Chapter 6
Longitudinal Models for Count Data

In longitudinal studies for count data, a small number of repeated count responses along with a set of multidimensional covariates are collected from a large number of independent individuals. For example, in a health care utilization study, the number of visits to a physician by a large number of independent individuals may be recorded annually over a period of several years. Also, the information on the covariates such as gender, number of chronic conditions, education level, and age, may be recorded for each individual. For $i = 1, \ldots, K$, and $t = 1, \ldots, T$, let y_{it} denote the count response and $x_{it} = (x_{it1}, \ldots, x_{itp})'$ denote the p-dimensional covariate vector collected at time point t from the ith individual. Let β be the effect of x_{it} on y_{it}. Note that because $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are T repeated count responses from the same individual, it is most likely that they are autocorrelated. The scientific concern is to find β, the effects of the covariates on the repeated count responses, after taking their autocorrelations into account.

Note that there are situations in practice, where the covariates of the ith individual may be time independent. We denote such covariates by $\tilde{x}_i = (x_{i1}, \ldots, x_{ip})'$. This is a simpler special case of the general situation with time-dependent covariates x_{it}. In Section 6.1, we provide the marginal distributional properties of the count response variable Y_{it} under the general situation when corresponding covariates are time dependent. For simplicity, Section 6.2 discusses the estimation of β by pretending that the repeated count responses are independent, even though in reality they are autocorrelated. In Section 6.3, we provide several autocorrelation structures for the repeated count data for the special case with time-independent covariates. A unified generalized quasi-likelihood (GQL) approach is discussed in Section 6.4 for the estimation of the regression effects β after taking the stationary correlations of the data into account.

Note that stationary autocorrelation models can be generalized to the nonstationary cases in various ways. We consider two types of nonstationary models. First, we consider a class of nonstationary autocorrelation models where all models produce the same specified marginal mean and variance functions. These models are given in Section 6.5. The same section also contains the estimating equation for β after taking the nonstationary correlations into account. Second, in Section 6.6, we

B.C. Sutradhar, *Dynamic Mixed Models for Familial Longitudinal Data*,
Springer Series in Statistics, DOI 10.1007/978-1-4419-8342-8_6,

demonstrate that the stationary autocorrelation models discussed in Section 6.3 may be generalized to a nonstationary class of models where these models may produce different marginal means and variances along with different correlation structures. The inferences for the regression effects β, after taking the nonstationary correlation structure of the repeated data into account are discussed in details, including the model misspecification effects. Note that in the stationary case, model selection is not necessary for the estimation of the regression effects, whereas model selection becomes an important issue in the nonstationary case. This model selection problem is also discussed in Section 6.6 for the second type of nonstationary autocorrelation models. A data example is considered in Section 6.7 to illustrate both correlation model selection and estimation of the parameters.

6.1 Marginal Model

Suppose that each of the count response variables $Y_{i1},\ldots,Y_{it},\ldots,Y_{iT}$ for the ith $(i=1,\ldots,K)$ follows the well-known Poisson distribution with a suitable mean parameter. Let $\mu_{it}=\exp(x'_{it}\beta)$ denote the mean of the Poisson distribution for Y_{it}. In the form of exponential density, one may then write the marginal distribution of Y_{it} as

$$f(y_{it})=\exp[\{y_{it}\theta_{it}-a(\theta_{it})\}+b(y_{it})] \tag{6.1}$$

[Nelder and Wedderburn (1972)], with

$$\theta_{it}=x'_{it}\beta,;a(\theta_{it})=\exp(\theta_{it}),\ \text{and}\ b(y_{it})=\log(\frac{1}{y_{it}!}).$$

We denote this marginal Poisson distribution as $Y_{it}\sim \text{Poi}(\mu_{it})$. For an auxiliary parameter s, by using the moment generating function (m.g.f.) of Y_{it} [see (4.9), also Exercise (4.5)] given by

$$M_{Y_{it}}(s)=E[\exp(sY_{it})]=\exp[a(s+\theta_{it})-a(\theta_{it})], \tag{6.2}$$

one may obtain the basic properties such as the first four moments of the marginal distribution (6.1) as in the following lemma.

Lemma 6.1 The first four moments of Y_{it} under the exponential family density (6.1) are given by

$$\begin{aligned}
\mu_{it}=[Y_{it}]&=a'(\theta_{it})\\
\sigma_{itt}=\text{var}[Y_{it}]&=a''(\theta_{it})\\
\tilde{\delta}_{itt}=E[Y_{it}-\mu_{it}]^3&=a'''(\theta_{it})\\
\tilde{\phi}_{itttt}=E[Y_{it}-\mu_{it}]^4&=a''''(\theta_{it})+3\sigma^2_{itt},
\end{aligned} \tag{6.3}$$

where $a'(\theta_{it})$, $a''(\theta_{it})$, $a'''(\theta_{it})$, and $a''''(\theta_{it})$ are, respectively, the first−, second−, third−, and the fourth-order derivatives of $a(\theta_{it})$ with respect to θ_{it}.

In the present longitudinal setup, the repeated count responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are most likely to be correlated, and these correlations, unlike the familial correlations developed through random effects in Chapter 4, should reflect the time effects. Some suitable modelling for this type of time effects based correlations is discussed in Section 6.3 for the cases when covariates are stationary (i.e., time independent), and in Sections 6.5 and 6.6 when covariates are nonstationary (i.e., time dependent). Note that if one is, however, interested to obtain only a consistent estimate for β as opposed to a consistent as well as efficient estimate, then, the repeated responses may be treated as independent and the marginal distribution (6.1) or the marginal properties in Lemma 6.1 may be exploited to construct suitable estimating equations to achieve such a goal. In the following section, we discuss three standard marginal model based estimation techniques that use either the marginal density in (6.1) or only the first two moments from Lemma 6.1.

6.2 Marginal Model Based Estimation of Regression Effects

Method of Moments (MM): Irrespective of the cases whether the repeated counts $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are independent or autocorrelated, one may always obtain the moment estimate of β by solving the moment equation

$$\sum_{i=1}^{K} \sum_{t=1}^{T} [x_{it}(y_{it} - a'(\theta_{it}))] = 0, \tag{6.4}$$

where $a'(\theta_{it}) = \mu_{it} = \exp(x'_{it}\beta)$ for Poisson y_{it}. By writing $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})' : T \times 1$; $\mu_i = (\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT})' : T \times 1$; and $X_i = (x_{i1}, \ldots, x_{it}, \ldots, x_{iT})' : T \times p$, the moment equation (6.4) may be re-expressed as

$$\sum_{i=1}^{K} [X'_i(y_i - \mu_i)] = 0. \tag{6.5}$$

Let the moment estimator of β, the root of the moment equation (6.5), be denoted by $\hat{\beta}_M$. This root may be obtained by using the iterative equation

$$\hat{\beta}_M(r+1) = \hat{\beta}_M(r) + \left[\sum_{i=1}^{K} X'_i A_i X_i\right]^{-1}_{(r)} \left[\sum_{i=1}^{K} X'_i(y_i - \mu_i)\right]_{(r)}, \tag{6.6}$$

where $A_i = \text{diag}[a''(\theta_{it})] = \text{diag}[\sigma_{itt}]$, and $[\cdot]_{(r)}$ denotes that the expression within the brackets is evaluated at $\beta = \hat{\beta}_M(r)$, the rth iterative value for $\hat{\beta}_M$. Note that because (6.5) is an unbiased estimating equation for the zero vector, $\hat{\beta}_M$ is a consistent estimator. Furthermore, because K individuals are chosen independently, by using multivariate central limit theorem [Mardia, Kent and Bibby (1979, p. 51)] it follows from (6.6) that $K^{\frac{1}{2}}(\hat{\beta}_M - \beta)$ is asymptotically multivariate Gaussian with zero mean vector and covariance matrix V_M given by

$$V_M = \mathrm{limit}_{K\to\infty} K \left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1} \left[\sum_{i=1}^{K} X_i' A_i^{1/2} C_i A_i^{1/2} X_i\right] \left[\sum_{i=1}^{K} X_i' A_i X\right]^{-1}, \qquad (6.7)$$

where C_i is the true correlation matrix of y_i, which may be unknown. This covariance matrix V_M may, however, be estimated by using the sandwich type estimator

$$\hat{V}_M = \mathrm{limit}_{K\to\infty} K \left[\sum_{i=1}^{K} X_i' A_i X_i\right]^{-1} \left[\sum_{i=1}^{K} X_i' (y_i - \mu_i)(y_i - \mu_i)' X_i\right] \left[\sum_{i=1}^{K} X_i' A_i X_i\right]^{-1} \qquad (6.8)$$

[see for example, Liang and Zeger (1986, p. 15)].

Quasilikelihood (QL) Method : Note that when there is a functional relationship between the mean and the variance of the response, Wedderburn (1974) [see also McCullagh (1983)] proposed a QL approach for independent data which exploits both mean and the variance in estimating the regression effects β. The QL estimating equation for β is given by

$$\sum_{i=1}^{K}\sum_{t=1}^{T}\left[\frac{\partial a'(\theta_{it})}{\partial \beta} \frac{(y_{it} - a'(\theta_{it}))}{\mathrm{var}(y_{it})}\right] = 0, \qquad (6.9)$$

where the $\mathrm{var}(Y_{it}) = a''(\theta_{it})$ is a function of the mean parameter $a'(\theta_{it}) = \mu_{it}$. In the Poisson case, for example,

$$\mathrm{var}(Y_{it}) = a''(\theta_{it}) = a'(\theta_{it}) = \mu_{it} = \exp(x_{it}'\beta).$$

Notice that there is no difference between this QL estimating equation (6.9) and the MM estimating equation (6.4).

We remark, however, that as opposed to the independence case, in a practical situation one would also exploit the correlation properties of the repeated responses in generalizing the QL estimating equation (6.9), but the MM approach will still use the estimating equation (6.5). Thus, in the longitudinal setup, the generalized QL approach will yield a different estimate for β than the MM approach.

Marginal Likelihood (ML) Method: It is true that the repeated counts

$$y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$$

are autocorrelated. If the correlations are, however, ignored, that is, the repeated responses are treated to be independent, then one may maximize the marginal likelihood function to obtain an independence assumption based 'working' likelihood estimate of β. By (6.1), the log of the marginal likelihood function of β is given by

$$\log L(\beta) = \sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})], \tag{6.10}$$

yielding the likelihood equation for β as

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it} - a'(\theta_{it})]\frac{\partial \theta_{it}}{\partial \beta} = 0. \tag{6.11}$$

Because, $\theta_{it} = x'_{it}\beta$, this likelihood equation is the same as the MM equation (6.4). Thus it is clear that all three approaches, namely, the MM, QL and ML methods yield the same estimate for β. All three approaches yield a consistent estimate for this regression effect.

6.3 Correlation Models for Stationary Count Data

Note that a marginal model based estimation approach may not yield an efficient regression estimate. Obtaining an efficient estimate will require exploitation of the joint probability or correlation model for the repeated count data. In this section, we discuss this issue, for a simpler situation when covariates of an individual are time independent. Note that this situation can arise in some longitudinal studies such as in a longitudinal clinical study where, for example, the number of weekly asthma attacks is recorded as the responses over a small period such as four weeks of time. Here, it is likely that the covariate information such as gender, education level, and number of other chronic diseases of the individual will remain the same for each week for the duration of the study over four weeks. This is, however, true that the repeated responses will still be correlated due to the influence of time, the time being a stochastic factor. In the end, it is of main interest to find the effects of the covariates on the responses after taking the correlations of the responses into account.

Recall that $\tilde{x}_i = (x_{i1}, \ldots, x_{ip})'$ denote the time-independent covariate vector for the ith individual. For this time-independent covariate, the mean and the variance of y_{it} may be written, following Lemma 6.1, as

$$E[Y_{it}] = \text{var}[Y_{it}] = \tilde{\mu}_i = \exp(\tilde{x}'_i\beta), \tag{6.12}$$

yielding the mean vector and the diagonal matrix of the variances as

$$\mu_i = \tilde{\mu}_i\mathbf{1},\ A_i = \text{diag}(\sigma_{itt}) = \text{diag}(\tilde{\mu}_i), \tag{6.13}$$

where $\mathbf{1}$ is the $T \times 1$ unit vector.

As far as the correlation structures for the repeated counts $y_{i1}, \ldots, y_{iT}$ are concerned, it was speculated in some of the original studies such as in Liang and Zeger (1986) that the correlations of the repeated data may follow Gaussian type such as autoregressive order 1 (AR(1)), moving average order (1) (MA(1)), or exchangeable (equicorrelations) correlation structures. But, as it is not easy to know the underlying true correlation structure, these authors have used a ‘working’ correlation structure based generalized estimating equations (GEE) approach for the efficient estimation of the regression effects. We discuss this GEE approach and its serious limitations in Section 6.4.

We now provide three correlation models [Sutradhar (2003), McKenzie (1988)] that yield the speculated AR(1), MA(1), and equicorrelation structures for repeated count data. In fact, these three low-order models are easily extendable to other possible higher-order models such as AR(2), MA(2), and ARMA(1,1) models.

6.3.1 Poisson AR(1) Model

Let $y_{i1} \sim \text{Poi}(\tilde{\mu}_i)$, where $\tilde{\mu}_i = \exp(\tilde{x}_i'\beta)$ as in (6.12). Furthermore, for $t = 2,\ldots,T$, let the response y_{it} at time t be related to $y_{i,t-1}$ at time $t-1$ as

$$y_{it} = \rho * y_{i,t-1} + d_{it}, \tag{6.14}$$

[McKenzie (1988), Sutradhar (2003)] where it is assumed that for given $y_{i,t-1}$, $\rho * y_{i,t-1}$ denotes the so-called binomial thinning operation (McKenzie, 1988). That is,

$$\begin{aligned} \rho * y_{i,t-1} &= \sum_{j=1}^{y_{i,t-1}} b_j(\rho) \\ &= z_{i,t-1}, \text{say}, \end{aligned} \tag{6.15}$$

with $\Pr[b_j(\rho) = 1] = \rho$ and $\Pr[b_j(\rho) = 0] = 1 - \rho$. Furthermore, it is assumed in (6.14) that $d_{it} \sim P(\tilde{\mu}_i(1-\rho))$ and is independent of $z_{i,t-1}$.

It then follows that each y_{it} satisfying the model (6.14) has marginally Poisson distribution with parameters as in (6.12). Also by direct calculation, it can be shown that

$$\begin{aligned} E[Y_{it}] &= E_{Y_{i,t-1}} E[Y_{it}|Y_{i,t-1}] = \tilde{\mu}_i \\ \text{var}[Y_{it}] &= E_{Y_{i,t-1}} \text{var}[Y_{it}|Y_{i,t-1}] + \text{var}_{Y_{i,t-1}} E[Y_{it}|Y_{i,t-1}] = \tilde{\mu}_i. \end{aligned} \tag{6.16}$$

Next, by similar calculations as in (6.16), for lag $\ell = 1,\ldots,T-1$, it can be shown from (6.14) that $E(Y_{it}Y_{i,t-\ell}) = \tilde{\mu}_i^2 + \tilde{\mu}_i\rho^{\ell}$, yielding the lag ℓ correlation between y_{it} and $y_{i,t-\ell}$, say $c^*_{i,(t-\ell)t}(\rho)$, as

$$\begin{aligned} \text{corr}(Y_{it}, Y_{i,t-\ell}) &= c^*_{i,(t-\ell)t}(\rho) \\ &= \rho^{\ell}, \end{aligned} \tag{6.17}$$

which is the same as lag ℓ correlation under the Gaussian AR(1) autocorrelation structure. But, the ρ parameter under the present AR(1) model (6.14) must satisfy the range restriction $0 \leq \rho \leq 1$, whereas in the Gaussian AR(1) structure ρ lies in the range $-1 < \rho < 1$.

6.3.2 Poisson MA(1) Model

For a scale parameter ρ, let

$$d_{it} \overset{\text{iid}}{\sim} \text{Poi}\left(\frac{\tilde{\mu}_i}{1+\rho}\right), \text{ for } t = 0, 1, \ldots, T,$$

where $\tilde{\mu}_i = \exp(\tilde{x}_i' \beta)$, $t = 0$ being an initial time. Next suppose that the response y_{it} is related to the d_{it} as

$$y_{it} = \rho * d_{i,t-1} + d_{it}, \text{ for } t = 1, \ldots, T, \tag{6.18}$$

where $\rho * d_{i,t-1} = \sum_{j=1}^{d_{i,t-1}} b_j(\rho)$ is the binomial thinning operation similar to (6.15). By similar calculations as in the AR(1) process, one obtains

$$\begin{aligned} E[Y_{it}] = \text{var}[Y_{it}] &= \tilde{\mu}_i \\ \text{corr}(Y_{it}, Y_{i,t-\ell}) &= c^*_{i,(t-\ell)t}(\rho) \\ &= \begin{cases} \rho/(1+\rho) & \text{for } \ell = 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6.19}$$

Note that the lag correlations in (6.19) have the same forms as in the Gaussian MA(1) correlation structure, except that in the present set up $0 \leq \rho \leq 1$, whereas under the Gaussian structure $-1 < \rho < 1$.

6.3.3 Poisson Equicorrelation Model

Suppose that y_{i0} is a Poisson variable with the mean parameter $\tilde{\mu}_i = \exp(\tilde{x}_i' \beta)$. Also suppose that

$$d_{it} \overset{\text{iid}}{\sim} \text{Poi}(\tilde{\mu}_i(1-\rho)) \text{ for all } t = 1, \ldots, T.$$

By similar arguments as for the AR(1) and MA(1) processes, one can show that y_{it} given by

$$y_{it} = \rho * y_{i0} + d_{it} \tag{6.20}$$

also follows the Poisson distribution (i.e., $y_{it} \sim \text{Poi}(\tilde{\mu}_i)$, yielding the marginal properties

$$E[Y_{it}] = \text{var}[Y_{it}] = \tilde{\mu}_i = \exp(\tilde{x}_i' \beta). \tag{6.21}$$

Note that these marginal properties may also be computed directly by using the model (6.20). As far as the product moments properties are concerned, it can be shown that

$$\begin{aligned}\text{corr}(Y_{it}, Y_{i,t-\ell}) &= c^*_{i,(t-\ell)t}(\rho) \\ &= \rho, \end{aligned} \tag{6.22}$$

for all $\ell = 1, 2, \ldots, T-1$, with $0 \leq \rho \leq 1$ instead of $-(1/T-1) \leq \rho \leq 1$ under the Gaussian equicorrelation model.

For convenience, we summarize the means, variances, and correlations for all three stationary correlation models, as in Table 6.1.

Table 6.1 A class of stationary correlation models for longitudinal count data and basic properties.

Model	Dynamic Relationship	Mean, Variance, & Correlations
AR(1)	$y_{it} = \rho * y_{i,t-1} + d_{it}, t = 2, \ldots$	$E[Y_{it}] = \mu_{i\cdot}$
	$y_{i1} \sim Poi(\mu_{i\cdot})$	$\text{var}[Y_{it}] = \mu_{i\cdot}$
	$d_{it} \sim Poi(\mu_{i\cdot}(1-\rho)), t = 2, \ldots$	$\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho_\ell$
		$= \rho^\ell$
MA(1)	$y_{it} = \rho * d_{i,t-1} + d_{it}, t = 1, \ldots$	$E[Y_{it}] = \mu_{i\cdot}$
	$d_{i0} \sim Poi(\mu_{i\cdot}/(1+\rho))$	$\text{var}[Y_{it}] = \mu_{i\cdot}$
	$d_{it} \sim Poi(\mu_{i\cdot}/(1+\rho)), t = 1, \ldots$	$\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho_\ell$
		$= \begin{cases} \frac{\rho}{1+\rho} & \text{for } \ell = 1 \\ 0 & \text{otherwise,} \end{cases}$
EQC	$y_{it} = \rho * y_{i1} + d_{it}, t = 2, \ldots$	$E[Y_{it}] = \mu_{i\cdot}$
	$y_{i1} \sim Poi(\mu_{i\cdot})$	$\text{var}[Y_{it}] = \mu_{i\cdot}$
	$d_{it} \sim Poi(\mu_{i\cdot}(1-\rho)), t = 2, \ldots$	$\text{corr}[Y_{it}, Y_{i,t+\ell}] = \rho_\ell$
		$= \rho$

6.4 Inferences for Stationary Correlation Models

6.4.1 Likelihood Approach and Complexity

As opposed to the marginal likelihood estimation by (6.10), it is natural that under the correlation models (6.14), (6.18), and (6.20), the likelihood construction would be complicated. This is because under these models, the likelihood function is given by

$$L(\beta, \rho) = \Pi_{i=1}^K [f(y_{i1}) \Pi_{t=2}^T f(y_{it}|y_{i,t-1})], \tag{6.23}$$

where $f(y_{i1}) = \exp(-\tilde{\mu}_i)\tilde{\mu}_i^{y_{i1}}/y_{i1}!$ is the Poisson density with $\tilde{\mu}_i = \exp(\tilde{x}_i'\beta)$, under all three models, but the conditional densities $f(y_{it}|y_{i,t-1})$ would have different forms under different models. For example, under the stationary AR(1) model

(6.14), the conditional density has the form given by

$$f(y_{it}|y_{i,t-1}) = \sum_{s=1}^{\min(y_{it},y_{i,t-1})} \frac{(y_{i,t-1})!}{s!(y_{i,t-1}-s)!}\rho^s(1-\rho)^{y_{i,t-1}-s}\frac{\exp(-\tilde{\mu}_i)\tilde{\mu}_i^{y_{it}-s}}{(y_{it}-s)!}, \quad (6.24)$$

yielding by (6.23) a complex likelihood, which is not easy to maximize with regard to the desired parameters β and ρ.

