Biometrika (2009), **96**, 2, *pp.* 293–306
doi: 10.1093/biomet/asp005
Advance Access publication 21 April 2009

Generalized method of moments estimation for linear regression with clustered failure time data

BY HUI LI

School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China
li_hui@bnu.edu.cn

AND GUOSHENG YIN

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.
gsyin@mdanderson.org

SUMMARY

We propose a generalized method of moments approach to the accelerated failure time model with correlated survival data. We study the semiparametric rank estimator using martingale-based moments. We circumvent direct estimation of correlation parameters by concatenating the moments and minimizing a quadratic objective function. We establish the consistency and asymptotic normality of the parameter estimators, and derive the limiting distribution of the objective function. We carry out simulation studies to examine the finite-sample properties of the method, and demonstrate its substantial efficiency gain over the conventional method. Finally, we illustrate the new proposal with an example from a diabetic retinopathy study.

Some key words: Accelerated failure time model; Asymptotic normality; Correlated survival data; Estimation efficiency; Moment condition; Rank estimation; Semiparametric model.

1. INTRODUCTION

Multivariate failure-time data often arise in biomedical and clinical studies, where the underlying correlations among the observed failure times may be artificially or naturally induced through clustering. For instance, in litter-matched mice experiments or family based genetic research, observations from subjects within the same cluster cannot be assumed to be independent, and ignoring the correlations would cause estimation efficiency loss. The proportional hazards model (Cox, 1972) has been generalized to accommodate correlated survival data (Wei et al., 1989; Lee et al., 1992). As an important alternative, the accelerated failure time model directly characterizes the covariate effects on accelerating or decelerating survival. It formulates a linear relationship between the logarithm of the failure time and covariates. Estimation and inference under the accelerated failure time model are often based on least-squares or rank estimators (Prentice, 1978; Buckley & James, 1979; Tsiatis, 1990; Ying, 1993; Jin et al., 2003).

When the survival data are correlated, Lin & Wei (1992) and Lee et al. (1993) adopted a working independence assumption for the least-squares and rank estimators under the marginal linear regression model. A natural way of gaining efficiency is to incorporate the correlation matrix in the estimation procedure. However, the true correlation structure is usually unknown. Gray (2003) studied weighted estimating equations for linear regression analysis of clustered survival data. More recently, Jin et al. (2006) developed rank-based monotone estimating functions based on the Gehan weight (Gehan, 1965), which guarantees a unique solution, as a global minimum, via linear programming. Under the accelerated failure time model, the variance-covariance matrix of the parameter estimators typically depends on the baseline hazard. Various resampling methods have been proposed for overcoming the difficulty of nonparametric functional estimation of the baseline hazard, but these are computationally intensive (Parzen et al., 1994; Jin et al., 2001).

The usual estimating equation can be viewed as a special case of the generalized method of moments, which has been extensively studied in econometrics (Hansen, 1982; Pakes & Pollard, 1989; Newey, 2004; Hall, 2005). Hansen (1982) established a comprehensive framework for the generalized method of moments and provided rigorous justification and asymptotics for the estimator. Pakes & Pollard (1989) derived asymptotic theories for the simulated method of moments when the function in the moment condition may be discontinuous. Qu et al. (2000) applied the generalized method of moments to longitudinal studies, and showed its advantages over generalized estimating equations (Liang & Zeger, 1986). Lai & Small (2007) proposed to analyze longitudinal data with three different types of time-dependent covariates based on the generalized method of moments. Since it combines moments, the generalized method of moments is parsimonious and useful for constructing efficient estimators, particularly when the efficiency bound is complicated and moment conditions are relatively easy to obtain. Under some regularity conditions (Chamberlain, 1987), the generalized method of moments estimator achieves the semiparametric efficiency bound in the sense of Bickel et al. (1993). In contrast to the generalized linear model, in which the variance component is usually a function of the mean parameter (McCullagh & Nelder, 1989, Ch. 2), rank estimation based on the generalized method of moments under the accelerated failure time model does not have this elegant structure. It thus makes estimation and inference much more challenging from both numerical and theoretical perspectives. Without needing to specify the correlation structure and the baseline hazard, we take a linear expansion of the weight matrix over a set of commonly used basis matrices, and develop rank-based moments using martingale properties.

2. RANK ESTIMATION

For $i = 1, \ldots, n$ and $k = 1, \ldots, K_i$, let T_{ik} and C_{ik} be the failure and censoring times for the kth subject in the ith cluster. We assume that T_{ik} is conditionally independent of C_{ik} given the p-dimensional bounded covariates Z_{ik}. Denote the observed time by $X_{ik} = \min(T_{ik}, C_{ik})$ and the censoring indicator by $\Delta_{ik} = I(T_{ik} \leqslant C_{ik})$, where $I(\cdot)$ is the indicator function. Observations from the same cluster may be dependent but exchangeable. We allow the cluster size to vary by setting $\log(X_{ik})$ to be $-\infty$, with taking $\log(X_{ik}) = -10^6$ being satisfactory in numerical studies, $\Delta_{ik} = 0$ and $Z_{ik} = 0$, when T_{ik} is missing for cluster i.

The marginal accelerated failure time model is given by

$$\log(T_{ik}) = \beta_0^{\mathrm{T}} Z_{ik} + \varepsilon_{ik},$$

where the distribution of the errors $(\varepsilon_{i1}, \ldots, \varepsilon_{iK_i})^{\mathrm{T}}$ is unknown and unspecified. For each i, $(\varepsilon_{i1}, \ldots, \varepsilon_{iK_i})$ are potentially dependent and share a common marginal distribution. For any i and j, let $K = \min(K_i, K_j)$. Then $(\varepsilon_{i1}, \ldots, \varepsilon_{iK})$ and $(\varepsilon_{j1}, \ldots, \varepsilon_{jK})$ have the same distribution with a continuous and bounded density on $(-\infty, \tau + \xi]^K$, for fixed truncation points $\tau > 0$ and $\xi > 0$. We compute the residuals, $e_{ik}(\beta) = \log(X_{ik}) - \beta^{\mathrm{T}} Z_{ik}$, based on which we define the at-risk process $Y_{ik}(u, \beta) = I\{e_{ik}(\beta) \geqslant u\}$ and the counting process $N_{ik}(u, \beta) = \Delta_{ik} I\{e_{ik}(\beta) \leqslant u\}$.

