Tweedie Class Log-linear Models for Longitudinal Data with Random Effects and Internal Correlation

Peter Xue-Kun Song
Department of Mathematics and Statistics, York University,
Toronto, Ontario M3J 1P3, Canada (song@pascal.math.yorku.ca)
and
Wenxin Jiang
Department of Statistics, Northwestern University,
Evanston, Illinois 60208, U.S.A. (wjiang@nwu.edu) *

June 30, 1998

Summary

A class of log-linear models for longitudinal data with random effects and within-subject correlation (or internal correlation) is studied. The variance-mean relation at the conditional level is taken from the class of Tweedie models (Jørgensen, 1997), accommodating a variety of continuous and discrete longitudinal data. The parameters are estimated from the estimating equation approach where only the first two moments are specified. Asymptotic tests are proposed to check the significance of the internal correlation. A simulation study is performed to demonstrate the importance of modeling the internal correlation. Finally we apply our method to analyse the epilepsy data (Thall and Vail, 1990), and found that modeling the internal correlation, in addition to the inclusion of random effects, may be appealing to achieve a proper representation of the correlation structure.

1 Introduction

It is commonly assumed in regression models with random effects (e.g., Laird and Ware, 1982) that repeated responses over time are conditionally independent given the random effects.

* *Key words:* Estimating equations; Internal correlation; Log-linear models; Longitudinal data; Random effects; Tweedie class.

Such an assumption is imposed mainly for the sake of mathematical simplicity in modeling and inference. It is well-known that the random effects induce marginal intra-correlation for each of subjects over time. However, is the assumption of *conditional* independence true, so that the responses at one time and another, after *conditioning* on the random effects, are no longer correlated? This question motivates us to look for statistical models and testing procedures for the within-subject correlation given random effects. We shall refer to this type of correlation as the *internal correlation* (IC). Note that the possibility of modeling IC has been discussed in the context of the linear models with random effects by Laird and Ware (1982). Assume that each vector of responses, $\mathbf{y}_i$, is composed of a mean function $\boldsymbol{\mu}_i$ which incorporates both fixed and random effects, as well as an error term $\boldsymbol{\varepsilon}_i$ which has a covariance matrix Σ. The conditional independence assumption corresponds to a diagonal covariance matrix Σ and hence the components of $\mathbf{y}_i$ are uncorrelated conditional on the random effects; while to represent IC, off-diagonal entries of Σ needs to be modeled. One such model with AR(1) structure is thoroughly investigated by Chi and Reinsel (1989) for normal responses, and a data example did show that there exists internal correlation in addition to the one induced by the random effects. Also see Diggle and Verbyla (1998).

Apparently, the issue of whether or not there exists IC would arise in any non-linear regression models with random effects. Theoretically, negligence of IC leads to misspecification of covariance structure if IC in fact exists. This results in inconsistent parameter estimation due to biased estimating equations and hence inconsistent variance component estimators. It is noted that in nonlinear regression models, the regression coefficients in the first moment specification can also be inconsistently estimated, since the first moment of the responses unconditional on the random effects depends also on the variance parameters.

Note also that the misspecified covariance structure will lead to imprecision in prediction– this is illustrated in Section 4. Without acknowledging IC, a future prediction of the response of a subject tends to be made without fully utilizing the information from the history.

In spite of all these potential problems of neglecting IC, there seems to be inadequately little discussion on IC in the existing literature of non-linear regression models. This paper intends to focus on a commonly-used class of non-linear models, namely the log-linear models. Thall and Veil (1990) proposed a log-linear model for repeated counts in which two types of random effects were introduced: one is time-specific and the other is subject-specific. However it can be easily verified that their model still virtually produces zero internal correlation. Similar to Thall and Veil (1990), Breslow and Clayton (1993, Page 16) used a log-linear mixed model to analyse a longitudinal data where a random effect that depends on both the subject and the time was included. This model is slightly more general than Thall and Veil's because it involves a subject-time interactive random effect whereas Thall and Veil express the two types of effects separately. Nevertheless Breslow and Clayton assume the conditional independence given all random effects including the interactive one, which leads again to a situation with zero internal correlation.

We introduce a class of log-linear models for longitudinal data where, conditional on the random effects, the variance-mean relations of the cross-sectional responses follow the class of Tweedie models (Jørgensen 1997, Chapter 4), including Poisson, gamma and inverse

Gaussian as special cases. In general, it is of interest to develop a statistical theory that can be used to model IC and check the assumption of conditional independence. We use the method of estimating equations based on moments for parameter estimation. Such a method was broadly applied in the analysis of longitudinal data in which the joint probability models are difficult to specify. See for example Zeger et al. (1988), Prentice and Zhao (1991) and Thall and Veil (1990). Some alternative approaches such as the MCMC algorithm (Zeger and Karim, 1991) and the approximate inference (Breslow and Clayton, 1993) work only if the conditional independence is assumed. This is because in these methods the explicit forms of the joint densities are needed, which, however, are not available when the conditional independence is violated.

We develop an asymptotic Wald's test to test the significance of IC, which is shown numerically in a simulation study. Finally, we illustrate our method with the epileptic seizures data (Thall and Veil, 1990) which was previously analysed by, for example, Thall and Veil (1990), Breslow and Clayton (1993) and Diggle, Liang and Zeger (1994).

This paper is organized as follows. The model specification is presented in Section 2, including the first two moments needed for the construction of estimating equations. Section 3 develops an estimating equation approach to parameter estimation. Prediction and model diagnosis are discussed in Section 4 and Section 5. The final two Sections 6 and 7 contain a simulation study and an analysis of epilepsy data.

2 Model specification

For the i-th of n subjects involved in a longitudinal study, a univariate time series $\{y_{it}, t = 1, \ldots, T_i\}$ is observed together with vectors $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ of explanatory variables associated with the fixed and random effects, respectively. We assume that given a q-dimensional vector of random effects $\mathbf{b}_i$, y_{it} are correlated over time with mean $\mu_{it}^b = \mathrm{E}(y_{it}|\mathbf{b}_i)$, variance $\mathrm{var}(y_{it}|\mathbf{b}_i) = \phi v_k\left(\mu_{it}^b\right)$ and autocorrelation $\mathrm{corr}(y_{it}, y_{is}|\mathbf{b}_i) = \rho_i(t,s)$, where $v_k(\cdot)$ is a Tweedie variance function of the form $v_k(\xi) = \xi^k$ which determines the variance-mean relation. A log-linear mixed model is given by

$$\log\left(\mu_{it}^b\right) = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i$$

where $\boldsymbol{\beta} \in \mathcal{R}^p$. The random effects $\mathbf{b}_1, \ldots, \mathbf{b}_n$ are assumed to be mutually independent normal variates with mean $\mathbf{0}$ and variance-covariance matrix $R(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown r-vector of parameters.