In the following section we provide an alternative efficient approach for the estimation of the parameters of the models.

6.4.2 GQL Approach

Recall that under the independence assumption, one can solve the quasi-likelihood [QL; Wedderburn (1974)] estimating equation (6.9) for β, but this will be an inefficient estimate given that the repeated responses are now assumed to follow either the AR(1) correlation model (6.14) with correlation structure (6.17), MA(1) correlation model (6.18) with correlation structure (6.19), or equicorrelation model (6.20) with correlation structure as in (6.22). Note that all three correlation structures given in (6.17), (6.19), and (6.22), may be represented by a general autocorrelation matrix of the form

$$C_i^*(\rho) = (c^*_{i,(t-\ell)t}(\rho)) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{bmatrix}, \quad (6.25)$$

[Sutradhar and Das (1999, Section 3)], where for $\ell = 1,\ldots,T-1$, ρ_ℓ represents the lag ℓ autocorrelation. For example, the AR(1) model based autocorrelation structure (6.17) may be represented by this correlation matrix $C_i^*(\rho)$ (6.25) by using $\rho_\ell = \rho^\ell$. Similarly, when one uses $\rho_1 = \rho/(1+\rho)$ and $\rho_2 = \rho_3 = \ldots = \rho_{T-1} = 0$, in (6.25), it produces the MA(1) correlation structure (6.19); and for $\rho_\ell = \rho$ for all $\ell = 1,\ldots,T-1$, $C_i^*(\rho)$ matrix in (6.25) represents the correlations under the equicorrelations structure (6.22).

It is therefore clear that if it is assumed that the repeated counted responses follow one of the AR(1), MA(1), or equi-correlation models, then one may estimate the regression effects under any of these three models by simply estimating this common $C_i^*(\rho)$ matrix in (6.25) and then using this estimated correlation matrix in a proper estimating equation for the regression effects β. Because $C_i^*(\rho)$ is the true correlation matrix for any of the three models, Sutradhar (2003, Section 3) proposed a generalized quasi-likelihood approach that generalizes the independence assumption based QL (6.9) approach of Wedderburn (1974) to the general stationary correlation setup. The GQL estimating equation for β is given by

$$\sum_{i=1}^{K} X_i' A_i \Sigma_i^{*-1}(\rho)(y_i - \mu_i) = 0, \tag{6.26}$$

where $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$, with $C_i^*(\rho)$ as the true stationary correlation structure for any of the AR(1), MA(1), or equicorrelation models. Note that in (6.26), $\mu_i = \tilde{\mu}_i \mathbf{1}$, $A_i = \text{diag}(\sigma_{itt}) = \text{diag}(\tilde{\mu}_i)$, as in (6.13), $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT})'$ is the $T \times 1$ vector of repeated counts for the ith individual, and $X_i' = [\tilde{x}_i, \ldots, \tilde{x}_i] : p \times T$ is the corresponding matrix of stationary covariates with $\tilde{x}_i = (x_{i1}, \ldots, x_{ip})'$ as the p-dimensional time-independent covariate vector as in (6.12).

Note that the GQL estimating equation (6.26) may be solved for β when ρ (i.e., all lag correlations $\rho_1, \ldots, \rho_\ell, \ldots, \rho_{T-1}$) is known. It is, however, not necessary to know the specific form for the correlation matrix $C_i^*(\rho)$, as this form in (6.25) is general which is valid under any of the three correlation structures (6.17), (6.19) and (6.22). In practice ρ is unknown, therefore the lag correlations can be consistently estimated by using the well-known method of moments. For $\ell = |u - t|$, $u \neq t$, $u, t = 1, \ldots, T$, the moment estimator for ρ_ℓ, the autocorrelation of lag ℓ, has the formula

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K} \sum_{t=1}^{T-\ell} \tilde{y}_{it} \tilde{y}_{i,t+\ell} / K(T-\ell)}{\sum_{i=1}^{K} \sum_{t=1}^{T} \tilde{y}_{it}^2 / KT}, \tag{6.27}$$

[Sutradhar and Kovacevic (2000, eqn. (2.18)); Sutradhar (2003)], where $\tilde{y}_{it}$ is the standardized residual, defined as $\tilde{y}_{it} = (y_{it} - \mu_{it}) / \{\sigma_{itt}\}^{1/2}$. Note that under the present stationary correlation models for the repeated count data $\mu_{it} = \sigma_{itt} = \tilde{\mu}_i$ as in (6.12) and (6.13).

Let $\hat{\beta}_{GQL}$ denote the GQL estimator of β which is obtained by solving (6.26) after using $\hat{\rho}_\ell$ from (6.27) for ρ_ℓ. Note that because the left-hand side of the GQL estimating equation in (6.26) is an unbiased estimating function for the zero vector, $\hat{\beta}_{GQL}$, the root of the equation (6.26) is a consistent estimator for β.

6.4.2.1 Asymptotic Distribution of the GQL Estimator

Note that $\hat{\beta}_{GQL}$ may be obtained from (6.26) by using the iterative equation

$$\begin{aligned} \hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) &+ \left[\sum_{i=1}^{K} X_i' \Sigma_i^{*-1}(\rho) X_i \right]_{(r)}^{-1} \\ &\times \left[\sum_{i=1}^{K} X_i' \Sigma_i^{*-1}(\rho)(y_i - \mu_i) \right]_{(r)}, \end{aligned} \tag{6.28}$$

where $[\cdot]_{(r)}$ denotes that the expression within the brackets is evaluated at $\beta = \hat{\beta}_{GQL}(r)$, the rth iterative value for $\hat{\beta}_{GQL}$. Because $y_1, \ldots, y_i, \ldots, y_K$ are independent, by using the central limit theorem, it then follows from (6.28) that as $K \to \infty$, $(\hat{\beta}_{GQL} - \beta)$ has the p-dimensional multivariate normal distribution with mean vector

0 and $p \times p$ covariance matrix V^* given by

$$V^*(\hat{\beta}_{GQL}) = \lim_{K\to\infty} \left\{ \sum_{i=1}^{K} X_i^T A_i^{1/2} C_i^{*-1}(\rho) A_i^{1/2} X_i \right\}^{-1}. \tag{6.29}$$

Note that this asymptotic distribution is given here for known ρ. This result, however, holds even when $\hat{\rho}$ is used for ρ. This is because it can be shown that $\hat{\rho}_\ell$ from (6.27) converges in probability to ρ_ℓ for all $\ell = 1, \ldots, T-1$.

6.4.2.2 'Working' Independence Assumption Based GQL Estimation

It is known that if one is interested in obtaining only a consistent estimator for β, this can be achieved by solving the GQL estimating equation (6.26) by pretending that the repeated responses are independent even though they are actually correlated following any of the three models (6.14), (6.18), or (6.20). Thus, we obtain a 'working' independence assumption based GQL estimate by solving

$$\sum_{i=1}^{K} X_i' A_i \Sigma_i^{*-1}(\rho)(y_i - \mu_i)|_{\rho=0} = \sum_{i=1}^{K} X_i'(y_i - \mu_i) = 0. \tag{6.30}$$

Note that this estimating equation is in fact the QL estimating equation (6.9) due to Wedderburn (1974), which is also the same as the MM estimating equation (6.5). This QL estimating equation is simpler to solve than the GQL (6.26) equation and this provides the consistent estimate for β.

Let $\hat{\beta}(I)$ denote the solution of (6.30). This estimator is the same as the MM estimator $\hat{\beta}_{MM}$ obtained from (6.5), therefore its asymptotic distribution is given by (6.7). Thus, $\hat{\beta}(I)$ has the asymptotic variance

$$V^*(\hat{\beta}(I)) = \mathrm{limit}_{K\to\infty} \left[\sum_{i=1}^{K} X_i' A_i X \right]^{-1} \left[\sum_{i=1}^{K} X_i' \Sigma_i^*(\rho) X_i \right] \left[\sum_{i=1}^{K} X_i' A_i X \right]^{-1}, \tag{6.31}$$

where $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$.

6.4.2.3 Efficiency of the Independence Assumption Based Estimator

Similar to the correlated linear model case [Amemiya (1985, Section 6.1.3)], a comparison of (6.31) with (6.29) shows that the independence assumption based estimator $\hat{\beta}(I)$ always has the less than or the same efficiency asthe GQL estimator $\hat{\beta}_{GQL}$. We provide a numerical example below to illustrate this efficiency issue.

The percentage efficiency of the uth $(u = 1, \ldots, p)$ component of the $\hat{\beta}(I)$ estimator, for example, is defined as

$$\mathrm{eff}(\hat{\beta}_u(I)) = \frac{\mathrm{var}(\hat{\beta}_{u,GQL})}{\mathrm{var}(\hat{\beta}_u(I))} \times 100, \tag{6.32}$$

where $\mathrm{var}(\hat{\beta}_{u,GQL})$ and $\mathrm{var}(\hat{\beta}_u(I))$ are the uth diagonal elements of the covariance matrices $V^*(\hat{\beta}_{GQL})$ (6.29) and $V^*(\hat{\beta}(I))$ (6.31), respectively. Let us take $p = 2$ for simplicity so that the Poisson mean and the variance μ_{it} for the ith $(i = 1, \ldots, K)$ at time t $(t = 1, \ldots, T)$, has the formula $\tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ under any of the three stationary models (6.14), (6.18), or (6.20). Let us consider $K = 100$, and three values of $T = 5, 10$, and 15. As far as the time-independent stationary covariates are considered, we choose

$$x_{it1} = \tilde{x}_{i1} = 1.0, \text{ for all } i = 1, \ldots, K, \text{ and } t = 1, \ldots, T,$$

and

$$x_{it2} = \tilde{x}_{i2} = \begin{cases} -1 & \text{for } t = 1, \ldots, T;\ i = 1, \ldots, K/4 \\ 0 & \text{for } t = 1, \ldots, T;\ i = (K/4)+1, \ldots, K/2 \\ 0 & \text{for } t = 1, \ldots, T;\ i = (K/2)+1, \ldots, 3K/4 \\ 1 & \text{for } t = 1, \ldots, T;\ i = (3K/4)+1, \ldots, K; \end{cases}$$

Next to compute the covariance matrices $V^*(\hat{\beta}_{GQL})$ (6.29) and $V^*(\hat{\beta}(I))$ (6.31), we need to construct the X_i and A_i matrices by

$$X_i = [\tilde{x}_{i1}\mathbf{1}_T, \tilde{x}_{i2}\mathbf{1}_T], \text{ and } A_i = \mathrm{diag}[\tilde{\mu}_i] \ : T \times T.$$

We also need to specify the correlation matrix $C_i^*(\rho)$. We choose all three correlation models AR(1), MA(1), and exchangeable correlation structures given by (6.17), (6.19), and (6.22), respectively. Note that because the lag 1 correlations under the AR(1) (6.17) and equicorrelations (6.22) structures are given as $\rho_1 = \rho$, we choose, for example, $\rho = 0.3$ and 0.7 under both AR(1) and equi-correlation structures. But, as the lag 1 correlation under the MA(1) structure has to satisfy the range $0 < \rho_1 = \rho/(1+\rho) < 0.5$, we choose, for example, two values of $\rho = 0.25$ and 0.67, yielding the lag 1 correlations $\rho_1 = 0.2$ and 0.4, respectively.

For $\beta_1 = \beta_2 = 1.0$, and for the selected values of ρ, the efficiencies of $\hat{\beta}(I)$ as compared to $\hat{\beta}_{GQL}$ are given in Table 1.

The results of Table 6.2 show that as expected the independence assumption based GQL estimator $\hat{\beta}(I)$ obtained by solving (6.30) always has less or the same efficiency as compared to the true correlation structure based GQL estimator $\hat{\beta}_{GQL}$ obtained by solving (6.26).

Table 6.2 Percentage relative efficiency of $\hat{\beta}_1(I)$ and $\hat{\beta}_2(I)$ to the generalized estimators $\hat{\beta}_{1,GQL}$ and $\hat{\beta}_{2,GQL}$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for AR(1), MA(1), and Equi-correlation structures, for $\mu_{it} = \tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ with $\beta_1 = \beta_2 = 1$

	AR(1)			MA(1)			EQC		
T	ρ	$\hat{\beta}_1(I)$	$\hat{\beta}_2(I)$	ρ	$\hat{\beta}_1(I)$	$\hat{\beta}_2(I)$	ρ	$\hat{\beta}_1(I)$	$\hat{\beta}_2(I)$
5	0.3	98	98	0.25	99	99	0.30	100	100
	0.49	96	96				0.49	100	100
	0.7	95	95	0.67	97	97	0.7	100	100
10	0.3	99	99	0.25	99	99	0.3	100	100
	0.49	96	96				0.49	100	100
	0.7	93	93	0.67	98	98	0.7	100	100
15	0.3	99	99	0.25	100	100	0.3	100	100
	0.49	97	97				0.49	100	100
	0.7	93	93	0.67	99	99	0.7	100	100

6.4.2.4 Performance of the GQL Estimation: A Simulation Example

Suppose that the repeated count responses follow either of the three stationary, namely AR(1)(6.17), MA(1) (6.19), or equicorrelation (6.22) structures. In estimating the regression effects β, the GQL approach does not, however, require us to know the specific correlation structure. What is needed here is: first consider that the repeated data for the ith individual has the autocorrelation matrix $C_i^*(\rho)$ (6.25) which in fact is a valid matrix not only for the above three correlation structures but also for any higher-order such as AR(2) and MA(2) correlation structures. Second, estimate this general autocorrelation matrix consistently and use the estimate in the GQL estimating equation (6.26) for β. This prompts the following two-step estimation.

Step 1. First, we solve the estimating equation for β (6.26) iteratively by (6.28), using starting values zero for longitudinal correlations and small positive or negative values for the regression parameters.
Step 2. This interim estimate of β from step 1 is then used in (6.27) to obtain the estimate of the autocorrelation matrix $C_i^*(\rho)$ in (6.25), which is used in turn in (6.28) to compute the new β estimate. This cycle of iterations continues until convergence.

To examine the performance of the above two-step based GQL estimation, we now consider a simulation study. Suppose that we follow the Poisson AR(1) model (6.14) and generate $T = 4$ repeated count observations for each of $K = 100$ independent individuals. As far as the covariates are concerned, we choose $p = 2$ time-independent covariates for each of these 100 individuals, given by

$$x_{it1} = \begin{cases} -1 & \text{for } t = 1,\ldots,T;\ i = 1,\ldots,K/4 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/4)+1,\ldots,K/2 \\ 0 & \text{for } t = 1,\ldots,T;\ i = (K/2)+1,\ldots,3K/4 \\ 1 & \text{for } t = 1,\ldots,T;\ i = (3K/4)+1,\ldots,K; \end{cases}$$

and

$$x_{it2} = z_i^* mbox for\, t = 1,\ldots,T;\ i = 1,\ldots,K,$$

where z_i^* is a standard normal quantity. In this problem, $\beta = (\beta_1, \beta_2)'$ denotes the effects of the two covariates on the repeated counts.

Note that even though the data are generated following the AR(1) model (6.14), the GQL approach does not, however, require this model to be known for the estimation of β. This is because the GQL estimating equation (6.26) is developed based on a general autocorrelation structure $C_i(\rho^*)$, which accommodates all three AR(1) (6.17), MA(1) (6.19), and exchangeable (6.22) correlation structures. Further note that for $T = 4$, this general autocorrelation structure has three lag correlations, namely, ρ_1, ρ_2, and ρ_3, to estimate, by using the formula (6.27) as explained in Step 2 above. It would be interesting to see how these three estimates behave in estimating the three lag correlations ρ, ρ^2, and ρ^3, for the AR(1) model that generated the data. Next these correlation estimates are used in step 1 to estimate β by solving the GQL estimating equation (6.26). For a selected set of parameter values, namely $\beta_1 = \beta_2 = 0.0$, and $\rho = 0.6,\ 0.8$, the simulation is repeated 500 times. The average and standard error of the 500 estimates for each parameter are given in Table 6.3. In the table, these estimates are referred to as the simulated mean (SM) and simulated standard error (SSE). The estimated standard errors (ESE) of the regression estimates are also computed. This is done by using the asymptotic covariance formula for $V^*(\hat{\beta}_{GQL})$ given in (6.29).

Table 6.3 Simulated means, simulated standard errors, and estimated standard errors of the GQL estimates for regression and autocorrelation coefficients for selected values of the true correlation parameter under the Poisson AR(1) process with $T = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, based on 500 simulations.

		Estimates				
AR(1) Correlation (ρ)	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.6	SM	-0.003	-0.001	0.595	0.352	0.203
	SSE	0.085	0.049	0.061	0.088	0.108
	ESE	0.086	0.050			
0.8	SM	0.000	0.003	0.791	0.626	0.496
	SSE	0.096	0.056	0.043	0.070	0.098
	ESE	0.098	0.057			

The results in Table 6.3 clearly show that the two-step based GQL approach estimates all parameters very well. For example, when $\rho = 0.8$, the lag correlation estimates are $0.791,\ 0.626,$ and $0.496,$ whereas the true AR(1) based lag correlations are $\rho = 0.8,\ \rho^2 = 0.64,$ and $\rho^3 = 0.512$. Similarly, the GQL approach estimates for $\beta_1 = \beta_2 = 0$ are $0.000,\ 0.003$. Furthermore, for this $\rho = 0.8$ case, the ESE of the regression estimates , that is, $0.098,$ and $0.0.57$ appear to be very close to the SSEs 0.096 and 0.056, respectively.

In Tables 6.4 and 6.5 below, we show similar results with regard to the performance of the GQL approach when data are generated under the MA(1) (6.18) and exchangeable (6.20) correlation models, respectively, by using the same covariates as in the AR(1) case.

Table 6.4 Simulated means, simulated standard errors, and estimated standard errors of the GQL estimates for regression and autocorrelation coefficients for selected values of the true correlation parameter under the Poisson MA(1) process with $T = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, based on 500 simulations.

		Estimates				
ρ (MA(1) Correlation (ρ_1))	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.25 (0.2)	SM	0.002	0.002	0.191	-0.006	0.004
	SSE	0.083	0.063	0.058	0.073	0.100
	ESE	0.081	0.063			
0.67 (0.4)	SM	-0.004	-0.004	0.396	-0.005	-0.004
	SSE	0.085	0.069	0.059	0.074	0.097
	ESE	0.088	0.070			

Table 6.5 Simulated means, simulated standard errors, and estimated standard errors of the GQL estimates for regression and autocorrelation coefficients for selected values of the true correlation parameter under the Poisson equicorrelation process with $T = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, based on 500 simulations.

		Estimates				
Equi-correlation (ρ)	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
0.6	SM	-0.006	-0.005	0.587	0.587	0.587
	SSE	0.119	0.096	0.064	0.065	0.088
	ESE	0.118	0.093			
0.8	SM	-0.009	-0.009	0.790	0.790	0.789
	SSE	0.131	0.101	0.043	0.041	0.059
	ESE	0.130	0.103			

6.4.3 GEE Approach and Limitations

In order to gain efficiency over the independence assumption based regression estimator $\hat{\beta}(I)$ (6.30), in the generalized estimating equations approach [Liang and Zeger (1986)], one solves a 'working' correlation matrix, $R(\alpha)$, based estimating equation

$$\sum_{i=1}^{K} X_i' A_i {V^*}_i^{-1}(\hat{\alpha})(y_i - \mu_i) = 0, \tag{6.33}$$

where $V_i^*(\alpha) = A_i^{1/2} R(\alpha) A_i^{1/2}$ is the working covariance matrix of y_i, α being an $s \times 1$ vector of parameters which fully characterizes $R(\alpha)$. Note that the GEE in (6.33) appears to be similar to the GQL estimating equations in (6.26), but they are quite different. Also, in (6.33), $\hat{\alpha}$ is obtained by solving a 'working' correlation model based moment equation. The data used in such a moment equation follow a different but true correlation structure, thus it is inappropriate to assume that $\hat{\alpha}$ converges to α [Crowder (1995)]. In view of this anomaly, any efficiency computations by using $\hat{\alpha}$ for α in the formula for the covariance matrix of the GEE estimator obtained from (6.33) [Liang and Zeger (1986)] would be incorrect.

