Ann Inst Stat Math (2018) 70:215–247
https://doi.org/10.1007/s10463-016-0590-9

# Inferences in semi-parametric dynamic mixed models for longitudinal count data

**Nan Zheng**[1] · **Brajendra C. Sutradhar**[1]

Received: 21 February 2016 / Revised: 15 August 2016 / Published online: 28 November 2016


**Abstract** This paper considers a semi-parametric mixed model for longitudinal counts under the assumption that for conditional on a common random effect over time the repeated count responses of an individual follow a Poisson AR(1) (auto-regressive order 1) non-stationary correlation structure. A step-by-step estimation approach is developed which provides consistent estimators for the non-parametric function, regression parameters, variance of the random effects, and auto-correlation structure of the model. Proofs for the consistency properties of the estimators along with their convergence rates are derived. A simulation study is conducted to examine first the estimation effects on parameters when the non-parametric function is ignored, and then an overall estimation study is carried out in the presence of the non-parametric function by including its estimation as well.

**Keywords** Consistency · Dynamic relationship for repeated counts · Generalized quasi-likelihood · Longitudinal correlations · Overdispersion of main interest · Parametric and non-parametric functions · Random effects and their variance · Regression effects of main interest · Semi-parametric model and estimation

## 1 Introduction

In a longitudinal setup, there are situations where repeated count responses such as the yearly number of visits to a physician, along with a set of primary (main) covariates are collected from a large number of independent individuals over a small period of time. Suppose that $t_{ij}$ denotes the time at which the $j$th ($j = 1, \ldots, n_i$) count response is

✉ Brajendra C. Sutradhar
bsutradh@mun.ca

[1] Department of Mathematics and Statistics, Memorial University, St. John's, NL A1C5S7, Canada

recorded from the $i$th ($i = 1, \ldots, K$) individual so that $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ denotes the $n_i \times 1$ vector of repeated count responses for the $i$th individual. Also suppose that $y_{ij}$ is influenced by a fixed and known $p-$dimensional time-dependent covariate vector $x_{ij}(t_{ij})$, and $\beta : p \times 1$ measures this influence. In this setup, it is likely that the repeated counts $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ from the same individual $i$ will be correlated. It is of interest to examine the effect ($\beta$) of the covariates on the responses after taking the correlations of the repeated responses into account. For this type of longitudinal count data analysis one may, for example, refer to Sutradhar (Section 3, 2003) [see also Sutradhar (Chapter 6, 2003)]. These authors have illustrated their longitudinal fixed effects model by analyzing a health care utilization (HCU) data set consisting of a repeated number of yearly physician visits of an individual along with his/her covariates such as gender, education level and age.

There also exist at least two generalizations of the above longitudinal fixed effects model. Some authors such as Sutradhar et al. (2016) (see also Severini and Staniswalis 1994; Zeger and Diggle 1994; Sneddon and Sutradhar 2004; Lin and Carroll 2001, 2006; You and Chen 2007; Warriyar and Sutradhar 2014) provide ($a$) a generalization of the fixed effects model to a longitudinal semi-parametric setup where it is assumed that in addition to the primary covariates ($x_{ij}(\cdot)$), certain secondary covariates are also needed to explain the mean and variances of the repeated counts. To be specific, suppose that $q_{ij}(t_{ij})$ is a scalar secondary covariate collected at time $t_{ij}$ which is not of direct interest but influences the response through a suitable non-parametric function $\gamma(q_{ij})$, where, for example, for a yearly data with $j$ as the $j$th year, $t_{ij}$ may refer to a day within $j$th year as the recording time for the secondary covariate $q_{ij}$ and be defined as $t_{ij} \equiv [(j-1)+\{\text{numeric day of the event occurrence for individual } i\} \times \{\text{number of days in year } j\}^{-1}]$ for $j = 1, \ldots, n_i; i = 1, \ldots, K$, which makes both $t_{ij}$ and $q_{ij}$ dense for $i = 1, \ldots, K$. For example, if the $i$th individual visited the physician on January 15th of the first calendar year ($j = 1$) of the longitudinal study, then $t_{i1}$ would be $t_{i1} = [0 + 15/365]$ instead of writing $t_{i1} = 1$, the year 1 of the study. In this longitudinal semi-parametric setup, a correlation model for repeated count data is formulated as: $y_{i1} \sim \text{Poisson}(\tilde{m}_{i1})$ with $\tilde{m}_{i1} = E[Y_{i1}] = \exp(x'_{i1}(t_{i1})\beta + \gamma(q_{i1}(\cdot))) = \text{Var}[Y_{i1}]$, and for $j = 2, \ldots, n_i$ it is assumed that the repeated counts follow an AR(1) (auto-regressive order 1) type dynamic model

$$y_{ij} = \rho * y_{i,j-1} + d_{ij} = \sum_{s=1}^{y_{i,j-1}} b_s(\rho) + d_{ij}, \tag{1}$$

where $\Pr[b_s(\rho) = 1] = \rho$ and $\Pr[b_s(\rho) = 0] = 1 - \rho$, with $\rho$ as the correlation index parameter; $d_{ij} \sim \text{Poisson}[\tilde{m}_{ij} - \rho\tilde{m}_{i,j-1}]$, with $\tilde{m}_{ij} = \exp(x'_{ij}(t_{ij})\beta + \gamma(q_{ij}))$ for $j = 2, \ldots, n_i$. In addition, $d_{ij}$ and $y_{i,j-1}$ are assumed to be independent. This semi-parametric dynamic model (1) for count data has been recently studied by Sutradhar et al. (2016) including the inferences for the non-parametric function $\gamma(\cdot)$ and the regression effects $\beta$. Note that this semi-parametric longitudinal fixed model (1) may be treated as a generalization of the AR(1) type longitudinal fixed effect models studied by Sutradhar (2010, Eq. (14)) and Sutradhar (2011, Chapter 6), for example. Furthermore, we remark that $\rho * y_{i,j-1}$ in (1) generates a count by adding $y_{i,j-1}$ binary responses, and this count-generating operation is referred to as the binary thinning operation. More specifically, model (1) is written for the counts such that the previous count $y_{i,j-1}$ generates a count, namely $\rho * y_{i,j-1}$, which after addition to an error count produces the count $y_{ij}$ for the present time $j$. Notice that the dynamic model (1) is similar to, but different from, the well-known auto-regressive order 1 (AR(1)) model for Gaussian (linear) data. For Gaussian responses, the model is written as $y_{ij} = \rho y_{i,j-1} + \epsilon_{ij}$ with $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$, which for $j \neq k$ provides a correlation structure $\text{corr}(y_{ij}, y_{ik}) = \rho^{|j-k|}$. See, for example, the study by Sutradhar (2011, Section 2.1) for details. But, unlike the Gaussian AR(1) model, model (1) produces a correlation structure (see Eq. (11)) in time-dependent covariates, and the correlation structure contains the decay function $\rho^{|j-k|}$ as well.

Some authors such as Montalvo (1997), Wooldridge (1999), Sutradhar and Bari (2007), Winkelmann (2008), Sutradhar (2011, Chapter 8), and Sutradhar et al. (2014) have studied ($b$) an alternative generalization of the longitudinal fixed model. Instead of considering a secondary covariate, these authors assumed that the repeated counts may be influenced by an individual random effect, say, $\tau_i^*$ for the $i$th individual in addition to the standard primary covariates. Thus, they generalized the longitudinal fixed effects model for repeated counts to the mixed model setup. Let $\tau_i^* \overset{iid}{\sim} N\left(0, \sigma_\tau^2\right)$, and $\tau_i = \tau_i^*/\sigma_\tau$. In this setup, the repeated responses $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ are modeled conditional on $\tau_i$, through a dynamic relationship given by

$$y_{ij}|\tau_i = \rho * y_{i,j-1}|\tau_i + d_{ij}|\tau_i, \quad j = 2, \ldots, n_i, \tag{2}$$

where it is assumed that $y_{i1}|\tau_i \sim \text{Poisson}(m_{i1}^*(\beta, \sigma_\tau|\tau_i))$, and for $j = 2, \ldots, n_i$, $y_{i,j-1}|\tau_i \sim \text{Poisson}(m_{i,j-1}^*)$, and $d_{ij}|\tau_i \sim \text{Poisson}(m_{ij}^* - \rho m_{i,j-1}^*)$ with $m_{ij}^*(\beta, \sigma_\tau|\tau_i) = \exp(x_{ij}'(t_{ij})\beta + \sigma_\tau \tau_i)$ for $j = 1, \ldots, n_i$. In (2), conditional on $\tau_i$, $d_{ij}$ and $y_{i,j-1}$ are independent. Furthermore, similar to model (1), for a given count $y_{i,j-1}$, $\rho * y_{i,j-1} = \sum_{k=1}^{y_{i,j-1}} b_k(\rho)$ is a binomial thinning operation, where $b_k(\rho)$ stands for a binary random variable with $\Pr[b_k(\rho) = 1] = \rho$ and $\Pr[b_k(\rho) = 0] = 1 - \rho$. It then follows that conditional on $\tau_i$, $y_{ij}$ follows a Poisson distribution with mean and variance given by $m_{ij}^*$ for all $j = 1, \ldots, n_i$, and the lag $(k - j)$ correlation has the formula $\text{Corr}(Y_{ij}, Y_{ik}|\tau_i) = \rho^{k-j}\sqrt{m_{ij}^*/m_{ik}^*}$ for $j < k$. The formulas for the unconditional mean, variance and correlations are available in Sutradhar and Bari (2007), for example. See also the Eqs. (6)–(9) below for similar formulas.

In practice, it may happen that in addition to the primary and secondary covariates used to construct the semi-parametric fixed model (1), the repeated responses of an individual may also be influenced by an individual latent effect as in model (2). To illustrate the need for this type of combined model, we refer to a longitudinal semi-parametric fixed model (Sutradhar et al. 2016) fitted to the HCU data to examine the effect of the primary covariates gender and education level, and of the secondary covariate age, on the number of yearly physician visits. Here, primary covariates were fitted through a parametric regression function and the secondary covariate was fitted non-parametrically. Note, however, that there remains a possibility that the physician visit may also be influenced by certain latent individual effects. The analysis of this type of longitudinal responses affected by both random effects and non-parametric functions is, however, not adequately addressed in the literature. To address this issue, in this paper, we propose a semi-parametric longitudinal mixed model for repeated count data collected from a large number of independent individuals. We remark that the proposed model may be treated as a generalization of the GLLMM (2) (generalized linear longitudinal mixed model) to the semi-parametric setup. Thus, it may be referred as the semi-parametric GLLMM (SGLLMM) for count data. This SGLLMM model along with its basic properties is described in Sect. 2. The step-by-step estimation for the non-parametric function, regression, overdispersion, and longitudinal correlation index parameters is given in Sect. 3. The asymptotic properties of the estimators are also discussed in this section. The results of an extensive simulation study on the performance of the estimation approaches for the proposed semi-parametric dynamic mixed model are reported in Sect. 4. Some concluding remarks are provided in Sect. 5.

## 2 Proposed SGLLMM for count data and its basic properties

In this section, we extend the GLLMM (2) for count data to the semi-parametric setup. For the purpose, we add a non-parametric function $\gamma(q_{ij})$ (see model (1)) to the linear predictor $x'_{ij}(t_{ij})\beta + \sigma_\tau \tau_i$ in the mixed model (2). Note that in many longitudinal studies, primary covariates are collected at $t_{ij} \equiv j$. Thus, for convenience one may use $x_{ij}(j)$ for $x_{ij}(t_{ij})$. Also when convenient we use simply $x_{ij}$ for $x_{ij}(j)$. Let

$$\begin{aligned}\mu^*_{ij}(\beta, \sigma_\tau, \gamma(\cdot)|\tau_i) &= \exp\{x'_{ij}(j)\beta + \sigma_\tau \tau_i + \gamma(q_{ij})\}, \quad \text{for all} \quad j = 1, \ldots, n_i \\ &\equiv \mu^*_{ij}.\end{aligned}$$

By combining models (1) and (2), we now write the proposed semi-parametric mixed model (SMM), an alternative shorter reference for SGLLMM, as

$$\begin{aligned} y_{i1}|\tau_i &\sim \text{Poisson}(\mu^*_{i1}), \\ y_{ij}|\tau_i &= \rho * y_{i,j-1}|\tau_i + d_{ij}|\tau_i, \quad \text{for} \quad j = 2, \ldots, n_i, \text{ with} \\ d_{ij}|\tau_i &\sim \text{Poisson}(\mu^*_{ij} - \rho\mu^*_{i,j-1}), \end{aligned} \tag{3}$$

where conditional on $\tau_i$, $d_{ij}$ and $y_{i,j-1}$ are independent. Notice that models (2) and (3) are similar but different. The difference lies in the fact that the SMM in (3) implies that $y_{ij}$, conditional on $\tau_i$, has the marginal Poisson distribution, that is, $y_{ij}|\tau_i \sim \text{Poi}(\mu^*_{ij})$, whereas under the mixed model (2), $y_{ij}|\tau_i \sim \text{Poi}(m^*_{ij})$ with $m^*_{ij} = \exp(x'_{ij}(j)\beta + \sigma\tau_i)$ which is free from non-parametric functions $\gamma(q_{ij})$. More specifically, the SMM (3) provides the conditional means and variances as

$$E[Y_{ij}|\tau_i] = \mu^*_{ij} = \exp\{x'_{ij}(j)\beta + \sigma_\tau \tau_i + \gamma(q_{ij})\} = \text{var}[Y_{ij}|\tau_i]. \tag{4}$$

Furthermore, for $j < k$, the conditional covariance between $y_{ij}$ and $y_{ik}$ under the SMM (3) can be derived as

$$\begin{aligned}
\mathrm{cov}(Y_{ij}, Y_{ik}|\gamma(\cdot), \tau_i) &= \mathrm{E}(Y_{ij}Y_{ik}|\gamma(\cdot), \tau_i) - \mathrm{E}(Y_{ij}|\gamma(\cdot), \tau_i)\mathrm{E}(Y_{ik}|\gamma(\cdot), ) \\
&= \mathrm{E}_{Y_{ij}}[Y_{ij}\mathrm{E}_{Y_{i,j+1}}\{\ldots \mathrm{E}_{Y_{i,k-1}}(\mathrm{E}(Y_{ik}|y_{i,k-1}, y_{i,k-2}, \ldots, y_{i,j+1}))\}] \\
&\quad - \mu^*_{ij}(\beta, \sigma_\tau, \gamma(q_{ij}), \tau_i)\mu^*_{ik}(\beta, \sigma_\tau, \gamma(q_{ik}), \tau_i) \\
&= \rho^{k-j}\mu^*_{ij}(\beta, \sigma_\tau, \gamma(q_{ij}), \tau_i) = \sigma_{i,jk}(\boldsymbol{\beta}, \sigma_\tau, \gamma(\cdot), \tau_i, \rho). \qquad (5)
\end{aligned}$$

After some algebras, one obtains the basic properties of the SMM (3) such as the unconditional mean, variance and pairwise covariances as follows.