There are various ways to model the internal autocorrelation function $\rho_i(t,s)$. A simple model based on the first order autoregressive assumption (AR(1)) states that $\rho_i(t,s) = \alpha^{|t-s|}$ for all subjects. This can be used for testing the conditional independence assumption $\alpha = 0$ and will be used in the simulation and the data analysis later. It is also possible to allow the internal correlation to be different among subjects, via $\rho_i(t,s) = \alpha_i^{|t-s|}$, where the parameter α_i may be further modeled as a function of a g-dimensional vector of covariates $\mathbf{u}_i$ of the form

$$\alpha_i = h\left(\mathbf{u}_i^\top \boldsymbol{\gamma}\right). \tag{1}$$

An example of $h(\cdot)$ which maps $\mathcal{R}$ to $(-1,1)$ is the hyperbolic tangent function $\mathrm{th}(x) = (e^x - e^{-x})/(e^x + e^{-x})$. If ρ_i is known positive, another popular choice for ρ_i would be the exponential correlation model or the Gaussian correlation model, being special cases of the stationary correlation $\rho_i = \rho_i(|t-s|)$. See Diggle et al. (1994, P.83–84).

It follows immediately from (11) in the Appendix that the marginal expectations, denoted by $m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta})$, equal to

$$m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \mathrm{E}(y_{it}) = \exp\left(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \frac{1}{2}\mathbf{z}_{it}^\top R(\boldsymbol{\theta})\mathbf{z}_{it}\right), \tag{2}$$

and the second moments, denoted by $q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi)$, are given as follows. Let

$$A_{its}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \exp\left\{\frac{k}{2}(\mathbf{x}_{it} + \mathbf{x}_{is})^\top \boldsymbol{\beta} + \frac{k^2}{8}\mathrm{tr}R(\boldsymbol{\theta})(\mathbf{z}_{it} + \mathbf{z}_{is})(\mathbf{z}_{it} + \mathbf{z}_{is})^\top\right\}, \tag{3}$$

and in particular for $k = 2$,

$$A_{its}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \exp\left\{(\mathbf{x}_{it} + \mathbf{x}_{is})^\top \boldsymbol{\beta} + \frac{1}{2}\mathrm{tr}R(\boldsymbol{\theta})(\mathbf{z}_{it} + \mathbf{z}_{is})(\mathbf{z}_{it} + \mathbf{z}_{is})^\top\right\}.$$

Then,

$$q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) = \mathrm{E}(y_{it}y_{is}) = \phi\rho_i(t,s)A_{its}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\theta}) + A_{its}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\theta}). \tag{4}$$

Refer to the Appendix for the derivation of (4) in detail. Note that both formulas of the first and second order moments above are valid for the general stationary models of IC.

Based on these first and second order moments, a system of unbiased estimating equations is obtained for parameter estimation in Section 3. We consequently develop a statistical method of testing the hypothesis $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ for the significance of internal correlation.

We would like to close this section by commenting on the Tweedie class of models. It is well known that the shape parameter $k \in (-\infty, 0] \cup [1, \infty)$ characterizes the types of distributions, including Gaussian ($k = 0$), Poisson ($k = 1$), gamma ($k = 2$) and inverse Gaussian ($k = 3$).

3 Parameter inference

In this section the discussions are restricted to model (1). Let $\Theta = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top, \phi)^\top$ be the grand vector of parameters to be estimated. In our notation, $\dim(\Theta) = p + g + r + 1$, where $p = \dim(\boldsymbol{\beta})$, $g = \dim(\boldsymbol{\gamma})$, $r = \dim(\boldsymbol{\theta})$, $1 = \dim(\phi)$. We now form a system of unbiased estimating equations and define our estimate of Θ as the solution to these equations. The explicit expressions of all the derivatives appearing in this section are presented in the Appendix. Set

$$\sum_{i=1}^{n}\sum_{t=1}^{T_i} C_{ivt}\left\{y_{it} - m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta})\right\} = 0,$$

for $v = 1, \ldots, p$, and set

$$\sum_{i=1}^{n} \sum_{t \le s} \tilde{B}_{iuts} \{y_{it} y_{is} - q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi)\} = 0,$$

for $u = 1, \ldots, g + r + 1$.

According to Crowder (1987), the optimal choice of C_{ivt} is $C_i = D_i^\top K_i^{-1}$, a $p \times T_i$ matrix with the (v, t)-th element C_{ivt}, $v = 1, \ldots, p; t = 1, \ldots, T_i$, where the (t, s) elements of K_i and D_i are given respectively by

$$[K_i]_{t,s} = q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) - m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta}) m_{is}(\boldsymbol{\beta}, \boldsymbol{\theta}), \quad [D_i]_{t,s} = \frac{\partial m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \beta_s}.$$

$\tilde{B}_{ivts}$ may be simply chosen to be

$$\tilde{B}_{ivts} = \frac{\partial}{\partial \eta_v} q_{its}(\boldsymbol{\beta}, \boldsymbol{\eta}), \ \ v = 1, \ldots, g + r + 1$$

where $\boldsymbol{\eta} = (\boldsymbol{\gamma}, \boldsymbol{\theta}, \phi)$.

We now express the above equations in matrix notation. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT_i})^\top$ and $\tilde{\mathbf{y}}_i = (y_{i1} y_{i1}, \ldots, y_{i1} y_{iT_i}, y_{i2} y_{i2}, \ldots, y_{iT_i} y_{iT_i})^\top$, containing all cross-product terms of $\mathbf{y}_i$ without repetition. Let $\mathbf{m}_i = \mathrm{E}(\mathbf{y}_i)$ and $\mathbf{q}_i = \mathrm{E}(\tilde{\mathbf{y}}_i)$. Then the system of unbiased estimating equations may be re-written in the form

$$\sum_{i=1}^{n} \begin{pmatrix} C_i & \mathbf{0} \\ \mathbf{0} & B_i \end{pmatrix} \left\{ \begin{pmatrix} \mathbf{y}_i \\ \tilde{\mathbf{y}}_i \end{pmatrix} - \begin{pmatrix} \mathbf{m}_i \\ \mathbf{q}_i \end{pmatrix} \right\} = \mathbf{0} \tag{5}$$

where for each $i = 1, \ldots, n$, B_i is a $(g + r + 1) \times \frac{1}{2} T_i (T_i + 1)$ matrix with the j-th row vector given by

$$[B_i]_{j\cdot} = (\tilde{B}_{ij11}, \ldots, \tilde{B}_{ij1T_i}, \tilde{B}_{ij22}, \ldots, \tilde{B}_{ij2T_i}, \ldots, \tilde{B}_{ijT_iT_i})$$

which is a $\frac{1}{2} T_i (T_i + 1)$-dimensional vector without repetition.