The linear rank-estimating equation takes the form

$$n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{-\infty}^{\tau}\phi(u,\beta)\{Z_{ik}-\bar{Z}(u,\beta)\}\,dN_{ik}(u,\beta)=0, \tag{1}$$

where $\phi(u,\beta)$ satisfies Condition 5 in Ying (1993, pp. 90–1), and

$$\bar{Z}(u,\beta)=\frac{\sum_{i=1}^{n}\sum_{k=1}^{K_i}Y_{ik}(u,\beta)Z_{ik}}{\sum_{i=1}^{n}\sum_{k=1}^{K_i}Y_{ik}(u,\beta)}.$$

If $\phi(u,\beta)=1$, (1) reduces to the log-rank statistic; if $\phi(u,\beta)=n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}Y_{ik}(u,\beta)$, it becomes the Gehan statistic; and if $\phi(u,\beta)$ is the marginal Kaplan–Meier estimator based on $\{(e_{ik}(\beta),\Delta_{ik}),\ i=1,\ldots,n;\ k=1,\ldots,K_i\}$, it corresponds to the Peto–Prentice generalization of the Wilcoxon statistic (Peto & Peto, 1972; Prentice, 1978). The empirical estimator of the square-integrable martingale is $\hat{M}_{ik}(t,\beta)=N_{ik}(t,\beta)-\int_{-\infty}^{t}Y_{ik}(u,\beta)d\hat{\Lambda}_0(u,\beta)$, where

$$\hat{\Lambda}_0(t,\beta)=\int_{-\infty}^{t}\frac{\sum_{i=1}^{n}\sum_{k=1}^{K_i}dN_{ik}(u,\beta)}{\sum_{i=1}^{n}\sum_{k=1}^{K_i}Y_{ik}(u,\beta)}.$$

We can rewrite (1) in the vector form

$$n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\tau}Z_i\Phi_i(u,\beta)\,d\hat{M}_i(u,\beta)=0, \tag{2}$$

where $\hat{M}_i(u,\beta)=\{\hat{M}_{i1}(u,\beta),\ldots,\hat{M}_{iK_i}(u,\beta)\}^{\mathrm{T}}$, $Z_i=(Z_{i1},\ldots,Z_{iK_i})$, and $\Phi_i(u,\beta)$ is a K_i-diagonal matrix with elements $\phi(u,\beta)$.

The estimator obtained from (2), although still consistent, may not be efficient, as it completely ignores the correlation information. A natural way of enhancing estimation efficiency is by incorporating a weight matrix to account for the within-cluster correlation; that is,

$$n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\tau}Z_i\Phi_i^{1/2}(u,\beta)R^{-1}(\alpha)\Phi_i^{1/2}(u,\beta)\,d\hat{M}_i(u,\beta)=0, \tag{3}$$

where $R(\alpha)$ is the unknown correlation matrix. Regardless of the complexity of $R^{-1}(\alpha)$, it may be represented as

$$R^{-1}(\alpha)=\sum_{l=1}^{m}\alpha_l C_{(l)},$$

where $(\alpha_1,\ldots,\alpha_m)$ are unknown constants, and $(C_{(1)},\ldots,C_{(m)})$ are a set of known basis matrices. The linear span of $C_{(l)}$ provides an adequate approximation of the true correlation structure. For example, if $R(\alpha)$ is an exchangeable matrix, then $R^{-1}(\alpha)=\alpha_1C_{(1)}+\alpha_2C_{(2)}$, where $C_{(1)}=I$, the identity matrix, and $C_{(2)}$ has 0 on the diagonal and 1 elsewhere; and, if $R(\alpha)$ is a first-order autoregressive AR(1) correlation matrix, $R^{-1}(\alpha)=\alpha_1C_{(1)}+\alpha_2C_{(2)}+\alpha_3C_{(3)}$, where $C_{(1)}=I$, $C_{(2)}$ is of the sandwich form with two main off-diagonals of 1 and 0 elsewhere, and $C_{(3)}$ is a matrix with the $(1,1)$ and (K,K) elements equal to 1 and 0 elsewhere.

Since martingale integrals have zero-mean, the expanded moment conditions are

$$n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\tau}Z_i\Phi_i^{1/2}(u,\beta)C_{(l)}\Phi_i^{1/2}(u,\beta)\,d\hat{M}_i(u,\beta)=0\quad(l=1,\ldots,m). \tag{4}$$

The unknown coefficients α_l can take any real values, which, however, do not need to be estimated in the generalized method of moments procedure. Furthermore, the α_l form an implicit weighting scheme that automatically adjusts the contribution of $C_{(l)}$ to the entire correlation structure. The proposed method is completely different from the usual weighted estimating equation; see for example Cai & Prentice (1995) and Gray (2003). Traditionally, one would solve the weighted estimating equation (3) directly, and this can be computationally difficult. In our case, we concatenate the moments in (4) and import them into a minimization procedure in such a way that we do not need to estimate $R(\alpha)$.

