Supplementary Materials to Transformations of Covariates for Longitudinal Data

WESLEY K. THOMPSON AND MINGE XIE

Department of Statistics, Rutgers University, Piscataway, NJ 08855

HELENE R. WHITE

Center of Alcohol Studies, Rutgers University, Piscataway, NJ 08855

Appendix A

Special Treatment of Discontinuity in Box-Tidwell and Fractional Polynomial Transformations

Suppose the linear predictor η_{ij} is related to covariate $x_{ij} > 0$ through the Box-Tidwell power transformation $x_{ij}^{(\gamma)}$, i.e., $\eta_{ij} = \beta_0 + \beta_1 x_{ij}^{(\gamma)}$. As a function of γ, $x^{(\gamma)}$ has a discontinuity at $\gamma = 0$, so that 0 belongs to the set $\mathcal{F}$ given in condition (C1). If the current estimate $\gamma^{(curr)} = 0$, an adjustment needs to be made to the iterative estimation procedure as given in Section 3. Using the first order Taylor's expansion about $\gamma^{(curr)} = 0$, we have

$$\eta_{ij} \;=\; \beta_0 + \beta_1 x_{ij}^{\gamma} \simeq \beta_0 + \beta_1 x_{ij}^{\gamma^{(curr)}} + \beta_1 x_{ij}^{\gamma^{(curr)}} \log x_{ij}(\gamma - \gamma^{(curr)}) = \beta_0' + \beta_1' \log x_{ij},$$

where $\beta_0' = \beta_0 + \beta_1$ and $\beta_1' = \beta_1(\gamma - \gamma^{(curr)})$. Clearly, the first step of the estimation procedure when $\gamma^{(curr)} = 0$ consists of fitting the first Taylor's expansion with re-parameterized β_0' and β_1'. Since we only can recover β_0' and β_1', but not β_0 and β_1, from this first step, we use the second order Taylor's expansion in the second step:

$$\begin{aligned}\eta_{ij} &\simeq \beta_0 + \beta_1 x_{ij}^{\gamma^{(curr)}} + \beta_1 x_{ij}^{\gamma^{(curr)}} \log x_{ij}(\gamma - \gamma^{(curr)}) + \beta_1 x_{ij}^{\gamma^{(curr)}} (\log x_{ij})^2 (\gamma - \gamma^{(curr)})^2/2 \\ &= \beta_0' + \beta_1' \log x_{ij} + \beta_1' (\log x_{ij})^2 \gamma/2.\end{aligned}$$

Replacing the terms in (3.2) and (3.3) respectively with the last lines in the above first and second order expansions, one can obtain an updated estimate $\gamma^{(new)}$ of γ.

The same situation exists for fitting fractional polynomials when the estimates for two powers $\gamma_1^{(curr)}$ and $\gamma_2^{(curr)}$ are equal. For example, suppose the model for η_{ij} is a fractional polynomial of second degree, or

$$\eta_{ij} = \begin{cases} \beta_0 + \beta_1 x_{ij}^{(\gamma_1)} + \beta_2 x_{ij}^{(\gamma_2)} & ,\gamma_1 \neq \gamma_2 \\ \beta_0 + \beta_1 x_{ij}^{(\gamma_1)} + \beta_2 x_{ij}^{(\gamma_1)} \log x_{ij} & , \gamma_1 = \gamma_2. \end{cases}$$

If $\gamma_1^{(curr)} = \gamma_2^{(curr)}$, by adopting the backfitting algorithm, we can proceed by fixing one of the power terms, say $\gamma_1^{(curr)}$, and expanding the second term around $\gamma_1^{(curr)}$:

$$\begin{aligned} \eta_{ij} &\simeq \beta_0 + \beta_1 x_{ij}^{\gamma_1^{(curr)}} + \beta_2 x_{ij}^{\gamma_1^{(curr)}} + \beta_2 x_{ij}^{\gamma_1^{(curr)}} \log x_{ij} (\gamma_2 - \gamma_1^{(curr)}) \\ &= \beta_0 + (\beta_1 + \beta_2) x_{ij}^{\gamma_1^{(curr)}} + \beta_2 (\gamma_2 - \gamma_1^{(curr)}) x_{ij}^{\gamma_1^{(curr)}} \log x_{ij} = \beta_0 + \beta_1'' x_{ij}^{\gamma_1^{(curr)}} + \beta_2'' x_{ij}^{\gamma_1^{(curr)}} \log x_{ij}, \end{aligned}$$