**Unconditional means and variances**

Since $\tau_i \overset{iid}{\sim} N\,(0, 1)$, the semi-parametric mixed model (SMM) (3) provides the unconditional means and variances of $y_{ij}$ $(j = 1, \ldots, n_i)$ as

$$\begin{aligned}
\mu_{ij} \equiv \mu_{ij}(\beta, \sigma_\tau, \gamma(\cdot)) &= E_{SMM}[Y_{ij}] = \exp\{x'_{ij}\beta + \gamma(q_{ij})\}E\,(\exp(\sigma_\tau\tau_i)) \\
&= \exp\left\{x'_{ij}\beta + \frac{\sigma^2_\tau}{2} + \gamma(q_{ij})\right\}, \qquad (6)
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{ijj} \equiv \sigma_{ijj}(\beta, \sigma_\tau, \gamma(\cdot)) &= \mathrm{Var}_{SMM}\left[Y_{ij}\right] \\
&= \mathrm{Var}[E(Y_{ij}|\tau_i)] + E[\mathrm{Var}(Y_{ij}|\tau_i)] = \mathrm{Var}[\mu^*_{ij}] + E[\mu^*_{ij}] \\
&= \exp\{2x'_{ij}\beta + 2\gamma(q_{ij})\}\mathrm{Var}(\exp(\sigma_\tau\tau_i)) + \mu_{ij} \\
&= \left[\exp\left\{x'_{ij}\beta + \gamma(q_{ij}) + \frac{\sigma^2_\tau}{2}\right\}\right]^2 (\exp(\sigma^2_\tau) - 1) + \mu_{ij} \\
&= \mu_{ij} + \mu^2_{ij}[\exp(\sigma^2_\tau) - 1], \qquad (7)
\end{aligned}$$

respectively, because

$$\begin{aligned}
\mathrm{Var}[\exp(\sigma_\tau\tau_i)] &= E[\exp(2\sigma_\tau\tau_i)] - [E\{\exp(\sigma_\tau\tau_i)\}]^2 \\
&= \exp(2\sigma^2_\tau) - \exp(\sigma^2_\tau) = \exp(\sigma^2_\tau)(\exp(\sigma^2_\tau) - 1).
\end{aligned}$$

**Unconditional covariances**

Notice from (5) that conditional on $\tau_i$, the covariance between $y_{ij}$ and $y_{ik}$ $(j < k)$ is given by $\mathrm{Cov}(Y_{ij}, Y_{ik}|\tau_i) = \rho^{k-j}\mu^*_{ij}$. Consequently, for $j < k$, the unconditional covariance between $y_{ij}$ and $y_{ik}$ has the form

$$\begin{aligned}
\sigma_{ijk} &\equiv \sigma_{ijk}(\beta, \sigma_\tau, \rho, \gamma(q_0)) \\
&= \mathrm{Cov}_{SMM}(Y_{ij}, Y_{ik}) = E[\mathrm{Cov}(Y_{ij}, Y_{ik}|\tau_i)] + \mathrm{Cov}[E(Y_{ij}|\tau_i), E(Y_{ik}|\tau_i)] \\
&= E[\rho^{k-j}\mu^*_{ij}] + \mathrm{Cov}[\exp(x'_{ij}\beta + \gamma(q_{ij}) + \sigma_\tau\tau_i), \exp(x'_{ik}\beta + \gamma(q_{ik}) + \sigma_\tau\tau_i)] \\
&= \rho^{k-j}\mu_{ij} + \mu_{ij}\mu_{ik}(\exp(\sigma^2_\tau) - 1), \qquad (8)
\end{aligned}$$

leading to the lag $(k - j)$ correlations as

$$\begin{aligned}\mathrm{Corr}_{SMM}(Y_{ij}, Y_{ik}) &= \frac{\sigma_{ijk}}{\sqrt{\sigma_{ijj}\sigma_{ikk}}} \\ &= \frac{\mu_{ij}\rho^{k-j} + \mu_{ij}\mu_{ik}(\exp(\sigma_\tau^2) - 1)}{[\{\mu_{ij} + \mu_{ij}^2(\exp(\sigma_\tau^2) - 1)\}\{\mu_{ik} + \mu_{ik}^2(\exp(\sigma_\tau^2) - 1)\}]^{\frac{1}{2}}}. \qquad (9)\end{aligned}$$

In addition, since in model (3) $d_{ij}|\tau_i \sim \mathrm{Poi}(\mu_{ij}^* - \rho\mu_{i,j-1}^*)$ with $(\mu_{ij}^* - \rho\mu_{i,j-1}^*) \geq 0$, the correlation index parameter $\rho$ must now satisfy the range restriction $0 \leq \rho < \min[1, \mu_{ij}^*/\mu_{i,j-1}^*]$, which is the same as

$$0 \leq \rho < \min[1, \nu_{ij}^*/\nu_{i,j-1}^*] \text{ for } j = 2, \ldots, n_i \text{ and } i = 1, \ldots, K, \qquad (10)$$

where $\nu_{ij}^* = \exp(x_{ij}'\beta + \gamma(q_{ij}))$.

Notice from (9) that under the proposed model, unlike the longitudinal fixed model, the correlation index parameter value $\rho = 0$ does not imply that the responses are uncorrelated. The repeated responses are uncorrelated only when both $\rho = 0$ and $\sigma_\tau^2 = 0$. However, since in the mixed model $\sigma_\tau^2 > 0$, the pairwise responses are positively correlated irrespective of the case whether $\rho$ is zero or not. One, therefore, has to be careful while estimating the regression effects $\beta$ and non-parametric function $\gamma(q_{ij}(t_{ij}))$ using any GEE(I) (independence assumption-based generalized estimating equation) approach. In fact, even though one can attempt to use $\rho = 0$ for initial estimation of these parameters and functions, but using $\sigma_\tau^2 = 0$ for their estimation would produce inconsistent estimates because $\sigma_\tau^2$ is involved in the mean function (6) along with $\beta$ and $\gamma(q_{ij})$. As opposed to the semi-parametric longitudinal fixed model (Severini and Staniswalis 1994; Lin and Carroll 2001, 2006; You and Chen 2007; Warriyar and Sutradhar 2014) this is a major additional estimation problem in the present semi-parametric longitudinal mixed model case. These estimation issues are further discussed in Sect. 3.

Note that there are situations in practice where one may need to develop the semi-parametric mixed models for the analysis of longitudinal linear and binary data. This may be done using appropriate dynamic models for such linear and binary data. In general, the dynamic models are different for linear, binary and count data. For example, in a linear case, following Warriyar and Sutradhar (2014) one may write an AR(1)-type SMM as

$$\begin{aligned} y_{i1} &= x_{i1}'(1)\beta + \sigma_\tau\tau_i + \gamma(q_{i1}) + \epsilon_{i1} \\ y_{ij} &= \{x_{ij}'(j)\beta + \gamma(q_{ij})\} + \rho^*(y_{i,j-1} - x_{i,j-1}'(j-1)\beta - \gamma(q_{i,j-1})) \\ &\quad + \sigma_\tau\tau_i + \epsilon_{ij}, \text{ for } j = 2, \ldots, n_i, \qquad (11)\end{aligned}$$

where $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$, (say). Notice that this dynamic model in (11) is quite different than the dynamic model (3) for the repeated counts. Similarly, in the binary case, one may write a semi-parametric dynamic mixed model by generalizing the so-called binary dynamic mixed logit (BDML) model studied by Sutradhar (2011, Eq. (9.27)), for example. To be specific, using similar notation as in (3), the SMM for binary responses may have the form

$$\begin{aligned} Pr[Y_{i1} = 1|\gamma(\cdot), \tau_i] &= \frac{\exp(x'_{i1}(t_{i1})\beta + \gamma(q_{i1}) + \sigma_\tau \tau_i)}{1 + \exp(x'_{i1}(t_{i1})\beta + \gamma(q_{i1}) + \sigma_\tau \tau_i)}, \\ Pr[Y_{ij} = 1|y_{i,j-1}, \gamma(\cdot), \tau_i] & \\ = \frac{\exp(x'_{ij}(t_{ij})\beta + \theta y_{i,j-1} + \gamma(q_{ij}) + \sigma_\tau \tau_i)}{1 + \exp(x'_{ij}(t_{ij})\beta + \theta y_{i,j-1} + \gamma(q_{ij}) + \sigma_\tau \tau_i)} & \; j = 2, \ldots, n_i, \end{aligned} \tag{12}$$

where $\theta$ is a dynamic dependence parameter which is quite different than $\rho$ in (3) and $\rho^*$ in (11), and $\gamma(\cdot)$ is the non-parametric function involved in the semi-parametric regression function. Further remark that because the correlation models (11) and (12) are different than that of (3), the inferences for the SMMs (11) and (12) will be different as well. We now turn back to our SMM (3) for the count data and develop its inferences in the next section.

## 3 Quasi-likelihood estimation for the proposed semi-parametric GLLMM

In this section, we develop a quasi-likelihood estimation approach which provides consistent estimates for all parameters and the non-parametric function involved in the SGLLMM. Note that this estimation approach has been used by some authors (Severini and Staniswalis 1994; Lin and Carroll 2001, 2006; Warriyar and Sutradhar 2014) for the SGLLFM (semi-parametric generalized linear longitudinal fixed model). Severini and Staniswalis (1994) and Lin and Carroll (2001) (see also Zeger and Diggle 1994) refer to their estimation approach as the semi-parametric generalized estimating equation (SGEE) approach which does not need any specification of the underlying longitudinal correlation structure. However, there has been many studies showing that independence assumption-based GEE (GEE(I)) approach may produce more efficient regression estimates at times than arbitrary 'working' correlation-based GEE approach. See, for example, the study by Sutradhar (2010, Section 3.1) (see also Sutradhar and Das 1999) in the context of GLLFM for count data. This efficiency behavior undermines the use of the GEE or SGEE approaches. Thus, we do not discuss the GEE approaches any further in the present paper. Instead, we assume that the repeated count data are generated following the AR(1) Poisson mixed model (3)-based correlation structure (9) and consequently use the true correlation structure-based semi-parametric generalized quasi-likelihood (GQL) approach for the estimation of the main regression effects (of the primary covariates) and the overdispersion parameter (Sutradhar and Bari 2007; Sutradhar 2011, Chapter 8). Next because the non-parametric function and the longitudinal correlations are of secondary interest, we estimate them using the simpler SQL (semi-parametric QL) and SMM (semi-parametric method of moments), respectively, as opposed to the SGQL approach. These estimation approaches are discussed in the following subsections.

### 3.1 Independence assumption-based QL estimation for the non-parametric function $\gamma(\cdot)$

The non-parametric function $\gamma(q_{ij})$ has to be estimated for all $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$, where $q_{ij}$ is a secondary covariate collected at time $t_{ij}$. Thus, it is equivalent to estimate $\gamma(q_0)$, say, where $q_0 \equiv q_{ij}$ for any given $i$ and $j$. Notice that because $K$ is large in the present setup and $t_{ij}$ can be dense, this in general leads $q_{ij}$ to be dense as well. However, $n_i$ is a total equi-spaced time period for the $i$th individual. In the GLLFM setup, some authors such as Lin and Carroll (2001) (see also Severini and Staniswalis 1994) have considered a study with $q_{ij}(t_{ij}) = t_{ij}$ and estimated the non-parametric function in fixed time, that is $\gamma(t_{ij})$, using a 'working' correlation structure-based estimating equation approach where the estimating equation was apparently constructed using the repeated responses $\{y_{i,t_{ij}}, j = 1, \ldots, n_i\}$ with correlation matrices of dimension $n_i \times n_i$ instead of $t_{in_i} \times t_{in_i}$. In fact, these authors went on further using $n_i = n$, say, for estimating their so-called $n \times n$ unstructured 'working' common correlation matrix. This undermines their correlation approach which was apparently claimed to have constructed for the dense time point-based responses. In this paper, we have made this issue clear by defining secondary covariates $q_{ij}$ at time $t_{ij}$ but the primary covariates $x_{ij}$ are recorded at time points $j = 1, \ldots, n_i$. Our approach also indicates that because $t_{ij}$ or $q_{ij}$ are simply fixed covariates, for consistent estimation of $\gamma(q_{ij})$, which is of secondary interest, it would be enough to use an independent assumption-based estimating equation, whereas the main regression parameter (effect of primary covariates) $\beta$ would be estimated consistently and as efficiently as possible using correlation structure-based estimating equation. Furthermore, unlike the existing fixed regression models, we also need to consistently and efficiently estimate the other main overdispersion parameter $\sigma_\tau^2$ involved in the present mixed model (3).