3. Generalized method of moments

3·1. *Estimation procedure*

Instead of estimating α_l, we split up the moment conditions corresponding to $C_{(l)}$ ($l = 1, \ldots, m$). In this case, there are more estimating equations than unknown parameters. If we define $W_i^{(l)}(u, \beta) = \Phi_i^{1/2}(u, \beta) C_{(l)} \Phi_i^{1/2}(u, \beta)$, and $w_{ijk}^{(l)}(u, \beta)$ is its (j, k)th element, then let $a_{ik}^{(l)}(u, \beta) = \sum_{j=1}^{K_i} Z_{ij} w_{ijk}^{(l)}(u, \beta)$ and

$$\bar{a}^{(l)}(u, \beta) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{K_i} Y_{ik}(u, \beta) a_{ik}^{(l)}(u, \beta)}{\sum_{i=1}^{n} \sum_{k=1}^{K_i} Y_{ik}(u, \beta)}.$$

We construct an extended moment in the form $S_n(\beta) = n^{-1} \sum_{i=1}^{n} U_i(\beta)$, where

$$U_i(\beta) = \begin{pmatrix} \sum_{k=1}^{K_i} \int_{-\infty}^{\tau} \{a_{ik}^{(1)}(u, \beta) - \bar{a}^{(1)}(u, \beta)\} \, d\hat{M}_{ik}(u, \beta) \\ \vdots \\ \sum_{k=1}^{K_i} \int_{-\infty}^{\tau} \{a_{ik}^{(m)}(u, \beta) - \bar{a}^{(m)}(u, \beta)\} \, d\hat{M}_{ik}(u, \beta) \end{pmatrix}.$$

The quadratic objective function is given by $Q_n(\beta) = S_n^{\mathrm{T}}(\beta) \Sigma_n^{-1}(\beta) S_n(\beta)$, where it makes intuitive sense to choose $\Sigma_n(\beta)$ as the empirical estimate of the variance-covariance matrix for $n^{1/2} S_n(\beta)$,

$$\Sigma_n(\beta) = n^{-1} \sum_{i=1}^{n} U_i(\beta) U_i^{\mathrm{T}}(\beta) - S_n(\beta) S_n^{\mathrm{T}}(\beta).$$

We obtain $\hat{\beta}$ by minimizing $Q_n(\beta)$, i.e., $\hat{\beta} = \arg\min_{\beta} Q_n(\beta)$. To minimize $Q_n(\beta)$, we can use the Nelder–Mead simplex algorithm, which does not require any derivatives or continuity of the target function (Nelder & Mead, 1965; Press et al., 1989). In the original work of Hansen (1982), a two-stage estimation procedure was provided: at the jth iteration, one can obtain $\hat{\beta}_j$ by minimizing $S_n^{\mathrm{T}}(\beta) \Sigma_n^{-1}(\hat{\beta}_{j-1}) S_n(\beta)$, with the estimator from the previous iteration plugged into $\Sigma_n^{-1}(\hat{\beta}_{j-1})$ so that the only unknown parameter is in $S_n(\beta)$. Equivalently, we can minimize $S_n^{\mathrm{T}}(\beta) \Sigma_n^{-1}(\beta) S_n(\beta)$ directly with respect to β (Hansen et al., 1996), and this is the procedure we used in our numerical studies.

3·2. *Asymptotic properties*

Let $G_n(\beta) = \{G_{(1)}^{\mathrm{T}}(\beta), \ldots, G_{(m)}^{\mathrm{T}}(\beta)\}^{\mathrm{T}}$, where $G_{(l)}(\beta) = \int_{-\infty}^{\tau} A_{(l)}(u, \beta) \, d\lambda_0(u)$, $\lambda_0(u)$ is the baseline hazard function and

$$A_{(l)}(u, \beta) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} Y_{ik}(u, \beta) \{a_{ik}^{(l)}(u, \beta) - \bar{a}^{(l)}(u, \beta)\} Z_{ik}^{\mathrm{T}} \quad (l = 1, \ldots, m).$$

Let $\mathcal{G}(\beta) = \{\mathcal{G}^{\rm T}_{(1)}(\beta), \ldots, \mathcal{G}^{\rm T}_{(m)}(\beta)\}^{\rm T}$, where $\mathcal{G}_{(l)}(\beta)$ is the limit of $G_{(l)}(\beta)$, and let $\Sigma(\beta)$ be the limit of $\Sigma_n(\beta)$.

THEOREM 1. *Under Assumptions* A1–A7 *in the Appendix,* $\hat\beta$ *converges to* β_0*, in probability.*

THEOREM 2. *Under Assumptions* A1–A7 *in the Appendix,*

$$n^{1/2}(\hat\beta - \beta_0) \to N\left[0, \{\mathcal{G}^{\rm T}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0)\}^{-1}\right],$$

in distribution.

The limiting distribution of $Q_n(\beta)$ can be used for statistical inference. More generally, when β contains nuisance parameters, the following theorem lays out the theoretical foundation for the asymptotic distribution of the quadratic function.

THEOREM 3. *Suppose that* $\beta = (\gamma^{\rm T}, \theta^{\rm T})^{\rm T}$*, where the* q*-dimensional parameter* γ *is of interest, and* θ *is a* $(p-q)$*-dimensional nuisance parameter. Under the null hypothesis* $H_0 : \gamma = \gamma_0$*, and the local alternative hypothesis* $H_1 : \gamma = \gamma_0 + n^{-1/2}\epsilon$*, if we define* $\tilde\theta = \arg\min_\theta Q_n(\gamma_0, \theta)$ *and* $(\hat\gamma, \hat\theta) = \arg\min_{(\gamma,\theta)} Q_n(\gamma, \theta)$*, then, in distribution,*

$$n\{Q_n(\gamma_0, \tilde\theta) - Q_n(\hat\gamma, \hat\theta)\} \to \chi^2_q(c),$$

a chi-squared distribution with q *degrees of freedom and noncentrality parameter* $c = \epsilon^{\rm T}(J_{\gamma\gamma} - J_{\gamma\theta}J^{-1}_{\theta\theta}J_{\theta\gamma})\epsilon$*, where* $J_{\gamma\gamma}$*,* $J_{\gamma\theta}$ *and* $J_{\theta\theta}$ *correspond to the block matrices in* $\mathcal{G}^{\rm T}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0)$*.*

The proofs of the three theorems are outlined in the Appendix. As special cases without any nuisance parameters, under the null hypothesis, the quadratic function $nQ_n(\hat\beta)$ has the asymptotic distribution of χ^2_{mp-p}, with $m > 1$. Moreover, in distribution, $nQ_n(\beta_0) \to \chi^2_{mp}$ and $n\{Q_n(\beta_0) - Q_n(\hat\beta)\} \to \chi^2_p$, as $n \to \infty$. These chi-squared tests are closely related to the usual likelihood ratio tests, and $\mathcal{G}^{\rm T}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0)$ behaves like Fisher's information matrix.