Let $\hat{\beta}_G$ be the solution for β based on (6.33). Next suppose that $\hat{\alpha}$ converges to α_0, which must be a function of the true correlation parameter (ρ). In order to examine the correlation misspecification effects on the efficiency of $\hat{\beta}_G$, Sutradhar and Das (1999) have suggested using this α_0 in the formula for the covariance matrix of $\hat{\beta}_G$. Thus, $K^{1/2}(\hat{\beta}_G - \beta)$ is now asymptotically multivariate Gaussian with zero mean vector and covariance matrix V_G given by

$$\begin{aligned} V_G = \lim_{K\to\infty} K \left(\sum_{i=1}^{K} X_i' A_i^{1/2} R^{-1}(\alpha_0) A_i^{1/2} X_i \right)^{-1} \\ \times \left\{ \sum_{i=1}^{K} X_i' A_i^{1/2} R^{-1}(\alpha_0) C_i(\rho) R^{-1}(\alpha_0) A_i^{1/2} X_i \right\} \\ \times \left\{ \sum_{i=1}^{K} X_i' A_i^{1/2} R^{-1}(\alpha_0) A_i^{1/2} X_i \right\}^{-1}, \end{aligned} \tag{6.34}$$

where $C_i(\rho)$ is the true correlation matrix, as given in (6.25).

6.4.3.1 Efficiency of the GEE Based Estimator Under Correlation Structure Mis-specification

As far as the correlation models are concerned, we consider the same three stationary Poisson correlation models as we took for Section 6.4.2.3. Note that similar to (6.32), the percentage efficiency of the uth $(u = 1, \ldots, p)$ component of the $\hat{\beta}_G$

estimator, for example, is defined as

$$\mathrm{eff}(\hat{\beta}_{u,G}) = \frac{\mathrm{var}(\hat{\beta}_{u,TR})}{\mathrm{var}(\hat{\beta}_{u,G})} \times 100, \tag{6.35}$$

where $\mathrm{var}(\hat{\beta}_{u,TR})$ is the uth diagonal element of the covariance matrix of the true correlation structure based estimator V^*_{TR} computed by (6.29) using the true correlation structure for $C^*_i(\rho)$, and $\mathrm{var}(\hat{\beta}_{u,G})$ is the uth diagonal element of the covariance matrix V_G given in (6.34). For the purpose, we first show how to compute α_0 under possible model mis-specifications, and then compute the efficiencies.

(i) Computation of α_0 Under True AR(1) Correlation Structure For EQC Working Correlation Structure

Under the working exchangeable correlation structure, $\hat{\alpha}$ satisfies the estimating equation

$$\sum_{i=1}^{K}\sum_{t\neq u}^{T}(\tilde{y}_{it}\tilde{y}_{iu} - \alpha) = 0, \tag{6.36}$$

where $\tilde{y}_{it} = (y_{it} - \mu_{it})/\{\sigma_{itt}\}^{1/2}$, as in (6.27), with $\mu_{it} = \sigma_{itt} = \tilde{\mu}_i = \exp(\tilde{x}'_i\beta)$ for the present stationary case. Note that for the true AR(1) correlation structure, $E(\tilde{y}_{it}\tilde{y}_{iu}) = \rho^{|t-u|}$ with $0 < \rho < 1$. This shows that $\hat{\alpha}$ obtained from (6.36), if it exists, will converge to α_0 satisfying

$$\alpha_0 = 2\rho\{T - (1-\rho^T)/(1-\rho)\}/T(T-1)(1-\rho). \tag{6.37}$$

For example, when $\rho = 0.7$ the equation (6.37) yields $\alpha_0 = 0.52$, 0.35 and 0.26 for $T = 5$, 10, and 15, respectively.

Now to compute the efficiency of the 'working' equicorrelation structure based GEE estimator $\hat{\beta}_G$, when in fact the repeated counts truly follow the AR(1) correlation structure, we need to put AR(1) based $C_i(\rho)$ and EQC based $R(\alpha_0)$ in (6.34), for example, with $\alpha_0 = 0.52$ when $\rho = 0.7$ for $T = 5$. The efficiencies for selected ρ and for the selected design covariates as in Section 6.4.2.4 for $T = 5,\ 10,\ 15$, are shown in Table 6.6.

For MA(1) Working Correlation Structure

For the working MA(1) correlation structure, we solve

$$\sum_{i=1}^{K}\sum_{t=1}^{T-1}\tilde{y}_{it}\tilde{y}_{i(t+1)} - K(T-1)\alpha = 0, \tag{6.38}$$

to obtain $\hat{\alpha}$. If $\hat{\alpha}$ exists, then in this case $\hat{\alpha}$ will converge to $\alpha_0 = \rho$, because, under the true AR(1) structure, $E(\tilde{y}_{it}\tilde{y}_{i(t+1)}) = \rho$. Note, however, that although in the present case ρ can take any value from 0 to 1, we can use only the range

Table 6.6 Percentage relative efficiency of $\hat{\beta}_{1,G}$ and $\hat{\beta}_{2,G}$ to the true correlation structure based estimators $\hat{\beta}_{1,T}(=\hat{\beta}_{1,GQL})$ and $\hat{\beta}_{2,T}(=\hat{\beta}_{2,GQL})$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for AR(1) structure, for $\mu_{it}=\tilde{\mu}_i=\exp(\tilde{x}_{i1}\beta_1+\tilde{x}_{i2}\beta_2)$ with $\beta_1=\beta_2=1$

True Correlation Structure AR(1)

Working Structure	MA(1)				EQC			
T	ρ	α_0	$\hat{\beta}_{1,MA(1)}$	$\hat{\beta}_{2,MA(1)}$	ρ	α_0	$\hat{\beta}_{1,EQC}$	$\hat{\beta}_{2,EQC}$
5	0.3	0.3	100	100	0.3	0.15	98	98
	0.49	0.49	95	95	0.7	0.52	95	95
10	0.3	0.3	100	100	0.3	0.08	99	99
	0.49	0.49	98	98	0.7	0.35	93	93
15	0.3	0.3	100	100	0.3	0.06	99	99
	0.49	0.49	97	97	0.7	0.26	93	93

$0<\rho(=\alpha_0)<0.5$ for the efficiency computation. This is because in the GEE approach ρ is unknown and the working correlation α can range from -0.5 to 0.5 only. This is clear from the formula of V_G in (6.34), where one cannot use $R^{-1}(\alpha_0)$ beyond the range $-0.5<\alpha<0.5$, as $R(\alpha)$ has the MA(1) correlation structure. In view of this we have chosen $\rho=0.3$ and 0.49 for our efficiency computations. These efficiencies are also reported in Table 6.6, for $T=5$, 10, and 15.

(ii) Computation of α_0 Under True MA(1) Correlation Structure For AR(1) Working Correlation Structure

Let $c_{i,ut}$ be the (u,t) element of the true correlation matrix $C_i(\rho)$. For MA(1) true correlation structure, $c_{i,ut}=\rho_1=\rho(1+\rho)$ if $|t-u|=1$, and $c_{i,ut}=0$ otherwise, where ρ_1 denotes the lag-1 correlation. Under this structure, ρ_1 satisfies $-0.5\leq\rho_1\leq 0.5$.

Now consider the working AR(1) correlation matrix. Here $r_{i,ut}=\alpha^{|t-u|}$ for $u,t=1,\ldots,T$. If we base the estimation again on the average correlation, the estimating equation

$$\sum_{i=1}^{K}\sum_{u<t}^{T}(\tilde{y}_{it}\tilde{y}_{iu}-\alpha^{|t-u|})=0 \tag{6.39}$$

results, giving $\hat{\alpha}$; a simple moment estimator for α, see also Crowder (1995), where $\tilde{y}_{iu}$ and $\tilde{y}_{it}$ are the standardized residuals defined as in (6.36). Because

$$E\left\{\sum_{u<t}^{T}\tilde{y}_{it}\tilde{y}_{iu}\right\}=(T-1)\rho_1=(T-1)\frac{\rho}{1+\rho}$$

under the MA(1) correlation structure, it follows from (6.39) that α_0 is in fact the solution of

$$\alpha_0(1-\alpha_0)^{-1}\{T-(1-\alpha_0^T)/(1-\alpha_0)\}-(T-1)\rho_1=0. \tag{6.40}$$

Therefore, if $\hat{\alpha}$ exists, $\hat{\alpha}$ will converge in probability to α_0, α_0 being related to ρ through (6.40). For example, when $\rho_1 = 0.4$, that is, $\rho = 0.67$, the α_0 values are approximately 0.31, 0.30, and 0.29 for $T = 5$, 10, and 15 respectively. For selected values of ρ, the efficiencies of $\hat{\beta}_G$ for the MA(1) versus AR(1) correlation structures, are shown in Table 6.7.

Table 6.7 Percentage relative efficiency of $\hat{\beta}_{1,G}$ and $\hat{\beta}_{2,G}$ to the true correlation structure based estimators $\hat{\beta}_{1,TR}(=\hat{\beta}_{1,GQL})$ and $\hat{\beta}_{2,TR}(=\hat{\beta}_{2,GQL})$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for MA(1) structure, for $\mu_{it} = \tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ with $\beta_1 = \beta_2 = 1$

True Correlation Structure MA(1)								
Working Structure	AR(1)				EQC			
T	ρ	α_0	$\hat{\beta}_{1,AR(1)}$	$\hat{\beta}_{2,AR(1)}$	ρ	α_0	$\hat{\beta}_{1,EQC}$	$\hat{\beta}_{2,EQC}$
5	0.25	0.17	100	100	0.25	0.08	99	99
	0.67	0.31	99	99	0.67	0.16	97	97
10	0.25	0.17	100	100	0.25	0.04	99	99
	0.67	0.30	100	100	0.67	0.08	98	98
15	0.25	0.17	100	100	0.25	0.04	99	99
	0.67	0.29	100	100	0.67	0.05	98	98

For EQC Working Correlation Structure

For the working exchangeable correlation matrix $R(\alpha)$, one writes $r_{i,ut} = \alpha$ for all u,t except for $u = t$. We must have $-\{1/(T-1)\} \le \alpha \le 1$ for $R(\alpha)$ to be a positive definite matrix, where T is the dimension of the $R(\alpha)$ matrix. It then follows that the moment estimator $\hat{\alpha}$ [see also Crowder (1995] for α is given by

$$\hat{\alpha} = \sum_{i=1}^{K}\sum_{u\neq t}^{T} \hat{r}_{i(ut)}/KT(T-1)$$

$$= \sum_{i=1}^{K}\sum_{u\neq t}^{T} \tilde{y}_{iu}\tilde{y}_{it}/KT(T-1). \tag{6.41}$$

Because $C_i(\rho)$ has the MA(1) correlation structure,

$$E(\hat{\alpha}) = \{KT(T-1)\}^{-1}2K(T-1)\rho_1 = 2\rho_1/T = \frac{2\rho}{T(1+\rho)}. \tag{6.42}$$

Thus, if $\hat{\alpha}$ exists, then $\hat{\alpha}$ converges to $\alpha_0 = 2\rho_1/T$. Therefore, to compute the efficiency of $\hat{\beta}_G$, we use the true $\rho_1 = \rho/(1+\rho)$ for $C_i(\rho)$ and $\alpha_0 = 2\rho_1/T$ for $R(\alpha_0)$ in V_G given in (6.34). For example, with $T = 5$ and $\rho = 0.67$, we use $\alpha_0 = 0.16$ in $R(\alpha_0)$. The efficiencies for selected values of ρ are shown in Table 6.7.

(iii) Computation of α_0 Under True Equicorrelation (EQC) Structure

For AR(1) Working Correlation Structure:

For the working AR(1) correlation structure, the estimating equation for α remains the same as (6.39). However, as $E(\tilde{y}_{iu}\tilde{y}_{it}) = \rho$ under the true exchangeable correlation structure, $\hat{\alpha}$ obtained from (6.39), if it exists, converges to α_0, now satisfying the equation

$$\alpha_0(1-\alpha_0)^{-1}\{T-(1-\alpha_0^T)/(1-\alpha_0)\} - T(T-1)\rho/2 = 0. \qquad (6.43)$$

Here $\rho \geq -1/(T-1)$. Consequently, we use only positive ρ values for efficiency computations. For example, when $\rho = 0.7$ is used in (6.43), α_0 is 0.83, 0.90, and 0.93 for $T = 5$, 10, and 15 respectively. Now the efficiencies of AR(1) 'working' structure based $\hat{\beta}_G$, when EQC is the true correlation structure, are shown in Table 6.8, for the selected values of ρ.

Table 6.8 Percentage relative efficiency of $\hat{\beta}_{1,G}$ and $\hat{\beta}_{2,G}$ to the true correlation structure based estimators $\hat{\beta}_{1,TR}(=\hat{\beta}_{1,GQL})$ and $\hat{\beta}_{2,TR}(=\hat{\beta}_{2,GQL})$, respectively, with true stationary correlation matrix $C_1^*(\rho)$ for EQC structure, for $\mu_{it} = \tilde{\mu}_i = \exp(\tilde{x}_{i1}\beta_1 + \tilde{x}_{i2}\beta_2)$ with $\beta_1 = \beta_2 = 1$.

EQC True Correlation Structure

Working Structure	AR(1)				MA(1)			
T	ρ	α_0	$\hat{\beta}_{1,AR(1)}$	$\hat{\beta}_{2,AR(1)}$	ρ	α_0	$\hat{\beta}_{1,MA(1)}$	$\hat{\beta}_{2,MA(1)}$
5	0.3	0.49	96	96	0.3	0.3	99	99
	0.7	0.83	95	95	0.49	0.49	92	92
10	0.3	0.65	95	95	0.3	0.3	99	99
	0.7	0.90	94	94	0.49	0.49	98	98
15	0.3	0.74	94	94	0.3	0.3	100	100
	0.7	0.93	93	93	0.49	0.49	98	98

For MA(1) Working Correlation Structure

For the working MA(1) correlation structure, the estimating equation for α is given by (6.38). Because $E(\tilde{y}_{it}\tilde{y}_{i(t+1)}) = \rho$ for the true exchangeable correlation structure, it follows from (6.38) that $\hat{\alpha}$, if it exists, converges to $\alpha_0 = \rho$. The efficiencies of $\hat{\beta}_G$ for the exchangeable versus MA(1) correlation structure are also shown in Table 6.8, for selected values of ρ.

Note that when the efficiencies displayed in Tables 6.6 – 6.8 under correlation structure misspecification are compared with those in Table 6.2 computed for the independence assumption based regression estimators, it is seen that in some cases, especially when EQC is the true correlation structure, the $\hat{\beta}(I)$ appears to be equally or more efficient than the GEE based estimator $\hat{\beta}_G$. For this reason, as Sutradhar and Das (1999) [see also Sutradhar (2003)] argued, there is no guarantee that the

GEE approach can provide more efficient estimates than the simpler MM estimates obtained from (6.6) or QL estimates obtained from (6.9).

6.5 Nonstationary Correlation Models

In Section 6.3, we provided three stationary correlation models for longitudinal count data. In Section 6.4, we discussed various estimation techniques including the GEE and GQL approaches, for the estimation of the regression effects. Note that in the GEE approach, the selection of a suitable 'working' correlation structure out of these three or other possible correlation structures is left to the user. It was shown in Section 6.4 [see also Sutradhar and Das (1999)] that the use of such a 'working' correlation structure may in reality produce a less efficient estimate for the regression effect β than the 'independence' assumption based estimate. As a remedy, Sutradhar (2003) has suggested using a general (robust) autocorrelation structure that accommodates the above three stationary correlation structures as special cases. Thus, as demonstrated in Section 6.4.2.3 (see Table 6.2), if the data follow this class of Gaussian type stationary correlation structure, then the solution of a generalized quasi-likelihood equation, following Sutradhar (2003), always produces consistent and efficient estimates.

There, however, remains a concern that it may not be reasonable to use a stationary correlation structure when it is known that the covariates are time dependent. In Section 6.5.1, we provide three nonstationary correlation models as a generalization of the stationary AR(1), MA(1), and EQC structures, discussed in Section 6.3. These models produce the same mean and variance functions, and different correlation structures, under both stationary and nonstationary conditions. Under the assumption that the repeated count data follow one of these three possible nonstationary models, in Section 6.5.2, we discuss the estimation of the parameters under all three models. In Section 6.6.1, we deal with more nonstationary autocorrelation models that belong to the same autocorrelation class as that of Section 6.5, but now the marginal means and variances can be different under different models. In Section 6.6.2 we provide a model selection criterion based on the principle of minimum error sum of squares. A simulation study is conducted in Section 6.6.3 to examine the performances of the estimates under the true as well as misspecified models. Also, the simulation study in the same section justifies the model selection criterion. In Section 6.7, a real-life data example is discussed both for model selection as well as estimation of the regression effects and the correlation parameters.

6.5.1 Nonstationary Correlation Models with the Same Specified Marginal Mean and Variance Functions

6.5.1.1 Nonstationary AR(1) Models

Suppose that y_{i1} follows the Poisson distribution with mean parameter $\mu_{i1} = \exp(x'_{i1}\beta)$; that is, $y_{i1} \sim \text{Poi}(\mu_{i1} = \exp(x'_{i1}\beta))$, and for $t = 2,\ldots,T$, y_{it} relates to $y_{i,t-1}$ through the dynamic relationship

$$y_{it} = \rho * y_{i,t-1} + d_{it}, \text{ for } t = 2,\ldots,T, \tag{6.44}$$

where

$$\rho * y_{i,t-1} = \sum_{s=1}^{y_{i,t-1}} b_s(\rho),$$

with $Pr[b_s(\rho) = 1] = \rho$ and $Pr[b_s(\rho) = 0] = 1-\rho$. Also suppose that

$$y_{i,t-1} \sim \text{Poi}(\mu_{i,t-1}), \text{ and } d_{it} \sim \text{Poi}(\mu_{it} - \rho\mu_{i,t-1}),$$

with $\mu_{it} = e^{x'_{it}\beta}$, and d_{it} and $y_{i,t-1}$ are independent. After some algebra, it may be shown that this model (6.44) yields the means and the variances as

$$E(Y_{it}) = \text{var}(Y_{it}) = \mu_{it} = e^{x'_{it}\beta}, \tag{6.45}$$

and for $u < t$ with $t = 2,\ldots,T$, nonstationary (ns) correlations, say $c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho)$, as

$$\begin{aligned} corr(Y_{iu}, Y_{it}) &= c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho) \\ &= \rho^{t-u}\sqrt{\frac{\mu_{iu}}{\mu_{it}}}, \end{aligned} \tag{6.46}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu_{it}}{\mu_{i,t-1}}\right], \; t = 2,\cdots,T. \tag{6.47}$$

Stationary Correlation Structure: Note that in the stationary case, that is, when the covariates are time independent such as $x_{it} = \tilde{x}_i$ for all $t = 1,\ldots,T$, the means and variances given by (6.45) and the correlation matrix given by (6.46) become stationary. In particular, the nonstationary correlations given by (6.46) reduce to the covariates free stationary correlations

$$c^*_{i,ut}(\rho)) = (\rho^{|t-u|}), \text{ for all } u \neq t, u,t = 1,\ldots,T, \tag{6.48}$$

which is same as the correlation in (6.17) derived under the stationary correlation model (6.14).