Equation (5) can be expressed in the form

$$\Psi(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \psi_i(\Theta) = \frac{1}{n} \sum_{i=1}^{n} W_i \{Y_i - Q_i(\Theta)\} = \mathbf{0} \tag{6}$$

where $W_i = \mathrm{blockdiag}(C_i, B_i)$, $Y_i = (\mathbf{y}_i^\top, \tilde{\mathbf{y}}_i^\top)^\top$ and $Q_i = (\mathbf{m}_i^\top, \mathbf{q}_i^\top)^\top$. We define the solution to equation (6) as the estimate of Θ, denoted by $\widehat{\Theta}$. This generates estimators for the subvectors, i.e., $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\phi}$, for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$ and ϕ, respectively.

It follows from the standard theory of estimating equation that under some mild conditions, the estimate $\widehat{\Theta}$ is consistent and asymptotically normal as the number of subject n tends to ∞, namely, if Θ is the true parameter,

$$\sqrt{n}(\widehat{\Theta} - \Theta) \xrightarrow{d} N(\mathbf{0}, \Sigma), \ \text{ as } n \to \infty.$$

with $\Sigma = \lim_n nJ^{-1}$ where J is the Godambe information matrix given by $J = S^\top V^{-1} S$. Here V is the variability matrix given by

$$V = \mathrm{E}\left(\Psi\Psi^\top\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{E}\left(\psi_i\psi_i^\top\right),$$

and S is the sensitivity matrix given by

$$S = \mathrm{E}\left(\nabla_\Theta \Psi\right) = -\frac{1}{n}\sum_{i=1}^{n} W_i \nabla_\Theta Q_i(\Theta).$$

Therefore the asymptotic variance for $\widehat{\Theta}$ is

$$\mathrm{Avar}(\widehat{\Theta}) = J^{-1} = S^{-1} V S^{-\top}$$

where V may be estimated by

$$\widehat{V} = \frac{1}{n^2}\sum_{i=1}^{n} W_i (Y_i - Q_i)(Y_i - Q_i)^\top W_i^\top$$

and the unknown parameter Θ is substituted by its estimate.

Now we consider the inference on IC. Let Σ_γ be the submatrix of Σ corresponding to the subvector of parameter $\boldsymbol{\gamma}$. It follows that

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} N(\mathbf{0}, \Sigma_\gamma), \text{ as } n \to \infty.$$

Let $\widehat{\alpha}_i = h\left(\mathbf{u}_i^\top \widehat{\boldsymbol{\gamma}}\right)$. By Slutsky's theorem we obtain

$$\sqrt{n}(\widehat{\alpha}_i - \alpha_i) \xrightarrow{d} N\left(0, G_i^\top \Sigma_\gamma G_i\right), \text{ as } n \to \infty,$$

where

$$G_i = \frac{\partial}{\partial \boldsymbol{\gamma}} h\left(\mathbf{u}_i^\top \boldsymbol{\gamma}\right) = h'\left(\mathbf{u}_i^\top \boldsymbol{\gamma}\right)\mathbf{u}_i.$$

Hence an asymptotic $100(1-\alpha)\%$ confidence interval for α_i is

$$\widehat{\alpha}_i \pm z(\alpha/2) s.e.(\widehat{\alpha}_i),$$

where the standard error $s.e.(\widehat{\alpha}_i)$ is approximately equal to $n^{-1/2}|h'\left(\mathbf{u}_i^\top \widehat{\boldsymbol{\gamma}}\right)|\sqrt{\mathbf{u}_i^\top \hat{\Sigma}_\gamma \mathbf{u}_i}$ where $\hat{\Sigma}_\gamma$ is a sample estimate of Σ_γ, and $z(\alpha/2)$ is the normal quantile with upper tail probability $\alpha/2$. For the choice of the hyperbolic tangent link function, the corresponding standard error is given by

$$n^{-1/2}\left\{\mathrm{csh}\left(\mathbf{u}_i^\top \widehat{\boldsymbol{\gamma}}\right)\right\}^{-1}\sqrt{\mathbf{u}_i^\top \hat{\Sigma}_\gamma \mathbf{u}_i}$$

where $\mathrm{csh}(x) = \left(e^x + e^{-x}\right)/2$.

The Wald statistic $\mathcal{W} = \widehat{\boldsymbol{\gamma}}\hat{\Sigma}_\gamma^{-1}\widehat{\boldsymbol{\gamma}}$ approximately follows a χ^2-distribution with g degrees of freedom, under the null hypothesis $\boldsymbol{\gamma} = \mathbf{0}$. If this hypothesis is rejected, then the internal correlation is statistically significant.

4 Prediction

Suppose there is a new subject, labeled as $i = 0$, scheduled to be observed at $t = 1, \ldots, T_0$. Let $\mathbf{y}_0 = (y_{01}, \ldots, y_{0T_0})^\top$. Decompose $\mathbf{y}_0 = (\mathbf{y}_{01}^\top, \mathbf{y}_{02}^\top)^\top$, where $\mathbf{y}_{01}$ is the observed history, and $\mathbf{y}_{02}$ is the future response to be predicted. The model of $\mathbf{y}_0$ follows that of the subjects $i = 1, \ldots, n$. Let $\mathbf{m}_0 = \mathrm{E}\mathbf{y}_0 = (m_{01}, \ldots, m_{0T_0})$, with the t-th element being

$$[\mathbf{m}_0]_t = \mathrm{E}y_{0t} = m_{0t}(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

and let $\Sigma_0 = \mathrm{var}(\mathbf{y}_0)$ with the (t, s)-th element being

$$[\Sigma_0]_{ts} = \mathrm{cov}(y_{0t}, y_{0s}) = q_{0ts}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) - m_{0t}(\boldsymbol{\beta}, \boldsymbol{\theta}) m_{0s}(\boldsymbol{\beta}, \boldsymbol{\theta})$$

for $s, t = 1, \ldots, T_0$. The parameters in $\mathbf{m}_0$ and Σ_0 can be estimated by the procedure in the previous section, based on the data from the subjects 1 to n, with part or all of their observation histories. Decompose $\mathbf{m}_0$ and Σ_0 as, respectively,

$$\begin{aligned} \mathbf{m}_0 &= (\mathbf{m}_{01}^\top, \mathbf{m}_{02}^\top)^\top \text{ and} \\ \Sigma_0 &= \begin{pmatrix} \Sigma_{011} & \Sigma_{012} \\ \Sigma_{021} & \Sigma_{022} \end{pmatrix} \end{aligned}$$

Then the BLP (best linear predictor) $\hat{\mathbf{y}}_{02}$ of $\mathbf{y}_{02}$ based on $\mathbf{y}_{01}$ is

$$\hat{\mathbf{y}}_{02} = \mathbf{m}_{02} + \Sigma_{021}\Sigma_{011}^{-1}(\mathbf{y}_{01} - \mathbf{m}_{01}) \tag{7}$$

with mean square prediction error

$$\mathrm{MSE}(\hat{\mathbf{y}}_{02}) = \mathrm{E}\{(\hat{\mathbf{y}}_{02} - \mathbf{y}_{02})^\top(\hat{\mathbf{y}}_{02} - \mathbf{y}_{02})\} = \mathrm{tr}\left\{\Sigma_{022} - \Sigma_{021}\Sigma_{011}^{-1}\Sigma_{012}\right\}.$$

Note that the BLP is automatically unbiased ($\mathrm{E}\left(\hat{\mathbf{y}}_{02} - \mathbf{y}_{02}\right) = \mathbf{0}$).

