011		1.0119	1.0091	0.5010
		(0.0589)	(0.0782)	(0.1663)		(0.0590)	(0.0785)	(0.1670)
500		1.0101	1.0101	0.4999		1.0101	1.0101	0.5001
		(0.0392)	(0.0409)	(0.1330)		(0.0401)	(0.0402)	(0.1334)
60	$\nu = 1$	1.0001	1.0109	0.9985	$\nu = 1$	1.0101	1.0101	1.1005
		(0.0781)	(0.0718)	(0.1995)		(0.0785)	(0.0720)	(0.2001)
100		1.0008	0.9991	1.0010		0.9991	1.0002	1.0013
		(0.0615)	(0.0531)	(0.1101)		(0.0617)	(0.0537)	(0.1115)
500		0.9991	1.0111	0.9999		1.0001	1.0101	1.0003
		(0.0380)	(0.0421)	(0.0995)		(0.0410)	(0.0422)	(0.1004)
60	$\nu = 1.5$	1.0111	1.0110	1.485	$\nu = 1.5$	1.1110	1.0110	1.4888
		(0.0723)	(0.0820)	(0.1010)		(0.0733)	(0.0825)	(0.1015)
100		0.9988	1.0001	1.5011		0.9981	0.9902	1.4990
		(0.0433)	(0.0555)	(0.0536)		(0.0438)	(0.0560)	(0.0543)
500		0.9990	1.0103	1.4950		1.0001	1.0101	1.4958
		(0.0166)	(0.0325)	(0.0316)		(0.0169)	(0.0330)	(0.0336)

for progabide, the baseline seizure rates denoted by x_{it3}, the age of the person denoted by x_{it4}, the interaction effect between treatment and baseline seizure rates x_{it5} and the visit x_{it6}. Note that x_{it6} is the only time-dependent covariate, which makes the mean parameter non-stationary. The mean parameter of the Com–Poisson distribution for the ith person is given by $\lambda_{it} = \exp(x_{it}^T\beta)$ with $x_{it} = (x_{it1}, x_{it2}, \ldots, x_{it5}, x_{it6})^T$ for $t = 1, \ldots, 4$. Here, β is the 6×1 vector of regression parameters.

Both methods yield reliable estimates of the regression and overdispersion parameters. The standard errors under GQL-I and GQL-II illustrate that GQL-I is slightly more efficient than GQL-II, thus confirming the pattern in the simulation studies. The treatment parameter under

TABLE 2. *Simulated mean estimates of the parameters and standard errors under designs 1 and 2 for GQL-I and GQL-II under non-stationary set-up.*