6.5.1.2 Nonstationary MA(1) Models

To generalize the stationary MA(1) model [Sutradhar (2003)] to the nonstationarity case, we consider the dynamic relationship

$$\begin{aligned} y_{i1} &\sim \text{Poi}(\mu_{i1} = \exp(x'_{i1}\beta)) \\ y_{it} &= \rho * d_{i,t-1} + d_{it}, \;\; \text{for } t = 2, \ldots, T, \end{aligned} \tag{6.49}$$

where

$$d_{it} \overset{iid}{\sim} \text{Poi}\left[\sum_{j=0}^{t-1} (-\rho)^j \mu_{i,t-j}\right] \;\; \text{for all} \;\; t = 1, \ldots, T.$$

After some algebra, this model yields the same means and variances as in (6.45) derived under the AR(1) model. Furthermore, it can be shown that the correlations are given by

$$\text{corr}(Y_{iu}, Y_{it}) = c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho) = \begin{cases} \frac{\rho\{\sum_{j=0}^{\min(u,t)-1} (-\rho)^j \mu_{i,\min(u,t)-j}\}}{\sqrt{\mu_{iu}\mu_{it}}} & \text{for } |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6.50}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \rho_{i20}, \ldots, \rho_{it0}, \ldots, \rho_{iT0}\right], \tag{6.51}$$

where ρ_{it0} is the solution of $\sum_{j=0}^{t-1}(-\rho)^j \mu_{i,t-j} = 0$. Note that this range restriction may allow only a narrow range for the ρ parameter.

Stationary Correlation Structure: Note that in the stationary case, the means and the variances have the form $\mu_{it} = \mu_{i\cdot} = \exp(\tilde{x}'_i\beta)$ for all $t = 1, \ldots, T$. Furthermore, by (6.50), the limiting correlations when $\min(u,t) \to \infty$ have the formula

$$c^*_{i,ut}(\rho) = \text{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho\{\sum_{j=0}^{\infty}(-\rho)^j = \frac{\rho}{1+\rho} & \text{for} |u-t| = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6.52}$$

which is free from the time-dependent covariates. This stationary correlation is the same as the correlation in (6.19) derived under the stationary MA(1) model (6.18).

6.5.1.3 Nonstationary EQC Models

To generate a nonstationary equicorrelations model, we consider

$$\begin{aligned} y_{i1} &\sim \text{Poi}(\mu_{i1} = \exp(x'_{i1}\beta)) \\ y_{it} &= \rho * y_{i1} + d_{it}, \;\; \text{for } t = 2, \ldots, T, \end{aligned} \tag{6.53}$$

where d_{it} is assumed to be distributed as

$$d_{it} \sim \text{Poi}(\mu_{it} - \rho\mu_{i1})$$

with $\mu_{it} = e^{x'_{it}\beta}$. Also it is assumed that d_{it} for $t = 2, \ldots, T$, are independent of y_{i1}. It then follows that $E(Y_{it}) = \text{var}(Y_{it}) = \mu_{it} = e^{x'_{it}\beta}$ as in the AR(1) and MA(1) cases, for all $t = 1, \ldots, T$, and for $u < t$,

$$\text{cov}(Y_{iu}, Y_{it}) = \rho\mu_{i1}, \tag{6.54}$$

yielding the nonstationary correlation structure

$$\text{corr}(Y_{iu}, Y_{it}) = c^{(ns)}_{i,ut}(x_{iu}, x_{it}, \rho) = \frac{\rho\mu_{i1}}{\sqrt{\mu_{iu}\mu_{it}}}, \tag{6.55}$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu_{it}}{\mu_{i1}}\right], \ t = 2, \ldots, T.$$

Stationary Correlation Structure: Note that when covariates are time independent, that is, $x_{it} = \tilde{x}_i$ for all $t = 1, \ldots, T$, the nonstationary correlations in (6.55) reduce to the stationary correlations in (6.22) derived under the stationary exchangeable correlation model (6.20).

For convenience, we summarize the means, variances, and correlations for all three nonstationary correlation models, as in Table 6.9.

Table 6.9 A class of nonstationary correlation models for longitudinal count data and basic properties.

Model	Dynamic Relationship	Mean, Variance and Correlations
AR(1)	$y_{it} = \rho * y_{i,t-1} + d_{it}, t = 2, \ldots, T$ $y_{i1} \sim Poi(\mu_{i1})$ $d_{it} \sim Poi(\mu_{it} - \rho\mu_{i,t-1}), t = 2, \ldots, T$	$E[Y_{it}] = \mu_{it}$ $\text{var}[Y_{it}] = \mu_{it}$ $\text{corr}[Y_{iu}, Y_{it}] = \rho^{(ns)}_{\lvert t-u \rvert}$ $= \rho^{\lvert t-u \rvert}\left[\frac{\mu_{iu}}{\mu_{it}}\right]^{\frac{1}{2}}$
MA(1)	$y_{it} = \rho * d_{i,t-1} + d_{it}, t = 2, \ldots, T$ $y_{i1} \sim Poi(\mu_{i1})$ $d_{it} \overset{iid}{\sim} Poi\left[\sum_{j=0}^{t-1}(-\rho)^j \mu_{i,t-j}\right] \ t = 1, \ldots, T$	$E[Y_{it}] = \mu_{it}$ $\text{var}[Y_{it}] = \mu_{it}$ $\text{corr}[Y_{iu}, Y_{it}] = \rho^{(ns)}_{\lvert u-t \rvert}$ $= \begin{cases} \frac{\rho\{\sum_{j=0}^{min(u,t)-1}(-\rho)^j \mu_{i,min(u,t)-j}\}}{\sqrt{\mu_{iu}\mu_{it}}} & \text{for } \lvert u-t \rvert = 1 \\ 0 & \text{otherwise,} \end{cases}$
EQC	$y_{it} = \rho * y_{i1} + d_{it}, t = 2, \ldots, T$ $y_{i1} \sim Poi(\mu_{i1})$ $d_{it} \sim P(\mu_{it} - \rho\mu_{i1}), t = 2, \ldots, T$	$E[Y_{it}] = \mu_{it}$ $\text{var}[Y_{it}] = \mu_{it}$ $\text{corr}[Y_{iu}, Y_{it}] = \rho^{(ns)}_{\lvert u-t \rvert}$ $= \frac{\rho\mu_{i1}}{\sqrt{\mu_{iu}\mu_{it}}}$

6.5.2 Estimation of Parameters

It follows from Sections 6.5.1.1 – 6.5.1.3 (see also Table 6.9) that all three nonstationary, namely AR(1), MA(1), and EQC, models have the same mean and variance structures. Their correlation structures are, however, different; that is, the nonstationary correlation matrix $C_i^{(ns)}(x_i,\rho) = (c_{i,ut}^{(ns)}(x_{iu},x_{it},\rho))$ is not the same under all three models. Suppose that the structure is identified (see Section 6.5.3 for an exploratory way for the model selection). Now assuming that we have a consistent estimate for ρ, say $\hat{\rho}$, we may obtain a consistent and highly efficient estimate for β by using the GQL approach that we provide below.

GQL Estimating Equation for β: Similar to the GQL estimation (6.26) for the stationary case, we now solve the GQL estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{(ns)^{-1}}(\hat{\rho})(y_i - \mu_i) = 0, \tag{6.56}$$

where $\mu_i = (\mu_{i1},\ldots,\mu_{it},\ldots,\mu_{iT})'$ is the mean vector of $y_i = (y_{i1},\ldots,y_{it},\ldots,y_{iT})'$ with

$$\begin{aligned} \mu_{it} &= \exp(x_{it}'\beta) \\ \Sigma_i^{(ns)}(\hat{\rho}) &= A_i^{1/2} C_i^{(ns)}(x_i,\hat{\rho}) A_i^{1/2}, \end{aligned} \tag{6.57}$$

where

$$A_i = \text{diag}[\sigma_{i11},\ldots,\sigma_{itt},\ldots,\sigma_{iTT}],$$

with $\sigma_{itt} = \exp(x_{it}'\beta)$. Furthermore, in (6.56), $\partial\mu_i'/\partial\beta = X_i'A_i$, with X_i as the $T \times p$ covariate matrix as defined earlier.

Let $\hat{\beta}_{GQL}$ denote the solution of (6.56) after using $\hat{\rho}$ computed under the selected model. Under mild regularity conditions one may then show that $\hat{\beta}_{GQL}$ has the asymptotic (as $K \to \infty$) normal distribution given by

$$K^{1/2}(\hat{\beta}_{GQL} - \beta) \sim N\left(0,\ K\left[\sum_{i=1}^{K} X_i' A_i \Sigma_i^{(ns)^{-1}} A_i X_i\right]^{-1}\right).$$

We now show how to compute $\hat{\rho}$ under all three models.

6.5.2.1 Estimation of ρ Parameter Under AR(1) Model

Moment Equation for ρ: Under the nonstationary AR(1) model (6.44), the moment estimate of ρ has the formula given by

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{t=2}^{T} \tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^{K}\sum_{t=1}^{T} \tilde{y}_{it}^2} \frac{KT}{\sum_{i=1}^{K}\sum_{t=2}^{T}[\mu_{i,t-1}/\mu_{it}]^{1/2}}, \tag{6.58}$$

where $\tilde{y}_{it} = [y_{it} - \mu_{it}]/\sqrt{\mu_{it}}$. Note that the formula for ρ given by (6.58) was obtained by equating the lag 1 sample autocorrelation with its population counterpart given by (6.46). Furthermore, $\hat{\rho}$ computed by (6.58) must satisfy the range restriction given in (6.47). This implies that if the value of $\hat{\rho}$ computed by (6.58) falls beyond the range shown in (6.47), we use the upper limit of ρ given in (6.47) as the estimate of ρ.

6.5.2.2 Estimation of ρ Parameter Under MA(1) Correlation Model

Note that unlike the formula for lag 1 correlations (6.46) under the AR(1) model, the formula for this lag 1 correlation given by (6.50) under the nonstationary MA(1) model (6.49) involves a complicated summation. Thus, it is convenient to solve the moment equation for ρ by using the Newton−Raphson iterative technique. To be specific, by writing the moment equation as

$$g(\rho) = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-1}\tilde{y}_{it}\tilde{y}_{i,t+1}/K(T-1)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\tilde{y}_{it}^2/KT} - \frac{\rho}{T-1}\sum_{u=1}^{T-1}\left[\frac{\sum_{j=0}^{u-1}(-\rho)^j\mu_{i,u-j}}{\sqrt{\mu_{iu}\mu_{i,u+1}}}\right] = 0, \tag{6.59}$$

we solve for ρ iteratively by using the Newton−Raphson iterative formula

$$\hat{\rho}(r+1) = \hat{\rho}(r) - \left[\{\frac{\partial g(\rho)}{\partial \rho}\}^{-1} g(\rho)\right]_{(r)},$$

where $[\cdot]_{(r)}$ denotes that the expression within brackets is evaluated at $\rho = \hat{\rho}(r)$, the rth iterative value of ρ. Note that $\hat{\rho}$ must satisfy the range restriction (6.51).

6.5.2.3 Estimation of ρ Parameter Under Exchangeable (EQC) Correlation Model

The moment estimating equation for the ρ parameter for the exchangeable model is quite similar to that of the AR(1) model. The difference between the two equations is that under the AR(1) process we have considered all lag 1 standardized residuals, whereas under the exchangeable model, one needs to use standardized residuals of all possible lags. Thus, following (6.58) for the AR(1) model, we write the moment formula for ρ under the exchangeable model as

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}}{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\tilde{y}_{it}^2}\ \frac{KT}{\sum_{i=1}^{K}\sum_{\ell=1}^{T-1}\sum_{t=1}^{T-\ell}\frac{\mu_{i1}}{[\mu_{it}\mu_{i,t+\ell}]^{\frac{1}{2}}}}, \tag{6.60}$$

where $\tilde{y}_{it} = [y_{it} - \mu_{it}]/\sqrt{\mu_{it}}$. Note that $\hat{\rho}$ must satisfy the range restriction in (6.55). This implies that if the value of $\hat{\rho}$ computed by (6.58) falls beyond the range shown in (6.55), we take $\hat{\rho}$ as the upper limit of ρ given in (6.55).

6.5.3 Model Selection

Note that in the stationary case it is not necessary to identify the correlation structure for the construction of the estimating equation (6.26) for β. This is because the estimating equation (6.26) is constructed based on a common correlation structure for $C_i^*(\rho)$ as given by (6.25) with ρ_ℓ estimated as

$$\hat{\rho}_\ell = \frac{\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\tilde{y}_{it}\tilde{y}_{i,t+\ell}/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}\tilde{y}_{it}^2/KT}, \tag{6.61}$$

(see also (6.27)) where $\tilde{y}_{it} = [y_{it} - \mu_{it}]/\sqrt{\sigma_{itt}}$. Nevertheless, if one would like to identify the stationary correlation structure for the purpose of forecasting or other reasons, this could be done by using the values of $\hat{\rho}_\ell$ for $\ell = 1,\ldots,T-1$. This is because one may show that

$$E[\hat{\rho}_\ell] = \rho_\ell,$$

approximately, and it is reasonable to use the values of $\hat{\rho}_\ell$ for $\ell = 1,\ldots,T-1$, to identify a stationary correlation structure.

As far as the identification of a nonstationary correlation structure is concerned, it appears that the values of $\hat{\rho}_\ell$ can still be used for such an identification. More specifically, simply compute the values of $\hat{\rho}_\ell$ by (6.61) and compare their pattern for best possible matching with those of $E[\hat{\rho}_\ell]$ under desired models for all possible values of $\rho = 0.0,\ 0.05,\ldots,0.90,\ 0.95$. Suppose that it is intended to find out whether the longitudinal count data follow one of the low-order, namely AR(1), MA(1), or EQC, models. To resolve such an issue, one would compute the $E[\hat{\rho}_\ell]$ under all these three models and select that model which produces a pattern for $\hat{\rho}_\ell$ similar to that of $E[\hat{\rho}_\ell]$.

For the longitudinal count data, the formulas for the expectations under the AR(1), MA(1), or EQC models are given by

$$\text{For } AR(1): E[\hat{\rho}_\ell] = \frac{\rho^\ell}{K(T-\ell)}\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\left[\frac{\mu_{it}}{\mu_{i,t+\ell}}\right]^{1/2} \quad \text{for } \ell = 1,\ldots,T-1 \tag{6.62}$$

$$\text{For } MA(1): E[\hat{\rho}_\ell] = \begin{cases} \frac{\rho}{K(T-\ell)}\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\left[\frac{\sum_{j=0}^{t-1}(-\rho)^j\mu_{i,t-j}}{\sqrt{\mu_{it}\mu_{i,t+\ell}}}\right] & \text{for } \ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.63}$$

$$\text{For } EQC: E[\hat{\rho}_\ell] = \frac{\rho}{K(T-\ell)}\sum_{i=1}^{K}\sum_{t=1}^{T-\ell}\left[\frac{\mu_{i1}}{\{\mu_{it}\mu_{i,t+\ell}\}^{\frac{1}{2}}}\right], \tag{6.64}$$

for $\ell = 1,\ldots,T-1$, where $\mu_{it} = \exp(x_{it}'\beta)$ for all $t = 1,\ldots,T$. Note that as far as the value of β is concerned for computing $\hat{\rho}_\ell$ by (6.61) and the expectations by (6.62) – (6.64), this may be obtained by solving the GQL estimating equation (6.26) under the 'working' independence assumption $\rho = 0.0$. This is because such an estimate is always consistent and one does not necessarily require an efficient estimate for β before the correlation structure is identified.

Further note that if the time dependent covariates are not so different over time, then the expected values in (6.62) – (6.64) would almost agree with the correlation pattern under the stationary case, described through (6.17), (6.19), and (6.22). To demonstrate this, we now examine empirically the pattern for $E[\hat{\rho}_\ell]$ under all three correlation models. For this purpose, we consider two time-dependent covariates as follows:

$$x_{it1} = \begin{cases} \frac{1}{2} & \text{for } t = 1,\ 2; \quad i = 1,\ldots,K/4 \\ 1 & \text{for } t = 3,\ 4; \quad i = 1,\ldots,K/4 \\ -\frac{1}{2} & \text{for } t = 1; \quad i = (K/4)+1,\ldots,3K/4 \\ 0 & \text{for } t = 2,\ 3; \quad i = (K/4)+1,\ldots,3K/4 \\ \frac{1}{2} & \text{for } t = 4; \quad i = (K/4)+1,\ldots,3K/4 \\ \frac{t}{8} & \text{for } t = 1,\ldots,4; \ i = (3K/4)+1,\ldots,K, \end{cases}$$

and

$$x_{it2} = \begin{cases} \frac{t-2.5}{8} & \text{for } t = 1,\ldots,4; \ i = 1,\ldots,K/2 \\ 0 & \text{for } t = 1,2; \quad i = (K/2)+1,\ldots,K \\ \frac{1}{2} & \text{for } t = 3,\ 4; \quad i = (K/2)+1,\ldots,K. \end{cases}$$

For $T = 4$ and $K = 100$, the values for $E[\hat{\rho}_\ell]$ computed by (6.62) – (6.64) for suitable values of ρ are displayed in Table 6.10.

It is clear from the results of the table that the $E[\hat{\rho}_\ell]$ for $\ell = 1,\ldots,T-1$, exhibit an exponentially decaying pattern under the nonstationary AR(1) model, whereas they exhibit a truncated pattern under the MA(1) model, and a constant pattern under the EQC model. These patterns are quite similar to those under the respective stationary correlation structure. Thus, it appears that in practice one may still exploit the values of $\hat{\rho}_\ell$ computed by (6.61) in order to diagnose the nonstationary correlation pattern. More specifically, because the values of $E[\hat{\rho}_\ell]$ for $\ell = 1,\ldots,T-1$, under the AR(1), MA(1), and EQC models exhibit three different patterns, and because the values of $\hat{\rho}_\ell$ computed from the data should reflect the pattern supported by the values of $E[\hat{\rho}_\ell]$, it is quite reasonable to examine the pattern generated by the values of $\hat{\rho}_\ell$ to diagnose the appropriate model.

Table 6.10 The pattern for $E[\hat{\rho}_\ell]$ for lag $\ell = 1,\ldots,T-1$, under AR(1), MA(1), and EQC correlation structures for longitudinal count data with selected values for the correlation index parameter ρ.

Correlation Structure								
AR(1)			MA(1)			EQC		
ρ	ℓ	$E[\hat{\rho}_\ell]$	ρ	ℓ	$E[\hat{\rho}_\ell]$	ρ	ℓ	$E[\hat{\rho}_\ell]$
0.3	1	0.282	0.1	1	0.089	0.3	1	0.251
	2	0.078		2	0.0		2	0.248
	3	0.022		3	0.0		3	0.248
0.5	1	0.469	0.2	1	0.168	0.5	1	0.417
	2	0.216		2	0.0		2	0.413
	3	0.103		3	0.0		3	0.412
0.6	1	0.563	0.3	1	0.239	0.6	1	0.502
	2	0.312		2	0.0		2	0.495
	3	0.178		3	0.0		3	0.494
0.68	1	0.638	0.4	1	0.306	0.7	1	0.587
	2	0.400		2	0.0		2	0.577
	3	0.259		3	0.0		3	0.577

6.6 More Nonstationary Correlation Models

6.6.1 Models with Variable Marginal Means and Variances

In this section, we demonstrate that as opposed to the nonstationary MA(1) model in (6.49), one may construct a different MA(1) model that produces the mean and the variance functions different from those produced by the nonstationary AR(1) (6.44) and EQC (6.53) models. These two latter models in (6.44) and (6.53) produce the mean and the variance as

$$E[Y_{it}] = \mathrm{var}[Y_{it}] = \exp(x_{it}'\beta). \tag{6.65}$$

We now construct an alternative MA(1) model to (6.49), and examine its mean, variance, and correlation structures.

6.6.1.1 Nonstationary MA(1) Models

Suppose that the non-stationary MA(1) model for the count responses has the same form, that is,

$$y_{it} = \rho * d_{i,t-1} + d_{it}, \tag{6.66}$$

as in (6.18) under the stationary case, but the model components are now assumed to satisfy the following distributional assumptions.

Assumption 1. For $t = 1,\ldots,T$, the discrete errors d_{it} follow the Poisson distribution as $d_{it} \sim P(\mu_{it}/(1+\rho))$, with $\mu_{it} = \exp(x'_{it}\beta)$.

Assumption 2. For all $t = 1,\ldots,T$, d_{it}s are independent.

Assumption 3. An initial discrete error $d_{i0} \sim P(\mu_{i0}/[1+\rho])$, where the choice of μ_{i0}, a function of some initial or past covariates, is left to the user. In the stationary case, $\mu_{i0} = \mu_{i1} = \cdots = \mu_{iT} = \mu_{i\cdot}$.