In quasi-likelihood (QL) approach for independent data (Wedderburn 1974), one explores the mean and the variance functions, variance being a function of mean such as in a GLM setup, to write a QL estimating equation for the parameter involved in the mean function. When the mean function involves a non-parametric function, one may estimate such a function by solving a kernel weight-based semi-parametric QL (SQL) estimating equation. For the estimation of $\gamma(q_0)$ in the present setup which influences the mean function $\mu_{ij}(\beta, \sigma_\tau, \gamma(q_0))$, the SQL estimating equation has the form

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(q_0)\frac{\partial \mu_{ij}(\beta, \sigma_\tau, \gamma(q_0))}{\partial \gamma(q_0)}\left(\frac{y_{ij} - \mu_{ij}(\beta, \sigma_\tau, \gamma(q_0))}{\sigma_{ijj}(\beta, \sigma_\tau, \gamma(q_0))}\right) = 0 \qquad (13)$$

(e.g., Carota and Parmigiani 2002; Warriyar and Sutradhar 2014), where $w_{ij}(q_0)$ is referred to as the so-called kernel weight defined as

$$w_{ij}(q_0) = p_{ij}\left(\frac{q_0 - q_{ij}}{b}\right) \Big/ \sum_{l=1}^{K}\sum_{u=1}^{n_l} p_{lu}\left(\frac{q_0 - q_{lu}}{b}\right) \qquad (14)$$

where $p_{ij}$ is the kernel density with $b$ as a suitable bandwidth parameter. Note that this SQL estimating Eq. (13) is different than the so-called 'working' correlation-based SGEE (semi-parametric GEE) used by Lin and Carroll (2001, 2006) (see also Severini and Staniswalis 1994). This is simpler than SGEE and it assures the consistency of the estimator, whereas even though SGEE is developed for efficient estimation it may, however, produce inefficient estimate than the 'working' independence-based SQL estimator (Sutradhar and Das 1999; Lin and Carroll 2001, Section 7).

With regard to the selection of the kernel density $p_{ij}(\cdot)$, it should be noted that there is, in fact, no unique choice for the selection of such a density. Some of the widely used kernel densities, for example, are the so-called Gaussian density given by

$$p_{ij}\left(\frac{q_0 - q_{ij}}{b}\right) = \frac{1}{\sqrt{2\pi}\, b} \exp\left\{-\frac{1}{2}\left(\frac{q_0 - q_{ij}}{b}\right)^2\right\}, \tag{15}$$

and the Epanechnikov kernel (Pagan and Ullah 1999, p. 28) with density

$$p_{ij}(\varphi) = \begin{cases} \frac{1}{4}(1-\varphi^2) & \text{for } |\varphi| \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad \text{with } \varphi = \frac{q_0 - q_{ij}}{b}. \tag{16}$$

In (14)–(16), $b$ is a suitable bandwidth parameter. Asymptotic results in Sect. 3.5 indicate that the consistent estimation for all parameters and functions under the present model requires $Kb^4 \to 0$ as $K \to \infty$. Thus, an appropriate choice of $b$ in the present setup should satisfy $b \propto K^{-\alpha}$ with $1/4 < \alpha \le 1/3$. Suppose that we choose $b = c_0 K^{-1/3.9}$. However, finding an analytical technique to choose an appropriate value for $c_0$ does not appear to be easy. In the simulation study in Sect. 4, we use $b = K^{-1/5}$ following Pagan and Ullah (1999), and Lin and Carroll (2001), for example. In some studies, this bandwidth parameter is determined by a local cross-validation technique (Sneddon and Sutradhar 2004, Section 3; Altman 1990). To examine the appropriateness of the choice $b = K^{-1/5}$ we apply a mini-max numerical procedure in Sect. 4 which appears to approximately justify the choice. We remark that this choice $b = K^{-1/5}$ should be equivalent to $b = c_0 K^{-1/3.9}$, found based on our asymptotic consideration. Its theoretical justification is, however, beyond the scope of the present study.

Now because $\partial\mu_{ij}(\beta, \sigma_\tau^2, \gamma(q_0))/\partial\gamma(q_0) = \mu_{ij}(\beta, \sigma_\tau^2, \gamma(q_0))$, the SQL estimating Eq. (13) can be further simplified as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(q_0)\left(\frac{y_{ij} - \mu_{ij}(\beta, \sigma_\tau^2, \gamma(q_0))}{1 + \mu_{ij}(\beta, \sigma_\tau^2, \gamma(q_0))(\exp(\sigma_\tau^2) - 1)}\right) = 0, \tag{17}$$

which, for given values of $\beta$ and $\sigma_\tau^2$, may be solved iteratively until convergence.

Notice that the estimate of $\gamma(q_0)$ from the SQL estimating Eq. (17) is a function of $\beta$ and $\sigma_\tau^2$. Hence, we denote the estimator of $\gamma(q_0)$ by $\hat{\gamma}(q_0; \beta, \sigma_\tau^2)$. The consistency property of this estimator is discussed in Sect. 3.5, and we study through simulations its finite sample properties along with the properties of other estimators in Sect. 4.

### 3.2 SGQL estimation of regression effects $\boldsymbol{\beta}$

Recall from (6)–(8) that if the non-parametric function $\gamma(q_0)$ was known, one then could construct the mean vector and covariance matrix of $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$, as

$$\begin{aligned} E(Y_i) &= \mu_i(\beta, \sigma_\tau^2, \gamma(\cdot)) \\ &= (\mu_{i1}(\beta, \sigma_\tau^2, \gamma(\cdot)), \ldots, \mu_{ij}(\beta, \sigma_\tau^2, \gamma(\cdot)), \ldots, \mu_{in_i}(\beta, \sigma_\tau^2, \gamma(\cdot)))', \end{aligned} \quad (18)$$

and

$$\mathrm{Cov}(Y_i) = \Sigma_i(\beta, \sigma_\tau^2, \rho, \gamma(\cdot)) = (\sigma_{ijk}(\beta, \sigma_\tau^2, \rho, \gamma(\cdot))) : n_i \times n_i, \quad (19)$$

respectively. However, it is clear from Sect. 3.1 that when $\gamma(q_{ij})$ are estimated by solving the SQL estimating Eq. (17), we obtain the estimator $\hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2)$ which contains unknown $\beta$ and $\sigma_\tau^2$. Consequently, the mean vector and the covariance matrix now have the forms

$$\begin{aligned} \bar{\mu}_i(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2)) = (\bar{\mu}_{i1}(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2)), \ldots, \bar{\mu}_{ij}(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2)), \ldots, \\ \bar{\mu}_{in_i}(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2)))' : n_i \times 1 \qquad \text{and} \quad (20) \\ \bar{\Sigma}_i(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\beta, \sigma_\tau^2)) = (\bar{\sigma}_{ijk}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\beta, \sigma_\tau^2))) : \quad n_i \times n_i, \quad (21) \end{aligned}$$

respectively. We now use these new notations from (20) and (21) and following Sutradhar (2003), for example, construct the semi-parametric GQL (SGQL) estimating equation for $\beta$ as

$$\begin{aligned} \sum_{i=1}^{K} \frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2))}{\partial \beta} \bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\beta, \sigma_\tau^2)) \\ \times (y_i - \bar{\mu}_i(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2))) = 0. \end{aligned} \quad (22)$$

For convenience, the computational formula for the derivative matrix $\frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2))}{\partial \beta}$ is given in Appendix A1.

The SGQL estimate of $\beta$ obtained by solving the estimating Eq. (22) will be denoted by $\hat{\beta}$. Its asymptotic and finite sample properties are discussed in Sects. 3.5.2 and 4, respectively.

### 3.3 SGQL estimation of the random effect variance $\sigma_\tau^2$

In general, the generalized method of moments (GMM) and the generalized quasi-likelihood (GQL) methods are popular for the estimation of the overdispersion index parameter $\sigma_\tau^2$ involved in the GLLMM (1) for repeated count data. However, it has been demonstrate by Rao et al. (2012) (see also Sutradhar 2011, Chapter 8, Table 8.2) in a linear longitudinal setup that the GQL approach produces more efficient estimate for this parameter as compared to the GMM approach. Furthermore, Sutradhar and

Bari (2007) has demonstrated that the GQL approach performs well in estimating this parameter in a longitudinal setup for count data. In this section, we generalize this GQL approach to the semi-parametric longitudinal setup. We also provide a normality (of count responses)-based SGQL (semi-parametric GQL) approximation.

### *3.3.1 SGQL estimation for $\sigma_\tau^2$ using raw squared responses*

Consider a vector of squared responses $u_i = [y_{i1}^2, \ldots, y_{ij}^2, \ldots, y_{in_i}^2]'$. Then a GQL estimating equation for $\sigma_\tau^2$ may be developed by minimizing the quadratic distance function

$$Q = (u_i - E[U_i])'\{\mathrm{Cov}[U_i]\}^{-1}(u_i - E[U_i]) \tag{23}$$

(Sutradhar and Bari 2007), where using $\bar{\mu}_{ij} \equiv \bar{\mu}_{ij}(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2))$ for $\mu_{ij} = E[Y_{ij}]$, one may compute the mean vector $E[U_i]$ using

$$E[Y_{ij}^2] = \bar{\lambda}_{ijj}(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot)) = \bar{\mu}_{ij} + \bar{\mu}_{ij}^2\, e^{\sigma_\tau^2}, \tag{24}$$

and the covariance matrix $\mathrm{Cov}(U_i)$ using

$$\mathrm{Var}(Y_{ij}^2) = \bar{\mu}_{ij}[1 + 7\bar{\mu}_{ij}\, \exp(\sigma_\tau^2) + 6\bar{\mu}_{ij}^2\, \exp(3\sigma_\tau^2) + \bar{\mu}_{ij}^3\, \exp(6\sigma_\tau^2)] - \bar{\lambda}_{ijj}^2, \tag{25}$$

and, for $j < k$,

$$\begin{aligned} \mathrm{Cov}(Y_{ij}^2, Y_{ik}^2) = {} & 2\rho^{2(k-j)}\bar{\mu}_{ij}^2\, \exp(\sigma_\tau^2) + 4\rho^{k-j}\, \bar{\mu}_{ik}\bar{\mu}_{ij}^2\, \exp(3\sigma_\tau^2) + 2\rho^{k-j}\, \bar{\mu}_{ij}^2\, \exp(\sigma_\tau^2) \\ & + 2\rho^{k-j}\, \bar{\mu}_{ik}\bar{\mu}_{ij}\, \exp(\sigma_\tau^2) + \rho^{k-j}\, \bar{\mu}_{ij} + \bar{\mu}_{ik}^2\bar{\mu}_{ij}^2\, \exp(6\sigma_\tau^2) + \bar{\mu}_{ik}\bar{\mu}_{ij}^2\, \exp(3\sigma_\tau^2) \\ & + \bar{\mu}_{ik}^2\bar{\mu}_{ij}\, \exp(3\sigma_\tau^2) + \bar{\mu}_{ik}\bar{\mu}_{ij}\, \exp(\sigma_\tau^2) - \bar{\lambda}_{ijj}\bar{\lambda}_{ikk}. \end{aligned} \tag{26}$$

Next, for convenience, using the notations

$$\begin{aligned} \bar{\lambda}_i(\beta, \sigma_\tau^2, \gamma(\cdot)) &= E(U_i) = E[Y_{i1}^2, \ldots, Y_{ij}^2, \ldots, Y_{in_i}^2]' \\ &= [\bar{\lambda}_{i11}(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot)), \ldots, \bar{\lambda}_{ijj}(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot)), \ldots, \bar{\lambda}_{in_in_i}(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))]', \end{aligned} \tag{27}$$

and $\bar{\Omega}_i = \mathrm{Cov}(U_i)$, we minimize $Q$ in (23) and obtain the SGQL estimating equation for $\sigma_\tau^2$ as

$$\sum_{i=1}^{K} \frac{\partial \bar{\lambda}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))}{\partial \sigma_\tau^2}\, \bar{\Omega}_i^{-1}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\cdot))(u_i - \bar{\lambda}_i(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))) = 0 \tag{28}$$

(Sutradhar 2004), where the derivative $\frac{\partial \bar{\lambda}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))}{\partial \sigma_\tau^2}$ has the formula as shown in Appendix A.2. The estimate of $\sigma_\tau^2$ obtained from (28) will be referred as the raw SGQL estimate.

### 3.3.2 SGQL estimation for $\sigma_\tau^2$ using corrected (CR) squared responses

For technical convenience, an alternative way to construct a GQL estimating equation for $\sigma_\tau^2$ would be exploiting the vectors of second-order corrected squared responses from the individuals. For the $i$th individual, let

$$g_i = [(y_{i1} - \bar{\mu}_{i1}(\cdot))^2, \ldots, (y_{ij} - \bar{\mu}_{ij}(\cdot))^2, \ldots, (y_{in_i} - \bar{\mu}_{in_i}(\cdot))^2]'$$

denote the second-order corrected squared response vector, with known $\bar{\mu}_{ij}(\cdot)$ (18) computed from the previous iteration under a suitable iterative scheme. Following (28), in this case, we write the SGQL estimating equation for $\sigma_\tau^2$ as

$$\sum_{i=1}^{K} \frac{\partial \bar{\sigma}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))}{\partial \sigma_\tau^2} \bar{\Omega}_{i,\mathrm{CR}}^{-1}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\cdot))(g_i - \bar{\sigma}_i(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))) = 0, \quad (29)$$

where

$$\bar{\sigma}_i = E(G_i) = (\bar{\sigma}_{i11}, \ldots, \bar{\sigma}_{ijj}, \ldots, \bar{\sigma}_{in_in_i})', \;\; \bar{\Omega}_{i,\mathrm{CR}} = \mathrm{Cov}(G_i), \quad (30)$$

with 'CR' indicating a 'corrected' response-based quantity, and by (24)

$$\bar{\sigma}_{ijj} = \bar{\mu}_{ij} + \bar{\mu}_{ij}^2(\exp(\sigma_\tau^2) - 1). \quad (31)$$

In (29),

$$\begin{aligned} \frac{\partial \bar{\sigma}_i'}{\partial \sigma_\tau^2} &= \frac{\partial}{\partial \sigma_\tau^2}(\bar{\sigma}_{i11}, \ldots, \bar{\sigma}_{ijj}, \ldots, \bar{\sigma}_{in_in_i}), \;\text{ with} \\ \frac{\partial \bar{\sigma}_{ijj}}{\partial \sigma_\tau^2} &= \frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} + 2\bar{\mu}_{ij}\left(\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2}\right)(\exp(\sigma_\tau^2) - 1) + \bar{\mu}_{ij}^2 \exp(\sigma_\tau^2), \end{aligned} \quad (32)$$

where the formula for $\partial \bar{\mu}_{ij}/\partial \sigma_\tau^2$ is given by (64) under Appendix A.2. Furthermore, we provide the formulas for the elements of the covariance matrix $\bar{\Omega}_{i,\mathrm{CR}} = \mathrm{Cov}(G_i)$ in Appendix A.3. The estimate of $\sigma_\tau^2$ obtained from (29) will be referred as the corrected (CR) SGQL estimate.