where $\beta_1'' = \beta_1 + \beta_2$ and $\beta_2'' = \beta_2(\gamma_2 - \gamma_1^{(curr)})$. The second step then consists of using the second order Taylor's expansion of the second term around $\gamma_1^{(curr)}$:

$$\begin{aligned} \eta_{ij} &\simeq \beta_0 + \beta_1 x_{ij}^{\gamma_1^{(curr)}} + \beta_2 x_{ij}^{\gamma_1^{(curr)}} + \beta_2 x_{ij}^{\gamma_1^{(curr)}} \log x_{ij} (\gamma_2 - \gamma_1^{(curr)}) + \beta_2 x_{ij}^{\gamma_1^{(curr)}} (\log x_{ij})^2 (\gamma_2 - \gamma_1^{(curr)})^2/2 \\ &= \beta_0 + \beta_1'' x_{ij}^{\gamma_1^{(curr)}} + \beta_2'' x_{ij}^{\gamma_1^{(curr)}} \log x_{ij} + \beta_2'' x_{ij}^{\gamma_1^{(curr)}} (\log x_{ij})^2 (\gamma_2 - \gamma_1^{(curr)})/2. \end{aligned}$$

From using the first and second Taylor's expansions with this reparameterization in the two-step estimation procedure, we can obtain updated estimate $\gamma_2^{(new)}$ of γ_2. We can then proceed in the same manner to update the estimate γ_1.

Appendix B

We sketch proofs of Theorems 1 and 2 under the assumption that the number of observations per cluster is bounded above by some fixed number t. Theorems 1 and 2 are proved under the conditions specified in B.1.

B.1 Conditions for Consistency and Asymptotic Normality

We make the following three regularity assumptions:

(I) $a'(\theta_{ij})$ and $u(\eta_{ij})$ are twice continuously differentiable, with $a''(\theta_{ij}) > 0$ for $\boldsymbol{\xi}$ in a neighborhood of $\boldsymbol{\xi}_0$.

(II) H_n and M_n are positive definite for large enough n, where $H_n = \sum_{i=1}^n D_i^T V_i^{-1} D_i$ and $M_n = \sum_{i=1}^n D_i^T V_i^{-1} \Sigma_i V_i^{-1} D_i$.

(III) $0 < c_1 \leq \zeta_{min} \leq \zeta_{max} \leq c_2 < \infty$, where $\zeta_{min} = \min_{1 \leq i \leq n}\{\lambda_{\min}(R_i^{-1})\}$ and $\zeta_{max} = \max_{1 \leq i \leq n}\{\lambda \max(R_i^{-1})\}$.

In **(II)**, D_i, V_i, and $\mathrm{cov}(\boldsymbol{Y}_i)$ are evaluated at $\boldsymbol{\xi}_0$. In **(III)**, $\lambda_{\min}(\cdot)$ (respectively $\lambda_{\max}(\cdot)$) denotes the smallest (respectively largest) eigenvalue.

The following conditions mirror those of Xie and Yang (2002), although ours are simpler because of the assumption that the cluster size is bounded:

(A) $\lambda_{\min}(H_n) \to \infty$.

(B) $\sup_{\boldsymbol{\xi} \in B_n(r)} \left\| H_n^{-1/2} \mathcal{D}(\boldsymbol{\xi}) H_n^{-1/2} - \mathrm{I} \right\| \xrightarrow{P} 0$, where $\mathcal{D}(\boldsymbol{\xi}) = -\partial U_n(\boldsymbol{\xi})/\partial \boldsymbol{\xi}^T$ and $B_n(r) = \left\{\boldsymbol{\xi} : \| H_n^{1/2}(\boldsymbol{\xi} - \boldsymbol{\xi}_0) \| \leq r\right\}$.

(C) Let $y_{ij}^* = (y_{ij} - \mu_{ij})/\sigma_{ij}$. Then there exists a $\delta > 0$ such that $\mathrm{E}(y_{ij}^*)^{(2+2/\delta)} \leq K$, for some constant K, for all i and j.