For $t = 1,\ldots,T$, by writing $z_{i,t-1} = \rho * d_{i,t-1}$, for convenience, one may now use the model (6.66) and compute the mean $\nu_{it} = E(Y_{it})$ and the variance $\sigma_{itt} = \mathrm{var}(Y_{it})$ as

$$\nu_{it} = E_{d_{i,t-1}}E[z_{i,t-1}] + E[d_{it}] = [\rho\mu_{i,t-1} + \mu_{it}]/(1+\rho), \tag{6.67}$$

and

$$\begin{aligned}\sigma_{itt} &= \mathrm{var}_{d_{i,t-1}}E[z_{it}|d_{i,t-1}] + E_{d_{i,t-1}}\mathrm{var}[z_{it}|d_{i,t-1}] + \mathrm{var}[d_{it}] \\ &= \mathrm{var}_{d_{i,t-1}}[\rho d_{i,t-1}] + E_{d_{i,t-1}}[\rho(1-\rho)d_{i,t-1}] + [\mu_{it}/(1+\rho)] \\ &= [\rho\mu_{i,t-1} + \mu_{it}]/(1+\rho), \end{aligned} \tag{6.68}$$

respectively. Thus, it is clear that for $t = 1,\ldots,T$, y_{it} has the mean ν_{it} and the variance $\sigma_{itt} = \nu_{it}$, which are, however, different from the mean and the variance functions given in (6.65) under the AR(1) and EQC models. Also, it is to be noted that the ρ parameter in the MA(1) model (6.66) must satisfy the range restriction $\max[-\mu_{it}/\mu_{i,t-1}] < \rho < 1$, for all i and t. Next by similar calculations as in the AR(1) model, it follows from (6.67) – (6.68) that under the MA(1) model, the ℓth $\ell = 1,\ldots,T-1$, lag autocorrelation is given by

$$\mathrm{corr}(Y_{it}, Y_{i,t-\ell}) = c^{(ns)}_{it,t-l}(x_i,\rho) = \left\{ \begin{array}{ll} [\rho\mu_{i,t-\ell}/(1+\rho)]/[\nu_{it}\nu_{i,t-\ell}]^{1/2} & \text{for } \ell = 1 \\ 0 & \text{for } \ell > 1. \end{array} \right\}, \tag{6.69}$$

which is nonstationary. This correlation structure is different from that (6.50) of the other MA(1) model (6.49).

Thus, under this alternative nonstationary MA(1) model (6.66), it is not only that the correlations are different from those of the AR(1) and EQC models, but the mean and the variances are also different.

6.6.2 Estimation of Parameters

Note that the three nonstationary models, namely AR(1), MA(1), and EQC introduced in Sections 6.5.1.1, 6.5.1.2, and 6.5.1.3, respectively, produce the same mean and variance functions but different correlation structures. In spite of their different correlation structures, the regression parameter β was estimated by solving the GQL estimating equation (6.56), which is unbiased for zero vector, irrespective of the model for the data. This happens because all three correlation models produce the same mean vector μ_i as given in (6.56). As opposed to Section 6.5, in Section 6.6 we now assume that the repeated count data are generated following either the AR(1) (6.44) or EQC (6.53) model from Section 6.5, or following the MA(1) model (6.66) introduced in Section 6.6.1.1. The MA(1) model (6.66) produces different mean and variance structure, thus it is no longer possible to use the estimating equation (6.56) for β to obtain consistent estimate, under the MA(1) model (6.66). This is, however, a valid equation to solve for β under the AR(1) and EQC models. Furthermore, for these two models (6.44) and (6.53), the ρ parameter is consistently estimated by (6.58) and (6.60), respectively.

In the next section, we demonstrate how to estimate β and ρ parameters of the MA(1) model (6.66).

6.6.2.1 GQL Estimation for Regression Effects β

We now fit the nonstationary MA(1) model (6.66) to the longitudinal count data. The mean and the variance structures under this model are given in (6.67) – (6.68), whereas the nonstationary correlation structure is given by (6.69).

Let

$$\nu_i = (\nu_{i1},\ldots,\nu_{it},\ldots,\nu_{iT})'$$

be the mean vector of y_i, where for $t = 1,\ldots,T$,

$$\nu_{it} = [\mu_{it} + \rho\mu_{i,t-1}]/(1+\rho)$$

by (6.67). For convenience, we assume that $\mu_{i0} = 0$. Furthermore, let $\Sigma_i^{(ns)}(\rho) = (\sigma_{iut})$ be the $T\times T$ covariance matrix of y_i, where

$$\sigma_{iut} = \begin{cases} \sigma_{itt}, & \text{if } u = t \\ \frac{\rho\mu_{iu}}{1+\rho}, & \text{if } u < t, \end{cases} \tag{6.70}$$

with σ_{itt} as in (6.68). It then follows that for known ρ, one may write the GQL estimating equation for β as

$$\sum_{i=1}^{K} \frac{\partial \nu_i'}{\partial\beta}\Sigma_i^{(ns)^{-1}}(\hat{\rho})(y_i - \nu_i) = 0, \tag{6.71}$$

which is a different estimating equation from that of under the AR(1) model (6.44) and EQC model (6.53).One may now solve (6.71) iteratively by using the Newton−Raphson algorithm. To be specific, (6.71) is solved for β iteratively by using

$$\hat{\beta}(r+1) = \hat{\beta}(r) + \left[\left\{\sum_{i=1}^{K}[(X_i^{'}A_i + Z_i^{'}B_i)\Sigma_i^{-1}(A_iX_i + B_iZ_i)]\right\}^{-1} \times \sum_{i=1}^{K}\left\{(X_i^{'}A_i + Z_i^{'}B_i)\Sigma_i^{-1}(y_i - \nu_i)\right\}\right]_{[r]}, \quad (6.72)$$

where

$$X_i' = (x_{i1}, \ldots, x_{it}, \ldots, x_{iT}), \quad Z_i' = (1_p, x_{i1}, \ldots, x_{i,T-1}),$$

$$A_i = \mathrm{diag}(\frac{\mu_{i1}}{1+\rho}, \frac{\mu_{i2}}{1+\rho}, \ldots, \frac{\mu_{it}}{1+\rho}, \ldots, \frac{\mu_{iT}}{1+\rho}),$$

$$B = \mathrm{diag}(0, \frac{\rho\mu_{i1}}{1+\rho}, \frac{\rho\mu_{i2}}{1+\rho}, \ldots, \frac{\rho\mu_{it}}{1+\rho}, \ldots, \frac{\rho\mu_{i,T-1}}{1+\rho}),$$

and $[.]_r$ denotes the fact that the expression within the brackets is evaluated at $\hat{\beta}(r)$. Let $\hat{\beta}_{GQL}$ denote the solution obtained from (6.72). Under mild regularity conditions it may be shown that $\hat{\beta}_{GQL}$ has the asymptotic (as $K \to \infty$) normal distribution given as

$$K^{\frac{1}{2}}(\hat{\beta}_{GQL} - \beta) \sim N\left(0,\ K\left[\sum_{i=1}^{K}(X_i'A_i + Z_i'B_i)\Sigma_i^{-1}(A_iX_i + B_iZ_i)\right]^{-1}\right). \quad (6.73)$$

6.6.2.2 Moment Estimation for the Correlation Parameter ρ

As far as the ρ parameter is concerned, we estimate this parameter consistently by using the well-known method of moments. For the purpose, we first observe under the MA(1) model that

$$E\left[\frac{(Y_{it} - \nu_{it})}{\sqrt{\nu_{it}}}\right]^2 = 1$$

$$E\left[\frac{(Y_{it} - \nu_{it})}{\sqrt{\nu_{it}}}\frac{(Y_{i,t-1} - \nu_{i,t-1})}{\sqrt{\nu_{i,t-1}}}\right] = \frac{\rho}{1+\rho}\frac{\mu_{i,t-1}}{\sqrt{\nu_{it}\nu_{i,t-1}}}. \quad (6.74)$$

Consequently, one may obtain a consistent estimator of ρ by solving the moment equation

$$\frac{a(\rho)}{b(\rho)} = \frac{\rho}{1+\rho} c(\rho), \tag{6.75}$$

where

$$a(\rho) = \frac{1}{K(T-1)} \sum_{i=1}^{K} \sum_{t=2}^{T} \frac{(Y_{it} - \nu_{it})}{\sqrt{\nu_{it}}} \frac{(Y_{i,t-1} - \nu_{i,t-1})}{\sqrt{\nu_{i,t-1}}}$$
$$b(\rho) = \frac{1}{KT} \sum_{i=1}^{K} \sum_{t=1}^{T} \left[\frac{(Y_{it} - \nu_{it})}{\sqrt{\nu_{it}}} \right]^2,$$

and

$$c(\rho) = \frac{1}{K(T-1)} \sum_{i=1}^{K} \sum_{t=2}^{T} \frac{\mu_{i,t-1}}{\sqrt{\nu_{it} \nu_{i,t-1}}}. \tag{6.76}$$

Note that unlike solving for ρ by (6.58) under the AR(1) process or by (6.60) under the EQC model, solving (6.75) for ρ under the MA(1) model is complicated as ν_{it} contains ρ for all $t = 1, \ldots, T$. One may, however, obtain an approximate solution, based on an iterative technique by using an initial value of ρ, say ρ_0, in all ν_{it}, and solving (6.75) for ρ as

$$\rho_1 = \frac{a(\rho_0)}{b(\rho_0)c(\rho_0) - a(\rho_0)}. \tag{6.77}$$

Next one may improve the estimate of ρ by using ρ_1 in place of ρ_0 in (6.75). That is, the new solution of ρ is obtained as

$$\rho_2 = \frac{a(\rho_1)}{b(\rho_1)c(\rho_1) - a(\rho_1)}. \tag{6.78}$$

This iteration continues until convergence.

6.6.3 Model Selection

Under the assumption that the longitudinal count data follow either the nonstationary AR(1) (6.44) or EQC (6.53) model described in Section 6.5, we have estimated their common regression parameter by (6.56), and their correlation parameter ρ was estimated by (6.58) and (6.60), respectively. Next, for the estimation of the parameters of the MA(1) model (6.66), we have used the GQL approach (6.71) for β estimation, and the moment estimating equation (6.75) for the estimation of the ρ parameter. Now the question arises, which model to recommend for use in practice? We consider a lag 1 model fitting approach to answer this question. Note that this model selection approach is different from that we have used in Section 6.5.3. One of the reasons for this difference in model selection approaches is that in Section 6.5 we have considered models with the same mean functions, whereas in this section we have considered models with different mean functions. To be more specific, when the models do not agree for the mean functions, it is better to fit them to the data separately and then see which model fits the data best. Thus, in this section, we fit a model M (say) to the data and simply compute the error sum of squares, G_M, under the model M, defined by

$$G_M = \sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it} - \hat{y}_{it}(M)]^2, \tag{6.79}$$

and recommend that model with the smallest value of the error sum of squares. In (6.79), $\hat{y}_{it}(M)$ denotes the fitted value of y_{it} under the model M.

The formula for $\hat{y}_{it}(M)$under each of the three models are as follows.

When Nonstationary AR(1) Model (6.44) Is Fitted

$$\hat{y}_{it} = \begin{cases} \hat{\mu}_{it} & \text{for } t = 1 \\ \hat{\mu}_{it} + \hat{\rho}\{y_{i,t-1} - \hat{\mu}_{i,t-1}\} & \text{for } t = 2,\ldots,T, \end{cases} \tag{6.80}$$

with $\hat{\mu}_{it} = \exp(x'_{it}\hat{\beta})$, where $\hat{\beta}$ is obtained by solving the GQL estimating equation (6.56) and $\hat{\rho}$ is obtained as the moment estimate by using (6.58).

When Non-stationary MA(1) Model (6.66) is Fitted

$$\hat{y}_{it} = \begin{cases} \frac{\hat{\mu}_{it}}{1+\hat{\rho}} & \text{for } t = 1 \\ \frac{\hat{\mu}_{it}+\hat{\rho}\hat{\mu}_{i,t-1}}{1+\hat{\rho}} & \text{for } t = 2,\ldots,T, \end{cases} \tag{6.81}$$

with $\hat{\mu}_{it} = \exp(x'_{it}\hat{\beta})$, but $\hat{\beta}$ is obtained by solving the GQL estimating equation (6.71) and $\hat{\rho}$ is obtained as the moment estimate by solving (6.75). Note that estimating equations in (6.71) and (6.75) under the MA(1) model are similar to but different from the AR(1) based estimating equations (6.56) and (6.58), respectively.

When Nonstationary Exchangeable or Equicorrelation (EQC) Model (6.53) Is Fitted

$$\hat{y}_{it} = \begin{cases} \hat{\mu}_{it} & \text{for } t = 1 \\ \hat{\mu}_{it} + \hat{\rho}\{y_{i1} - \hat{\mu}_{i1}\} & \text{for } t = 2,\ldots,T, \end{cases} \tag{6.82}$$

with $\hat{\mu}_{it} = \exp(x'_{it}\hat{\beta})$, where $\hat{\beta}$ and $\hat{\rho}$ are obtained by solving the GQL (6.56) and moment estimating equation (6.60).

6.6.4 Estimation and Model Selection: A Simulation Example

We now consider a simulation study and examine the performance of the GQL estimation approach discussed in Section 6.6.2. We also examine the performance of the mean squared errors (MSEs) based model selection approach discussed in Section 6.6.3. We demonstrate here that if a misspecified model is used, then the GQL approach may lead to inconsistent estimates for the regression effects causing a serious inference problem. This happens when the mean and the variance functions of the true model are different from those of the so-called 'working' or misspecified model.

6.6.4.1 Simulated Estimates Under the True and Misspecified Models

To choose a simulation design, we take $p = 2$ and $\beta_1 = \beta_2 = 0.5$. With regard to the correlation index parameter, we consider two cases, one with moderately large $\rho = 0.5$ and the other with large $\rho = 0.75$. Next we choose $K = 300$, where K is the number of independent individuals. As far as the values of the covariates are concerned, we consider two time-dependent covariates given in Section 6.5.3.

Next, for a selected value of K, and ρ, we simulate the longitudinal responses $y_{i1}, \ldots, y_{iT}$, following a true, say AR(1) or exchangeable correlation model as described in Section 6.5.1, or the MA(1) model as described in Section 6.6.1. We consider 1000 simulations. In each simulation, we then estimate the parameters β_1, β_2, and ρ, by using the formulas for all three processes as discussed in Section 6.6.2. The simulated mean and the simulated standard error of the estimates are reported in Table 6.11.

The results in Table 6.11 clearly indicate that fitting a 'working' nonstationary model can be extremely dangerous. For example, when the longitudinal data are generated, say following the MA(1) model, and also the estimates are obtained by fitting the MA(1) model, the GQL estimates appear to perform very well. The GQL estimates computed based on either the AR(1) or EQC model, however, appear to be far off from the true parameter values. To be specific, when $\rho = 0.75$, the true MA(1) based GQL estimates for $\beta_1 = 0.5$ and $\beta_2 = 0.5$ are 0.491 with standard error 0.175, and 0.499 with standard error 0.175, respectively. These estimates are very close to the true values. Similarly, the moment estimate for $\rho = 0.75$ is found to be 0.749 with small standard error 0.064, which indicates superb performance of the GQL approach provided the true model is used for the estimation. On the contrary, when AR(1) model is used as the 'working' model, the regression estimates are found to be -1.016 and 1.709 for true $\beta_1 = \beta_2 = 0.5$. It is clear that these estimates are complete nonsense. Similar results hold for ρ estimation. The AR(1) based moment estimate for $\rho = 0.75$ is found to be 1.000, which is also highly biased. Note that these results are not surprising. This is because unlike under the stationary models [Liang and Zeger (1986), Sutradhar (2003)], the mean and variance structures under different correlation models may be different.

Table 6.11 The simulated means and the simulated standard errors of the estimates of the regression and the correlation index parameters under both true and 'working' nonstationary AR(1), MA(1), and EQC (equicorrelations) models for longitudinal count data, with true $\beta_1 = \beta_2 = 0.5$, for $K = 300$ individuals, and a selected value of ρ, based on 1000 simulations.

			True Nonstationary Correlation Model					
			AR(1)		MA(1)		EQC	
Working Model	True ρ	Parameters	SM	SSE	SM	SSE	SM	SSE
AR(1)	0.60	β_1	0.499	0.111	−0.159	0.370	0.502	0.125
		β_2	0.494	0.103	1.171	0.306	0.491	0.116
		ρ	0.599	0.033	0.847	0.076	0.504	0.044
	0.75	β_1	0.499	0.094	−1.016	0.279	0.504	0.114
		β_2	0.503	0.087	1.790	0.232	0.499	0.104
		ρ	0.749	0.029	1.000	0.004	0.696	0.042
MA(1)	0.60	β_1	0.477	0.138	0.483	0.178	0.360	0.130
		β_2	0.388	0.133	0.506	0.177	0.601	0.129
		ρ	0.386	0.031	0.598	0.062	0.249	0.039
	0.75	β_1	0.481	0.127	0.491	0.175	0.368	0.122
		β_2	0.367	0.125	0.499	0.175	0.611	0.121
		ρ	0.452	0.028	0.749	0.064	0.291	0.042
EQC	0.60	β_1	0.498	0.126	0.215	0.278	0.498	0.110
		β_2	0.496	0.111	0.875	0.253	0.498	0.097
		ρ	0.521	0.042	0.717	0.080	0.597	0.044
	0.75	β_1	0.497	0.115	0.777	0.446	0.498	0.090
		β_2	0.500	0.097	1.618	0.350	0.500	0.080
		ρ	0.655	0.038	0.966	0.054	0.749	0.041

Remark that because the AR(1) and EQC models produce the same mean and the variance functions, the estimates under model misspecification do not vary too much but the standard errors tend to be larger under the misspecified models [Sutradhar and Das (1999)]. For example, when the data are generated following the AR(1) model, the AR(1) model based estimates for β_1, β_2, and ρ, have the standard errors 0.094, 0.087, 0.029, whereas the EQC model based corresponding standard errors are 0.115, 0.097, 0.038, confirming inefficient estimation under the 'working' correlation models.

In summary, when the longitudinal data follow a nonstationary correlation model, the effect of selecting a 'working' model with different mean and variance functions can be very serious. Thus, it is important to identify the true model to fit the data.

6.6.4.2 Model Selection

Note that it is practical to attempt to fit a possible low-order correlation model to given longitudinal data. But it may not be easy to identify the actual correlation structure for the data, especially when the data may follow one of the three nonstationary correlation models discussed in the paper. We thus recommend fitting all three models initially to the given data and compute the G_M statistic defined in

(6.79) under all three fitted models. One may then choose the model which produces the smallest value of the statistic G_M. The simulation results reported in Table 6.12 appear to support this technique of model selection.

Table 6.12 The simulated error sum of squares (ESS) under both true and 'working' nonstationary AR(1), MA(1), and EQC (equi-correlations) models for longitudinal count data, with true $\beta_1 = \beta_2 = 0.5$, for $K = 300$ individuals, and a selected value of ρ, based on 1000 simulations.

	True nonstationary Correlation Model			
		AR(1)	MA(1)	EQC
Selected ρ	Working Model	ESS	ESS	ESS
0.60	AR(1)	0.967	1.378	1.180
	MA(1)	1.281	1.138	1.158
	EQC	1.053	1.347	1.012
0.75	AR(1)	0.788	1.450	1.046
	MA(1)	1.249	1.120	1.145
	EQC	0.919	1.425	0.856

For example, when the data were generated following the nonstationary AR(1) model (6.44) with $\rho = 0.75$, the simulated average values of the G_M statistic computed by using the fitted values based on AR(1) (6.80), MA(1) (6.81), and EQC (6.82) models are found to be 0.788, 1.450, and 1.046, respectively. It is then clear that when the data follow the AR(1) model and the AR(1) model is fitted, the G_M statistic has the smallest value. Similar results hold under the other two models too.

6.7 A Data Example: Analyzing Health Care Utilization Count Data

We now consider an illustration for the application of the nonstationary correlation models for repeated count data discussed in Section 6.6, by analyzing the health care utilization data, earlier studied by Sutradhar (2003), for example. This dataset, provided in Appendix 6A, is a part of the longitudinal dataset collected by the General Hospital of the city of St. John's, Canada. To be specific, here we consider the longitudinal count data that contain the complete records for 144 individuals for four years $(n = 4)$ from $1985 - 1988$. The number of visits to a physician by each individual during a given year was recorded as the response, and this was repeated for four years. Also, the information on four covariates, namely, gender, number of chronic conditions, education level, and age, were recorded for each individual. Note that as the responses are counts, it is appropriate to assume that the response variable, marginally, follows the Poisson distribution, and the repeated counts recorded for four years will be longitudinally correlated. Along the lines of

Liang and Zeger (1986) we assume that the data may follow any of the low-order correlations such as AR(1), MA(1), or EQC models discussed in Section 6.6. Note that because these models produce different mean and the variance structures, they must be fitted by using these varied mean, variance, and correlation structures for the purpose of obtaining consistent and efficient estimates for the regression effects β and the correlation index parameter ρ.