### 3.3.3 Normal approximation-based SGQL estimation using squared corrected responses

The normality-based SGQL estimating equation would be the same as that of (29) constructed based on corrected squared responses except that the fourth-order moment matrix $\bar{\Omega}_{i,\mathrm{CR}}$ is now replaced with a normality-based fourth-order moment matrix, say $\bar{\Omega}_{i,\mathrm{CRN}}$. Thus, in notation of (29),

$$\mathrm{Cov}_N(G_i) = \bar{\Omega}_{i,\mathrm{CRN}}, \quad (33)$$

where the elements of this matrix are computed from the normality-based fourth-order product moments formula

$$
\begin{aligned}
& E_N[(Y_{ij} - \bar{\mu}_{ij})(Y_{ik} - \bar{\mu}_{ik})(Y_{il} - \bar{\mu}_{il})(Y_{im} - \bar{\mu}_{im})] \\
& \quad = \bar{\sigma}_{ijk}\bar{\sigma}_{ilm} + \bar{\sigma}_{ijl}\bar{\sigma}_{ikm} + \bar{\sigma}_{ijm}\bar{\sigma}_{ikl}.
\end{aligned} \tag{34}
$$

for $i = 1, \ldots, K$ and $1 \leq j, k, l, m \leq n_i$. For example, under normality,

$$
\begin{aligned}
\mathrm{Var}[(Y_{ij} - \bar{\mu}_{ij})^2] &= E_N[(Y_{ij} - \bar{\mu}_{ij})^4] - \bar{\sigma}^2_{ijj} \\
&= 3\bar{\sigma}^2_{ijj} - \bar{\sigma}^2_{ijj} = 2\bar{\sigma}^2_{ijj},
\end{aligned} \tag{35}
$$

by (34). Notice that the normality assumption for count responses $\{y_{ij}, j = 1, \ldots, n_i\}$ makes the higher order moments calculation simpler. Remark that Prentice and Zhao (1991) and Zhao and Prentice (1990) have used such normality approximation to compute higher order moments for binary repeated data. This approximation appears to work well for repeated count data in the GLM setup (Sutradhar 2011, Chapter 8), whereas in this section we have considered its use in the semi-parametric longitudinal mixed model setup.

For completeness, using the notations from (33) to (35), we now write the desired normality-based SGQL estimating equation for $\sigma^2_\tau$ as

$$
\sum_{i=1}^{K} \frac{\partial \bar{\sigma}'_i(\beta, \sigma^2_\tau, \hat{\gamma}(\cdot))}{\partial \sigma^2_\tau} \bar{\Omega}^{-1}_{iC,N}(\beta, \sigma^2_\tau, \rho, \hat{\gamma}(\cdot))(g_i - \bar{\sigma}_i(\beta, \sigma^2_\tau, \hat{\gamma}(\cdot))) = 0, \tag{36}
$$

which is solved iteratively until convergence. The solution for $\sigma^2_\tau$ obtained from (36) will be referred as the normality-based corrected (CRN) SGQL estimate.

### 3.4 Moment estimation for the longitudinal correlation index parameter $\rho$

The estimation of the regression parameter $\beta$ discussed in Sect. 3.2 and of overdispersion parameter $\sigma^2_\tau$ discussed in Sect. 3.3 require the longitudinal correlation index parameter $\rho$ to be known. We show in this section that the $\rho$ parameter can be estimated by solving an unbiased moment equation that leads to a consistent estimator. Notice from (7) and (8) that the variances and the lag 1 covariances of the repeated counts under the present model have the formulas

$$
\begin{aligned}
E[(Y_{ij} - \mu_{ij})^2] &= \sigma_{ijj} = \mu_{ij} + \mu^2_{ij}(\exp(\sigma^2_\tau) - 1), \quad \text{and} \\
E[(Y_{ij} - \mu_{ij})(Y_{i,j+1} - \mu_{i,j+1})] &= \sigma_{i,j,j+1} = \rho\mu_{ij} \\
&\quad + \mu_{ij}\mu_{i,j+1}(\exp(\sigma^2_\tau) - 1),
\end{aligned} \tag{37}
$$

respectively. Let $\tilde{y}_{ij} = (y_{ij} - \bar{\mu}_{ij})/(\bar{\sigma}_{ijj})^{1/2}$. As shown in Appendix A.4, the moment estimator of the correlation index parameter involved in (37) has the formula

$$\hat{\rho} = \frac{\bar{a}_1 - \bar{b}_1}{\bar{g}_1}, \tag{38}$$

where

$$\bar{a}_1 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i - 1} \tilde{y}_{ij} \tilde{y}_{i,j+1} / \sum_{i=1}^{K} (n_i - 1)}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} \tilde{y}_{ij}^2 / \sum_{i=1}^{K} n_i} \tag{39}$$

$$\bar{g}_1 = \sum_{i=1}^{K} \sum_{j=1}^{n_i - 1} \bar{\mu}_{ij} (\bar{\sigma}_{ijj} \bar{\sigma}_{i,j+1,j+1})^{-1/2} \Big/ \sum_{i=1}^{K} (n_i - 1) \tag{40}$$

and

$$\bar{b}_1 = (\exp(\sigma_\tau^2) - 1) \sum_{i=1}^{K} \sum_{j=1}^{n_i - 1} \bar{\phi}_{ij} \bar{\phi}_{i,j+1} \Big/ \sum_{i=1}^{K} (n_i - 1), \tag{41}$$

with $\bar{\phi}_{ij} = \bar{\mu}_{ij}/(\bar{\sigma}_{ijj})^{1/2}$.

Note that the estimation of the non-parametric function $\gamma(\cdot)$ (Sect. 3.1), regression effects $\beta$ (Sect. 3.2), overdispersion component $\sigma_\tau^2$ (Sect. 3.3), and the longitudinal correlation index parameter $\rho$ by (38) is carried out in cycles of iteration until convergence.

## 3.5 Asymptotic results

### *3.5.1 Asymptotic properties of the SQL estimator of $\gamma(\cdot)$*

Note that the SQL estimating Eq. (13) (see also (17)) is an extension of the well-known QL estimating equation (Wedderburn 1974). This estimating equation, which is free of $\rho$, is written by exploiting the means and the variances of the responses, variance being a function of the mean in the present GLMM setup, by treating the repeated responses of an individual as independent. The non-parametric function $\gamma(q_{\ell u})$ has to be evaluated for all $u = 1, \ldots, n_l$; and $\ell = 1, \ldots, K$. For convenience, in (13), we have shown the estimation for $\gamma(q_0)$ for $q_0 \equiv q_{\ell u}$ for a selected value of $\ell$ and $u$. Recall from (17) that $\hat{\gamma}(q_0; \beta, \sigma_\tau^2)$ is an SQL estimator of $\gamma(q_0)$. It is shown in Appendix B.1 that

$$\hat{\gamma}(q_0; \beta, \sigma_\tau^2) - \gamma(q_0) = O(b^2) + o_p(1/\sqrt{K}), \tag{42}$$

where $b \propto K^{-\alpha}$ for a suitable value for $\alpha$. Consequently, for $\sqrt{K}$-consistency of $\hat{\gamma}(q_0; \beta, \sigma_\tau^2)$, we need to have $Kb^4 \to 0$ as $K \to \infty$, which requires $1/4 < \alpha \leq 1/3$.

*3.5.2 Asymptotic properties of the SGQL estimator of $\beta$*

Notice that in the SGQL estimating Eq. (22) for $\beta$, we have used $\hat{\gamma}(\beta, \sigma_\tau^2)$ for $\hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2)$ by suppressing $q_{ij}$ for notational simplicity. By the same token, $\hat{\gamma}(\beta, \sigma_\tau^2)$ used to define $\bar{\mu}_i(\cdot)$ and $\bar{\Sigma}_i$, refers to using all values of $\hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2)$ for $j = 1, \ldots, n_i$. Suppose we express all these $n_i$ values of the function as

$$\hat{\gamma}(q_i; \beta, \sigma_\tau^2) \equiv [\hat{\gamma}(q_{i1}; \beta, \sigma_\tau^2), \ldots, \hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2), \ldots, \hat{\gamma}(q_{in_i}; \beta, \sigma_\tau^2)].$$

Then by (22), for true $\beta$, define

$$\begin{aligned} D_K(\beta) = \frac{1}{K} \sum_{i=1}^{K} & \frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \beta} \bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2), \rho) \\ & \times [Y_i - \bar{\mu}_i(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))]. \end{aligned}$$

Because the SGQL estimator $\hat{\beta}$ of $\beta$ obtained from (22) satisfy $D_K(\hat{\beta}) = 0$, a linear Taylor expansion about true $\beta$ provides

$$D_K(\beta) + (\hat{\beta} - \beta) D_K'(\beta) + o_p(1/\sqrt{K}) = 0, \tag{43}$$

yielding

$$\begin{aligned} \hat{\beta} - \beta &= -[D_K'(\beta)]^{-1}[D_K(\beta) + o_p(1/\sqrt{K})] \\ &= F_K^{-1}(\beta) D_K(\beta) + o_p(1/\sqrt{K}), \end{aligned} \tag{44}$$

where

$$\begin{aligned} F_K(\beta) = \frac{1}{K} \sum_{i=1}^{K} & \frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \beta} \bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2), \rho) \\ & \times \frac{\partial \bar{\mu}_i(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \beta'}. \end{aligned}$$

Next, for a suitable covariance matrix $V_\beta$ as defined in Appendix B.2, it is shown in the same appendix that

$$\sqrt{K}\{\hat{\beta} - \beta - O(b^2)\} \to N(0, V_\beta) \tag{45}$$

as$K \to \infty$. Further because $b \propto K^{-\alpha}$, it follows from (45) that for $\sqrt{K}$-consistency of $\hat{\beta}$, we need to have $Kb^4 \to 0$ as $K \to \infty$, which happens when $1/4 < \alpha \leq 1/3$ (see Lin and Carroll 2001, for example, for upper limit).

*3.5.3 Asymptotic properties of the SGQL estimator of $\sigma_\tau^2$*

Notice that the SGQL estimating Eq. (22) for $\beta$ has the form

$$\sum_{i=1}^{K} \frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2))}{\partial \beta} \bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\beta, \sigma_\tau^2))(y_i - \bar{\mu}_i(\beta, \sigma_\tau^2, \hat{\gamma}(\beta, \sigma_\tau^2))) = 0,$$

whereas the SGQL estimating Eq. (28) for $\sigma_\tau^2$ has a similar but different form given by

$$M_K = \sum_{i=1}^{K} \frac{\partial \bar{\lambda}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))}{\partial \sigma_\tau^2} \bar{\Omega}_i^{-1}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\cdot))(u_i - \bar{\lambda}_i(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))) = 0. \quad (46)$$

One of the big differences between these estimating equations lies in the fact that even though the non-parametric function estimate $\hat{\gamma}(\cdot)$ involved in both equations is a function of the first-order response $\{y_{ij}\}$ (see Eq. 17), the estimating equation for $\beta$ is constructed using the first-order response vector $y_i$, whereas $u_i$ used to construct the estimating equation for $\sigma_\tau^2$ is a vector of second-order (squared) responses. This difference has to be accommodated when asymptotic properties of the SGQL estimator of $\sigma_\tau^2$ is derived following the asymptotic properties of $\hat{\beta}$ given in the last section.

Because the SGQL estimator $\hat{\sigma}_\tau^2$ of $\sigma_\tau^2$ obtained from (46) satisfies $M_K(\hat{\sigma}_\tau^2) = 0$, a linear Taylor series expansion, similar to (44) for $\beta$ estimation, about $\sigma_\tau^2$ provides

$$\sqrt{K}\{\hat{\sigma}_\tau^2 - \sigma_\tau^2\} = L_K^{-1}\{\sqrt{K}\, M_K\} + o_p(1), \quad (47)$$

where

$$L_K = \frac{1}{K} \sum_{i=1}^{K} \frac{\partial \bar{\lambda}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))}{\partial \sigma_\tau^2} \bar{\Omega}_i^{-1}(\beta, \sigma_\tau^2, \rho, \hat{\gamma}(\cdot)) \frac{\partial \bar{\lambda}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(\cdot))}{\partial \sigma_\tau^2}. \quad (48)$$

Next for a variance quantity $V_{\sigma_\tau^2}$ defined in Appendix B.3, it is shown under the same appendix that

$$\sqrt{K}\{\hat{\sigma}_\tau^2 - \sigma_\tau^2 - O(b^2)\} \xrightarrow{D} N(0, V_{\sigma_\tau^2}) \quad \text{as } K \to \infty. \quad (49)$$

Similar to the condition for the $\sqrt{K}$-consistency of $\hat{\beta}$, for $\sqrt{K}$-consistency of $\hat{\sigma}_\tau^2$, we need to have $O(b^2) = Kb^4 \to 0$ as $K \to \infty$, that is, $1/4 < \alpha \leq 1/3$.