We now consider the prediction of $\mathbf{y}_{02}$ in the situation where the internal correlation is mistakenly neglected. Let $\tilde{\mathbf{y}}_{02}$ be a "BLP" calculated by (7) under the misspecified covariance matrix. Clearly, $\mathrm{MSE}(\tilde{\mathbf{y}}_{02}) \geq \mathrm{MSE}(\hat{\mathbf{y}}_{02})$, since $\tilde{\mathbf{y}}_{02}$ is linear in $\mathbf{y}_{01}$, and $\hat{\mathbf{y}}_{02}$ is the true BLP. This implies that the negligence of IC could degrade the precision of prediction. An example of this for normal responses can be seen in Fig. 2 of Chi and Reinsel (1989).

5 Model diagnosis

We develop diagnostic tools to check the assumptions of the variance-mean relation, the log-linear link, and the autoregressive structure for IC.

The random effect $\mathbf{b}_i$ is estimated by BLUP, the best linear unbiased predictor, given the set of observations on the i-th subject. Since the expectation of $\mathbf{b}_i$ is zero, we immediately have

$$\widehat{\mathbf{b}}_i = \mathrm{cov}(\mathbf{b}_i, \mathbf{y}_i)\mathrm{var}^{-1}(\mathbf{y}_i)(\mathbf{y}_i - \mathbf{m}_i),$$

and the mean squared error is

$$\mathrm{MSE}(\widehat{\mathbf{b}}_i) = \mathrm{E}\left\{(\hat{\mathbf{b}}_i - \mathbf{b}_i)^\top (\hat{\mathbf{b}}_i - \mathbf{b}_i)\right\} = \mathrm{tr}\left\{\mathrm{var}(\mathbf{b}_i) - \mathrm{cov}(\mathbf{b}_i, \mathbf{y}_i)\mathrm{var}^{-1}(\mathbf{y}_i)\mathrm{cov}^\top(\mathbf{b}_i, \mathbf{y}_i)\right\}.$$

It is easy to show that

$$\mathrm{cov}(\mathbf{b}_i, \mathbf{y}_i) = R(\boldsymbol{\theta}) Z_i \mathrm{diag}(m_{i1}, \ldots, m_{iT_i})$$

where $Z_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT_i})$, a $q \times T_i$ matrix. Consequently the estimate of the conditional mean μ_{it}^b is given by

$$\widehat{\mu}_{it}^{\hat{b}} = \exp\left(\mathbf{x}_{it}^\top \widehat{\boldsymbol{\beta}} + \mathbf{z}_{it}^\top \widehat{\mathbf{b}}_i\right).$$

Now consider a conditional error defined by

$$\varepsilon_{it} = y_{it} - \mu_{it}^b, \ t = 1, \ldots, T_i, \ i = 1, \ldots, n.$$

Clearly $\mathrm{E}(\varepsilon_{it} | \mathbf{b}_i) = 0$ and

$$\mathrm{var}(\varepsilon_{it} | \mathbf{b}_i) = \phi\left(\mu_{it}^b\right)^k = \phi \exp\left\{k(\mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i)\right\}.$$

The sample counterpart of ε_{it} is $e_{it} = y_{it} - \widehat{\mu}_{it}^{\hat{b}}$, and the standardized version is then

$$\overline{e}_{it} = e_{it} / \sqrt{\widehat{\mathrm{var}}(\varepsilon_{it} | \hat{\mathbf{b}}_i)} = e_{it} / \sqrt{\hat{\phi}\left(\hat{\mu}_{it}^{\hat{b}}\right)^k}.$$

Checking the assumption of the variance-mean relation may be done by the plot of residuals $\overline{e}_{it}$ versus the log fitted values $\log \widehat{\mu}_{it}^{\hat{b}}$. See for example Jørgensen et al. (1996). In the ideal situation, all points should randomly scatter around the horizontal line at zero. Any departure from this, for example a megaphone shape, suggests a violation of the assumption on the variance-mean relation.

An informal check for the log-link function can be developed by generalizing McCullagh and Nelder's (1989) plot of the adjusted dependent variable z against the linear predictor $\hat{\eta}$. For our set-up, define

$$z_{it}^{b_i} = \log\left(\mu_{it}^{b_i}\right) + \frac{y_{it} - \mu_{it}^{b_i}}{\mu_{it}^{b_i}}$$

and $\eta_{it}^{b_i} = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{z}_{it}^\top \mathbf{b}_i$. Clearly $\mathrm{E}\left(z_{it}^{b_i} | \mathbf{b}_i\right) = \log\left(\mu_{it}^{b_i}\right) = \eta_{it}^{b_i}$. If the log-linearity holds, then the plot of $\hat{z}_{it}^{\hat{b}_i}$ against $\hat{\eta}_{it}^{\hat{b}_i}$ should show a straight line.

To check the autoregressive function assumption for the internal correlation, we first define

$$y_{it}^* = \frac{y_{it}}{\sqrt{\mathrm{var}(y_{it} | \mathbf{b}_i)}} \tag{8}$$

such that $\mathrm{var}(y_{it}^* | \mathbf{b}_i) = 1$. Assume that now IC follows AR(2), rather than AR(1) which we used in the model, then

$$\mathrm{corr}(y_{it}^*, y_{it-1}^* | \mathbf{b}_i) = \frac{\pi_1}{1 - \pi_2}, \quad \mathrm{corr}(y_{it}^*, y_{it-2}^* | \mathbf{b}_i) = \frac{\pi_1^2}{1 - \pi_2} + \pi_2 \tag{9}$$

where π_1 and π_2 are the first and second partial autocorrelation coefficients. Note that our previous AR(1) model assumes $(\pi_1, \pi_2) = (\alpha_i, 0)$ for subject i. Define

$$e_{it}^{(1)} = y_{it}^* - \pi_1 y_{it-1}^* \tag{10}$$

to be the AR(1) residual for subject i at time t. It follows from the Proposition in Appendix that if the AR(1) assumption is true, i.e. if $\pi_2 = 0$, then $\mathrm{corr}(e_{it}^{(1)}, e_{it-1}^{(1)} | \mathbf{b}_i) = 0$. This property can be used for diagnosis of AR(1) assumption, via a scatterplot of $\hat{e}_{it}^{(1)}$ against $\hat{e}_{it-1}^{(1)}$ for all i and t. This check can be done alternatively using the autocorrelation function (ACF) plot of the $\hat{e}_{it}^{(1)}$'s. Here

$$\hat{e}_{it}^{(1)} = \hat{y}_{it}^* - \hat{\pi}_1 \hat{y}_{it-1}^*,$$

where we define

$$\hat{y}_{it}^* = y_{it} / \sqrt{\hat{\phi} \left(\hat{\mu}_{it}^{b}\right)^k},$$

and $\hat{\pi}_1$ is in fact equal to $\hat{\alpha}_i$ in our model, which is the estimate of α_i defined in equation (1).