Following the notations used in Sections 6.5 and 6.6, the four covariates for the ith ($i = 1,\ldots,K = 144$) individual at time t ($t = 1,\ldots,4$) are denoted by $x_{it1}, x_{it2}, x_{it3},$ and x_{it4} respectively. The first covariate geneder was coded as 0 for female and 1 for male. Thus, at any time t, $x_{it1} = 0$ if the ith individual is female, otherwise $x_{it1} = 1$. Similarly, the number of chronic diseases was coded as $x_{it2} = 0$ for the absence of chronic disease for the ith individual at time t, and $x_{it2} = 1$ if the ith individual had 1 or more chronic diseases at time t. The third covariate, education level, x_{it3}, was coded as 1 for less than high school, and 0 for high school or higher education. The last covariate, x_{it4}, represents the age of the individual. The effects of these covariates are denoted by $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$, so that the mean of the count response for the ith individual at a time point t is given by (6.65) under the nonstationary AR(1) and EQC structures, and by (6.67) under the nonstationary MA(1) model. In all these mean functions $x_{it} = (x_{it1}, x_{it2}, x_{it3}, x_{it4})'$.

We now apply the GQL estimation methodology discussed in Section 6.6. By using the response data y_{it} and x_{it} vector for all $i = 1,\ldots,144$, individuals and over $t = 1,\ldots,4$, years, we obtain the estimate of β and ρ from Section 6.5.1.1 under the nonstationary AR(1), from Section 6.6.1.1 under the MA(1), and similarly from Section 6.5.1.3 under the EQC models. These results along with the standard errors of the estimates of β computed by using the asymptotic covariance matrices from these three sections, are reported in Table 6.13.

Table 6.13 Comparison of the estimates of the regression and the correlation parameters under the nonstationary AR(1), MA(1), and EQC (equicorrelations) models in fitting the health care utilization data.

Nonstationary Correlation Models

	AR(1)		MA(1)		EQC	
Parameters	EST	SE	EST	SE	EST	SE
Gender effect (β_1)	−0.223	0.060	−0.179	0.054	−0.204	0.065
Chronic effect (β_2)	0.374	0.072	0.363	0.065	0.341	0.078
Education effect (β_3)	−0.428	0.074	−0.400	0.066	−0.390	0.081
Age effect(β_4)	0.029	0.001	0.031	0.001	0.029	0.001
ρ	0.554	–	0.769	–	0.529	–
$\rho_y(1)$	0.546	–	0.486	–	0.521	–
G_M	14.20	–	20.46	–	15.34	

As far as the selection of a model from these three lower-order models is concerned, we have computed the fitted residual squared distance G_M by (6.79) under all three models and reported them in the same Table 6.13. As the G_M statistic has the lowest value 14.20 under the AR(1) structure, we chose the AR(1) model to interpret the estimates.

As the first covariate gender was coded as 1 for male and 0 for female, it follows from (6.65) and (6.67) that the negative value of $\hat{\beta}_1 = -0.223$ suggests that the females made more visits to the physician as compared to the males. The positive values of $\hat{\beta}_2 = 0.374$ and $\hat{\beta}_4 = 0.029$ suggest that individuals having one or more chronic diseases or individuals belonging to the older age group pay more visits to the physicians, as expected. The third covariate education level was coded as 1 for less than high school, 0 for higher education. The effect of the education level on the physician visits was found to be $\hat{\beta}_3 = -0.428$. This negative estimate shows that highly educated individuals pay more visits as compared to individuals with a low level of education. One of the reasons for this type of behavior of this covariate may be that the individuals with a high-level education (more than high school) are more concerned about their health condition as compared to the individuals with low-level education.

Note that the standard errors of the regression estimates under the AR(1) model were found to be

$$\text{s.e.}(\hat{\beta}_1) = 0.060, \quad \text{s.e.}(\hat{\beta}_2) = 0.072, \quad \text{s.e.}(\hat{\beta}_3) = 0.074, \quad s.e.(\hat{\beta}_4) = 0.001.$$

As these standard errors are quite small as compared to the corresponding values of the regression estimates, all four covariates appear to have significant effects on the physician visits. Further note that the standard errors of the estimates under the MA(1) model appear to be smaller than the corresponding standard errors under the AR(1) model. Nevertheless, the estimates under the MA(1) model cannot be trusted as it is evident from the simulation study (see Table 6.11) that they can be highly biased when the data really follow the AR(1) model. Here the data as mentioned earlier appear to follow the AR(1) model with the smallest G_M value.

6.8 Models for Count Data from Longitudinal Adaptive Clinical Trials

In a clinical trial study with human subjects, it is highly desirable that one use certain data-dependent treatment allocation rules which exploit accumulating past information to assign individuals to treatments so that more study subjects are assigned to the better treatment. For example, consider a clinical trial study to examine the performance of a new treatment for asthma prevention. Suppose that one individual patient is assigned to one of the treatments in an adaptive way and number of asthma attacks for a week is recorded. Here the number of asthma attacks for a week may be considered to follow a Poisson distribution. Once the outcome of the first individual is known, the treatment for the second individual may be decided based on the outcome of the first individual as well as the covariate information of the individual.

Similarly, a treatment is assigned to the third individual based on the outcomes of the past two individuals and their covariate information. This adaptive procedure continues for a large number of weeks, say for 100 weeks for the treatment of 100 individuals. Note that 100 or more weeks is a reasonable duration for the completion of an intensive clinical trial study. Here, the purpose is to determine the effects of the treatments after treating a large proportion of subjects by the better treatment.

Note that there are many clinical studies including the aforementioned asthma study where it may be necessary to record the count responses repeatedly over a small period of time, from a patient based on the same assigned treatment, assignment of treatment being done in a longitudinal adaptive way. For example, for the asthma problem, it may be better to collect responses from a patient weekly for a period of $T = 4$ weeks, say, where the responses will be longitudinally correlated. As far as the treatment assignment is concerned, the assignment of the treatment to the third patient, for example, will be benefitted from the first week's response of the second patient, and the first and second weeks' responses from the first patient, and so on. The main purpose of this section is to discuss such longitudinal count data collected from a clinical trial study based on a suitable adaptive design. For the purpose, following Sutradhar and Jowaheer (2006), we first provide two longitudinal adaptive designs in Section 6.8.1. In Section 6.8.2, we demonstrate through a simulation study that the longitudinal adaptive designs discussed in Section 6.8.1 indeed allocate more patients to a better treatment. The overall treatment effects and the effects of other possible covariates are consistently and efficiently estimated in Section 6.8.3 by using a weighted GQL (WGQL) approach, based on the complete data collected from all patients during the study. We remark here that the WGQL approach indicates that the longitudinal adaptive design weights responsible for the collection of the longitudinal count data are incorporated in the so-called GQL approach discussed in the previous sections.

6.8.1 Adaptive Longitudinal Designs

Autocorrelated Poisson Model Conditional on Design Weights: Suppose that K independent patients will be treated in the clinical study and T longitudinal count responses will be collected from each of them. Also, for simplicity, let there be two treatments A and B to treat these patients and A is the better treatment between the two. Next suppose that δ_i refers to the selection of the treatment for the ith $(i = 1, \ldots, K)$ patient, and

$$\delta_i = \begin{cases} 1, & \text{if } i\text{th patient is assigned to } A \\ 0, & \text{if } i\text{th patient is assigned to } B \end{cases}$$

with

$$\Pr(\delta_i = 1) = w_i \;\text{ and }\; \Pr(\delta_i = 0) = 1 - w_i. \tag{6.83}$$

Here w_i refers to the better treatment selection probability for the ith patient. Now to construct a longitudinal adaptive design one needs to derive the formulas for the selection probabilities $w_i (i = 1, \ldots, K)$ so that in the long run more patients are treated by A.

Note that the value of δ_i determines the treatment by which the ith patient will be treated. Now suppose that conditional on δ_i, y_{it} denotes the count response recorded from the ith patient at time $t (t = 1, \ldots, T)$, and x_{it} denotes the p-dimensional covariate vector corresponding to y_{it}, defined as

$$\begin{aligned} x_{it} &= (\delta_i, x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})' \\ &= (\delta_i, x_{it}^{*'})', \end{aligned} \tag{6.84}$$

where $x_{it}^* = (x_{it2}, \ldots, x_{itu}, \ldots, x_{itp})'$ denote the $p - 1 \times 1$ vector of covariates such as prognostic factors (e.g., age, chronic conditions, and smoking habit) for the ith patient available at time point t. Thus, for $i = 2, \ldots, K$, the distribution of δ_i, that is, the formula of w_i, will depend on $\{\delta_1, \ldots, \delta_{i-1}\}$ and available responses y_{kv} $(k = 1, \ldots, i-1; 1 \leq v \leq T)$ along with their corresponding covariate vector x_{kv}. For $i = 1$, w_1 is assumed to be known.

As far as the availability of the repeated responses is concerned, we assume that for all $i = 1, \ldots, K$, once δ_i becomes known, the repeated count responses from the ith patient will be available following a Poisson distribution with conditional mean and variance (conditional on δ_i) given by

$$E(Y_{it} | \delta_i, x_{it}^*) = \text{var}(Y_{it} | \delta_i, x_{it}^*) = \exp(\theta_{it}), \tag{6.85}$$

where $\theta_{it} = x_{it}' \beta$, with $x_{it} = (\delta_i, x_{it}^{*'})'$. Also we assume that the pairwise longitudinal correlations between two repeated count responses are given by

$$\begin{aligned} \text{corr}[(Y_{it}, Y_{iv}) | \delta_i, x_{it}^*, x_{iv}^*] &= \rho_{|t-v|}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho) \\ &= c_{i,tv}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho), \end{aligned} \tag{6.86}$$

where $c_{i,tv}^{(ns)}(\delta_i, x_{it}^*, x_{iv}^*, \rho)$ has the formulas given by (6.46), (6.50), and (6.55) under the nonstationary AR(1), MA(1), and EQC models, respectively. It then follows by (6.85) and (6.86) that the conditional (on δ_i) covariance between y_{it} and y_{iv} is given by

$$\text{cov}[(Y_{it}, Y_{iv}) | \delta_i, x_{it}^*, x_{iv}^*] = \rho_{|t-v|}^{(ns)} \{\exp(\theta_{it} + \theta_{iv})\}^{\frac{1}{2}}.$$

Note, however, that for simplicity we use the stationary correlations based covariance matrix given by

$$\begin{aligned} \text{cov}[(Y_{it}, Y_{iv}) | \delta_i, x_{it}^*, x_{iv}^*] &\simeq c_{i,tv}^*(\rho) \{\exp(\theta_{it} + \theta_{iv})\}^{1/2} \\ &= \rho_{|t-v|} \{\exp(\theta_{it} + \theta_{iv})\}^{\frac{1}{2}}. \end{aligned} \tag{6.87}$$

6.8.1.1 Simple Longitudinal Play-the-Winner (SLPW) Rule to Formulate w_i

Note that in the cross-sectional setup, i.e., when $T = 1$ there exist a number of options to formulate the adaptive design weights w_i for $i = 1,\ldots,K$. For example, we refer to the

(i) randomized play the winner (RPW) rule [Zelen (1969); Wei and Durham (1978); Wei et al. (1990)],
(ii) random walks rule [Durham and Flournoy (1994)],
(iii) group sequential test [Jennison and Turnbull (2001)], and
(iv) optimum biased coin designs [Pocock and Simon(1975); Smith (1984); Atkinson (1999)].

The purpose of these designs is to assign a better treatment to an incoming patient based on the past outcomes of the experiment as well as the covariate information. Note that even if there are controversies [Royall 1991; Farewell, Viveros, and Sprott (1993)] about the usefulness of the play the winner rule, this seems to be the only design which was applied by some investigators [see, e.g., Tamura et al (1994); Rosenberger (1996)]. In this section, following Sutradhar, and Jowaheer (2006) [see also Sutradhar, Biswas, and Bari (2005)] we discuss a SLPW design to deal with longitudinal count data.

Note that as w_i is the probability of selection of the better treatment for the ith patient, it is convenient to compute w_i by considering two types of balls in an urn, the first type being the indicator for the selection of the better treatment A and the second type for the other treatment. The two types of balls are added to the urn as follows.

(a) As in the beginning we have no reason to believe that any particular treatment is better than the other, we take the initial urn composition in a 50:50 fashion. Thus, the urn will have two types of balls, say α balls of each type at the outset, and the probability that the first patient will be treated by treatment A is 0.5; that is, $\Pr(\delta_1 = 1) = w_1 = 0.5$. For simplicity one may use $\alpha = 1$.

(b) Suppose that at the selection stage of the ith patient $\{y_{rt}\}$ denote all available responses for $r = 1,\ldots,i-1$ and $1 \le t \le \min(T, i-r)$. The range of t here depends on the value of r. For example, for the selection time of the ith $(i = 2,\ldots,K)$ patient, $t = 1$ when $r = i-1$. Similarly $t = 1,2$ for $r = i-2$. Also suppose that at this selection stage we take all these available responses into account and for a suitable τ value and for specific available response y_{rt}, we add τ balls of the same kind by which the patient was treated if $y_{rt} \le m_0^*$, and add τ balls of the opposite kind in the urn if $y_{rt} > m_0^*$. Here m_0^* is a threshold value of the responses so that any patient with response less than this may be thought to belong to the success group. By the same token, if the response exceeds this threshold value, the patient may be thought to belong to the failure group. Thus, at this stage, we add τ balls for each and every available response. In general τ can be small such as $\tau = 2$, or 4.

(c) On top of the past responses, it may also be sensible to take into account the condition of certain covariates which, along with the treatment (A or B) were responsible for yielding those past responses y_{rt}. For a suitable quantity u_{rt} defined such that a large value of u_{rt} implies the prognostic factor based on a less serious condition of the rth $(r = 1,\ldots,i-1)$ past patient, $G - u_{rt}$ balls of the same kind by which the rth patient was treated and u_{rt} balls of the opposite kind are added, at the treatment selection stage for the ith patient, where $[0,\ G]$ is the domain of u_{rt}.

The above scheme described through **(a)** to **(c)**, produces the selection probabilities $w_i(i = 2,\ldots,K)$ for the cases $2 \leq i \leq T$ as in Exercise 6.4, and for $i > T$ as in Exercise 6.5.

6.8.1.2 Bivariate Random Walk (BRW) Design

Note that in the cross-sectional setup, apart from the randomized play-the-winner rule, there exist some alternative adaptive designs such as the random walk rule [see, e.g., Temple (1981), and Storer (1989)] to collect and analyze the clinical trial data. These random walk rules are variants of the familiar up-and-down rules [Anderson, McCarthy, and Tukey (1946), Derman (1957)]. For example, in the two treatment case, if the $(i-1)$th $(i = 2,\ldots,K,)$ patient is assigned to treatment A, then the ith patient will be assigned to treatment A with probability p_i, and to treatment B with probability q_i, such that $p_i + q_i = 1$. The parameters p_i and q_i depend on the previous patient's response and some random event, such as the result of a biased coin flip.

Remark that in the SLPW design in the previous section, the design weight w_i was mainly dependent on the responses of the individuals $1,2,\ldots,i-1$, as well as on the conditions of their covariates. Consequently, the construction of any random walk type of rules must be based on past responses as well as covariates. As in the previous section, suppose that a greater value of u_{rt} implies a better condition of the rth past patient and it was a more favorable condition of the patient to treat. By the same token, a smaller value of u_{rt} means that the patient was serious. Now to make sure that this better or serious covariate condition of the past patient does not influence the selection of the treatment for the present ith patient, and also to make sure that the past better response (say, a low value of the response such as $y_{rt} \leq y_0$) gets more weight for the assignment of the patient to the better treatment, one may use a bivariate probability structure given by

$$Pr(u_{rt} \leq u_0, y_{rt} \leq y_0) = p_{rt},\ Pr(u_{rt} \leq u_0, y_{rt} > y_0) = q_{rt},$$

$$Pr(u_{rt} > u_0, y_{rt} \leq y_0) = q_{rt},\ Pr(u_{rt} > u_0, y_{rt} > y_0) = h_{rt},$$

so that $p_{rt} + 2q_{rt} + h_{rt} = 1.0$. Here the parameters are chosen such that $p_{rt} > q_{rt} > h_{rt}$. Note that the bivariate probability structure arises from the consideration of using the past responses and the covariate condition of the patients.

The design weights w_i under this BRW rule are given in Exercise 6.6 for the case $2 \leq i \leq T$, and in Exercise 6.7 for the case $i > T$.

6.8.2 Performance of the SLPW and BRW Designs For Treatment Selection: A Simulation Study

In the last two sections, we have discussed how to construct the longitudinal adaptive design weights represented by w_i for the selection of a better treatment for the ith patient, for all $i = 2, \ldots, K$. We now conduct an empirical study to examine the performance of w_i under both SLPW and BRW designs.

To evaluate w_i under the SLPW design, we use the following steps.

Step 1. Parameter Selection: Clinical Design Parameters

$$\alpha = 1.0, \ ; \ G = 3.0, \ \text{ and } \ \tau = 2 \text{ and } 4.$$

Longitudinal Response Model Parameters

$$K = 100 \text{ subjects}, \ \ p = 3 \text{ covariates}, \ \ \beta_1 = 0.5, 1.00; \ \ \beta_2 = 0.5; \ \ \beta_3 = 0.25,$$

along with Poisson AR(1) responses for $T = 4$ time points with correlation index parameter $\rho = 0.9$. Also, use threshold count $m_0^* = 8$.

Note that the $p = 3$ covariates are denoted by $x_{it} = (\delta_i, x_{it2}, x_{it3})'$. Here δ_i is the treatment selection for the ith patient. Suppose that x_{it2} and x_{it3} are both non-stochastic covariates. Let $x_{it2} = 0, 1, \ldots, 5$ denote the number of chronic diseases for the ith patient at the entry time to the clinical experiment, and $x_{it3} = 1, 2, \ldots, 6$ be the age group of the ith patient. These two covariates are virtually time independent. We generate these covariates as

$$x_{it2} \sim \text{Binomial}(5, p = 0.9)$$

$$z_{it3} \sim \text{Uniform}(20, 80),$$

for all $i = 1, \ldots, K$, and $t = 1, \ldots, T$, and then assign

$$x_{it3} = \begin{cases} 1 & \text{for } 20 \le z_{it3} < 30 \\ 2 & \text{for } 30 \le z_{it3} < 40 \\ 3 & \text{for } 40 \le z_{it3} < 50 \\ 4 & \text{for } 50 \le z_{it3} < 60 \\ 5 & \text{for } 60 \le z_{it3} < 70 \\ 6 & \text{for } 70 \le z_{it3} \le 80. \end{cases}$$

Step 2. Generate Correlated Responses for First Individual: First using $w_1 = \frac{1}{2}$, generate δ_1 such that $Pr[\delta_1 = 1] = w_1$. Now for $i = 1$, that is, for the first patient, use

$$x_{11} = [\delta_1, x_{111}, x_{112}]'$$

and generate y_{11} following

$$y_{11} \sim \mathrm{Poi}(\mu_{11} = \exp(x'_{11}\beta).$$

Next use the stationary Poisson AR(1) model (6.14), that is,

$$y_{1t} = \rho * y_{1,t-1} + d_{11},$$

to generate the remaining three responses, namely $y_{12}, y_{13},$ and y_{14}.

Step 3. Generation of the nonstochastic *u*-Variable: Next to generate w_2, one depends on the y_{11} just generated and also on a u-variable which is a function of the second and third covariates. We now define the nonstochastic u-variable, u_{it}, given by

$$u_{it} = \frac{2}{x_{it2} + 1} + \frac{1}{x_{it3}}$$

which ranges from 0.5 to 3. This aids the consideration of $G = 3$ under the SLPW design.

Step 4. Generation of w_i and δ_i for $i = 2, \ldots, K$: Use the formula for w_i from Exercise 6.4 and 6.5. The desired y_{it} values are generated following the model (6.14); that is,

$$y_{1t} = \rho * y_{1,t-1} + d_{1t}. \tag{6.88}$$

Step 5. Generate δ_i. Once w_i is computed, obtain δ_i such that $Pr[\delta_i = 1] = w_i$, and compute $\delta^* = \sum_{i=1}^{K} \delta_i$ in each simulation.

In a manner similar to that of the SLPW design, we now evaluate w_i under the BRW design. To compute w_i in the BRW design, one requires an upper limit for the u-variable, say $u_0 = 1$ and an upper limit for y_{rt}, say $y_0 = 8$ for all past rth individuals at time point $t = 1, \ldots, 4$. By using $\beta_1 = 1.0, \beta_2 = 0.25$, and $\beta_3 = 0$ we generate w_2 and other values of $w_i, i = 3, \ldots, 100$ by using the formulas from Exercise 6.6

and 6.7. For the BRW design we also use $p_{rt} = 0.75, q_{rt} = 0.10$, and $h_{rt} = 0.05$ as the bivariate probabilities depending on the past responses and the values of the u-variable.