3·3. *Estimation efficiency*

We examine the efficiency gain of our estimator $\hat\beta$ based on the generalized method of moments approach. Without loss of generality, we compare the estimation efficiency of the generalized method of moments procedure with only one correlation matrix, $C_{(1)}$, and that with additional basis matrices.

We first partition $\mathcal{G}(\beta_0)$ as $\mathcal{G}(\beta_0) = \{\mathcal{G}^{\rm T}_{(1)}(\beta_0), \tilde{\mathcal{G}}^{\rm T}(\beta_0)\}^{\rm T}$, where $\mathcal{G}_{(1)}(\beta_0)$ corresponds to the first basis matrix $C_{(1)}$ and $\tilde{\mathcal{G}}(\beta_0) = \{\mathcal{G}^{\rm T}_{(2)}(\beta_0), \ldots, \mathcal{G}^{\rm T}_{(m)}(\beta_0)\}^{\rm T}$ correspond to the remaining basis matrices. If we let $H(\beta_0) = \Sigma_{22}(\beta_0) - \Sigma_{21}(\beta_0)\Sigma^{-1}_{11}(\beta_0)\Sigma_{12}(\beta_0)$, then

$$\Sigma(\beta_0) = \begin{pmatrix} \Sigma_{11}(\beta_0) & \Sigma_{12}(\beta_0) \\ \Sigma_{21}(\beta_0) & \Sigma_{22}(\beta_0) \end{pmatrix}, \quad \Sigma^{-1}(\beta_0) = \begin{pmatrix} B_{11}(\beta_0) & B_{12}(\beta_0) \\ B_{21}(\beta_0) & B_{22}(\beta_0) \end{pmatrix},$$

where

$$B_{11}(\beta_0) = \Sigma^{-1}_{11}(\beta_0) + \Sigma^{-1}_{11}(\beta_0)\Sigma_{12}(\beta_0)H^{-1}(\beta_0)\Sigma_{21}(\beta_0)\Sigma^{-1}_{11}(\beta_0),$$
$$B_{12}(\beta_0) = B^{\rm T}_{21}(\beta_0) = -\Sigma^{-1}_{11}(\beta_0)\Sigma_{12}(\beta_0)H^{-1}(\beta_0),$$
$$B_{22}(\beta_0) = H^{-1}(\beta_0).$$

Since $\Sigma(\beta_0)$ is positive definite, as is $H(\beta_0)$, we can write $H(\beta_0) = L^{\mathrm{T}}(\beta_0)L(\beta_0)$. Then, some algebraic manipulation gives

$$\mathcal{G}^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0) = \mathcal{G}^{\mathrm{T}}_{(1)}(\beta_0)\Sigma_{11}^{-1}(\beta_0)\mathcal{G}_{(1)}(\beta_0) + \{\mathcal{G}^{\mathrm{T}}_{(1)}(\beta_0)\Sigma_{11}^{-1}(\beta_0)\Sigma_{12}(\beta_0)L^{-1}(\beta_0) - \tilde{\mathcal{G}}^{\mathrm{T}}(\beta_0)L^{-1}(\beta_0)\}^{\otimes 2}.$$

Since $\{\mathcal{G}^{\mathrm{T}}_{(1)}(\beta_0)\Sigma_{11}^{-1}(\beta_0)\mathcal{G}_{(1)}(\beta_0)\}^{-1}$ is the covariance matrix of $\hat{\beta}$ obtained by only incorporating $C_{(1)}$, it follows that $\{\mathcal{G}^{\mathrm{T}}_{(1)}(\beta_0)\Sigma_{11}^{-1}(\beta_0)\mathcal{G}_{(1)}(\beta_0)\}^{-1} - \{\mathcal{G}^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0)\}^{-1}$ is positive semidefinite.

3·4. *Variance estimation*

The variance-covariance matrix of $\hat{\beta}$ involves the baseline hazard function of the error, which is difficult to estimate reliably through nonparametric methods. Various resampling techniques have been investigated for variance estimation that circumvent nonparametric estimation; see for example Parzen et al. (1994) and Jin et al. (2001). To overcome the intensive computation of resampling procedures, we take a variance decomposition method based on a linear expansion of $S_n(\beta)$ around β_0, motivated by Huang (2002). Through a Cholesky decomposition, we decompose the estimated variance-covariance matrix of $S_n(\beta)$ at $\beta = \hat{\beta}$, i.e., $n^{-1}\Sigma_n(\hat{\beta}) = D^{\mathrm{T}}D$, where $D = (d_1, \ldots, d_{mp})$. Then the perturbed version of the generalized method of moments estimator $\tilde{\beta}_j$ satisfies

$$n^{-1}\sum_{i=1}^{n} U_i(\tilde{\beta}_j) = d_j \quad (j = 1, \ldots, mp).$$

Once again, we can obtain $\tilde{\beta}_j$ by minimizing $\tilde{Q}_n(\beta_j) = \tilde{S}_n^{\mathrm{T}}(\beta_j)\tilde{\Sigma}_n^{-1}(\beta_j)\tilde{S}_n(\beta_j)$, where $\tilde{S}_n(\beta_j) = n^{-1}\sum_{i=1}^{n}\tilde{U}_i(\beta_j)$, $\tilde{U}_i(\beta_j) = U_i(\beta_j) - d_j$ and $\tilde{\Sigma}_n(\beta_j) = n^{-1}\sum_{i=1}^{n}\tilde{U}_i(\beta_j)\tilde{U}_i^{\mathrm{T}}(\beta_j) - \tilde{S}_n(\beta_j)\tilde{S}_n^{\mathrm{T}}(\beta_j)$. Finally, if we define $\eta = (\tilde{\beta}_1 - \hat{\beta}, \ldots, \tilde{\beta}_{mp} - \hat{\beta})$, then $\eta\eta^{\mathrm{T}}$ is a consistent estimator for the variance-covariance matrix of $\hat{\beta}$. The validity of the proposed variance estimation can be established by taking a linear expansion of the estimating function. For each j, we can show that $\tilde{\beta}_j = \hat{\beta} + \Gamma(\beta_0)d_j + o_p(n^{-1/2} + \|\tilde{\beta}_j - \hat{\beta}\|)$, where $\Gamma(\beta_0) = \{\mathcal{G}^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0)\}^{-1}\mathcal{G}^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)$.