(D) $\tau_n = \max_{1 \leq i \leq n} \left\{\lambda \max(H_n^{-1/2} D_i^T V_i^{-1} D_i H_n^{-1/2})\right\} \to 0$.

Remark: Note that conditions **(A)** through **(D)** are satisfied under a variety of circumstances. One set of easily verifiable conditions is as follows:

(R) (i) $0 < |x_{ijk}| < C$ for all i, j, k, for some constant C, or x_{ijk} from a compact set bounded away from 0.

(ii) $\lambda_{\min}(\sum_{i=1}^n Z_i^T Z_i) \to \infty$.

That condition **(R)** implies **(A)** through **(D)** can be directly verified as a special case of Corollary 1, Section 5 of Xie and Yang (2002). Note that the condition **(R)** is essentially condition **($\boldsymbol{R_c^*}$)** of Fahrmeir and Kaufmann (1985).

B.2 Proofs of Theorems

Proof of Theorem 1: By regularity conditions **(I)** and **(II)** and non-singularity of $\mathcal{D}(\boldsymbol{\xi})$, $M_n^{-1/2} U_n(\boldsymbol{\xi})$ is a continuous, injective function from $B_n(r)$ to its image. Let $E_n =$

$\left\{\left\|M_n^{-1/2}U_n(\boldsymbol{\xi}_0)\right\| \leq \inf_{\xi\in\delta B_n(r)} \left\|M_n^{-1/2}\{U_n(\boldsymbol{\xi}) - U_n(\boldsymbol{\xi}_0)\}\right\|\right\}$, where $\delta B_n(r)$ be the boundary of the sphere $B_n(r)$. On the set E_n, there exists a solution $\hat{\boldsymbol{\xi}}_n \in B_n(r)$ to $U_n(\boldsymbol{\xi}) = 0$.

For $\boldsymbol{\xi} \in \delta B_n(r)$, there exists a $\bar{\boldsymbol{\xi}} \in B_n(r)$ such that

$$M_n^{-1/2}\left\{U_n(\boldsymbol{\xi}) - U_n(\boldsymbol{\xi}_0)\right\} = M_n^{-1/2}\mathcal{D}(\bar{\boldsymbol{\xi}})(\boldsymbol{\xi}_0 - \boldsymbol{\xi}) \xrightarrow{P} M_n^{-1/2}H_n(\boldsymbol{\xi}_0 - \boldsymbol{\xi}).$$

Moreover,

$$\left\|M_n^{-1/2}H_n(\boldsymbol{\xi}_0 - \boldsymbol{\xi})\right\| \geq \lambda_{\min}(M_n^{-1/2}H_n^{1/2})\left\|H_n^{1/2}(\boldsymbol{\xi}_0 - \boldsymbol{\xi})\right\| = c_n r,$$

where $c_n = \lambda_{\min}(M_n^{-1/2}H_n^{1/2}) > c$ for some $c > 0$ by condition **(III)**. Thus by Chebyshev's inequality,

$$P(E_n) \geq P(\left\|M_n^{-1/2}U_n(\boldsymbol{\xi}_0)\right\| \leq cr) \geq 1 - \frac{p}{cr},$$

where p is the dimension of $\boldsymbol{\xi}$. By choosing r appropriately, we can make this last quantity as close to 1 as we wish. This proves $\mathrm{P}\{U_n(\hat{\boldsymbol{\xi}}_n) = \mathbf{0}\} \rightarrow 1$. Since $\hat{\boldsymbol{\xi}}_n \in B_n(r)$ and $\lambda_{\min}(H_n) \to \infty$, it follows that $\hat{\boldsymbol{\xi}}_n \xrightarrow{P} \boldsymbol{\xi}_0$.