*3.5.4 Consistency of the moment estimator of $\rho$*

For $Y^*_{ij} = (Y_{ij} - \mu_{ij})/(\sigma_{ijj})^{1/2}$, it follows that

$$E({Y^*_{ij}}^2) = 1 \quad \text{for all } i \text{ and } j \tag{50}$$

$$\Rightarrow \quad E\left[\sum_{j=1}^{n_i}({Y^*_{ij}}^2 - 1)\right] = 0 \quad \text{for all } i = 1, \ldots, K$$

and

$$E\left[\left(\frac{Y_{ij} - \mu_{ij}}{\sqrt{\sigma_{ijj}}}\right)\left(\frac{Y_{i,j+1} - \mu_{i,j+1}}{\sqrt{\sigma_{i,j+1,j+1}}}\right)\right] = \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}} \tag{51}$$

$$\Rightarrow \quad E\left[\sum_{j=1}^{n_i-1}\left(Y^*_{ij}Y^*_{i,j+1} - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right)\right] = 0 \quad \text{for all } i = 1, \ldots, K.$$

Now because $E\left[\left(\sum_{j=1}^{n_i-1}\left[Y^*_{ij}Y^*_{i,j+1} - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]\right)^2\right]$ in (51) and $E\left[\left(\sum_{j=1}^{n_i}\left[{Y^*_{ij}}^2 - 1\right]\right)^2\right]$ in (50) are all functions of $\mu_{ij}$, $\sigma^2_\tau$ and $\rho$, they are bounded under the assumption that $\mu_{ij}$ and $n_i$ are all bounded. Thus, for a sufficiently large but finite $m_0$, one may write

$$E\left[\left(\sum_{j=1}^{n_i-1}\left[Y^*_{ij}Y^*_{i,j+1} - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]\right)^2\right] < m_0,$$

$$\text{and also } E\left[\left(\sum_{j=1}^{n_i}[{Y^*_{ij}}^2 - 1]\right)^2\right] < m_0, \tag{52}$$

for all $i = 1, \ldots, K$. When this condition in (52) is satisfied, and because $Y_{ij}$'s are independent for different $i$, it follows from the law of large numbers for independent random variables (Breiman 1968, Theorem 3.27) that

$$\frac{\sum_{i=1}^K\sum_{j=1}^{n_i-1}\left(y^*_{ij}y^*_{i,j+1} - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right)}{\sum_{i=1}^K(n_i-1)} \xrightarrow{P} 0$$

$$\Rightarrow \frac{\sum_{i=1}^K\sum_{j=1}^{n_i-1} y^*_{ij}y^*_{i,j+1}}{\sum_{i=1}^K(n_i-1)} = \frac{\rho\sum_{i=1}^K\sum_{j=1}^{n_i-1}\frac{\mu_{ij}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^K(n_i-1)}$$

$$+ \frac{(\exp(\sigma^2_\tau) - 1)\sum_{i=1}^K\sum_{j=1}^{n_i-1}\frac{\mu_{ij}\mu_{i,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^K(n_i-1)} + o_p(1), \tag{53}$$

and

$$\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij}^{*\,2}-1)}{\sum_{i=1}^{K}n_i} \xrightarrow{P} 0$$
$$\Rightarrow \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}y_{ij}^{*\,2}}{\sum_{i=1}^{K}n_i} = 1+o_p(1). \tag{54}$$

Next, dividing (53) by (54), we obtain

$$a_1(1+o_p(1)) = \rho\, g_1 + b_1 + o_p(1) \Rightarrow \tilde{\rho} = \frac{a_1-b_1}{g_1} = \rho + o_p(1) \quad \text{as } K \to \infty,$$

where $a_1$, $b_1$ and $g_1$ are given by (69), (72) and (71), respectively under Appendix A.4. So

$$\tilde{\rho} = \frac{a_1-b_1}{g_1} \xrightarrow{P} \rho \quad \text{as } K \to \infty. \tag{55}$$

Note that this consistency result in (55) remains valid when $\gamma(\cdot)$ in $\mu_{ij}$s is replaced by its consistent estimate $\hat{\gamma}(\cdot)$. Thus, following (55), $\hat{\rho}$ in (38) is consistent for $\rho$.

## 4 A simulation study

We generate four sets of data under the proposed SGLLMM (3) as follows:
**Step 1. Parameters selection:** We consider the following four sets of parameter values.
**Set 1:** $(\beta_1, \beta_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 0.5$, $\rho = 0.5$;
**Set 2:** $(\beta_1, \beta_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 0.5$, $\rho = 0.8$;
**Set 3:** $(\beta_1, \beta_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 1.0$, $\rho = 0.5$;
**Set 4:** $(\beta_1, \beta_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 1.0$, $\rho = 0.8$.

**Step 2. Primary covariate selection:** For the primary covariate selection, we choose $n_i = 4$ equi-spaced time points for all $i = 1, \ldots, K$, with $K = 100$. Next, because $\beta = (\beta_1, \beta_2)'$ is the effect of two time $(j)$-dependent primary covariates, we choose these covariates as

$$x_{ij1}(j) = \begin{cases} \frac{1}{2} & \text{for } i = 1, \ldots, 25 \text{ and } j = 1, 2 \\ 1 & \text{for } i = 1, \ldots, 25 \text{ and } j = 3, 4 \\ -\frac{1}{2} & \text{for } i = 26, \ldots, 75 \text{ and } j = 1 \\ 0 & \text{for } i = 26, \ldots, 75 \text{ and } j = 2, 3 \\ \frac{1}{2} & \text{for } i = 26, \ldots, 75 \text{ and } j = 4 \\ \frac{j}{2n_i} & \text{for } i = 76, \ldots, 100 \text{ and } j = 1, 2, 3, 4 \end{cases}$$

and

$$x_{ij2}(j) = \begin{cases} \frac{j-2.5}{2n_i} & \text{for } i = 1, \ldots, 50 \text{ and } j = 1, 2, 3, 4 \\ 0 & \text{for } i = 51, \ldots, 100 \text{ and } j = 1, 2 \\ \frac{1}{2} & \text{for } i = 51, \ldots, 100 \text{ and } j = 3, 4. \end{cases}$$

Note that these covariate values are also available from Sutradhar (2010, p. 188). These values are chosen to reflect the variable time dependence for the different groups of individuals. Thus, the choice is quite general. One may choose other specific covariates depending on the situations.

**Step 3. Random effects generation:** The random effects $\tau_i$ for $i = 1, \ldots, 100$, are generated from $N(0, 1)$ distribution.

**Step 4. Secondary covariate selection:** For a given $i (i = 1, \ldots, 100)$, we choose a value for $q_{ij}$ from a uniform ($U$) distribution, namely

$$q_{ij} \sim U[j - 0.5, j + 0.5], \tag{56}$$

for $j = 1, \ldots, n_i = 4$. Note that for each $j = 1, \ldots, 4$, the interval $[j-0.5, j+0.5]$ was divided into 25 (alternatively it could be 50 or 100, and so on) equi-spaced points allowing one value to be chosen from 25 values. Thus, altogether $n_i = 4$ values were chosen from $j$-related four intervals. This was independently repeated for $K = 100$ individuals. Consequently, these 400 values are expected to be dense and they reflect the time dependence. One may select other distribution to choose $q_{ij}$. In practice, these values are available as a part of the data.

**Step 5. Non-parametric function selection:** We chose, for example, a quadratic non-parametric function given by

$$\gamma(q_{ij}) = 0.3 + 0.2\left(q_{ij} - \frac{n_i + 1}{2}\right) + 0.05\left(q_{ij} - \frac{n_i + 1}{2}\right)^2 \tag{57}$$

with $n_i = 4$, where $q_{ij}$ is generated by (56). Remark that in practice this non-parametric function influencing $y_{ij}$ would be unknown.

**Step 6. Data generation:** The repeated counts $\{y_{ij}, j = 1, \ldots, n_i; i = 1, \ldots, K\}$ are then generated by the SGLLMM (3).

Note that when data are generated under the present SGLLMM (or SMM in brief) (3) following the aforementioned six steps, but one ignores the presence of non-parametric function in the model and makes an attempt to estimate the parameters ($\beta$, $\sigma^2_\tau$, and $\rho$) by treating the data as though they were generated from the GLLMM, the estimates are bound to be biased. We examine the performance of such naive GQL (NGQL) estimators by repeating the data generation 1000 times and computing the simulated mean (SM), simulated standard error (SSE), and simulated mean squared error (SMSE) of the NGQL estimates for $\beta$ and $\sigma^2_\tau$, and moment estimate of $\rho$. The parameter values and their simulated estimates are shown in Table 1.

**Table 1** Simulated means (SMs), simulated standard errors (SSEs) and mean squared errors (MSEs) of NGQL estimates (ignoring the presence of non-parametric function) of regression parameters $\beta$ and random effects variance $\sigma_\tau^2$ under non-stationary AR(1) correlation model (3) for selected values of correlation index parameter $\rho$ with $K = 100$, $n_i = 4$; based on 1000 simulations

| True $\beta = (\beta_1, \beta_2)'$ | $\sigma_\tau^2$ | $\rho$ | Quantity | $\hat{\beta}_{1,\text{NGQL}}$ | $\hat{\beta}_{2,\text{NGQL}}$ | $\hat{\sigma}^2_{\tau,\text{NGQL}}$ | $\hat{\rho}_{\text{Moment}}$ |
|---|---|---|---|---|---|---|---|
| $\beta = (0.5, 0.5)'$ | 0.5 | 0.5 | SM | 0.9483 | 1.1209 | 0.6829 | 0.1850 |
| | | | SSE | 0.1199 | 0.1778 | 0.1468 | 0.1640 |
| | | | MSE | 0.2153 | 0.4171 | 0.0550 | |
| | | 0.8 | SM | 0.9638 | 1.1134 | 0.6836 | 0.5261 |
| | | | SSE | 0.0995 | 0.1550 | 0.1445 | 0.1540 |
| | | | MSE | 0.2250 | 0.4003 | 0.0545 | |
| | 1.0 | 0.5 | SM | 0.9669 | 1.0962 | 1.1364 | 0.1100 |
| | | | SSE | 0.1227 | 0.1751 | 0.2656 | 0.1662 |
| | | | MSE | 0.2331 | 0.3861 | 0.0890 | |
| | | 0.8 | SM | 0.9704 | 1.0957 | 1.1461 | 0.3337 |
| | | | SSE | 0.1031 | 0.1469 | 0.2960 | 0.2288 |
| | | | MSE | 0.2319 | 0.3764 | 0.1088 | |

As expected, the results in Table 1 show that the estimates of $\beta$ and $\sigma_\tau^2$ are highly biased. For example, when $\rho = 0.5$, for the true regression parameter $\beta = (0.5, 0.5)'$ and random effects variance $\sigma_\tau^2 = 0.5$, the estimated values of $\beta$ and $\sigma_\tau^2$ are found to be $(0.9483, 1.1209)'$ and 0.6829, respectively. The estimate for $\rho = 0.5$ was found to be 0.185. Clearly all of these naive estimates computed by ignoring the non-parametric function $\gamma(q_{ij})$ are useless, and hence one must take the non-parametric function $\gamma(q_{ij})$ into account in estimating these regression, overdispersion and correlation index parameters. This will require the consistent estimation of the non-parametric function as well, which was discussed in Sect. 3.1.

We now examine the performance of the proposed semi-parametric estimation approach discussed in Sect. 3 for the estimation of the non-parametric function $\gamma(q_{ij})$, and the parameters ($\beta, \sigma_\tau^2,$ and $\rho$). The overdispersion parameter $\sigma_\tau^2$ was estimated using squared response-based exact, corrected squared (CR) response-based CR-exact, and CR-normal (CRN) techniques as discussed in Sect. 3.3. We also examine the performance of the SGQL approach by pretending that the correlation index parameter $\rho$ is zero. Remark that in the present setup, $\rho = 0$ does not mean the repeated responses are independent. The independence follows when both $\rho = 0$ and $\sigma_\tau^2 = 0$. All estimates (simulated mean, SM) along with their standard errors (SSE) and mean square errors (MSE) are obtained based on 1000 simulations. The results are provided in Table 2 for $\beta$, $\sigma_\tau^2$ and $\rho$ parameters. The SQL estimate for the non-parametric function $\gamma(\cdot)$ is displayed in Fig. 1. Note that this estimate uses $\sigma_\tau^2$ estimated by the exact weight matrix discussed above. As discussed in Sect. 3.1, the bandwidth parameter $b$ for $\gamma(\cdot)$ estimation is chosen as $K^{-1/5}$. We have also verified this bandwidth choice by an ad hoc mini-max approach discussed in Appendix C.

Figure 1 shows that the SQL approach estimates the true non-parametric curve well. The estimated curve almost coincides with the true curve when overdispersion index

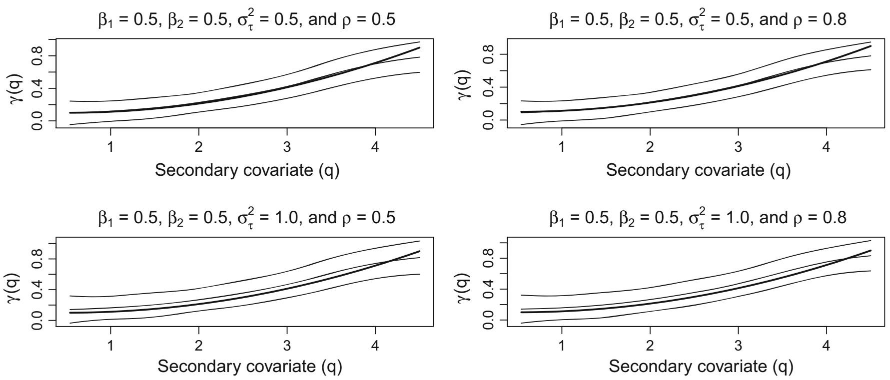


**Fig. 1** The plot for $\gamma(\cdot)$ estimation for the approach with $\sigma_\tau^2$ estimated by the exact weight matrix given in Sect. 3.3.1. The thick curve is for the true $\gamma(\cdot)$ function value. The *thinner curves* are for the estimated $\gamma(\cdot)$ value and one standard error. The bandwidth $b = K^{-1/5}$

parameter is small, that is, $\sigma_\tau^2 = 0.5$. This holds for small and large correlation index parameter ($\rho$) values. The curve estimate is not so satisfactory when $\sigma_\tau^2 = 1.0$. This happens because $\sigma_\tau^2 = 1.0$ produces large overdispersion in the data and as the results of Table 2 show, the estimates $\hat{\sigma}_\tau^2$ are slightly biased when $\sigma_\tau^2 = 1.0$.