In general checking the normality assumption for random effects in the context of nonlinear regression models is difficult as the distribution of $\hat{\mathbf{b}}_i$ is determined by that of the response variables and can be very skewed, even if the $\mathbf{b}_i$'s are normal. The plot of $\hat{\mathbf{b}}_i$ against the quantiles of the standard normal usually shows a curvature pattern, which does not necessarily imply the violation of the normality assumption. A formal check might be developed by extending the method of Davidian and Gallant (1993) based on modeling the density of the random effects by a flexible family of smooth functions which are the products of normal densities with squared polynomials. This is beyond the scope of the present paper, and we do not pursue it further here.

In the following sections of simulation study and data analysis, we will be focusing on the Poisson-type variance-mean relation, corresponding to the member $k = 1$ of the Tweedie class. A natural choice of k is often suggested by the type of responses. For example, for count data, we could try the Poisson-type model ($k = 1$), while for non-negative continuous responses, we could try the gamma-type model ($k = 2$). A formal choice of k could be performed based on a technique similar to that discussed in McCullagh and Nelder (1989, P.400), based on repeating analyses for different k's and choosing among the k's that fit the data satisfactorily. A goodness-of-fit criteria $\log(H)$, discussed by Thall and Veil (1990), could be used in conjunction with our estimating equation approach. The drawback is that re-fitting the data for a number of k's can be computationally intensive.

6 Simulation

To evaluate the impact of IC in modeling, we conducted a simple simulation study based on a log-linear model given as follows:

$$y_{it} \sim \mathrm{Poisson}(\mu_{it}^{b_i}), \; b_i \sim N(0, \sigma^2), i = 1, \ldots, n, t = 1, \ldots, T$$

where

$$\log \mu_{it}^{b_i} = \beta_0 + \beta_1 t/T + b_i$$

and $\rho_i(t,s) = \text{corr}(y_{it}, y_{is}|b_i) = \alpha^{|t-s|}$, $\alpha \in (-1,1)$. Note that in this setting, $\phi = 1$, $k = 1$, $R(\boldsymbol{\theta}) = \sigma^2$, $\mathbf{u}_i = 1$, $\mathbf{z}_{it} = 1$, $\mathbf{x}_{it} = (x_{t,1}, x_{t,2})^\top = (1, t/T)^\top$ and $\alpha = \text{th}(\gamma)$.

The algorithm used to generate longitudinal counts with IC is the following.

STEP 1: Sampling $b_1, \ldots, b_n$ *i.i.d.* from $N(0, \sigma^2)$

STEP 2: Calculate $\mu_{it} = \exp(\beta_0 + \beta_1 t/T + b_i)$

STEP 3: For each $i = 1, \ldots, n$, generate a time series of counts with given AR(1) autocorrelation and Poisson margins with mean vector $(\mu_{i1}, \ldots, \mu_{iT})$. This step was done by invoking an algorithm proposed by Song (1997).

In the simulation, the true parameters are $\beta_0 = 0$, $\beta_1 = 0.5$, $\sigma^2 = 0.5$, $\alpha = 0.5$, and $n = 100$, $T = 10$. The focus of the study is on the comparison of the consequences of the negligence and inclusion of IC. Table 1 reports the numerical results of estimates and standard errors under assumptions of zero-IC (or genuine random effects model) and of the presence of IC.

Table 1

Estimates and standard errors for parameters in the simulation models.

	Presence of IC		IC=0		
parameter	estimate	stand. err.	estimate	stand. err.	True value
β_0	0.1469	0.1289	0.3710	0.0597	0.0000
β_1	1.2885	0.3774	1.2201	0.0916	0.5000
α	0.8760	0.1627	—	—	0.5000
σ^2	0.4093	0.6995	0.2800	0.2061	0.5000

Clearly the estimates of β_1 in the two models do not differ much and are both significant at significant level 0.05. Both seem to be relatively big in comparison to the true value. We suspect that this discrepancy is partly caused by the outliers (we found three outliers in the simulated data), and partly due to the finite sample bias.

Since the p-value of the hypothesis $H_0 : \alpha = 0$ is virtually zero, we reject the null hypothesis and conclude that the data have a significant internal correlation conditional on the random effects. Note that the analysis based on the 0-IC model presented a distorted picture of the truth: It not only assumes in the first place a zero internal correlation, but also obtains an insignificant estimate of the random effects variance. We would be led to conclude that the responses have no temporal correlation, whether caused by IC or by the random effects, which is against the truth.

For IC model, in contrast to the hypothesis $H_0 : \alpha = 0$ (in favour of the genuine random effect model), the hypothesis $H_0 : \sigma^2 = 0$ (in favour of the model with IC) has a p-value equal to 0.7208. This implies that for our simulated data, neglecting IC is much more unsatisfactory than neglecting the random effects in modeling. Similar findings are also discussed in Chi and Reinsel (1989).

It is noted that most point estimates of parameters in the model with IC are closer to the true parameters than those in the model with zero IC; however this is not the case for standard errors. This is probably due to the fact that the model with IC has one more parameter.

Finally, to illustrate the methods of model diagnosis in Section 5, we demonstrate two diagnostic plots to assess the assumptions of log-linearity and AR(1) correlation. The plot for checking the variance-mean relation was made but not shown here since it is very similar to the one presented in Jørgensen et al. (1996).

Figure 1 shows a scatterplot of the adjusted dependent variable $\hat{z}_{it}^{\hat{b}_i}$ against the predictor $\hat{\eta}_{it}^{\hat{b}_i}$, in which the dashed line stands for the least-squares fitted line and the solid line for the diagonal line. Apparently the two lines are rather close, confirming the log-linearity assumption.

The ACF plot of AR(1) residuals $\hat{e}_{it}^{(1)}$ is shown in Figure 2, and it clearly indicates that these residuals are uncorrelated, agreeing with the AR(1) assumption that was used to generate the data for the simulation study.

Figure 1 and Figure 2 are about here.