To show asymptotic Normality, let $Z_{n,i} = \boldsymbol{\lambda}^T M_n^{-1/2} D_i^T V_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$, for any fixed $\boldsymbol{\lambda}^T\boldsymbol{\lambda} = 1$. Conditions **(C)** and **(D)** ensure that $\sum_{i=1}^n E\left\{Z_{n,i}^2 I(Z_{n,i}^2 > \epsilon^2)\right\} \to 0, \forall\epsilon > 0$. Here, $I(E)$ is an indicator function, equal to 1 if set E is true and equal 0 otherwise. Thus, by the Lindeberg central limit theorem, $Z_n = \sum_{i=1}^n Z_{n,i} \xrightarrow{\mathcal{L}} N(0,1)$, i.e., $M_n^{-1/2}U_n(\boldsymbol{\xi}_0) \xrightarrow{\mathcal{L}} N(0, I)$. Furthermore, by Taylor's expansion, we have $H_n^{-1/2}U_n(\boldsymbol{\xi}_0) = H_n^{-1/2}\mathcal{D}(\bar{\boldsymbol{\xi}})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) = H_n^{-1/2}\mathcal{D}(\bar{\boldsymbol{\xi}})H_n^{-1/2}H_n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$, for some $\bar{\boldsymbol{\xi}}$ in $B_n(r)$. By condition **(B)**, $H_n^{-1/2}U_n(\boldsymbol{\xi}_0)$ is asymptotically equivalent to $H_n^{1/2}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$. Hence, together with $M_n^{-1/2}U_n(\boldsymbol{\xi}_0) \xrightarrow{\mathcal{L}} N(0, I)$, we can conclude that $(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \xrightarrow{\mathcal{L}} N(0, H_n^{-1}M_nH_n^{-1})$.

Proof of Theorem 2: Note that from the proof of Theorem 1, $M_n^{-1/2}U_n(\boldsymbol{\xi}_0) \xrightarrow{\mathcal{L}} \mathrm{N}(0, I)$.

Thus for the working score function T_S we have

$$T_S = n^{-1}U_n(\boldsymbol{\xi}_0)^T W_\xi U_n(\boldsymbol{\xi}_0) = U_n(\boldsymbol{\xi}_0)^T H_n^{-1} U_n(\boldsymbol{\xi}_0) \xrightarrow{\mathcal{L}} \sum_{k=1}^{p} a_k\chi_{1,k}^2,$$

where the $\chi_{1,k}^2$ are independent χ^2 variates, and the weights $a_1 \leq a_2 \leq \ldots \leq a_p$ are the eigenvalues of $M_n^{1/2}H_n^{-1}M_n^{1/2}$, or equivalently, the eigenvalues of $H_n^{-1}M_nH_n^{-1}H_n = V_\xi W_\xi^{-1}$. This proves Theorem 2.

Appendix C

Simulation studies

We carry out two simulation studies in this section. Our objective is to evaluate the performance of the estimates $\hat{\gamma}$ and the coverage percentages of the (asymptotic) 95% confidence intervals, calculated by the iterative procedure outlined in Section 3. In the first study, we also evaluate the performance of the working score test proposed in Section 4.

The data sets in the first study follow the Gaussian marginal model

$$Y_{ij} = \beta_0 + \beta_1 x_{ij}^{(\gamma)} + \epsilon_{ij}, \quad i = 1, \ldots, 300, \quad j = 1, \ldots, 6, \tag{C.1}$$

where the $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$. The ϵ_{ij} are generated with two types of within-subject correlation structure: exchangeable and first-order autoregressive (AR-1). For exchangeable correlation structure, the within-subject correlation $\mathrm{corr}(Y_{ij_1}, Y_{ij_2}) = \rho$ for $j_1 \neq j_2$; for AR-1 correlation structure, $\mathrm{corr}(Y_{ij_1}, Y_{ij_2}) = \rho^{|j_1 - j_2|}$. For each of the two types of correlation structure, we consider $3 \times 3 = 9$ settings for simulation, which correspond to the combinations of $\gamma = .5$, 0, or -1 and $\rho = 0$, 0.5, or 0.8. Under each setting, the covariates x_{ij} are set as $x_{ij} = j, j = 1, \ldots, 6$, then standardized to have unit variance. The regression parameters are $\beta_0 = 0$ and $\beta_1 = 2$, and the variance σ^2 is fixed at .5.