Next, the results from Table 2 show that the main regression parameters $\beta_1 = \beta_2 = 0.5$ are estimated very well by the proposed SGQL approach irrespective of the SGQL approaches (Approx ($\rho = 0$), Exact, CR-exact or CR-normal) used for the estimation of $\sigma_\tau^2$. This estimation pattern holds whether correlation index $\rho$ is small (0.5) or large (0.8). For example, for large $\rho = 0.8$ and small $\sigma_\tau^2 = 0.5$ (estimated by exact weight-based approach), the SGQL estimates of $\beta = (\beta_1, \beta_2)' \equiv (0.5, 0.5)'$ are $(0.4940, 0.4844)'$ with MSEs $(0.0165, 0.0516)'$. The estimates are similar even when $\sigma_\tau^2$ is large (1.0). Specifically the estimates are $(0.4922, 0.4701)'$ with MSEs $(0.0163, 0.0471)'$. As far as the estimation of correlation parameters $\sigma_\tau^2$ and $\rho$ is concerned, the SGQL approaches for $\sigma_\tau^2$ and the method of moments for $\rho$ work well whether $\sigma_\tau^2$ is small or large. For example, when $\rho = 0.8$, for $\sigma_\tau^2 = 0.5$, the CR-normal weight-based SGQL approach produces the estimate $\hat{\sigma}_\tau^2 = 0.474$ with MSE 0.0163, and the method of moments yields $\rho$ estimate as 0.770 with SSE 0.0870. For large $\sigma_\tau^2 = 1.0$, the CR-normal weight-based SGQL approach produces the estimate $\hat{\sigma}_\tau^2 = 0.8841$ with MSE 0.0720, and the method of moments yields $\rho$ estimate as 0.740 with SSE 0.1446. Thus, the simulation study suggests that the proposed estimation approaches perform very well whether the overdispersion is small or relatively large.

## 5 Concluding remarks

In the past, many authors such as Breslow and Clayton (1993), Breslow and Lin (1995), and Lin and Breslow (1996) have studied the inference techniques for the parameters involved in the generalized linear mixed models (GLMMs) for binary and count data. To be specific, Breslow and Clayton (1993), among others, have used a best linear unbiased prediction (BLUP) analog estimation approach (also known as the

**Table 2** Simulated means (SMs), simulated standard errors (SSEs) and mean squared errors (MSEs) of the SGQL estimates of regression parameters $\beta$ and the exact, CR-exact, CRN weight matrix-based SGQL estimates for the random effects variance $\sigma_\tau^2$ under non-stationary AR(1) correlation model (3) for selected values of correlation index parameter $\rho$ with $K = 100$, $n_i = 4$; based on 1000 simulations

| True $\beta = (\beta_1, \beta_2)'$ | $\sigma_\tau^2$ | $\rho$ | Method | Quantity | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}_\tau^2$ | $\hat{\rho}$ |
|---|---|---|---|---|---|---|---|---|
| $\beta = (0.5, 0.5)'$ | 0.5 | 0.5 | Exact | SM | 0.4947 | 0.4881 | 0.4899 | 0.4594 |
| | | | | SSE | 0.1576 | 0.2907 | 0.1252 | 0.1254 |
| | | | | MSE | 0.0249 | 0.0846 | 0.0158 | |
| | | | Approx ($\rho = 0$) | SM | 0.4947 | 0.4876 | 0.4803 | 0.4710 |
| | | | | SSE | 0.1576 | 0.2909 | 0.1264 | 0.1180 |
| | | | | MSE | 0.0248 | 0.0847 | 0.0164 | |
| | | | CR-exact | SM | 0.4942 | 0.4872 | 0.4737 | 0.4771 |
| | | | | SSE | 0.1576 | 0.2909 | 0.1265 | 0.1221 |
| | | | | MSE | 0.0248 | 0.0847 | 0.0167 | |
| | | | CRN | SM | 0.4941 | 0.4868 | 0.4694 | 0.4824 |
| | | | | SSE | 0.1575 | 0.2908 | 0.1249 | 0.1162 |
| | | | | MSE | 0.0248 | 0.0847 | 0.0165 | |
| | | 0.8 | Exact | SM | 0.4940 | 0.4844 | 0.5013 | 0.7503 |
| | | | | SSE | 0.1283 | 0.2267 | 0.1269 | 0.0959 |
| | | | | MSE | 0.0165 | 0.0516 | 0.0161 | |
| | | | Approx ($\rho = 0$) | SM | 0.4941 | 0.4838 | 0.4792 | 0.7684 |
| | | | | SSE | 0.1284 | 0.2269 | 0.1291 | 0.0839 |
| | | | | MSE | 0.0165 | 0.0517 | 0.0171 | |
| | | | CR-exact | SM | 0.4937 | 0.4843 | 0.4784 | 0.7657 |
| | | | | SSE | 0.1281 | 0.2271 | 0.1265 | 0.0918 |
| | | | | MSE | 0.0164 | 0.0517 | 0.0165 | |
| | | | CRN | SM | 0.4936 | 0.4841 | 0.4739 | 0.7689 |
| | | | | SSE | 0.1280 | 0.2269 | 0.1250 | 0.0870 |
| | | | | MSE | 0.0164 | 0.0517 | 0.0163 | |
| | 1.0 | 0.5 | Exact | SM | 0.5032 | 0.4896 | 0.8823 | 0.4470 |
| | | | | SSE | 0.1660 | 0.2809 | 0.2263 | 0.1891 |
| | | | | MSE | 0.0275 | 0.0789 | 0.0650 | |
| | | | Approx ($\rho = 0$) | SM | 0.5021 | 0.4898 | 0.8845 | 0.4569 |
| | | | | SSE | 0.1665 | 0.2811 | 0.2346 | 0.1881 |
| | | | | MSE | 0.0277 | 0.0790 | 0.0683 | |
| | | | CR-exact | SM | 0.5017 | 0.4894 | 0.8768 | 0.4595 |
| | | | | SSE | 0.1665 | 0.2814 | 0.2327 | 0.1901 |
| | | | | MSE | 0.0277 | 0.0792 | 0.0693 | |
| | | | CRN | SM | 0.5017 | 0.4881 | 0.8743 | 0.4673 |
| | | | | SSE | 0.1662 | 0.2805 | 0.2336 | 0.1852 |
| | | | | MSE | 0.0276 | 0.0787 | 0.0703 | |

**Table 2** continued

| True $\beta = (\beta_1, \beta_2)'$ | $\sigma_\tau^2$ | $\rho$ | Method | Quantity | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\sigma}_\tau^2$ | $\hat{\rho}$ |
|---|---|---|---|---|---|---|---|---|
| | | 0.8 | Exact | SM | 0.4922 | 0.4701 | 0.8989 | 0.7221 |
| | | | | SSE | 0.1275 | 0.2152 | 0.2461 | 0.1545 |
| | | | | MSE | 0.0163 | 0.0471 | 0.0707 | |
| | | | Approx ($\rho = 0$) | SM | 0.4918 | 0.4686 | 0.8883 | 0.7368 |
| | | | | SSE | 0.1278 | 0.2152 | 0.2402 | 0.1456 |
| | | | | MSE | 0.0164 | 0.0472 | 0.0701 | |
| | | | CR-exact | SM | 0.4921 | 0.4690 | 0.8875 | 0.7331 |
| | | | | SSE | 0.1277 | 0.2152 | 0.2384 | 0.1500 |
| | | | | MSE | 0.0163 | 0.0472 | 0.0694 | |
| | | | CRN | SM | 0.4911 | 0.4685 | 0.8841 | 0.7399 |
| | | | | SSE | 0.1280 | 0.2156 | 0.2421 | 0.1446 |
| | | | | MSE | 0.0165 | 0.0474 | 0.0720 | |

The non-parametric function $\gamma(\cdot)$ is estimated by SQL approach and $\rho$ is estimated using method of moments in all cases. The bandwidth $b = K^{-1/5} = 0.3981072$

PQL (penalized quasi-likelihood approach)), where random family effects are treated to be the fixed effects and the regression and variance components of the GLMMs are estimated based on the so-called estimates of the random effects. As opposed to the BLUP analog approach of Breslow and Clayton (1993), Jiang and Zhang (2001) have suggested an improved method of moments. It, however, follows from Sutradhar (2004) that the estimators obtained based on the improved method of moments (IMM) may also be highly inefficient as compared to the estimators obtained based on a generalized quasi-likelihood (GQL) approach. The GQL estimators are consistent and highly efficient, the exact maximum likelihood estimators being fully efficient (i.e., optimal) which are, however, known to be cumbersome to compute, specially in the GLLMM (generalized linear longitudinal mixed model) setup. For GQL inferences in the GLLMM setup for count and binary data, we refer to Sutradhar et al. (2008, 2014), respectively, among others.

In this paper, we have extended the GLLMM for count data to the semi-parametric setup. The proposed model is referred to as the semi-parametric GLLMM (SGLLMM) (3) which has been discussed in Sect. 2. This SGLLMM may also be considered as a generalization of the semi-parametric generalized linear longitudinal model (SGLLM) studied, for example, by Severini and Staniswalis (1994), Lin and Carroll (2001), You and Chen (2007) and Warriyar and Sutradhar (2014), to the mixed model setup. We have developed an SQL approach for the estimation of the non-parametric function, an SGQL approach for the estimation of main regression and overdispersion parameters, whereas the longitudinal correlation index parameter has been estimated using the well-known method of moments. These estimation methods along with the asymptotic properties of the estimators are discussed in Sect. 3. A simulation study conducted in Sect. 4 indicates that the proposed estimation approaches perform very well whether the overdispersion parameter is small or large.

**Acknowledgements** We thank the Editor and two reviewers for their constructive and helpful comments on the previous version of the paper. This research was partially supported by a Grant from NSERC.

## Appendix A. Aids for analytical estimation

### Appendix A.1. Computation of the derivative matrix in the SGQL estimating Eq. (22) for $\beta$

Note that the computation of the derivative matrix $\frac{\partial \bar{\mu}_i'(\beta,\sigma_\tau^2,\hat{\gamma}(\beta,\sigma_\tau^2))}{\partial \beta}$ in (22) requires the formula for the derivative $\frac{\hat{\gamma}(\beta,\sigma_\tau^2)}{\partial \beta}$, whereas this derivative would have been zero if $\gamma(\cdot)$ was known. The exact formula for the gradient matrix $\frac{\partial \bar{\mu}_i'(\beta,\sigma_\tau^2,\hat{\gamma}(\beta,\sigma_\tau^2))}{\partial \beta}$ may be computed as

$$\begin{aligned}&\frac{\partial \bar{\mu}_i'(\beta,\sigma_\tau^2,\hat{\gamma}(\beta,\sigma_\tau^2))}{\partial \beta}\\ &\quad=\frac{\partial(\bar{\mu}_{i1}(\beta,\sigma_\tau^2,\hat{\gamma}(\beta,\sigma_\tau^2)),\ldots,\bar{\mu}_{in_i}(\beta,\sigma_\tau^2,\hat{\gamma}(\beta,\sigma_\tau^2)))}{\partial \beta},\end{aligned}\tag{58}$$

where for $j=1,\ldots,n_i$, one obtains

$$\begin{aligned}&\frac{\partial \bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_{ij};\beta,\sigma_\tau^2))}{\partial \beta}\\ &\quad=\bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_{ij};\beta,\sigma_\tau^2))\left[x_{ij}+\frac{\partial \hat{\gamma}(q_{ij};\beta,\sigma_\tau^2)}{\partial \beta}\right].\end{aligned}\tag{59}$$

Now to compute the derivative $\frac{\partial \hat{\gamma}(q_{ij};\beta,\sigma_\tau^2)}{\partial \beta}$ for (59), we turn back to the estimating Eq. (17) for $\gamma(q_0)$, and take its derivative with respect to $\beta$ and obtain

$$\begin{aligned}&\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(q_0)\left\{\frac{\left(\exp(\sigma_\tau^2)-1\right)y_{ij}+1}{\left[1+\bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_0;\beta,\sigma_\tau^2))\left(\exp(\sigma_\tau^2)-1\right)\right]^2}\right\}\\ &\quad\times\hat{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_0;\beta,\sigma_\tau^2))\left[x_{ij}+\frac{\partial \hat{\gamma}(q_0;\beta,\sigma_\tau^2)}{\partial \beta}\right]=0,\end{aligned}$$

yielding

$$\begin{aligned}&\frac{\partial \hat{\gamma}(q_0;\beta,\sigma_\tau^2)}{\partial \beta}\\ &=\frac{-\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(q_0)\left\{\frac{1+y_{ij}\left(\exp(\sigma_\tau^2)-1\right)}{\left[1+\bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_0;\beta,\sigma_\tau^2))\left(\exp(\sigma_\tau^2)-1\right)\right]^2}\right\}\bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_0;\beta,\sigma_\tau^2))\,x_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(q_0)\left\{\frac{1+y_{ij}\left(\exp(\sigma_\tau^2)-1\right)}{\left[1+\bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_0;\beta,\sigma_\tau^2))\left(\exp(\sigma_\tau^2)-1\right)\right]^2}\right\}\bar{\mu}_{ij}(\beta,\sigma_\tau^2,\hat{\gamma}(q_0;\beta,\sigma_\tau^2))}.\end{aligned}\tag{60}$$