I	Design 1	$\hat{\beta}_{0\,\mathrm{GQL-I}}$	$\hat{\beta}_{1\,\mathrm{GQL-I}}$	$\hat{\nu}_{\mathrm{GQL-I}}$	Design 2	$\hat{\beta}_{0\,\mathrm{GQL-II}}$	$\hat{\beta}_{1\,\mathrm{GQL-II}}$	$\hat{\nu}_{\mathrm{GQL-II}}$
60	$\nu=0.5$	1.0998	0.9996	0.5010	$\nu=0.5$	1.1102	1.0002	0.4998
		(0.0922)	(0.0990)	(0.2331)		(0.0935)	(0.0999)	(0.2365)
100		1.0065	1.1112	0.4997		1.0070	1.1111	0.4992
		(0.0786)	(0.0881)	(0.1933)		(0.0790)	(0.0888)	(0.1965)
500		1.0001	1.0001	0.5023		1.0001	1.0001	0.5025
		(0.0692)	(0.0665)	(0.1873)		(0.0699)	(0.0670)	(0.1883)
60	$\nu=1$	1.0001	1.0001	0.9995	$\nu=1$	0.9999	1.0001	1.0001
		(0.0583)	(0.0620)	(0.1492)		(0.0590)	(0.0626)	(0.1495)
100		1.0001	0.9999	1.0001		1.0001	1.0000	1.0003
		(0.0432)	(0.0530)	(0.1201)		(0.0435)	(0.0535)	(0.1210)
500		0.9995	1.0001	0.9999		0.9995	0.9991	1.0002
		(0.0399)	(0.0497)	(0.1105)		(0.0402)	(0.0498)	(0.1110)
60	$\nu=1.5$	1.0110	1.0012	1.5150	$\nu=1.5$	1.0115	1.0015	1.5140
		(0.0756)	(0.0835)	(0.1023)		(0.0764)	(0.0842)	(0.1025)
100		0.9991	1.0012	1.4981		1.0001	1.0002	1.4987
		(0.0642)	(0.0623)	(0.0871)		(0.0650)	(0.0640)	(0.0882)
500		0.9995	1.0001	1.4997		1.0001	1.0001	1.4998
		(0.0452)	(0.0561)	(0.0616)		(0.0462)	(0.0564)	(0.0621)
I	Design 2	$\hat{\beta}_{0\,\mathrm{GQL-I}}$	$\hat{\beta}_{1\,\mathrm{GQL-I}}$	$\hat{\nu}_{\mathrm{GQL-I}}$	Design 2	$\hat{\beta}_{0\,\mathrm{GQL-II}}$	$\hat{\beta}_{1\,\mathrm{GQL-II}}$	$\hat{\nu}_{\mathrm{GQL-II}}$
60	$\nu=0.5$	1.1001	0.9999	0.4990	$\nu=0.5$	1.1003	0.9993	0.4986
		(0.0636)	(0.0995)	(0.1910)		(0.0642)	(0.0999)	(0.1925)
100		0.9995	1.0001	0.5001		1.1010	1.0001	0.5004
		(0.0593)	(0.0671)	(0.1713)		(0.0601)	(0.0682)	(0.1736)
500		1.0105	1.0001	0.4999		1.0110	0.9999	0.4995
		(0.0238)	(0.0520)	(0.1431)		(0.0241)	(0.0532)	(0.1440)
60	$\nu=1$	1.0001	1.0001	0.9995	$\nu=1$	1.0001	1.0001	1.0001
		(0.0181)	(0.0331)	(0.1289)		(0.0187)	(0.0338)	(0.1301)
100		1.0000	0.9996	0.9999		0.9996	0.9999	1.0001
		(0.0101)	(0.0111)	(0.1111)		(0.0110)	(0.0115)	(0.1119)
500		0.9998	1.0011	0.9999		1.0001	1.0012	1.0002
		(0.0098)	(0.0103)	(0.1023)		(0.0099)	(0.0110)	(0.1030)
60	$\nu=1.5$	0.9992	1.0010	1.499	$\nu=1.5$	0.9995	1.0010	1.5001
		(0.0823)	(0.0881)	(0.1129)		(0.0830)	(0.0890)	(0.1134)
100		0.9991	1.0016	1.5001		0.9999	0.9992	1.4996
		(0.0636)	(0.0651)	(0.0489)		(0.0640)	(0.0672)	(0.0590)
500		0.9995	1.0001	1.4994		1.0005	1.0001	1.4998
		(0.0321)	(0.0330)	(0.0410)		(0.0336)	(0.0338)	(0.0415)

both methods is negative, which indicates 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. In terms of computations, GQL-I is rather sensitive to the choice of initial values, as shown in simulation studies. In fact, for even small initial values of the regression and dispersion parameters, GQL-I yields ill-conditioned covariance matrices at several individual levels whilst for the same set of initial values, GQL-II performs satisfactorily well and converges smoothly to the above estimates. To solve this issue with GQL-I, we use the GQL-I estimates under the Poisson model as the initial values. This is quite feasible here given that there were only 59 individuals but, in a large cluster size, this procedure may lead to further computational problems. The only remark for GQL-II is that it is time consuming as there were two sets of iterations to be accomplished at each cluster size. Overall, in a real-life data application, GQL-I is not appealing as compared to GQL-II.