We execute the iterative estimation procedure outlined in Section 3 on 500 simulated data sets per setting. For each data set, we calculate the (asymptotic) 95% confidence interval for $\hat{\gamma}$. The resulting proportion of confidence intervals containing the true value of γ is given in the left-hand columns of Table A1. The coverage percentages are close to 95%, except that when ρ is large the working independence assumption leads to serious over-coverage. Also, the AR-1 working correlation gives slightly higher coverage than does the exchangeable working correlation when $\rho = 0.5$ or 0.8. The numbers in the italicized right-hand columns are the median estimates of $\hat{\gamma}$. The median values are quite close to their true values. We also calculate (but do not list in the table) the mean standard error for each setting. As the within-cluster correlation becomes higher, correctly modeling the covariance structure generally leads to shorter confidence interval estimates.

We have tested hypotheses H: $\gamma = 1$ vs. K: $\gamma \neq 1$ using the modified working score test statistic $T_S/\bar{a}$, with exchangeable working correlation. The value $\gamma = 1$ is away from the

Table A1: Box-Tidwell Transformation for Gaussian Response Models: Coverage and Median Estimates of γ

		AR-1						EX					
γ	ρ	IND		EX		AR-1		IND		EX		AR-1	
.5	0	.964	*.504*	.964	*.504*	.964	*.501*	.938	*.499*	.940	*.499*	.936	*.499*
	.5	.970	*.502*	.926	*.502*	.952	*.502*	.992	*.499*	.958	*.499*	.992	*.498*
	.8	1.000	*.500*	.960	*.500*	.972	*.501*	1.000	*.500*	.966	*.500*	.984	*.499*
0	0	.956	*.004*	.956	*.004*	.958	*.004*	.952	*.006*	.954	*.006*	.954	*.006*
	.5	.970	*.004*	.940	*.004*	.952	*.004*	.990	*.003*	.940	*.003*	.980	*.003*
	.8	1.000	*.003*	.966	*.003*	.970	*.003*	1.000	*.003*	.962	*.003*	.982	*.003*
-1	0	.950	*−1.004*	.952	*−1.004*	.950	*−1.004*	.936	*−.997*	.938	*−.997*	.934	*−.997*
	.5	.958	*−1.003*	.916	*−1.003*	.948	*−1.003*	.996	*−.999*	.958	*−.999*	.990	*−1.000*
	.8	.998	*−1.004*	.930	*−1.004*	.978	*−1.004*	1.000	*−1.000*	.960	*−1.000*	.984	*−1.000*

Simulation studies computing coverage percentages of 95% confidence intervals and median estimates of the transformation parameter γ under two settings: (i) true correlation is autoregressive-1 (AR-1), and (ii) true correlation is exchangeable (EX). On the second row, IND, EX, AR-1 indicate that the working correlations used in the parameter estimation are independent, exchangeable, autoregressive-1, respectively. The first two columns are the γ and ρ values used to simulate the data. The left-hand columns in the main body contain the coverage percentage of the (asymptotic) 95% CIs. The italicized right-hand columns are the median values of $\hat{\gamma}$.

true γ values used in the simulations. The working score test rejects the null hypothesis each time under the settings in these simulations, indicating very high power. This matches the results from the 95% confidence interval estimates, in that none of the confidence intervals contain the value $\gamma = 1$. To evaluate the rejection rates of the working score test when the null hypothesis is true, we have performed a separate set of simulations (not listed in the table) on 500 data sets simulated as in (C.1), with true $\gamma = 1$ and $\rho = 0$. The rejection rate, or proportion of test statistics $T_S/\bar{a} > \chi^2_{1,.95} = 3.84$, is 0.062, which is slightly higher than the expected rate of 0.05.

The second simulation study consists of binary responses with marginal mean modeled as

$$\mu_{ij} = P(Y_{ij} = 1 | x_{ij}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}, \quad i = 1, \ldots, 400, \; j = 1, \ldots, 20\,, \qquad \text{(C.2)}$$

where $\eta_{ij} = \beta_0 + \phi_2(x_{ij}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Here, $\phi_2(x_{ij}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ denotes a fractional polynomial of degree 2. The data are actually generated from a random intercept model,

$$P(Y_{ij} = 1 | x_{ij}, \boldsymbol{\beta}^*, \boldsymbol{\gamma}, b_i) = \frac{\exp(\eta^*_{ij})}{1 + \exp(\eta^*_{ij})}, \quad i = 1, \ldots, 400, \; j = 1, \ldots, 20\,, \qquad \text{(C.3)}$$