**Appendix A.2. Computation of the derivative vector in the raw SGQL estimating Eq. (28) for $\sigma_\tau^2$**

Recall the squared response vector $\bar{\lambda}_i \equiv \bar{\lambda}_i(\beta, \sigma_\tau^2, \gamma(\cdot))$ from (27) and write

$$\begin{aligned}
\frac{\partial \bar{\lambda}_i'}{\partial \sigma_\tau^2} &= \frac{\partial(\bar{\lambda}_{i11}, \ldots, \bar{\lambda}_{ijj}, \ldots, \bar{\lambda}_{in_in_i})}{\partial \sigma_\tau^2}, \text{ with} \\
\frac{\partial \bar{\lambda}_{ijj}}{\partial \sigma_\tau^2} &= \frac{\partial\left(\bar{\mu}_{ij} + \bar{\mu}_{ij}^2 \exp(\sigma_\tau^2)\right)}{\partial \sigma_\tau^2} \\
&= \frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} + 2\bar{\mu}_{ij}\left(\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2}\right)\exp(\sigma_\tau^2) + \bar{\mu}_{ij}^2 \exp(\sigma_\tau^2).
\end{aligned} \tag{61}$$

In (61),

$$\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} = \bar{\mu}_{ij}\left[\frac{1}{2} + \frac{\partial \hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2)}{\partial \sigma_\tau^2}\right]. \tag{62}$$

Next by labeling $\bar{\mu}_{ij}$ with $\bar{\mu}_{ij}(q_{ij})$, we obtain

$$\begin{aligned}
\frac{\partial \hat{\gamma}(q_0; \beta, \sigma_\tau^2)}{\partial \sigma_\tau^2} &= -\frac{1}{2} \\
&\quad - \exp(\sigma_\tau^2)\left[\frac{\sum_{i=1}^K \sum_{j=1}^{n_i} w_{ij}(q_0)\left\{\frac{y_{ij} - \bar{\mu}_{ij}(q_0)}{[1+\bar{\mu}_{ij}(q_0)(\exp(\sigma_\tau^2)-1)]^2}\right\}\bar{\mu}_{ij}(q_0)}{\sum_{i=1}^K \sum_{j=1}^{n_i} w_{ij}(q_0)\left\{\frac{1+y_{ij}(\exp(\sigma_\tau^2)-1)}{[1+\bar{\mu}_{ij}(q_0)(\exp(\sigma_\tau^2)-1)]^2}\right\}\bar{\mu}_{ij}(q_0)}\right],
\end{aligned} \tag{63}$$

yielding

$$\begin{aligned}
\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} &\equiv \frac{\partial \bar{\mu}_{ij}(q_{ij})}{\partial \sigma_\tau^2} \\
&= -\exp(\sigma_\tau^2)\left[\frac{\sum_{l=1}^K \sum_{u=1}^{n_l} w_{lu}(q_{ij})\left\{\frac{y_{lu} - \bar{\mu}_{lu}(q_{ij})}{[1+\bar{\mu}_{lu}(q_{ij})(\exp(\sigma_\tau^2)-1)]^2}\right\}\bar{\mu}_{lu}(q_{ij})}{\sum_{l=1}^K \sum_{u=1}^{n_l} w_{lu}(q_{ij})\left\{\frac{1+y_{lu}(\exp(\sigma_\tau^2)-1)}{[1+\bar{\mu}_{lu}(q_{ij})(\exp(\sigma_\tau^2)-1)]^2}\right\}\bar{\mu}_{lu}(q_{ij})}\right] \\
&\quad \times \bar{\mu}_{ij}(q_{ij}).
\end{aligned} \tag{64}$$

The formula for the derivative is obtained using (64) in (61).

**Appendix A.3. Computation of the elements of the covariance matrix $\bar{\Omega}_{iC}$ in the corrected (CR) SGQL estimating Eq. (29) for $\sigma_\tau^2$**

The formulas for the elements of $\bar{\Omega}_{iC}$ may be computed in a manner similar to that for the elements of $\bar{\Omega}_i$ in (28). To be specific, the variance and covariance elements of $\bar{\Omega}_{iC}$ have the formulas

$$\begin{aligned}\mathrm{Var}[(Y_{ij}-\bar{\mu}_{ij})^2] = {} & \bar{\mu}_{ij}^4(\exp(6\sigma_\tau^2)-4\exp(3\sigma_\tau^2)+6\exp(\sigma_\tau^2)-3)\\ & +\bar{\mu}_{ij}^3(6\exp(3\sigma_\tau^2)-12\exp(\sigma_\tau^2)+6)+\bar{\mu}_{ij}^2(7\exp(\sigma_\tau^2)-4)\\ & +\bar{\mu}_{ij}-\bar{\sigma}_{ijj}^2, \end{aligned} \tag{65}$$

and

$$\begin{aligned} & \mathrm{Cov}[(Y_{ij}-\bar{\mu}_{ij})^2,(Y_{ik}-\bar{\mu}_{ik})^2]\\ & = [\bar{\mu}_{ij}^2\bar{\mu}_{ik}(4\rho^{k-j}+1)+\bar{\mu}_{ij}\bar{\mu}_{ik}^2](\exp(3\sigma_\tau^2)-2\exp(\sigma_\tau^2)+1)\\ & \quad +2\rho^{k-j}\bar{\mu}_{ij}^2(\exp(\sigma_\tau^2)-1+\rho^{k-j}\exp(\sigma_\tau^2))+\bar{\mu}_{ij}\bar{\mu}_{ik}[2\rho^{k-j}(\exp(\sigma_\tau^2)-1)+\exp(\sigma_\tau^2)]\\ & \quad +\rho^{k-j}\bar{\mu}_{ij}+\bar{\mu}_{ij}^2\bar{\mu}_{ik}^2(\exp(6\sigma_\tau^2)-4\exp(3\sigma_\tau^2)+6\exp(\sigma_\tau^2)-3)-\bar{\sigma}_{ijj}\bar{\sigma}_{ikk}, \end{aligned} \tag{66}$$

respectively.

### Appendix A.4. Derivation for the moment estimator in (38) for the correlation index parameter $\rho$

Let $y_{ij}^* = (y_{ij}-\mu_{ij})/(\sigma_{ijj})^{1/2}$. It then follows from (6) to (8) that

$$E\left[\frac{\sum_{i=1}^K\sum_{j=1}^{n_i} y_{ij}^{*\,2}}{\sum_{i=1}^K n_i}\right] = 1, \tag{67}$$

and

$$\begin{aligned} E\left[\frac{\sum_{i=1}^K\sum_{j=1}^{n_i-1} y_{ij}^* y_{i,j+1}^*}{\sum_{i=1}^K (n_i-1)}\right] = {} & \frac{\rho\sum_{i=1}^K\sum_{j=1}^{n_i-1}\frac{\mu_{ij}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^K(n_i-1)}\\ & +\frac{\left(\exp(\sigma_\tau^2)-1\right)\sum_{i=1}^K\sum_{j=1}^{n_i-1}\frac{\mu_{ij}\mu_{i,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^K(n_i-1)}. \end{aligned} \tag{68}$$

We now exploit (67) and (68), more specifically we consider the ratio of the quantities within the square brackets in (67) and (68) and denote it by $a_1$ as

$$a_1 = \frac{\sum_{i=1}^K\sum_{j=1}^{n_i-1} y_{ij}^* y_{i,j+1}^* / \sum_{i=1}^K(n_i-1)}{\sum_{i=1}^K\sum_{j=1}^{n_i} y_{ij}^{*\,2} / \sum_{i=1}^K n_i} \tag{69}$$

We may then write a first-order approximate expectation as

$$E\left[a_1\right] = \rho g_1 + b_1, \tag{70}$$

where

$$g_1 = \sum_{i=1}^{K} \sum_{j=1}^{n_i - 1} \mu_{ij} (\sigma_{ijj} \sigma_{i,j+1,j+1})^{-1/2} / \sum_{i=1}^{K} (n_i - 1), \tag{71}$$

and

$$b_1 = (\exp(\sigma_\tau^2) - 1) \sum_{i=1}^{K} \sum_{j=1}^{n_i - 1} \phi_{ij} \phi_{i,j+1} / \sum_{i=1}^{K} (n_i - 1), \tag{72}$$

with $\phi_{ij} = \mu_{ij}/(\sigma_{ijj})^{1/2}$. Next by replacing $\mu_{ij}$, $\sigma_{ijj}$, and $\sigma_{i,j+1,j+1}$ with $\bar{\mu}_{ij}$, $\bar{\sigma}_{ijj}$, and $\bar{\sigma}_{i,j+1,j+1}$, respectively, one can compute $\bar{a}_1, \bar{b}_1$, and $\bar{g}_1$, from $a_1, b_1$, and $g_1$, respectively. Consequently, by (70), we write the moment estimator of $\rho$ as

$$\hat{\rho} = \frac{\bar{a}_1 - \bar{b}_1}{\bar{g}_1}, \tag{73}$$

which is (38).

## Appendix B. Aids for asymptotic results

### Appendix B.1. Consistency of the SQL estimator of the non-parametric function $\gamma(q_0)$

For notational simplicity, here we use $\mu_{ij}(q_0)$ for $\mu_{ij}(\beta, \sigma_\tau, \gamma(q_0))$. Now for known $\beta$ and $\sigma_\tau^2$, and for true mean $\mu_{ij} = \exp(x'_{ij}\beta + \frac{\sigma_\tau^2}{2} + \gamma(q_{ij}))$, a Taylor expansion of (17) gives

$$\begin{aligned} \hat{\gamma}(q_0; \beta, \sigma_\tau^2) - \gamma(q_0) &= A_K \\ &+ \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(q_0) \frac{\mu_{ij} - \mu_{ij}(q_0)}{1 + \mu_{ij}(q_0)\left(\exp(\sigma_\tau^2) - 1\right)}}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(q_0) \frac{\mu_{ij}(q_0)}{1 + \mu_{ij}(q_0)\left(\exp(\sigma_\tau^2) - 1\right)}} + o_p(1/\sqrt{K}) \end{aligned} \tag{74}$$

where

$$A_K = \frac{1}{B_K} \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} p_{ij}(q_0) \frac{y_{ij} - \mu_{ij}}{1 + \mu_{ij}(q_0)\left(\exp(\sigma_\tau^2) - 1\right)}$$

with $B_K = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} p_{ij}(q_0) \frac{\mu_{ij}(q_0)}{1 + \mu_{ij}(q_0)\left(\exp(\sigma_\tau^2) - 1\right)}$, and $p_{ij}(q_0)$ is the short abbreviation for $p_{ij}(\frac{q_0 - q_{ij}}{b})$ defined in (14). The computational details are omitted. The fact that $A_K$ has zero mean and bounded variance implies

$$A_K = O_p(1/\sqrt{K}), \tag{75}$$

according to Theorem 14.4-1 in Bishop et al. (2007). Furthermore, it can be shown that the second term in (74) is in the order of $O(b^2)$. Consequently, using (75) in (74), one obtains $\hat{\gamma}(q_0; \beta, \sigma_\tau^2) - \gamma(q_0) = O(b^2) + o_p(1/\sqrt{K})$ which is (42).

## Appendix B.2. Derivation of the asymptotic distribution of the SGQL estimator of the regression effects $\boldsymbol{\beta}$

Notice from (44) that

$$\begin{aligned}\lim_{K\to\infty} F_K &= E\left[\frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \beta}\, \bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2), \rho)\, \frac{\partial \bar{\mu}_i(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \beta'}\right] \\ &= F, \text{ say.}\end{aligned} \tag{76}$$

Let $Z_{1i} = \frac{\partial \bar{\mu}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \beta} \bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho)$, and $\bar{v}_{1i'}^{j'k'}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho)$ be the $(j', k')$th element of the inverse covariance matrix $\bar{\Sigma}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2), \rho)$. Also let $B_K(q_{i'k'})$ represent $B_K$ in (74) when $q_0$ is replaced by general $q_{i'k'}$, and

$$Z_{2i} = (Z_{2i1}, \cdots, Z_{2in_i})$$

where

$$\begin{aligned}Z_{2ij} = \sum_{i'=1}^{K} \sum_{j'=1}^{n_i} \sum_{k'=1}^{n_i} \frac{1}{B_K(q_{i'k'})} \frac{\partial \bar{\mu}_{i'j'}(\beta, \sigma_\tau^2, \hat{\gamma}(q_{i'j'}; \beta, \sigma_\tau^2))}{\partial \beta}\, \bar{v}_{1i'}^{j'k'}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho) \\ \mu_{i'k'}(\beta, \sigma_\tau^2, \gamma(q_{i'k'}))\, \frac{p_{ij}(q_{i'k'})}{1 + \mu_{ij}(q_{i'k'})(e^{\sigma_\tau^2} - 1)}.\end{aligned}$$

After some algebra, it then follows that $D_K(\beta)$ in (44) reduces to

$$D_K(\beta) = \frac{1}{K}\sum_{i=1}^{K}(Z_{1i} - Z_{2i})(Y_i - \mu_i) + O(b^2) + o_p(1/\sqrt{K}). \tag{77}$$

Hence, using (77) and (76) in (44), one obtains

$$\begin{aligned}\sqrt{K}\{\hat{\beta} - \beta\} = F^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}(Z_{1i} - Z_{2i})\,(Y_i - \mu_i) \\ + O(\sqrt{Kb^4}) + o_p(1).\end{aligned} \tag{78}$$

Next because $E[Y_i - \mu_i] = 0$, and $\text{cov}[Y_i] = \Sigma_i$, using Lindeberg–Feller central limit theorem (Amemiya 1985, Theorem 3.3.6) for independent random variables with non-identical distributions, one obtains

$$\sqrt{K}\{\hat{\beta} - \beta - O(b^2)\} \rightarrow N(0, V_\beta) \tag{79}$$

as in (45). In (79),

$$V_\beta = F^{-1} \frac{1}{K} \left[ \sum_{i=1}^{K} (Z_{1i} - Z_{2i})\, \Sigma_i\, (Z_{1i} - Z_{2i})' \right] F^{-1}.$$