Next, in each of 1000 simulations we generate binary values δ_i with corresponding probability w_i, where the w_i are generated as above except that $w_1 = 0.5$. In each simulation we then calculate $\delta^* = \sum_{i=1}^{100} \delta_i$. For different parameter values under two designs, the mean and standard error of δ^* are shown in Table 6.14.

Table 6.14 Simulated mean values and simulated standard errors of the total number of patients $\delta^* = \sum_{i=1}^{100} \delta_i$ receiving the better treatment (A) among $K = 100$ subjects under both SLPW and BRW designs, based on 1000 simulations.

Design	ρ		β_1	δ^* Mean	SE
SLPW	0.9	$\tau = 2$	0.50	62	4.90
			1.00	58	4.73
		$\tau = 4$	0.50	68	4.92
			1.0	61	4.79
BRW	0.9		0.50	61	4.74
			1.00	56	5.01

It is clear from Table 6.14 that the design weights w_i under both SLPW and BRW designs appear to perform well for the selected parameter values. In all cases, the design weights appear to help assign more patients to the better treatment. More specifically, for $\tau = 2$ and $\beta_1 = 0.50$, the SLPW design assigns on the average 62 patients out of 100 to the better treatment A. Similarly for $\beta_1 = 0.50$, the BRW design assigns 61 patients on the average to the better treatment A. Note that all these values of total number of patients receiving treatment A are significant as the standard errors of $\delta^* = \sum_{i=1}^{100} \delta_i$ are reasonably small in all cases. Remark that β_1 in both designs represent the treatment effect. In both SLPW and BRW designs, smaller values of the response variable y indicate that the treatment is better. For example, a fewer number of asthma attacks for an individual implies that the individual received the better treatment. This justification also follows, for example, from the formulas for w_i in Exercises 6.4 and 6.6. This is because as the threshold point m_0^* in the SLPW design and the cut point (y_0, u_0) in the BRW design are predetermined and fixed, the smaller values of the response variable y will produce many of $I(y_{rt}) \leq m_0^*$ as 1 in the formula for w_i in Exercise 6.4, and $\delta_{y_{rt}} p_{rt}$ in the formula for w_i in Exercise 6.6 will contribute significantly. Thus, the better treatment should produce smaller values of y in the present setup. This in turn means that the smaller values of β_1 should indicate the better treatment. Consequently, the formulation of the design weights for both SLPW and BRW designs appear to work well as more patients are seen to be assigned to treatment A when $\beta_1 = 0.5$ as compared to $\beta_1 = 1.0$.

6.8.3 Weighted GQL Estimation for Treatment Effects and Other Regression Parameters

In previous sections, the repeated count responses for the ith individual were represented by a vector $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT}]'$ with its mean vector μ_i, and covariance matrix $\Sigma_i^*(\rho) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$ (6.26) under the stationary correlation models or $\Sigma_i^{(ns)}(\rho) = A_i^{1/2} C_i^{(ns)}(x_i, \rho) A_i^{1/2}$ (6.56) under the nonstationary correlation models. It was, however, demonstrated in Sections 6.8.1 and 6.8.2 under the longitudinal adaptive clinical trial setup, that a treatment is selected first for the ith individual based on adaptive design weight w_i, and then the responses are collected. To reflect this operation, we now denote the response vector as

$$y_i(w_i) = [y_{i1}(w_i), \ldots, y_{it}(w_i), \ldots, y_{iT}(w_i)]'$$

and its mean vector and stationary correlations based covariance matrix, for example, by

$$\mu_i(w_{i0}), \text{ and } \Sigma_i^*(w_{i0}, \rho),$$

respectively, where w_{i0} is the limiting value of w_i, for example, $w_{i0} = E[w_i]$.

6.8.3.1 Formulas for $\mu_i(w_{i0})$, and $\Sigma_i^*(w_{i0}, \rho)$:

Construction of the Mean Vector $\mu_i(w_{i0})$ Let

$$z'_{it} = x'_{it}|_{\delta_i=1} = (1, x^{*'}_{it}), \text{ and } z^{*'}_{it} = x'_{it}|_{\delta_i=0} = (0, x^{*'}_{it}),$$

where $x^*_{it} = (x_{it2}, \ldots, x_{itp})'$. Also, define

$$\mu^*_{rt1} = \exp(z'_{rt}\beta), \text{ and } \mu^*_{rt2} = \exp(z^{*'}_{rt}\beta). \tag{6.89}$$

Now by taking the average over the distribution of δ_i, it follows from (6.85) that the unconditional mean of Y_{it}, that is, $\mu_{it}(w_{i0})$ has the formula given by

$$\begin{aligned} E(Y_{it}|x^*_{it}) &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\delta_2,\ldots,\delta_{i-1}} E(Y_{it}|\delta_i, \ldots, \delta_1) \\ &= w_{i0} \exp(z'_{it}\beta) + (1 - w_{i0}) \exp(z^{*'}_{it}\beta) \\ &= w_{i0}\mu^*_{it1} + (1 - w_{i0})\mu^*_{it2} \\ &= \mu_{it}(w_{i0}), \end{aligned} \tag{6.90}$$

where for $i = 1, \ldots, K$, w_{i0} is the expectation of w_i, with $w_i = Pr(\delta_i = 1|y_{H_{i-1}})$ as defined in Exercises 6.4 and 6.5 for the SLPW design, and in Exercises 6.6 and 6.7, for the BRW design. More specifically, for the SLPW design, w_{i0} can be computed for the case $2 \le i \le T$ as

$$\begin{aligned} w_{i0} &= E_{\delta_1} E_{\delta_2|\delta_1} \cdots E_{\delta_i|\delta_1,\delta_2,\ldots,\delta_{i-1}} E(\delta_i | y_{H_{i-1}}) \\ &= \frac{1}{2\alpha + \frac{1}{2} i(i-1)(G+\tau)} \\ &\quad \times \left[\alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} [\{(G - u_{rt}) + \tilde{\mu}_{rt1}\tau\} w_r \right. \\ &\quad \left. + \{u_{rt} + (1 - \tilde{\mu}_{rt2})\tau\}(1 - w_r)] \right], \end{aligned} \tag{6.91}$$

and for the case $i > T$ as

$$\begin{aligned} w_{i0} &= E_{\delta_1} E_{\delta_2|\delta_1} \ldots E_{\delta_i|\delta_1,\delta_2,\cdots,\delta_{i-1}} E(\delta_i | y_{H_{i-1}}) \\ &= \left\{ 2\alpha + (G+\tau)T \left(i - \frac{T+1}{2} \right) \right\}^{-1} \\ &\quad \times \left[\alpha + \sum_{r=1}^{i-T} \sum_{t=1}^{T} \{(G - u_{rt} + \tilde{\mu}_{rt1}\tau) w_r + (u_{rt} + (1 - \tilde{\mu}_{rt2})\tau)(1 - w_r)\} \right. \\ &\quad + \sum_{r=i-T+1}^{i-1} \sum_{t=1}^{i-r} \{((G - u_{rt}) + \tilde{\mu}_{rt1}\tau) w_r \\ &\quad \left. + (u_{rt} + (1 - \tilde{\mu}_{rt2})\tau)(1 - w_r)\} \right], \end{aligned} \tag{6.92}$$

with

$$\tilde{\mu}_{rt1} = \int_0^{m_0^*} f(y_{rt} | \theta_{rt} = z'_{rt}\beta) = \sum_{k=0}^{m_0^*} \frac{\exp(-\mu^*_{rt1})(\mu^*_{rt1})^k}{k!}$$

and

$$\tilde{\mu}_{rt2} = \int_0^{m_0^*} f(y_{rt} | \theta_{rt} = z^{*\prime}_{rt}\beta) = \sum_{k=0}^{m_0^*} \frac{\exp(-\mu^*_{rt2})(\mu^*_{rt2})^k}{k!},$$

where m_0^* is the threshold count as mentioned before.

Note that the computation of the unconditional mean vector $\mu_i(w_{i0})$ for the BRW design is similar to that of SLPW design, and hence omitted.

Construction of the Covariance Matrix $\Sigma_i^*(w_{i0}, \rho)$

Next, we construct the unconditional covariance matrix $\Sigma_i^*(\rho)$ of the Y_i vector as follows. Recall that given $\delta_1, \delta_2, \ldots, \delta_i$, or simply say, given δ_i, the conditional variance of Y_{it} and the conditional covariance between Y_{it} and Y_{iv} are given in (6.85) and (6.87), respectively. Now by similar arguments as for the construction of the mean vector, the unconditional covariance between Y_{it} and Y_{iv} may be computed as

$$
\begin{aligned}
\mathrm{cov}[(Y_{it},Y_{iv})|x^*_{it},x^*_{iv}] &= E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}[\mathrm{cov}(Y_{it},Y_{iv})|\delta_i] \\
&\quad + \mathrm{cov}_{\delta_1,\ldots,\delta_i}[E(y_{it}|\delta_i),E(y_{iv}|\delta_i)], \\
&= E_{\delta_1}E_{\delta_2|\delta_1}\ldots E_{\delta_i|\delta_1,\ldots,\delta_{i-1}}[\rho_{|t-v|}\{\exp[(\theta_{it}+\theta_{iv})'\beta]\}^{1/2}] \\
&\quad +\mathrm{cov}_{\delta_1,\ldots,\delta_i}\{\exp[(\theta_{it}+\theta_{iv})'\beta]\} \\
&= \rho_{|t-v|}\left[w_{i0}\{\mu^*_{it1}\mu^*_{iv1}\}^{1/2}+(1-w_{i0})\{\mu^*_{it2}\mu^*_{iv2}\}^{1/2}\right] \\
&\quad +w_{i0}\{\mu^*_{it1}\mu^*_{iv1}\}+(1-w_{i0})\{\mu^*_{it2}\mu^*_{iv2}\}-\mu_{it}(w_{i0}\mu_{iv}(w_{i0}) \\
&= \sigma^*_{ijk}(w_{i0},\rho),\ \text{say}, \qquad (6.93)
\end{aligned}
$$

where μ^*_{it1} and μ^*_{it2} are given as in (6.89), and $\mu_{it}(w_{i0})$ is given as in (6.90). For $t=v$, equation (6.93) yields the unconditional variance of y_{it} given by

$$
\mathrm{var}(Y_{it}|x^*_{it}) = \mu^*_{it}+\{w_{i0}\mu^{*2}_{it1}+(1-w_{i0})\mu^{*2}_{it2}\}-\mu^{*2}_{it}. \qquad (6.94)
$$

The construction of the covariance matrix $\Sigma^*_i(w_{i0},\rho)=(\sigma^*_{ijk}(w_{i0},\rho))$, say, is now completed by (6.93) and (6.94).

6.8.3.2 Weighted GQL Estimation of β

Note that $\beta=[\beta_1,\beta_2,\ldots,\beta_p]'$ is the effect of the covariate

$$
x_{it}=[\delta_i,x^{*'}_i t]'=[\delta_i,x_{it2},\ldots,x_{itp}]'
$$

on y_{it} for all $i=1,\ldots,K$, and $t=1,\ldots,T$, where y_{it} is now collected based on longitudinal adaptive design scheme and is represented by $y_{it}(w_i)$. Because $E[Y_i(w_i)]=\mu_i(w_{i0})$ by (6.90), and $\mathrm{var}[Y_i(w_i)]=\Sigma^*_i(w_{i0},\rho)$ by (6.93) and (6.94), similar to the construction of the GQL estimating equation (6.26) or (6.56), we may now construct a weighted GQL estimating equation for β given by

$$
\sum_{i=1}^{K}\frac{\partial\mu'_i(w_{i0})}{\partial\beta}\Sigma^{*-1}_i(w_{i0},\hat{\rho})(y_i(w_i)-\mu_i(w_{i0})=0. \qquad (6.95)
$$

where $\hat{\rho}$ is a consistent estimate of ρ, the longitudinal correlation index parameter of the model. Now, by treating the data as though they follow the stationary correlation structure, one may apply the MM and equate the sample auto-covariance to the autocovariance of the data given by (6.93) and obtain a moment estimate of ρ_ℓ ($\ell=|t-v|=1,\ldots,T-1$) as

$$
\hat{\rho}_\ell=\frac{N_1-N_2}{D}, \qquad (6.96)
$$

where

$$N_1 = \frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[(y_{it}-\mu_{it}(w_{i0}))(y_{iv}-\mu_{iv}(w_{i0}))/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[y_{it}-\mu_{it}(w_{i0})]^2/KT}$$

$$N_2 = -\frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}[w_{i0}\mu^*_{it1}\mu^*_{iv1}+(1-w_{i0})\mu^*_{it2}\mu^*_{iv2}-\mu_{it}(w_{i0})\mu_{iv}(w_{i0})]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[\mu_{it}(w_{i0})-\mu^2_{it}(w_{i0})+w_{i0}\mu^{*2}_{it1}+(1-w_{i0})\mu^{*2}_{it2}]/KT},$$

and

$$D = \frac{\sum_{i=1}^{K}\sum_{|t-v|=\ell}\left[w_{i0}\{\mu^*_{it1}\mu^*_{iv1}\}^{1/2}+(1-w_{i0})\{\mu^*_{it2}\mu^*_{iv2}\}^{1/2}\right]/K(T-\ell)}{\sum_{i=1}^{K}\sum_{t=1}^{T}[\mu_{it}(w_{i0})-\mu^2_{it}(w_{i0})+w_{i0}\mu^{*2}_{it1}+(1-w_{i0})\mu^{*2}_{it2}]/KT}.$$

For given $\hat{\rho}_\ell$ (a function of $\hat{\rho}$), the solution of (6.95) may easily be obtained by using the Newton−Rapson iterative equation.

$$\hat{\beta}_{(m+1)} = \hat{\beta}_{(m)} + \left[\sum_{i=1}^{K}\frac{\partial\mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})\frac{\partial\mu_i(w_{i0})}{\partial\beta'}\right]^{-1}_m \times\left[\sum_{i=1}^{K}\frac{\partial\mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})(y_i(w_i)-\mu_i(w_{i0}))\right]_m, \qquad (6.97)$$

where $\hat{\beta}_{(m)}$ is the value of β at the mth iteration and $[\cdot]_m$ denotes that the expression within brackets is evaluated at $\hat{\beta}_{(m)}$. Let $\hat{\beta}_{WGQL}$ be the solution of (6.97), which is consistent for β.

Under some mild regularity conditions, it may be shown from (6.97) that for large K, $\hat{\beta}_{WGQL}$ has an asymptotically p-dimensional normal distribution with mean β and covariance matrix $\mathrm{var}(\hat{\beta}_{WGQL})$ which may be consistently estimated by using the sandwich type estimator given by

$$\begin{aligned}\hat{\mathrm{var}}(\hat{\beta}_{WGQL}) = &\left[\sum_{i=1}^{K}\frac{\partial\mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})\frac{\partial\mu_i(w_{i0})}{\partial\beta'}\right]^{-1}\\ &+\left[\sum_{i=1}^{K}\frac{\partial\mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})\frac{\partial\mu_i(w_{i0})}{\partial\beta'}\right]^{-1}\\ &\times\left[2\sum_{i<r}^{K}\frac{\partial\mu_i'(w_{i0})}{\partial\beta}\Sigma_i^{*-1}(w_{i0},\hat{\rho})(y_i-\mu_i(w_{i0}))\right.\\ &\left.\times(y_r-\mu_r(w_{i0}))'\Sigma_r^{*-1}(w_{r0},\hat{\rho})\frac{\partial\mu_r(w_{r0})}{\partial\beta'}\right]\end{aligned}$$

$$\times \left[\sum_{i=1}^{K} \frac{\partial \mu_i'(w_{i0})}{\partial \beta} \Sigma_i^{*-1}(w_{i0}, \hat{\rho}) \frac{\partial \mu_i(w_{i0})}{\partial \beta'} \right]^{-1}. \tag{6.98}$$

Formula for the Derivative $(\partial \mu_i'(w_{i0}))/\partial \beta$ in (6.95)

As

$$\frac{\partial \mu_{it}(w_{i0})}{\partial \beta} = w_{i0}\mu_{it1}^* z_{it} + (1 - w_{i0})\mu_{it2}^* z_{it}^*,$$

the $p \times T$ matrix $\partial \mu_i'(w_{i0})/\partial \beta$ is computed as

$$\frac{\partial \mu_i'(w_{i0})}{\partial \beta} = w_{i0} Z_i' A_{i1} + (1 - w_{i0}) Z_i^{*'} A_{i2}, \tag{6.99}$$

where $Z_i' = (z_{i1}, \ldots, z_{it}, \ldots, z_{iT})$ and $Z_i^{*'} = (z_{i1}^*, \ldots, z_{it}^*, \ldots, z_{iT}^*)$ are $p \times T$ matrices, $A_{i1} = \text{diag}[\mu_{i11}^*, \ldots, \mu_{iT1}^*]$, and $A_{i2} = \text{diag}[\mu_{i12}^*, \ldots, \mu_{iT2}^*]$, with

$$\mu_{it1}^* = \exp(z_{it}'\beta), \;\; \mu_{it2}^* = \exp(z_{it}^{*'}\beta),$$

where $z_{it} = (1, x_{it}^{*'})'$ and $z_{it}^* = (0, x_{it}^{*'})'$, for all $t = 1, \ldots, T$.

Exercises

6.1. (Section 6.5.1.1) [Likelihood estimation for nonstationary AR(1) model]
Consider the nonstationary AR(1) model given by (6.44). Then demonstrate that similar to that (6.23) of the stationary AR(1) model (6.14), one may write the likelihood function for the model (6.44) as

$$L(\beta, \rho) = \Pi_{i=1}^{K}[f(y_{i1})\Pi_{t=2}^{T} f(y_{it}|y_{i,t-1})],$$

with

$$f(y_{it}|y_{i,t-1}) = \exp[-(\mu_{it} - \rho\mu_{i,t-1})]$$
$$\times \sum_{s=1}^{\min(y_{it}, y_{i,t-1})} \frac{(y_{i,t-1})! \rho^s (1-\rho)^{y_{i,t-1}-s} (\mu_{it} - \rho\mu_{i,t-1})^{y_{it}-s}}{s!(y_{i,t-1} - s)!(y_{it} - s)!}.$$

Now, argue that the likelihood estimation of β and ρ, is extremely complicated.

6.2. (Section 6.5.1.1) [Conditional moments for nonstationary AR(1) model]
Show either by using the conditional density from Exercise 6.1, or by direct computation from the model (6.44), that for $t = 2, \ldots, T$, the conditional mean and variance of y_{it} given $y_{i,t-1}$ have the formulas:

$$E[Y_{it}|y_{i,t-1}] = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1})$$
$$\text{var}[Y_{it}|y_{i,t-1}] = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1}) - \rho^2 y_{i,t-1}.$$

Next, verify that for $u < t$, the conditional covariance has the formula

$$\text{cov}[\{Y_{iu}, Y_{it}\}|y_{i,u-1}, y_{i,t-1}] = 0.$$

6.3. (Section 6.5.2) [Conditional GQL estimating equation]
Denote the conditional mean and the variance in Exercise (6.2) by $\mu^*_{it|t-1}$ and $\lambda_{itt|t-1}$, respectively. Let $\mu^* = [\mu_{i1}, \mu^*_{i2|1}, \ldots, \mu^*_{it|t-1}, \ldots, \mu^*_{iT|T-1}]'$ be the $T \times 1$ conditional mean vector, and $\Lambda_i = \text{diag}[\mu_{i1}, \lambda_{i22|1}, \ldots, \lambda_{itt|t-1}, \ldots, \lambda_{iTT|T-1}]$ is the $T \times T$ conditional covariance matrix of y_i. Then, similar to (6.56), argue that a consistent estimator of β can also be obtained by solving the conditional GQL estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \mu^{*\prime}}{\partial \beta} \Lambda_i^{-1}(\hat{\rho})(y_i - \mu^*) = 0,$$

where $\hat{\rho}$ is obtained by using (6.58) as in the unconditional estimation. Also, derive the formulas for the elements of the $p \times T$ derivative matrix $\partial \mu^{*\prime}/\partial \beta$. Comment on the relative efficiency of this conditional GQL estimator of β as compared to the unconditional GQL estimator obtained from (6.56).