7 Analysis of the epilepsy data

Now we apply our model to analyse Thall and Vail's (1990) data of epileptic seizures. The data were originally reported by Leppik et al. (1985), and previously analysed by Thall and Vail (1990), Breslow and Clayton (1993) and Diggle et al. (1994) using log-linear models with random effects but assuming zero internal correlation. Our re-analysis of these data aims at assessing the assumption of conditional independence made in the previous analyses. We adopt Diggle, Liang and Zeger's formulation (1994, Page 187), that is,

$$\log \mathrm{E}(y_{ij}|\mathbf{u}_i) = \beta_0 + \beta_1 \mathrm{Trt}_{ij} + \beta_2 \mathrm{Vst}_{ij} + \beta_3 \mathrm{Trt} \times \mathrm{Vst}_{ij} + u_{i1} + u_{i2}\mathrm{Vst}_{ij} + \log(t_{ij})$$

where

$$\mathrm{Vst}_{ij} = \mathrm{Vst}_j = \begin{cases} 1 & \text{for } j = 1, 2, 3 \text{ or } 4 \\ 0 & \text{for } j = 0 \end{cases}$$

$$\mathrm{Trt}_{ij} = \mathrm{Trt}_i = \begin{cases} 1 & \text{if } i\text{-th subject assigned to progabide group} \\ 0 & \text{if } i\text{-th subject assigned to placebo group} \end{cases}$$

$$t_{ij} = t_j = \begin{cases} 2 & \text{for } j = 1, 2, 3 \text{ or } 4 \\ 8 & \text{for } j = 0 \end{cases}$$

and $\mathbf{u}_i = (u_{i1}, u_{i2})$ are $i.i.d$ normal random effects with mean $(0, 0)$ and variance matrix Σ with elements

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

In previous analyses, it is known that the data appear to be overdispersed, and hopefully the overdispersion would be automatically accounted for through the inclusion of random effects. With the flexibility of our method, we actually include the overdispersion parameter, $\phi = e^{\gamma}$ in addition to the IC parameter and the random effects. By doing so, we can conduct a formal test on $H_0 : \phi = 0$ to inspect if the random effects have fully explained the overdispersion. Here we assume the exponential correlation model for IC, which takes the form $\exp(\alpha|j-k|) = \rho^{|j-k|}$, $\rho = \exp(\alpha)$, $\alpha \in (-\infty, 0]$.

We start with a model including only the random intercept u_{i1}. Two analyses were performed, one on the full data, the other with patient number 207 taken out, which is believed to be an outlier. Numerical results are tabulated in Table 2.

Table 2

Estimates and standard errors for Epilepsy data analysis.

	0-IC Model	IC Model	
Variable	Complete Data	Complete Data	Without 207
Intercept	0.9499 (0.1553)	1.1641 (0.1737)	1.1292 (0.1788)
Treatment	0.0265 (0.2219)	0.0173 (0.2119)	−0.0798 (0.1980)
Visit	0.1106 (0.1161)	0.1534 (0.1128)	0.1461 (0.1092)
Trt.Vst	−0.1020 (0.2135)	−0.1292 (0.2638)	−0.3868 (0.1714)
σ_{11}	0.7953 (0.2507)	0.2926 (0.1443)	0.3743 (0.1494)
ρ		0.8530 (0.0539)	0.7050 (0.2169)
ϕ		16.1592 (7.5178)	5.1588 (3.2295)
$\log(H)$	20.9037	29.5597	26.0338

Our estimates for the genuine random effects model (0-IC model) are close to the one Diggle et al. (1994) obtained but our standard errors are relatively bigger than theirs probably due to our different approach.

When the IC-model is fitted with the full data, the IC parameter, the overdispersion parameter, as well as the variance parameter σ_{11}^2 are all significant at level 0.05. This clearly indicates that the correlation structure is not adequately explained only by the random effects alone – there exists significant correlation pattern over time after conditioning on the random effect. The same story happens to the overdispersion, that is, the random effects appear to have played a limited role in modeling the overdispersion. However, note that the estimate of overdispersion drops off dramatically and is no longer significant when the outlier 207 is dropped.

To assess the goodness-of-fit for these models, we may apply the criterion of $\log(H)$ (Thall and Vail 1990) to our method of estimating equations. The larger the $\log(H)$ is, the better the fit, provided that the number of parameters is the same. The results are reported in Table 2. Based on this criterion, we are in favour of the IC model over the genuine random effect model.

We also fit the data without the outlier 207 by the model with both random effects u_{i1} and u_{i2}, and the results are reported in Table 3.

Table 3

Estimates and standard errors without patient 207.

Variable	0-IC Model	IC Model
Intercept	1.1343 (0.1582)	1.1387 (0.1681)
Treatment	−0.1080 (0.1937)	−0.0961 (0.1940)
Visit	−0.0218 (0.1020)	−0.0028 (0.1119)
Trt.Vst	−0.2987 (0.1718)	−0.3223 (0.1666)
σ_{11}	0.4265 (0.0700)	0.4000 (0.1037)
σ_{12}	0.0227 (0.0352)	0.0507 (0.0726)
σ_{22}	0.2199 (0.0896)	0.1450 (0.0920)
ρ		0.1956 (0.8706)
ϕ		2.4770 (2.1925)
$\log(H)$	34.1620	33.8194

It is evident that adding the second random effect in the genuine random effects model improves the goodness-of-fit of the 0-IC model as the value of $\log(H)$ increases from 20.9037 up to 33.5904 when the full data are used (details are not shown in Table 3). On the other hand, with both random effects u_{i1} and u_{i2} present, the inclusion of IC and overdispersion in modeling does not boost the value for the goodness-of-fit at all, and the parameters σ_{12}^2, σ_{22}^2, ρ, and ϕ become statistically insignificant. This indicates the problem of over-parameterization for the correlation structure depicted by the two sets of parameters $(\sigma_{12}^2, \sigma_{22}^2)$ and (ρ, ϕ). In other words, the correlation structure of the data can be modeled by σ_{11}^2 together with either $(\sigma_{12}^2, \sigma_{22}^2)$ or (ρ, ϕ) but not together with both. Selection between the two might be needed and carried out using the $\log(H)$ criteria. For this data, it seems that the model with $(\sigma_{12}^2, \sigma_{22}^2)$ fits the data better than the one with (ρ, ϕ). An opposite result could also occur in practice, as illustrated by an example of Chi and Reinsel (1989) and by our simulation study, which is more likely to occur when the number of observations over time is relatively large so that the serial correlation tends to carry certain stochastic pattern that cannot be easily dealt with by the random effects alone.

In conclusion, modeling IC is useful to assess the assumption of conditional independence and is appealing to achieve an appropriate representation of the correlation structure in regression models with random effects.

Acknowledgements

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada and the NSF grant DMS-9505799 of the United States.

References

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

Chi, E and Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association* **84**, 452–459.

Crowder, M. (1987). On linear and quadratic estimating function. *Biometrika*, **74**, 591–597.

Davidian, M. and Gallant, A.R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika* **80**, 475-488.

Diggle, P.J. and Verbyla, A.P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54**, 401-415.

Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

Donnelly, C.A., Larid, N.M. and Ware, J.H. (1995). Prediction and creation of smooth curves for temporally correlated longitudinal data. *Journal of the American Statistical Association* **90**, 984–989.

Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. and Sun, L. (1997). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statistics in Medicine* **15**, 823-836.

Jørgensen, B. (1997). *The Theory of Dispersion Models.* London: Chapman & Hall.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

Leppik, I. E., et al. (1985). A double-blind crossover evaluation of progabide in partial seizures. *Neurology* **35**, 285.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, second edition. London: Chapman & Hall.