4. SIMULATION STUDIES

In our simulation studies, the failure times were simulated from the marginal linear regression model, where $(\varepsilon_{i1}, \ldots, \varepsilon_{iK})^{\mathrm{T}}$ were generated from a multivariate normal distribution with zero-mean and an exchangeable covariance matrix of $I_K + \rho(1_K 1_K^{\mathrm{T}} - I_K)$, with $\rho = 0{\cdot}5$ and $0{\cdot}8$, I_K being the $K \times K$ identity matrix and 1_K a K-vector of ones. We included two independent covariates in the model: $Z_1 \sim \mathrm{Ber}(0{\cdot}5)$ and $Z_2 \sim \mathrm{Un}[0, 1]$, with the regression coefficients $\beta_1 = \beta_2 = 1$. Censoring times were generated from uniform distributions to yield desirable censoring percentages. The number of clusters was $n = 100$, and the cluster size was $K = 4$. We implemented the log-rank, Gehan and Wilcoxon linear rank statistics for comparison. For each configuration, we carried out 1000 simulations.

For each data realization, we computed the biases for the proposed estimators of β_1 and β_2, the standard errors, using the decomposition method, the standard deviations, characterizing the sampling variation over 1000 simulations, and the 95% confidence interval coverage rates. The estimation performance can be assessed via the mean squared error. Table 1 summarizes results for $m = 1$ with only one identity matrix, $C_{(1)} = I_4$, and for $m = 2$, with an additional basis matrix with

Table 1. *Simulation study of generalized method of moments rank estimation with cluster size $K = 4$ and exchangeable multivariate normal errors. The basis matrices are $C_{(1)} = I_4$ and $C_{(2)} = 1_4 1_4^{\mathrm{T}} - I_4$*

ρ	c%	Method	m	β_1 Bias	SE	SD ($\times 10^{-2}$)	MSE	CR	β_2 Bias	SE	SD ($\times 10^{-2}$)	MSE	CR
0·5	0	Log rank	1	−0·2	10·9	10·9	1·2	94·1	−0·3	18·7	18·5	3·4	94·8
			2	−0·1	9·0	9·1	0·8	93·6	−0·6	15·6	16·1	2·6	94·0
		Gehan	1	−0·2	10·2	10·0	1·0	95·0	−0·1	17·5	17·4	3·0	94·1
			2	−0·2	8·3	8·3	0·7	94·9	−0·6	14·3	14·9	2·2	94·1
		Wilcoxon	1	−0·2	10·2	10·1	1·0	94·9	−0·1	17·5	17·4	3·0	94·4
			2	−0·2	8·3	8·3	0·7	94·6	−0·6	14·3	14·9	2·2	94·4
	25	Log rank	1	−0·3	11·5	11·4	1·3	95·1	0·0	19·9	19·7	3·9	94·6
			2	−0·4	10·1	10·1	1·0	95·0	−0·1	17·6	17·6	3·1	94·1
		Gehan	1	−0·2	10·9	10·7	1·2	94·6	−0·1	18·7	18·7	3·5	94·2
			2	−0·3	9·3	9·3	0·9	94·4	−0·3	15·9	16·3	2·7	94·3
		Wilcoxon	1	−0·3	10·8	10·7	1·1	95·2	−0·1	18·5	18·6	3·5	94·3
			2	−0·3	9·2	9·2	0·8	94·5	−0·2	15·9	16·2	2·6	93·3
0·8	0	Log rank	1	0·0	10·6	10·7	1·1	93·7	0·1	18·4	18·2	3·3	94·7
			2	0·0	6·1	6·3	0·4	92·5	−0·4	10·5	10·8	1·2	92·9
		Gehan	1	−0·1	10·2	10·0	1·0	95·1	0·3	17·4	17·0	2·9	94·3
			2	−0·2	5·5	5·4	0·3	94·1	−0·4	9·3	9·9	1·0	92·5
		Wilcoxon	1	−0·1	10·2	10·0	1·0	95·4	0·3	17·4	17·0	2·9	94·6
			2	−0·2	5·5	5·4	0·3	94·1	−0·4	9·3	9·9	1·0	93·2
	25	Log rank	1	−0·2	11·5	11·5	1·3	94·2	0·0	19·8	19·6	3·8	93·8
			2	−0·4	8·1	8·3	0·7	93·3	−0·4	13·8	14·4	2·1	92·9
		Gehan	1	0·0	10·8	10·7	1·1	94·8	0·0	18·5	18·4	3·4	94·8
			2	−0·2	6·7	6·8	0·5	94·1	−0·3	11·5	12·0	1·4	92·8
		Wilcoxon	1	−0·1	10·8	10·7	1·1	94·8	0·1	18·5	18·3	3·3	95·6
			2	−0·2	6·7	6·8	0·5	94·0	−0·5	11·6	12·1	1·5	93·4

SE, average of estimated standard errors; SD, standard deviation; MSE, mean squared error; CR, 95% coverage rate; m, number of basis matrices.

diagonal elements of 0 and off-diagonal elements of 1, $C_{(2)} = 1_4 1_4^{\mathrm{T}} - I_4$. The biases are very small no matter which weight function is used. The standard errors are close to the standard deviations, indicating good performance of the variance estimation based on the Cholesky decomposition method. The 95% confidence interval coverage rates match the nominal level very well. Among the three different weight functions, the Gehan and Wilcoxon weights appear to behave slightly better in terms of mean squared error and the 95% coverage rate. When the number of moments increased from $m = 1$ to 2, the former is substantially reduced. In particular, there is more efficiency gain in the scenarios with higher correlations and a lower rate of censoring.