6.4. (Section 6.8.1.1) [w_i for the case $2 \le i \le T$ under **SLPW** rule]
As the selection of the ith patient is made at the ith time point, by this time, the $(i-1)$th patient has yielded one response and $(i-2)$th patient has yielded two responses and so on. Use the rules **(a), (b),** and **(c)** from the Section 6.8.1.1 and argue that at this treatment selection stage for the ith patient, there are

$$n^*_{i-1} = 2\alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} (G + \tau) = 2\alpha + \frac{1}{2} i(i-1)(G + \tau)$$

balls in total in the urn. Also justify that among these balls, there are

$$n^*_{i-1,1}(y_{H_{i-1}}) = \alpha + \sum_{r=1}^{i-1} \sum_{t=1}^{i-r} [\delta_r\{(G - u_{rt}) + I[y_{rt} \le m^*_0]\tau\}$$
$$+ (1 - \delta_r)\{u_{rj} + I[y_{rt} > m^*_0]\tau\}]$$

balls of first type, where $y_{H_{i-1}}$ indicates the history of responses from the past $i-1$ patients. The number of second type of balls may be denoted by $n^*_{i-1,2}(y_{H_{i-1}})$. It then follows that for given $y_{H_{i-1}}$, the conditional probability that $\delta_i = 1$ is given by

$$w_i = Pr(\delta_i = 1|y_{H_{i-1}}) = n^*_{i-1,1}(y_{H_{i-1}})/n^*_{i-1}.$$

6.5. (Section 6.8.1.1) [w_i for the case $i > T$ under **SLPW** rule]
Argue that under this case, at the treatment selection stage for the ith patient, there

are

$$\tilde{n}_{i-1} = 2\alpha + \sum_{r=1}^{i-T}\sum_{t=1}^{T}(G+\tau) + \sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}(G+\tau)$$

balls in total in the urn. Also argue that among these balls, there are $\tilde{n}_{i-1,1}(y_{H_{i-1}})$ balls of first type, where

$$\begin{aligned}\tilde{n}_{i-1,1}(y_{H_{i-1}}) = \alpha &+ \sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_r\{(G-u_{rt})+I[y_{rt} \le m_0^*]\tau\} \\ &+(1-\delta_r)\{u_{rt}+I[y_{rt} > m_0^*]\tau\}] \\ &+ \sum_{r=1-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_r\{(G-u_{rt})+I[y_{rt} \le m_0^*]\tau\} \\ &+(1-\delta_r)\{u_{rt}+I[y_{rt} > m_0^*]\tau\}].\end{aligned}$$

Clearly, for this $i > T$ case, one may then evaluate the design weight w_i as

$$w_i = \frac{\tilde{n}_{i-1,1}(y_{H_{i-1}})}{\tilde{n}_{i-1}}.$$

6.6. (Section 6.8.1.2) [w_i for the case $2 \le i \le T$ under **BRW** rule]
Let $\delta_{u_{rt}} = 1$ for $u_{rt} \le u_0$ and $\delta_{u_{rt}} = 0$ otherwise. Similarly, let $\delta_{y_{rt}} = 1$ for $y_{rt} \le y_0$ and $\delta_{y_{rt}} = 0$ otherwise. Verify, in the fashion similar to that of Exercise 6.4 that under the BRW rule, the design weight w_i has the formula

$$w_i = \frac{\sum_{r=1}^{i-1}\sum_{t=1}^{i-r}[\delta_{u_{rt}}g(y_{rt})]+[(1-\delta_{u_{rt}})s(y_{rt})]}{\sum_{r=1}^{i-1}\sum_{t=1}^{i-r}(p_{rt}+2q_{rt}+h_{rt})},$$

where $g(y_{rt}) = \delta_{y_{rt}}p_{rt} + (1-\delta_{y_{rt}})q_{rt}$, and $s(y_{rt}) = \delta_{y_{rt}}q_{rt} + (1-\delta_{y_{rt}})h_{rt}$.

6.7. (Section 6.8.1.2) [w_i for the case $i > T$ under **BRW** rule]
For this case, make an argument similar to that of Exercise 6.5 for the SLPW design, and justify under the BRW rule, that w_i has the formula given by

$$\begin{aligned}w_i = &\frac{1}{0.5i(i-1)-0.5(i-T)(i-T-1)} \\ &\times[\sum_{r=1}^{i-T}\sum_{t=1}^{T}[\delta_{u_{rt}}g(y_{rt})]+[(1-\delta_{u_{rt}})s(y_{rt})] \\ &+\sum_{r=i-T+1}^{i-1}\sum_{t=1}^{i-r}[\delta_{u_{rt}}g(y_{rt})]+[(1-\delta_{u_{rt}})s(y_{rt})]],\end{aligned}$$

where $g(y_{rt})$ and $s(y_{rt})$ are defined as in Exercise 6.6.

References

1. Amemiya, T. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.
2. Anderson, T. W., McCarthy, P. J., & Tukey, J. W. (1946). Staircase Method of Sensitivity Testing. *Naval Ordinance Report*, 35 – 46. Princeton, NJ: Statistical Research Group.
3. Atkinson, A. C. (1999). Optimum biased-coin designs for sequential treatment allocation with covariate information. *Statist. Med.,* 18, 1741 – 1752.
4. Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika,* 82, 407 – 410.
5. Derman, C. (1957). Nonparametric up and down experimentation. *Ann. Math. Stat.*, 28, 795 – 798.
6. Durham, S. D. & Flournoy, N. (1994). Random walks for quantile estimation. In *Statistical Decision Theory and Related Topics V* (S. S. Gupta and J. O. Berger, eds.) 467 – 476. New York: Springer.
7. Farewell, V. T. , Viveros, R. , & Sprott, D. A. (1993). Statistical consequences of an adaptive treatment allocation in a clinical trial. *Canad. J. Statist.,* 21, 21 – 27.
8. Jennison, C. & Turnbull, B. W. (2001). Group sequential tests with outcome-dependent treatment assignment. *Sequential Anal.,* 20, 209 – 234.
9. Liang, K. Y. & Zeger, S . L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 78, 13 – 22.
10. Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.
11. McCullagh, P. (1983). Quasilikelihood functions. *Ann. Statist.* 11, 59 – 67.
12. McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advan. Appl. Probab.,* 20, 822 – 835.
13. Nelder, J. & Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. A135,* 370 – 384.
14. Pocock, S. J. & Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics,* 31, 103 – 115.
15. Rosenberger, W. F. (1996). New directions in adaptive designs. *Statist. Sci.,* 11, 137 – 149.
16. Royall, R. M. (1991). Ethics and statistics in randomized clinical trials (with discussion). *Statist. Sci.,* 6, 52 – 62.
17. Smith, R. L. S. (1984). Properties of bias coin designs in sequential clinical trials.*Annal. Statist.,* 12, 1018 – 1034.
18. Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statist. Sci.,* 18, 377 – 393.
19. Sutradhar, B.C., Biswas, A., & Bari, W. (2005). Marginal regression for binary longitudinal data in adaptive clinical trials. *Scand. J. Statist.,* 32, 93 – 113.
20. Sutradhar, B. C. & Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika,* 86, 459 – 465.
21. Sutradhar, B. C. & Jowaheer, V. (2006). Analyzing longitudinal count data from adaptive clinical trials: a weighted generalized quasi-likelihood approach. *J. Statist. Comput. Simul.,* 76, 1079 – 1093.
22. Sutradhar, B. C. & Kovacevic, M. (2000). Analyzing ordinal longitudinal survey data: Generalized estimating equations approach. *Biometrika,* 87, 837 – 848.
23. Storer, B, E, (1989). Design and analysis of phase 1 clinical trial. *Biometrics,*, 45, 925 – 937.
24. Tamura, R. N. , Faries, D. E. , Andersen, J. S., & Heiligenstein, J. H. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *J. Amer. Stat. Assoc.,* 89, 768 – 776.
25. Temple, R. (1981). Government view points of clinical trials. *Drug Inf. J.,* 16, 10 – 17.
26. Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika,* 61, 439 – 447.
27. Wei, L. J. & Durham, S. (1978). The randomized play-the-winner rule in medical trials. *J. Am. Statist. Assoc.,* 73, 840 – 843.

28. Wei, L. J., Smythe, R. T. Lin, D. Y., & Park, T. S. (1990). Statistical inference with data-dependent treatment allocation rules. *J. Am. Statist. Assoc.,* 85, 156 – 62.
29. Zelen, M. (1969). Play-the-winner rule and the controlled clinical trial. *J. Am. Statist. Assoc.,* 64, 131 – 46.

Appendix

Table 6A. Health care utilization data for six years from 1985 to 1990 collected by Health Science Center, Memorial University, St. John's, Canada. [Code: column 1 (C1)-Family identification; C2-Member identification; C3-Gender (1 for male, 2 for female); C4-Chronic disease status (0 for no chronic disease, 1 for 1 chronic disease and so on); C5-Education level (1 for less than high school, 2 for high school, 3 for university graduate, and 4 for post graduate); C6-Age at 1985; C7-C12-Number of physician visits from 1985 to 1990]

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
7	101	2	3	3	51.7	10	15	10	6	7	14
7	102	1	2	2	55.4	0	6	0	4	2	0
7	203	2	1	4	24.9	12	6	2	0	3	2
7	204	2	0	4	21.5	0	1	1	0	0	0
27	101	2	1	3	49.5	2	11	8	7	7	3
27	102	1	1	4	50.7	13	13	16	12	18	12
27	203	1	0	4	20.2	1	5	0	2	0	0
27	203	1	0	4	20.2	2	3	7	1	0	0
36	101	2	2	3	49.7	5	5	4	18	11	9
36	102	1	1	3	54.6	1	0	0	2	1	1
36	203	2	0	3	26.0	10	6	9	9	21	16
36	204	1	0	2	22.4	3	4	1	0	4	1
189	101	2	1	3	58.6	4	3	1	3	0	6
189	102	1	0	4	58.3	1	0	0	3	0	3
189	203	2	3	2	31.7	8	4	4	12	12	7
189	204	2	1	3	20.2	2	0	6	2	2	5
436	101	2	0	1	62.1	10	8	7	10	8	11
436	102	1	0	1	68.9	6	5	2	6	4	6
436	203	1	0	3	31.8	1	3	4	0	0	0
436	204	1	0	4	23.8	2	2	5	0	0	0
469	101	2	4	2	44.1	4	1	6	7	13	3
469	102	1	0	3	47.5	2	0	1	0	1	1
469	203	1	0	3	23.7	2	4	3	2	1	0
469	204	1	2	4	21.2	5	5	5	0	8	0
574	101	2	0	1	47.2	4	10	12	17	13	10
574	102	1	4	1	52.9	8	9	14	23	22	15
574	203	2	1	3	23.2	5	3	6	6	5	7
574	204	1	0	2	21.9	2	0	3	3	1	1
580	101	2	2	1	41.9	2	5	1	0	1	0
580	102	1	0	2	44.2	1	1	4	24	5	2
580	203	2	1	2	20.5	13	11	11	16	18	21
580	204	2	0	2	23	9	3	4	3	19	3
706	101	2	2	3	40.7	17	5	1	5	3	2
706	102	1	0	1	42.9	1	1	7	6	1	0
706	203	1	0	3	21.5	1	3	0	3	0	0
706	204	1	0	3	19.9	0	0	0	0	0	0

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
754	101	2	1	2	49.8	8	2	5	12	8	8
754	102	1	0	1	50.8	0	2	0	0	2	0
754	203	1	1	2	21.3	0	0	0	0	1	0
754	204	1	0	4	25.3	1	1	1	0	0	2
758	101	2	1	2	60.9	2	5	1	1	0	0
758	102	1	0	4	63.7	1	0	0	0	0	0
758	203	1	1	4	22.8	0	0	0	2	1	0
758	204	1	0	4	20.9	2	11	4	11	10	4
921	101	2	1	1	50.8	0	3	0	3	7	14
921	202	1	1	1	26.4	1	1	4	3	5	2
921	203	2	1	3	25.2	3	2	2	1	2	2
921	204	2	0	2	21.9	3	2	4	2	5	16
965	101	2	1	1	44.8	13	18	13	13	15	17
965	102	1	2	1	48.6	4	2	0	3	0	6
965	203	1	0	3	25	4	3	1	0	6	2
965	204	1	0	3	20.9	2	3	1	1	3	1
993	101	2	3	1	67.3	2	3	3	2	4	3
993	203	1	2	1	31.3	2	0	1	1	2	3
993	204	2	1	2	22	11	6	3	4	17	8
993	205	1	0	1	22.3	1	1	4	9	4	1
1054	101	2	0	2	41.1	1	11	3	5	2 4	9
1054	102	1	2	1	43.6	3	4	10	4	11	11
1054	203	2	1	4	22.2	4	2	3	4	14	11
1054	204	2	2	4	20.3	1	4	3	5	10	9
1120	101	2	3	1	52.7	2	9	2	1	9	7
1120	102	1	0	1	63.1	0	0	0	0	0	0
1120	203	2	0	4	32.2	12	7	27	11	5	13
1120	204	1	1	2	26	1	3	0	3	10	3
1269	101	2	0	4	56.1	1	3	1	9	10	14
1269	102	1	1	4	56.3	4	0	3	8	4	4
1269	203	1	0	4	22	2	0	2	0	2	1
1269	204	2	0	4	20.5	0	0	0	1	0	0
1333	101	2	1	1	50.9	2	2	1	0	0	0
1333	102	1	0	1	49.5	3	6	2	9	5	4
1333	203	2	0	3	22.6	0	0	0	0	0	0
1333	204	1	0	2	20.6	0	0	0	1	4	12
1344	101	2	2	1	46.4	0	0	0	2	2	3
1344	203	1	0	1	24	0	1	0	0	0	0
1344	204	1	0	1	28.8	0	0	0	0	0	0
1344	205	1	1	1	20.3	2	0	1	1	0	1
1361	101	2	0	1	71.6	4	7	9	8	3	8
1361	202	2	0	3	35.3	2	4	7	9	10	6

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1361	203	1	0	2	33	3	3	5	2	0	3
1361	204	1	0	3	27.4	1	1	2	2	2	3
1397	101	2	0	3	25.3	7	3	5	7	5	5
1397	102	1	1	1	53	2	4	5	6	6	3
1397	203	1	0	4	27.3	2	0	0	0	0	0
1397	204	1	0	3	22	12	1	2	2	4	4
1637	101	2	1	4	43.5	6	10	2	2	3	3
1637	102	1	1	4	47.4	0	3	4	1	0	0
1637	203	1	0	4	23.1	0	0	0	1	1	0
1637	204	1	1	4	21.7	1	2	2	4	5	2
1664	101	2	2	4	47.2	25	9	8	14	12	29
1664	102	1	2	2	49.2	4	3	9	0	10	4
1664	203	2	0	4	23.5	3	3	0	2	2	1
1664	204	1	1	4	22.3	1	1	0	0	0	0
1669	101	2	0	2	50.6	0	0	0	2	4	1
1669	202	2	0	3	24.7	7	5	5	12	7	6
1669	203	1	0	4	22.5	0	0	1	1	2	0
1669	204	1	0	2	20.9	0	0	1	0	0	3
1682	101	2	1	1	62.1	0	2	3	1	0	0
1682	102	1	4	1	65.2	7	0	0	0	0	0
1682	203	1	3	3	29	9	9	12	5	4	4
1682	404	2	4	1	74.9	13	17	16	15	14	10
1702	101	2	2	1	59.2	6	5	2	1	1	6
1702	102	1	2	1	64	0	0	0	0	0	2
1702	203	1	1	1	21.1	0	0	0	0	0	0
1702	304	2	3	1	85.2	6	7	8	6	24	0
1703	101	2	1	3	56.9	3	4	3	10	4	14
1703	202	1	0	4	25.5	0	0	0	0	0	0
1703	204	2	0	4	22.1	1	0	1	3	0	0
1703	305	2	1	2	80.5	5	7	4	8	4	8
1728	101	2	1	1	40.1	5	3	2	2	2	1
1728	102	1	4	3	51.5	12	13	10	7	22	19
1728	203	2	1	2	24.3	10	11	4	5	7	3
1728	204	1	0	3	20.4	3	2	3	2	2	2
1737	101	2	3	2	43.8	11	6	9	4	4	4
1737	102	1	1	4	44.1	6	0	8	1	0	8
1737	203	2	0	3	21.9	1	4	10	8	25	10
1737	204	1	0	4	22.9	0	0	0	0	0	0
1751	101	2	5	2	52	9	12	11	6	18	15
1751	102	1	0	1	55.5	0	0	2	0	1	0
1751	203	1	1	1	23.6	3	2	8	2	3	6
1751	204	1	0	1	22.6	1	8	3	2	1	3

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
1838	101	2	0	2	44.7	3	3	3	2	10	11
1838	102	1	1	1	46	3	1	2	2	0	3
1838	203	1	0	4	23.5	2	3	1	4	1	0
1838	404	2	1	1	76.4	0	0	7	5	8	4
1876	101	2	1	1	46.7	0	0	4	4	0	2
1876	102	1	1	3	51.1	2	10	10	16	10	6
1876	203	2	0	3	24.6	5	2	0	0	0	0
1876	205	2	4	4	21	2	1	1	2	3	5
1925	101	2	1	3	52.6	19	4	12	9	7	5
1925	102	1	0	2	60.2	4	15	13	5	1	7
1925	203	2	0	4	21.5	9	6	4	13	8	0
1925	204	1	0	4	23.2	0	0	1	0	0	0
1935	101	2	1	3	65.9	2	1	3	4	5	12
1935	102	1	1	1	67.6	9	6	7	8	7	7
1935	203	1	0	2	25.6	2	1	0	0	0	0
1935	204	2	0	3	38.4	4	2	4	9	17	18
2046	101	2	0	1	56.3	11	17	4	3	12	9
2046	202	1	0	1	33.4	0	0	0	0	0	0
2046	203	1	0	2	27.8	1	1	0	3	3	9
2046	204	2	0	3	25	0	3	4	5	5	8
2076	101	2	2	3	52	5	3	6	8	3	3
2076	102	1	1	1	53.8	2	0	3	7	6	2
2076	203	2	0	4	24.6	14	11	5	1	2	0
2076	204	1	3	3	31.4	2	1	4	3	4	14
41	102	1	0	1	54	0	0	0	0	0	0
41	203	2	0	4	22	2	2	2	9	7	0
41	204	1	0	4	23	3	2	2	4	7	0
101	101	2	1	1	62.8	2	0	0	0	1	0
101	102	1	5	1	65.9	2	2	5	10	7	2
101	203	1	1	3	24.2	0	0	0	0	0	0
129	101	2	3	1	56.3	10	14	7	9	9	13
129	102	1	1	1	57.1	9	15	8	10	13	2
129	204	1	0	4	21.6	1	1	4	1	0	0
208	102	1	0	4	50.5	0	0	0	7	11	12
208	203	1	0	4	25.3	0	1	1	1	4	1
208	204	1	0	3	23.8	1	1	1	1	0	1
219	101	2	4	1	62.5	11	17	8	18	23	17
219	203	2	1	1	40.4	9	4	2	6	4	2
219	204	2	1	4	21.3	5	2	1	4	0	0

Table Cont'd

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
522	102	1	0	1	51.2	1	5	7	7	9	6
522	203	2	1	3	21.6	11	7	3	8	20	19
522	204	2	1	2	24.4	12	7	19	6	12	7
605	101	2	1	1	58.2	2	6	0	2	0	2
605	102	1	0	1	58.6	0	0	0	0	0	1
605	203	1	1	2	21.3	0	0	0	0	1	2
622	203	1	0	1	25	0	0	0	0	0	0
622	204	2	0	1	30.5	0	0	0	0	0	0
622	205	2	0	1	22.4	3	5	0	10	23	18
731	101	2	1	3	50.2	4	5	3	8	13	11
731	204	1	0	4	24	0	0	3	3	0	0
731	205	1	1	4	21.9	3	2	5	1	5	0
1097	101	2	0	3	43	2	3	2	1	0	6
1097	102	1	1	4	49.1	3	0	3	2	2	2
1097	203	1	0	4	23.5	1	4	1	3	2	2
1689	101	2	0	1	44.9	3	7	5	16	7	8
1689	102	1	2	3	47.8	1	8	24	22	14	8
1689	204	1	3	2	21.6	6	8	3	2	6	4
1906	101	2	4	1	67.8	27	23	29	39	19	16
1906	202	2	0	2	47.5	2	0	4	5	9	8
1906	203	1	1	2	50.2	12	8	8	11	9	13

Springer

http://www.springer.com/978-1-4419-8341-1

Dynamic Mixed Models for Familial Longitudinal Data
Sutradhar, B.C.
2011, XVIII, 494 p., Hardcover
ISBN: 978-1-4419-8341-1