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

Song, P. X.-K. (1997). Generating dependent random numbers with given correlations and margins from exponential dispersion models. *Journal of Statistical Computation and Simulation* **56**, 317-335.

Thall, P.F. and Vail, S.C. (1990). Some covariance models for longitudinal data with overdispersion. *Biometrics* **46**, 657–671.

Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79–86.

Zeger, S.L., Liang, K.-Y. and Albert, P.S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049–1060.

Appendix

Here we will repeatedly use the following property of the multivariate normal moment generating function. For $\mathbf{b} \sim N(\mathbf{0}, R)$,

$$\mathrm{E}\left\{\exp\left(\mathbf{z}^\top \mathbf{b}\right)\right\} = \exp\left(\frac{1}{2}\mathbf{z}^\top R\mathbf{z}\right). \tag{11}$$

This immediately leads to the first moment in (2) by first conditioning on the normal random effect $\mathbf{b}_i$. Now we derive the marginal covariance $\mathrm{cov}(y_{it}, y_{is})$ useful for obtaining the second moment (4) in Section 2. Note that

$$\begin{aligned}
\mathrm{E}\mathrm{cov}(y_{it}, y_{is}|\mathbf{b}_i) &= \rho_i(t,s)\mathrm{E}\left\{\mathrm{var}(y_{it}|\mathbf{b}_i)\mathrm{var}(y_{is}|\mathbf{b}_i)\right\}^{1/2} \\
&= \phi\rho_i(t,s)\mathrm{E}\exp\left\{\frac{k}{2}(\mathbf{x}_{it}+\mathbf{x}_{is})^\top\boldsymbol{\beta} + \frac{k}{2}(\mathbf{z}_{it}+\mathbf{z}_{is})^\top \mathbf{b}_i\right\} \\
&= \phi\rho_i(t,s)\exp\left\{\frac{k}{2}(\mathbf{x}_{it}+\mathbf{x}_{is})^\top\boldsymbol{\beta} + \frac{k^2}{8}(\mathbf{z}_{it}+\mathbf{z}_{is})^\top R(\boldsymbol{\theta})(\mathbf{z}_{it}+\mathbf{z}_{is})\right\} \qquad (12) \\
&= \phi\rho_i(t,s)A^{(k)}_{its}(\boldsymbol{\beta},\boldsymbol{\theta}). \qquad (13)
\end{aligned}$$

Here $A^{(k)}_{its}$ is as given in (3). Also,

$$\mathrm{cov}\left(\mathrm{E}(y_{it}|\mathbf{b}_i), \mathrm{E}(y_{is}|\mathbf{b}_i)\right) = \exp\left(\mathbf{x}_{it}^\top\boldsymbol{\beta} + \mathbf{x}_{is}^\top\boldsymbol{\beta}\right)\mathrm{cov}\left(\exp\left(\mathbf{z}_{it}^\top\mathbf{b}_i\right), \exp\left(\mathbf{z}_{is}^\top\mathbf{b}_i\right)\right)$$

where

$$\begin{aligned}
&\mathrm{cov}\left(\exp\left(\mathbf{z}_{it}^\top\mathbf{b}_i\right), \exp\left(\mathbf{z}_{is}^\top\mathbf{b}_i\right)\right) = \mathrm{E}\exp\left\{(\mathbf{z}_{it}+\mathbf{z}_{is})^\top\mathbf{b}_i\right\} - \mathrm{E}\left\{\exp(\mathbf{z}_{it}^\top\mathbf{b}_i)\right\}\mathrm{E}\left\{\exp(\mathbf{z}_{is}^\top\mathbf{b}_i)\right\} \\
&= \exp\left\{\frac{1}{2}(\mathbf{z}_{it}+\mathbf{z}_{is})^\top R(\boldsymbol{\theta})(\mathbf{z}_{it}+\mathbf{z}_{is})\right\} - \exp\left\{\frac{1}{2}\mathbf{z}_{it}^\top R(\boldsymbol{\theta})\mathbf{z}_{it} + \frac{1}{2}\mathbf{z}_{is}^\top R(\boldsymbol{\theta})\mathbf{z}_{is}\right\} \\
&= \exp\left\{\frac{1}{2}\mathbf{z}_{it}^\top R(\boldsymbol{\theta})\mathbf{z}_{it} + \frac{1}{2}\mathbf{z}_{is}^\top R(\boldsymbol{\theta})\mathbf{z}_{is}\right\}\left\{\exp\left(\mathbf{z}_{it}^\top R(\boldsymbol{\theta})\mathbf{z}_{is}\right) - 1\right\}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathrm{cov}(y_{it}, y_{is}) &= \mathrm{E}\mathrm{cov}(y_{it}, y_{is}|\mathbf{b}_i) + \mathrm{cov}\left(\mathrm{E}(y_{it}|\mathbf{b}_i), \mathrm{E}(y_{is}|\mathbf{b}_i)\right) \\
&= \phi\rho_i(t,s)\exp\left\{\frac{k}{2}(\mathbf{x}_{it}+\mathbf{x}_{is})^\top\boldsymbol{\beta} + \frac{k^2}{8}(\mathbf{z}_{it}+\mathbf{z}_{is})^\top R(\boldsymbol{\theta})(\mathbf{z}_{it}+\mathbf{z}_{is})\right\} + \\
&\quad \exp\left(\mathbf{x}_{it}^\top\boldsymbol{\beta} + \mathbf{x}_{is}^\top\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}_{it}^\top R(\boldsymbol{\theta})\mathbf{z}_{it} + \frac{1}{2}\mathbf{z}_{is}^\top R(\boldsymbol{\theta})\mathbf{z}_{is}\right)\left\{\exp\left(\mathbf{z}_{it}^\top R(\boldsymbol{\theta})\mathbf{z}_{is}\right) - 1\right\} \\
&= \phi\rho_i(t,s)A^{(k)}_{its}(\boldsymbol{\beta},\boldsymbol{\theta}) + A^{(2)}_{its}(\boldsymbol{\beta},\boldsymbol{\theta}) - m_{it}(\boldsymbol{\beta},\boldsymbol{\theta})m_{is}(\boldsymbol{\beta},\boldsymbol{\theta}).
\end{aligned}$$

Some first derivatives for functions $m_{it}(\cdot)$ and $q_{its}(\cdot)$ with respect to Θ are given as follows. With regard to function m_{it}, for each $v = 1, \ldots, p$,

$$\frac{\partial}{\partial\beta_v}m_{it}(\boldsymbol{\beta},\boldsymbol{\theta}) = m_{it}(\boldsymbol{\beta},\boldsymbol{\theta})x_{itv},$$

and for each $v = 1, \ldots, r$,

$$\frac{\partial}{\partial \theta_v} m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{1}{2} m_{it}(\boldsymbol{\beta}, \boldsymbol{\theta}) \mathrm{tr} \left\{ \frac{\partial R(\boldsymbol{\theta})}{\partial \theta_v} \mathbf{z}_{it} \mathbf{z}_{it}^\top \right\}.$$

For function q_{its},

$$\frac{\partial}{\partial \beta_v} q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) = \left\{ \frac{k}{2} \phi \rho_i(t,s) A_{its}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\theta}) + A_{its}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\theta}) \right\} (x_{itv} + x_{isv}), \ v = 1, \ldots, p,$$

and if model (1) is used, then

$$\frac{\partial}{\partial \gamma_v} q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) = \phi \rho_i'(t,s) \left(\frac{\partial \alpha_i}{\partial \gamma_v} \right) A_{its}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\theta}), \ v = 1, \ldots, g,$$

where $\rho_i'(t,s) = |t-s| \alpha_i^{|t-s|-1}$.