where $\eta_{ij}^* = \beta_0^* + \phi_2(x_{ij}, \boldsymbol{\beta}^*, \boldsymbol{\gamma}) + b_i$. The b_i are random intercepts generated as independent normal random variables with variance σ_b^2. The connection between the regression coefficients $\boldsymbol{\beta}$ in the marginal model (C.2) and the fixed effects $\boldsymbol{\beta}^*$ in the random intercept model (C.3), is given by $\boldsymbol{\beta} \approx \boldsymbol{\beta}^*(1 + \sigma_b^2/1.702^2)^{-1/2}$; see, Xie, Simpson and Carroll, (2000) for details. Thus, model (C.2) is a close but not exact approximation to the generated data.

For the first subset of simulations in this study, we set the transformations powers to $\gamma_1 = 0$ (log transformation) and $\gamma_2 = 1$. We set the covariates x_{ij} as $x_{ij} = j, j = 1, \ldots, 20$, and then standardize them to have three times unit variance. The regression parameters $(\beta_0, \beta_1, \beta_2) = (-1, -3, 1)$ are chosen to give a reasonably large range of means for each cluster. For each run, σ_b^2 is fixed at 0 (i.e., no random intercept, which leads to uncorrelated responses) or 5 (which results in correlated observations within clusters). We perform the iterative estimation procedure on 500 simulated data sets using the exchangeable working correlation structure. For each simulation, a parameter estimate and its 95% confidence intervals are calculated. Table A2 gives the median estimates of $\hat{\gamma}_1$ and $\hat{\gamma}_2$, as well as the first and third quartiles. As can be seen from the table, the medians of the estimates are close to the true values of 0 and 1. The coverage given for each parameter is the proportion of times that the true value of γ_1 and γ_2 falls within the 95% confidence interval. Coverage percentages are again close to 95%. A small portion of the standard errors for confidence interval estimates are inflated considerably by multicollinearity when the estimated values of γ_1 and γ_2 are close to each other (around 0.5), but the majority of the estimated standard errors of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are not affected in this way.

The second subset of simulations in this study has power parameters set at $\gamma_1 = \gamma_2 = 1$. We set the covariates x_{ij} as $x_{ij} = j, j = 1, \ldots, 20$, then standardize them to have unit variance. The regression parameters $(\beta_0, \beta_1, \beta_2) = (2.5, -3.5, 2.5)$ are again chosen to give a reasonably large range of means per cluster, and σ_b^2 is fixed at 0 or 5. In this case, the parameters are estimated under the assumption that the powers are equal, so that only one estimate $\hat{\gamma}$ is obtained. The median estimates are quite close to 1, with coverage percentages close to 95%, as predicted.

Table A2: Fractional Polynomial Transformation for Binomial Response Models: Coverage and Median Estimates for $\boldsymbol{\gamma}$

				$\hat{\gamma}_1$				$\hat{\gamma}_2$		
γ_1	γ_2	σ_b^2	Coverage γ_1	$Q_{.25}$	Median	$Q_{.75}$	Coverage γ_2	$Q_{.25}$	Median	$Q_{.75}$
0	1	0	.944	-.285	-.011	.229	.966	.704	1.002	1.428
		5	.954	-.227	.021	.238	.972	.674	.994	1.338
1	1	0	.956	.982	.997	1.052	-	-	-	-
		5	.966	.989	1.006	1.063	-	-	-	-

Simulation studies computing coverage percentages of 95% confidence intervals and quartiles of estimates for the transformation parameters $\boldsymbol{\gamma}$. The first three columns are the $\boldsymbol{\gamma}$ and ρ values used to simulate the data. Coverage γ_1 (respectively coverage γ_2) is the proportion of the 95% confidence intervals for γ_1 (respectively γ_2) containing their true values. The $Q_{.25}$, Median and $Q_{.75}$ are the indicated quartiles of $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

References

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342-368.

Xie, M., Simpson, D.G., and Carroll, R. (2000). Random Effects in Censored Ordinal Regression: Latent Structure and Bayesian Approach. *Biometrics* **56** 376–383.

Xie, M. and Yang, Y (2002). Asymptotics for Generalized Estimating Equations with Large Cluster Sizes. *Ann. Statist.* (in press).