TABLE 3. *Number of non-convergent simulations in GQL-I and GQL-II under non-stationary set-up.*

I	Design 1 (stationary)	GQL-I	GQL-II	Design 1 (non-stationary)	GQL-I	GQL-II
60	$\nu = 0.5$	4892	30	$\nu = 0.5$	5600	10
100		3890	23		4030	22
500		2300	2		2334	4
60	$\nu = 1$	6712	15	$\nu = 1$	7400	13
100		5423	18		3192	11
500		6601	10		6671	13
60	$\nu = 1.5$	3421	20	$\nu = 1.5$	4521	14
100		4210	12		6561	11
500		2222	3		3131	12
I	Design 2 (stationary)	GQL-I	GQL-II	Design 2 (non-stationary)	GQL-I	GQL-II
60	0.5	4818	22	$\nu = 0.5$	5516	13
100		4620	14		2891	10
500		2331	12		2717	9
60	$\nu = 1$	7171	23	$\nu = 1$	7771	14
100		3451	12		4515	16
500		2101	12		2818	10
60	$\nu = 1.5$	4313	20	$\nu = 1.5$	5421	11
100		3411	11		4121	10
500		2313	10		2819	12

TABLE 4. *Epileptic data: estimates of the regression and dispersion parameters using GQL-I and GQL-II approaches.*

Method	INTC	TR	BR	Age	INTA	Visit	$\hat{\nu}$
GQL-I	0.4582	−0.2471	0.0027	0.0210	0.0011	0.5142	0.5222
	(0.4321)	(0.1521)	(0.0040)	(0.0109)	(0.0048)	(0.3121)	(0.4510)
GQL-II	0.4601	−0.2484	0.0032	0.0222	0.0012	0.5182	0.5261
	(0.4372)	(0.1555)	(0.0045)	(0.0111)	(0.0056)	(0.3201)	(0.4534)

8. *Conclusion*

With a surge in the range of applications for longitudinal count data with time-dependent covariates, there is a renewed interest in developing longitudinal non-stationary count models. This paper focuses on developing AR(1) non-stationary structure for Com–Poisson counts as it is a fast-growing distribution that can handle all forms of dispersion. The non-stationary Com–Poisson autocorrelation structure was constructed using a binomial thinning operation. As regards the estimation of regression and dispersion parameters under such set-up, a new GQL estimating function was constructed based on the MCGM algorithm. The MCGM algorithm was used purposely to approximate the inverse of the ill-conditioned covariance matrix following some major computational issues that were present under simple longitudinal stationary cases. This new algorithm was applied on simulated and real-life data applications, where it was clearly shown that the new GQL performs computationally well and yields reliable estimates of mean and dispersion parameters.

References

1. J. Castillo and M. Perez-Cassany, 'Weighted Poisson distributions for overdispersion and under-dispersion situations', *Ann. Inst. Statist. Math.* 50 (1998) 567–585.
2. V. Jowaheer and B. Sutradhar, 'Analysing longitudinal count data with overdispersion', *Biometrika* 89 (2002) 389–399.
3. V. Jowaheer and B. Sutradhar, 'Fitting lower order nonstationary autocorrelation models to the time series of Poisson counts', *WSEAS Trans. Math.* 4 (2005) 427–434.
4. N. Mamode Khan and V. Jowaheer, 'Comparing joint estimation and GMM adaptive estimation in COM–Poisson longitudinal regression model', *Comm. Statist. Simulation Comput.* 42 (2013) 755–770.
5. E. McKenzie, 'Autoregressive moving-average processes with negative binomial and geometric marginal distributions', *Adv. Appl. Probab.* 18 (1986) 679–705.
6. R. Prentice and L. Zhao, 'Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses', *Biometrics* 47 (1991) 825–839.
7. G. Shmueli, T. Minka, J. Borle and P. Boatwright, 'A useful distribution for fitting discrete data', *J. R. Stat. Soc. Ser. C Appl. Stat.* 54 (2005) 127–142.
8. B. Sutradhar and K. Das, 'On the efficiency of regression estimators in generalised linear models for longitudinal data', *Biometrika* 86 (1999) 459–465.
9. P. Thall and S. Vail, 'Some covariance models for longitudinal count data with overdispersion', *Biometrika* 46 (1990) 657–671.
10. Y. Wang and V. Carey, 'Working correlation structure misspecification, estimation and covariate design: implications for generalized estimating equations performance', *Biometrika* 90 (2003) no. 1, 29–41.

Naushad Mamode Khan
University of Mauritius
Reduit
Mauritius
n.mamodekhan@uom.ac.mu