## Appendix B.3. Derivation of the asymptotic distribution of the SGQL estimator of the overdispersion index parameter $\sigma_\tau^2$

Express $M_K$ in (46) as

$$M_K = M_{1K} - M_{2K}, \tag{80}$$

where

$$\begin{aligned} M_{1K} &= \frac{1}{K} \sum_{i=1}^{K} Q_{1i}[u_i - \lambda_i(\beta, \sigma_\tau^2, \gamma(q_i))] \\ M_{2K} &= \frac{1}{K} \sum_{i=1}^{K} Q_{1i}[\bar{\lambda}_i(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2)) - \lambda_i(\beta, \sigma_\tau^2, \gamma(q_i))], \end{aligned} \tag{81}$$

with

$$Q_{1i} = \frac{\partial \bar{\lambda}_i'(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2))}{\partial \sigma_\tau^2}\, \bar{\Omega}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho).$$

For $j, k = 1, \ldots, n_i$, let $v_{2i}^{jk}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho)$ be the $(j, k)$th element of the $n_i \times n_i$ inverse fourth-order moments matrix $\bar{\Omega}_i^{-1}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho)$. Now because

$\bar{\lambda}_i(\cdot)=[\bar{\lambda}_{i1}(\cdot), \ldots, \bar{\lambda}_{ij}(\cdot), \ldots, \bar{\lambda}_{in_i}(\cdot)]'$, and $\lambda_i(\cdot)=[\lambda_{i1}(\cdot), \ldots, \lambda_{ij}(\cdot), \ldots, \lambda_{in_i}(\cdot)]'$,

$M_{2K}$ in (81) may be expressed as

$$\begin{aligned} M_{2K} = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} & \frac{\partial \bar{\lambda}_{ij}(\beta, \sigma_\tau^2, \hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2))}{\partial \sigma_\tau^2} \\ & v_{2i}^{jk}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho)[\bar{\lambda}_{ik}(\beta, \sigma_\tau^2, \hat{\gamma}(q_{ik}; \beta, \sigma_\tau^2)) - \lambda_{ik}(\beta, \sigma_\tau^2, \gamma(q_{ik}))]. \end{aligned} \tag{82}$$

Next for

$$W_{ijk}^* = \frac{\partial \bar{\lambda}_{ij}(\beta, \sigma_\tau^2, \hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2))}{\partial \sigma_\tau^2}\, v_{2i}^{jk}(\beta, \sigma_\tau^2, \hat{\gamma}, \rho) \frac{\partial \bar{\lambda}_{ik}(\beta, \sigma_\tau^2, \gamma(q_{ik}))}{\partial \gamma(q_{ik})},$$

a Taylor expansion of $\bar{\lambda}_{ij}(\beta, \sigma_\tau^2, \hat{\gamma}(q_{ij}; \beta, \sigma_\tau^2))$ with respect to $\gamma(q_{ij})$ for all $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$, reduces $M_{2K}$ in (82) to

$$M_{2K} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i}\{W_{ijk}^*[\hat{\gamma}(q_{ik}; \beta, \sigma_\tau^2) - \gamma(q_{ik})] \\ + O_p([\hat{\gamma}(q_{ik}; \beta, \sigma_\tau^2) - \gamma(q_{ik})]^2)\}. \quad (83)$$

Further, using the formula for $\hat{\gamma}(q_{ik}; \beta, \sigma_\tau^2) - \gamma(q_{ik})$ from (42), $M_{2K}$ in (83) may be re-expressed as

$$M_{2K} = \frac{1}{K}\sum_{i=1}^{K} Q_{2i}(Y_i - \mu_i) + O(b^2) + o_p(1/\sqrt{K}), \quad (84)$$

where $Q_{2i} = (Q_{2i1}, \cdots, Q_{2ij}, \cdots, Q_{2in_i})'$ with

$$Q_{2ij} = \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_i}\sum_{k'=1}^{n_i}\frac{1}{B_K(q_{i'k'})}W_{i'j'k'}^*\frac{p_{ij}(q_{i'k'})}{1 + \mu_{ij}(q_{i'k'})(e^{\sigma_\tau^2} - 1)}.$$

By applying the result from (84) and (80), it now follows from (47) that

$$\sqrt{K}\{\hat{\sigma}_\tau^2 - \sigma_\tau^2\} = L^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K} Q_{1i}[u_i - \lambda_i(\beta, \sigma_\tau^2, \gamma(q_i))] \\ - L^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K} Q_{2i}(Y_i - \mu_i) + O(\sqrt{Kb^4}) + o_p(1), \quad (85)$$

where $L = \lim_{K\to\infty} L_K = E[L_K]$, $L_K$ being given by (48).

Further, using similar arguments as in (79), one may apply the Lindeberg–Feller central limit theorem for non-identically distributed random variables, and obtains

$$\sqrt{K}\left\{\hat{\sigma}_\tau^2 - \sigma_\tau^2 - O\left(b^2\right)\right\} \xrightarrow{D} N(0, V_{\sigma_\tau^2}) \quad \text{as } K \to \infty, \quad (86)$$

where

$$V_{\sigma_\tau^2} = L^{-1}\frac{1}{K}\sum_{i=1}^{K}[Q_{1i}\Omega_i(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2), \rho)Q_{1i}^T \\ + Q_{2i}\Sigma_i(\beta, \sigma_\tau^2, \hat{\gamma}(q_i; \beta, \sigma_\tau^2), \rho)Q_{2i}^T - 2Q_{1i}\mathrm{Cov}(U_i, Y_i)Q_{2i}^T]L^{-1}. \quad (87)$$

Similar to the condition for the $\sqrt{K}$-consistency of $\hat{\beta}$, for $\sqrt{K}$-consistency of $\hat{\sigma}_\tau^2$, we need to have $O(b^2) = Kb^4 \to 0$ as $K \to \infty$, that is, $1/4 < \alpha \leq 1/3$.

Note that in this section we have derived the asymptotic properties of $\hat{\sigma}_\tau^2$ which is obtained by applying the SGQL estimation approach using squared responses as discussed in Sect. 3.3.1. The derivation of the asymptotic properties for the (CR)SGQL and (CRN)SGQL estimators of $\sigma_\tau^2$ obtained by applying the SGQL approach from Sect. 3.3.2 or 3.3.3 will be similar and hence omitted to save space.

## Appendix C. A cross checking for bandwidth selection under the simulation study

It was indicated in Sect. 3.1 that in general the bandwidth parameter $b$ involved in the kernel weights for the estimation of the non-parametric function $\gamma(\cdot)$ is chosen as $b = K^{-1/5}$ (Pagan and Ullah 1999; Lin and Carroll 2001) which is related to our asymptotic choice $b = c_0 K^{-1/3.9}$, $c_0$ being an unknown constant. In our simulation study, in Sect. 4 we have considered $K = 100$ and hence chosen $b = K^{-1/5} \approx 0.40$. To validate this choice further, we computed the asymptotic mean square error of the non-parametric function estimator $\hat{\gamma}(q_0, b; \beta, \sigma_\tau^2, \rho)$ (see (17)) using

$$\text{MSE}(q_0, b; \beta, \sigma_\tau^2, \rho) = \text{Bias}^2(q_0, b; \beta, \sigma_\tau^2) + \text{Variance}(q_0, b; \beta, \sigma_\tau^2, \rho), \quad (88)$$

for known $\beta$ and $\sigma_\tau^2$. By a trial and error method we then choose the value of $b$ which minimizes $\max_{q_0} \text{MSE}(q_0, b; \beta, \sigma_\tau^2, \rho)$ using all possible values of $q_0$, that is, we obtain $b$ following the mini-max criterion

$$\min_b \max_{q_0} \text{MSE}(q_0, b; \beta, \sigma_\tau^2, \rho) \Rightarrow b.$$

Note that using (74) from Appendix B.1, we compute the approximate bias and variance terms in (88) as

$$\text{Bias}(q_0, b; \beta, \sigma_\tau^2) \simeq \frac{\sum_{i=1}^K \sum_{j=1}^{n_i} p_{ij}(q_0) \frac{\mu_{ij} - \mu_{ij}(q_0)}{1+\mu_{ij}(q_0)(\exp(\sigma_\tau^2)-1)}}{\sum_{i=1}^K \sum_{j=1}^{n_i} p_{ij}(q_0) \frac{\mu_{ij}(q_0)}{1+\mu_{ij}(q_0)(\exp(\sigma_\tau^2)-1)}},$$

and

$$\begin{aligned} &\text{Variance}(q_0, b; \beta, \sigma_\tau^2, \rho) \simeq \text{Var}(A_K) \\ &= \frac{1}{B_K^2} \frac{1}{K^2} \sum_{i=1}^K \left\{ \sum_{j=1}^{n_i} \left[ \frac{p_{ij}(q_0)}{1 + \mu_{ij}(q_0)(\exp(\sigma_\tau^2) - 1)} \right]^2 \sigma_{ijj} \right. \\ &\quad \left. + 2 \sum_{j=1}^{n_i - 1} \sum_{k=j+1}^{n_i} \left[ \frac{p_{ij}(q_0)}{1 + \mu_{ij}(q_0)(\exp(\sigma_\tau^2) - 1)} \right] \left[ \frac{p_{ik}(q_0)}{1 + \mu_{ik}(q_0)(\exp(\sigma_\tau^2) - 1)} \right] \sigma_{ijk} \right\}, \end{aligned}$$

respectively, where $\sigma_{ijj}$ and $\sigma_{ijk}$ are given in (7)–(8).

Using $q_0$ as any of the 100 equal-spaced points from 0.5 to 4.5 (see (56)), and using a set of trial values for $b$ as $b = 0.20, 0.22, \ldots, 0.66, 0.68$, the above mini-max criterion, for fixed $\beta = (\beta_1, \beta_2)' = (0.5, 0.5)'$, produced the estimate of $b$ as

$$b = 0.36, 0.36, 0.42, 0.42,$$

corresponding to

$$(\sigma_\tau^2, \rho) = (0.5, 0.5), (0.5, 0.8), (1.0, 0.5), (1.0, 0.8).$$

These values of $b$ are close to $b = K^{-1/5} = 0.40$ for $K = 100$, which supports the use of the general formula $b = K^{-1/5}$ where $K$ is the number of independent individuals in the study.

## References


Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, *85*, 749–758.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete Multivariate Analysis*. New York: Springer Science+Business Media LLC.

Breiman, L. (1968). *Probability*. Reading, Massachusetts: Addison-Wesley Pub. Co.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9–25.

Breslow, N. E., & Lin, X. (1995). Bias correction in generalized linear models with a single component of dispersion. *Journal of the American Statistical Association*, *82*, 81–92.

Carota, C., & Parmigiani, G. (2002). Semiparametric regression for count data. *Biometrika*, *89*, 265–281.

Chiou, J., & Muller, H. (2005). Estimated estimating equations: semiparametric inference for clustered and longitudinal data. *Journal of Royal Statistical Society, B*, *67*, 531–553.

Jiang, J., & Zhang, W. (2001). Robust estimation in generalized linear mixed models. *Biometrika*, *88*, 753–765.

Lin, X., & Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, *96*, 1045–1056.

Lin, X., & Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, *91*, 1007–1016.

Lin, X., & Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of Royal Statistical Society*, *68*, 69–88.

Montalvo, J. G. (1997). GMM estimation of count-panel-data models with fixed effects and predetermined instruments. *Journal of Business and Economic Statistics*, *15*, 82–89.

Pagan, A., & Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Prentice, R. L., & Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, *47*, 825–839.

Rao, R. P., Sutradhar, B. C., & Pandit, V. N. (2012). GMM versus GQL inferences in semi-parametric linear dynamic mixed models. *Brazilian Journal of Probability and Statistics*, *26*, 167–177.

Severini, T. A., & Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, *89*, 501–511.

Sneddon, G., & Sutradhar, B. C. (2004). On semiparametric familial-longitudinal models. *Statistics and Probability Letters*, *69*, 369–379.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science*, *18*, 377–393.

Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models. *Sankhya B: The Indian Journal of Statistics*, *66*(2), 261–289.

Sutradhar, B. C. (2010). Inferences in generalized linear longitudinal mixed models. *Canadian Journal of Statistics, special issue*, *38*, 174–196.

Sutradhar, B. C. (2011). *Dynamic Mixed Models for Familial Longitudinal Data*. New York: Springer.

Sutradhar, B. C., & Bari, W. (2007). On generalized quasilikelihood inference in longitudinal mixed model for count data. *Sankhya*, *69*, 671–699.

Sutradhar, B. C., & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika*, *86*, 459–465.

Sutradhar, B. C., Rao, R. P., & Pandit, V. N. (2008). Generalized method of moments versus generalized quasi-likelihood inferences in binary panel data models. *Sankhya B*, *70*, 34–62.

Sutradhar, B. C., Jowaheer, V., & Rao, R. P. (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.

Sutradhar, B. C., Warriyar, K. V., & V., Zheng, N., (2016). Inferences in semi-parametric dynamic models for repeated count data. *Australian and New Zealand Journal of Statistics*, *58*, 397–434.

Warriyar, K. V., & V., Sutradhar, B. C., (2014). Estimation with improved efficiency in semi-parametric linear longitudinal models. *Brazilian Journal of Probability and Statistics*, *28*, 561–586.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gaussian-Newton method. *Biometrika*, *61*, 439–447.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. New York: Springer.

Wooldridge, J. (1999). Distribution-free estimation of some non-linear panel data models. *Journal of Econometrics*, *90*, 77–97.

You, J., & Chen, G. (2007). Semiparametric generalized least squares estimation in partially linear regression models with correlated errors. *Journal of Statistical Planning and Inference*, *137*, 117–132.

Zeger, S. L., & Diggle, P. J. (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, *50*, 689–699.

Zhao, L. P., & Prentice, R. L. (1990). Correlated binary regression using a generalized quadratic model. *Biometrika*, *77*, 642–648.