For $v = 1, \ldots, r$,

$$\frac{\partial}{\partial \theta_v} q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) = \frac{1}{2} \left\{ \frac{k^2}{4} \phi \rho_i(t,s) A_{its}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\theta}) + A_{its}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\theta}) \right\} \mathrm{tr} \left\{ \frac{\partial R(\boldsymbol{\theta})}{\partial \theta_v} (\mathbf{z}_{it} + \mathbf{z}_{is})(\mathbf{z}_{it} + \mathbf{z}_{is})^\top \right\},$$

and

$$\frac{\partial}{\partial \phi} q_{its}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \phi) = \rho_i(t,s) A_{its}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\theta}).$$

Now we introduce the Proposition we referred to in Section 5.

Proposition:
Let y_{it}^* and $e_{it}^{(1)}$ be as defined in (8) and (10), respectively. Under the AR(2) model (9), we have

$$\begin{aligned}
&(i) \qquad \mathrm{var}(e_{it}^{(1)} | \mathbf{b}_i) = 1 - \frac{\pi_1^2(1+\pi_2)}{1-\pi_2}; \\
&(ii) \qquad \mathrm{cov}(e_{it}^{(1)}, e_{it-1}^{(1)} | \mathbf{b}_i) = \frac{\pi_1 \pi_2^2}{1-\pi_2}; \\
&(iii) \qquad \mathrm{corr}(e_{it}^{(1)}, e_{it-1}^{(1)} | \mathbf{b}_i) = \frac{\pi_1 \pi_2^2}{1 - \pi_2 - \pi_1^2(1+\pi_2)}.
\end{aligned}$$

Proof : Denote $V^b(\cdot) = \mathrm{var}(\cdot | \mathbf{b}_i)$ and $C^b(\cdot, \cdot) = \mathrm{cov}(\cdot, \cdot | \mathbf{b}_i)$.

(i)

$$\begin{aligned}
V^b(e_{it}^{(1)}) &= V^b(y_{it}^* - \pi_1 y_{it-1}^*) \\
&= V^b(y_{it}^*) + \pi_1^2 V^b(y_{it-1}^*) - 2\pi_1 C^b(y_{it}^*, y_{it-1}^*) \\
&= 1 + \pi_1^2 - 2\pi_1 \left(\frac{\pi_1}{1-\pi_2} \right) \\
&= 1 - \pi_1^2 \left(\frac{1+\pi_2}{1-\pi_2} \right).
\end{aligned}$$

(ii)

$$
\begin{aligned}
& C^b(e_{it}^{(1)}, e_{it-1}^{(1)}) = C^b(y_{it}^* - \pi_1 y_{it-1}^*, y_{it-1}^* - \pi_1 y_{it-2}^*) \\
= \quad & C^b(y_{it}^*, y_{it-1}^*) - \pi_1 C^b(y_{it-1}^*, y_{it-1}^*) - \pi_1 C^b(y_{it}^*, y_{it-2}^*) + \pi_1^2 C^b(y_{it-1}^*, y_{it-2}^*) \\
= \quad & \frac{\pi_1}{1-\pi_2} - \pi_1 - \pi_1 \left(\frac{\pi_1^2}{1-\pi_2} + \pi_2 \right) + \pi_1^2 \left(\frac{\pi_1}{1-\pi_2} \right) \\
= \quad & \pi_1 \left(\frac{\pi_2^2}{1-\pi_1} \right).
\end{aligned}
$$

(iii) Take $\text{corr}(e_{it}^{(1)}, e_{it-1}^{(1)}|\mathbf{b}_i) = C^b(e_{it}^{(1)}, e_{it-1}^{(1)})/\sqrt{V^b(e_{it}^{(1)})V^b(e_{it-1}^{(1)})}$, use (i) and (ii), we get

$$
\begin{aligned}
\text{corr}(e_{it}^{(1)}, e_{it-1}^{(1)}|\mathbf{b}_i) &= \frac{\pi_1 \pi_2^2/(1-\pi_1)}{1-\pi_1^2(1+\pi_2)/(1-\pi_1)} \\
&= \frac{\pi_1 \pi_2^2}{1-\pi_1-\pi_1^2(1+\pi_2)}.
\end{aligned}
$$

□

Remark:

The result in (iii) suggests that if the AR(1) assumption is true, i.e., if $\pi_2 = 0$, then $\text{corr}(e_{it}^{(1)}, e_{it-1}^{(1)}|\mathbf{b}_i) = 0$, as is claimed and used in Section 5 for testing the AR(1) assumption for IC.

Figure 1: Checking log-linearity

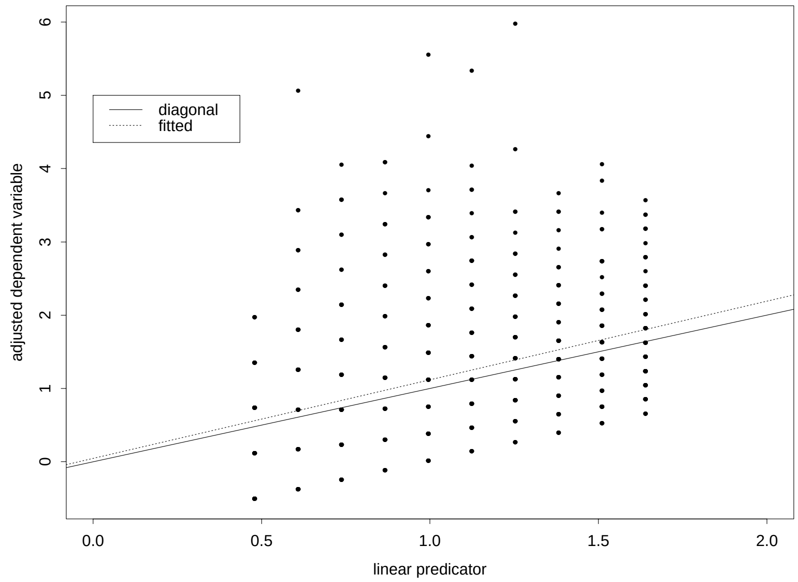

Figure 2: ACF function for checking AR(1) assumption.