In Table 2, we examine cases with more basis matrices by taking $m = 1, 2$ and 3, corresponding to $C_{(1)} = I_4$, $C_{(2)}$ of the sandwich form with two main off-diagonals of 1 and 0 elsewhere, and $C_{(3)}$ with the three main diagonal elements of 0 and 1 elsewhere, respectively. Clearly, there is an overall tendency for mean squared error to decrease as m increases. The efficiency gain is greater when m is increased from one to two than when m is increased from two to three. The biases of the proposed estimators are negligible. The standard errors are very close to the standard deviations, which provide a good approximation of the variability of the estimators. The 95% coverage rates are accurate. We can compare the efficiency gain between adding a correctly specified or a misspecified second basis matrix. For the cases with $m = 2$ in Tables 1 and 2, the data were generated with an exchangeable correlation structure for the error. The estimates in

Table 2. *Simulation study of generalized method of moments rank estimation with* 25% *censoring, cluster size* $K = 4$ *and exchangeable multivariate normal errors. The basis matrices are* $C_{(1)} = I_4$, $C_{(2)}$ *of the sandwich form with two main off-diagonals of* 1 *and* 0 *elsewhere, and* $C_{(3)}$ *with the three main diagonal elements of* 0 *and* 1 *elsewhere*

ρ	Method	m	β_1 Bias	SE	SD	MSE	CR	β_2 Bias	SE	SD	MSE	CR
					$(\times 10^{-2})$					$(\times 10^{-2})$		
0·5	Log rank	1	−0·3	11·5	11·4	1·3	95·1	0·0	19·9	19·7	3·9	94·6
		2	−0·3	10·8	10·6	1·1	94·5	0·2	18·9	19·0	3·6	94·1
		3	−0·2	10·4	10·2	1·0	93·9	−0·2	18·4	17·9	3·2	94·3
	Gehan	1	−0·2	10·9	10·7	1·2	94·6	−0·1	18·7	18·7	3·5	94·2
		2	−0·2	9·9	9·8	1·0	94·9	−0·1	17·3	17·3	3·0	93·7
		3	−0·1	9·5	9·4	0·9	94·2	0·0	16·4	16·4	2·7	94·1
	Wilcoxon	1	−0·3	10·8	10·7	1·1	95·2	−0·1	18·5	18·6	3·5	94·3
		2	−0·3	9·9	9·8	1·0	94·0	0·0	17·2	17·3	3·0	94·2
		3	−0·2	9·4	9·3	0·9	93·7	−0·2	16·4	16·4	2·7	93·9
0·8	Log rank	1	−0·2	11·5	11·5	1·3	94·2	0·0	19·8	19·6	3·8	93·8
		2	−0·4	9·1	9·3	0·9	93·4	0·1	16·0	16·6	2·7	93·2
		3	−0·4	8·2	8·4	0·7	94·4	−0·7	14·3	14·7	2·2	91·9
	Gehan	1	0·0	10·8	10·7	1·1	94·8	0·0	18·5	18·4	3·4	94·8
		2	−0·3	7·7	7·9	0·6	93·3	0·0	13·8	13·8	1·9	95·2
		3	−0·2	6·9	6·9	0·5	94·2	−0·4	12·0	12·1	1·5	94·2
	Wilcoxon	1	−0·1	10·8	10·7	1·1	94·8	0·1	18·5	18·3	3·3	95·6
		2	−0·3	7·8	7·9	0·6	94·1	0·0	13·9	14·0	2·0	94·2
		3	−0·2	6·9	7·0	0·5	93·9	−0·4	12·1	12·2	1·5	93·5

SE, average of estimated standard errors; SD, standard deviation; MSE, mean squared error; CR, 95% coverage rate; m, number of basis matrices.

Table 1 were based on the correct basis matrices, namely $C_{(1)} = I_4$ and $C_{(2)} = 1_4 1_4^{\mathrm{T}} - I_4$, while those in Table 2 were estimated using $C_{(1)} = I_4$ and $C_{(2)}$ consisting of two main off-diagonals of 1 and 0 elsewhere. Therefore, the efficiency gain with $m = 2$ in Table 2 is relatively smaller than that in Table 1. Once the third basis matrix is added in Table 2, we can see some additional efficiency gain.

To examine the robustness of our method, when $K = 4$, we simulated the errors from a heavier-tailed distribution: a multivariate t-distribution with three degrees of freedom, and covariance matrix $3I_4 + 3\rho(1_4 1_4^{\mathrm{T}} - I_4)$, $\rho = 0{\cdot}5$ and $0{\cdot}8$. The results summarized in Table 3 show good performance of our method, and are comparable to those with normal errors in Table 1. We conclude that the estimation procedure based on the generalized method of moments with $m > 1$ can enhance estimation efficiency over the working independence model with $m = 1$. When data are highly correlated and lightly censored, the efficiency gain can be substantial. However, because of possible redundancy, adding more basis matrices may not have further impact on the efficiency gain, after a certain number of basis matrices are included. In the numerical studies with $K = 4$, adding the third basis matrix still improves the efficiency, but only to a small extent.

The proposed method also performed well with varying cluster sizes. We also examined the limiting chi-squared distribution for the quadratic function with finite-sample sizes. In the case with $K = 2$, $\rho = 0{\cdot}5$, $n = 200$ and 25% censoring, we computed the quadratic functions at the true value β_0 and the estimator $\hat{\beta}$. We took $m = 2$ with $C_{(1)} = I_2$ and $C_{(2)} = 1_2 1_2^{\mathrm{T}} - I_2$. Quantile-quantile plots of the observed quantiles of $nQ_n(\beta_0)$ and $n\{Q_n(\beta_0) - Q_n(\hat{\beta})\}$ against the theoretical quantiles from $\chi^2_{(4)}$ and $\chi^2_{(2)}$, respectively, showed the desired linear patterns.

Table 3. *Simulation study of generalized method of moments rank estimation with* 25% *censoring, cluster size* $K = 4$*, where the errors are generated from a multivariate t-distribution. The basis matrices are* $C_{(1)} = I_4$ *and* $C_{(2)} = 1_4 1_4^{\mathrm{T}} - I_4$

			β_1					β_2				
ρ	Method	m	Bias	SE	SD	MSE	CR	Bias	SE	SD	MSE	CR
					($\times 10^{-2}$)					($\times 10^{-2}$)		
0·5	Log rank	1	0·5	14·8	14·7	2·2	94·3	1·5	25·3	26·3	7·0	93·4
		2	0·3	12·8	13·3	1·8	93·1	1·2	22·2	23·1	5·3	93·5
	Gehan	1	0·4	13·4	13·1	1·7	93·9	1·7	22·9	23·3	5·4	94·8
		2	0·2	11·3	11·2	1·3	95·2	1·1	19·4	19·6	3·8	93·6
	Wilcoxon	1	0·4	13·2	13·0	1·7	94·7	1·6	22·9	23·3	5·5	94·1
		2	0·3	11·2	11·3	1·3	94·6	1·1	19·5	19·6	3·9	94·1
0·8	Log rank	1	0·6	14·8	14·5	2·1	94·8	1·2	25·3	26·0	6·7	93·8
		2	0·2	10·2	10·8	1·2	92·6	0·6	17·6	18·2	3·3	92·8
	Gehan	1	0·6	13·3	12·7	1·6	95·2	1·6	22·8	22·8	5·2	94·2
		2	0·3	8·1	8·2	0·7	94·2	0·9	14·1	13·8	1·9	95·0
	Wilcoxon	1	0·6	13·1	12·6	1·6	94·4	1·5	22·5	22·7	5·2	93·6
		2	0·3	8·2	8·4	0·7	94·4	0·8	14·3	14·1	2·0	94·5

SE, average of estimated standard errors; SD, standard deviation; MSE, mean squared error; CR, 95% coverage rate; m, number of basis matrices.

5. Example

We applied our method to data from a study which was conducted by the Diabetic Retinopathy Study Research Group (1985). The primary objective of the study was to determine whether or not laser photocoagulation would help prevent severe visual loss from proliferative diabetic retinopathy. The outcome of interest was time to onset of blindness from the initiation of treatment, recorded in months. The 197 patients in this analysis represented a 50% simple random sample of the patients with high-risk diabetic retinopathy. Each patient had one eye randomized to laser treatment, while the other eye received no treatment. The prognostic factors included treatment, equal to 1 if treated and 0 if untreated, diabetic type, equal to 1 for adult diabetes and 0 for juvenile, age at diagnosis of diabetes and the risk of each eye. The risk scores could be different for the left and right eyes of each patient, ranging from 0·5 to 1 after being divided by 12. To assess the laser treatment effect while adjusting for other covariates, we fitted the marginal accelerated failure time model and implemented the generalized method of moments estimation procedure using one and two basis matrices, respectively. We took $C_{(1)} = I_2$, and $C_{(2)} = 1_2 1_2^{\mathrm{T}} - I_2$. We explored three different weight functions $\phi(u, \beta)$, corresponding to the log-rank, Gehan and Wilcoxon methods. Figure 1 shows different patterns for these three weight functions evaluated at $\hat{\beta}$ based on two basis matrices: both the Gehan and Wilcoxon methods give relatively smaller weights to the larger observations; the Gehan statistic down-weights the long-term survival times the most; and the log-rank method treats all the data equally. Table 4 shows that the treatment and eye-specific risk appeared to be statistically significant. Laser photocoagulation significantly prolonged time to onset of blindness, while the higher eye risk was associated with shorter vision survival. Patient age and diabetic type did not have much influence on the time to blindness. The generalized method of moments with two basis matrices yielded similar point estimates, but smaller variances, compared to those with a single-identity basis matrix. The estimates and inferences are consistent across the three different weight functions and the generalized method of moments procedures with either one or two basis matrices.

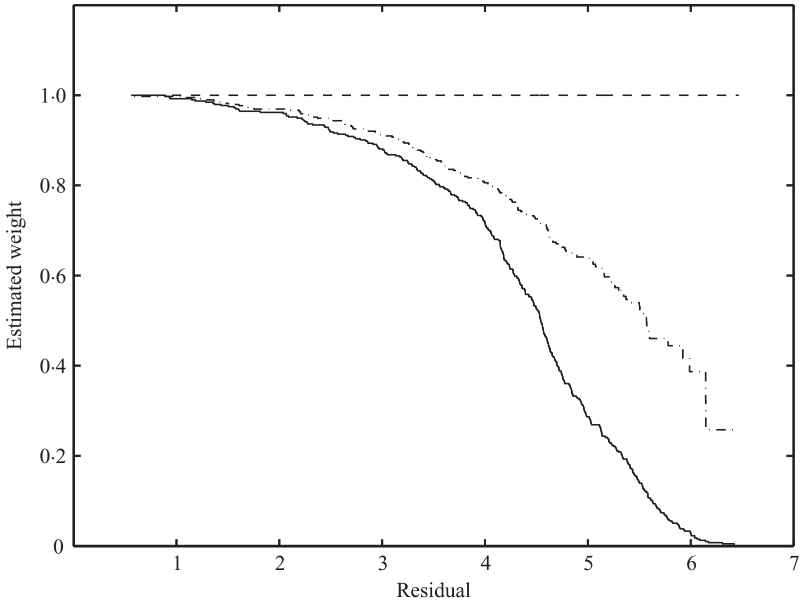

Fig. 1. Diabetic retinopathy study data. Plots of three different weight functions versus the estimated residuals, log rank (dashed), Wilcoxon (dash-dotted) and Gehan (solid).

Table 4. *Diabetic retinopathy study. Generalized method of moments rank estimates of covariate effects under the marginal accelerated failure time model with one and two basis matrices*

		Log rank			Gehan			Wilcoxon		
m	Covariate	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
1	Treatment	1·094	0·203	<0·001	0·987	0·219	<0·001	1·058	0·171	<0·001
	Diabetic type	0·189	0·396	0·633	0·340	0·415	0·413	0·236	0·382	0·537
	Age	−0·196	0·224	0·382	−0·164	0·243	0·500	−0·149	0·189	0·431
	Risk group	−2·090	0·908	0·021	−2·547	0·979	0·009	−2·466	0·860	0·004
2	Treatment	1·085	0·172	<0·001	1·053	0·202	<0·001	1·008	0·286	<0·001
	Diabetic type	0·163	0·389	0·675	0·272	0·469	0·562	0·215	0·392	0·583
	Age	−0·171	0·190	0·368	−0·124	0·228	0·587	−0·144	0·201	0·474
	Risk group	−2·323	0·953	0·015	−2·195	0·798	0·006	−2·244	0·828	0·007

SE, estimated standard error; m, number of basis matrices.

6. DISCUSSION

The generalized method of moments framework is not limited to linear regression models with multivariate censored data; it is readily applicable to other hazard-based survival models. The method is very attractive in situations in which it is difficult to obtain the likelihood, but where martingale-based moments are easier to construct. We can select the basis matrices by examining the reduction of the standard errors of the parameter estimates. If adding an extra basis matrix does not improve the estimation precision, this basis matrix may be redundant and thus can be disregarded. It would be interesting to allow the correlation matrix $R(\alpha)$ to be time-dependent, which would then require the basis matrices to depend on time.

Acknowledgement

We thank Professor D. M. Titterington, an associate editor and two referees for their insightful comments, which led to substantial improvements in this paper. We also thank Dr. Annie Qu for helpful discussions. This research was done while Hui Li was visiting the Department of Biostatistics at the M. D. Anderson Cancer Center.

Appendix

Technical details

We assume the following regularity conditions throughout the derivations.

Assumption A1. There exists $\xi > 0$, such that $\text{pr}\{\log(X_{ik}) - \beta_0^{\tau} Z_{ik} > \tau + \xi\} > 0$, for all i and k.

Assumption A2. The density of the error is bounded and continuous, and censoring times and covariates Z_{ik} are bounded.

Assumption A3. For $i = 1, \ldots, n$ and $l = 1, \ldots, m$, $W_i^{(l)}(u, \beta)$ is uniformly bounded.

Assumption A4. There exists an integrable function $g(u)$, such that the marginal baseline hazard function $\lambda_0(u)$ of the error satisfies

$$\left| \lambda_0(u + \epsilon) - \lambda_0(u) - \epsilon \frac{d\lambda_0(u)}{du} \right| \leqslant \epsilon^2 g(u),$$

for $u < \tau$, and $|\epsilon| \leqslant \xi$.

Assumption A5. There is a matrix $\mathcal{A}_{(l)}(u, \beta)$, for $l = 1, \ldots, m$, such that, in a neighbourhood $\mathcal{B}$ of β_0,

$$\sup_{\beta \in \mathcal{B}, u \leqslant \tau + \xi} \left\| A_{(l)}(u, \beta) - \mathcal{A}_{(l)}(u, \beta) \right\| \rightarrow 0,$$

in probability, and $\mathcal{G}_{(l)}(\beta_0) = \int_{-\infty}^{\tau} \mathcal{A}_{(l)}(u, \beta_0)\, d\lambda_0(u)$ is nonsingular.

Assumption A6. For $l = 1, \ldots, m$, there exists a continuous function $\bar{\mu}^{(l)}(u, \beta)$, such that

$$\sup_{\|\beta - \beta_0\| < \delta/\surd n} n^{-1/2} \left\| \sum_{i=1}^{n} \sum_{k=1}^{K_i} \int_{-\infty}^{\tau} \{\bar{a}^{(l)}(u, \beta) - \bar{\mu}^{(l)}(u, \beta)\}\, dM_{ik}(u, \beta) \right\| = o_p(1),$$

for any $\delta > 0$.

Assumption A7. There exists a deterministic positive definite matrix $\Sigma(\beta_0)$, such that, in probability,

$$\sup_{\|\beta - \beta_0\| < \delta/\surd n} \|\Sigma_n(\beta) - \Sigma(\beta_0)\| \rightarrow 0.$$

Most of these conditions are similar to those given by Tsiatis (1990) and Gray (2003).

Proof of Theorem 1. Without loss of generality, we consider the lth component of $S_n(\beta)$,

$$S_{(l)}(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K_i} \int_{-\infty}^{\tau} \left\{ a_{ik}^{(l)}(u, \beta) - \bar{a}^{(l)}(u, \beta) \right\} dM_{ik}(u, \beta).$$

Along the lines of similar arguments in Tsiatis (1990) and Gray (2003), $S_{(l)}(\beta)$ is asymptotically linear in a small neighbourhood of the true parameter β_0. For any $\delta > 0$,

$$\sup_{\|\beta - \beta_0\| < \delta/\surd n} n^{1/2} \|S_{(l)}(\beta) - S_{(l)}(\beta_0) - G_{(l)}(\beta_0)(\beta - \beta_0)\| = o_p(1).$$

Therefore, we have

$$n^{1/2}S_n(\beta) = n^{1/2}\begin{pmatrix} S_{(1)}(\beta) \\ \vdots \\ S_{(m)}(\beta) \end{pmatrix} = n^{1/2}\begin{pmatrix} S_{(1)}(\beta_0) + G_{(1)}(\beta_0)(\beta-\beta_0) \\ \vdots \\ S_{(m)}(\beta_0) + G_{(m)}(\beta_0)(\beta-\beta_0) \end{pmatrix} + o_p(1).$$

Following this route, we define a function $S_n^*(\beta)$ as a linear expansion around β_0, $S_n^*(\beta) = S_n(\beta_0) + G_n(\beta_0)(\beta-\beta_0)$, so that $S_n(\beta)$ is asymptotically equivalent to $S_n^*(\beta)$. We can write $S_n^*(\beta) = n^{-1}\sum_{i=1}^n U_i^*(\beta)$, where

$$U_i^*(\beta) = U_i(\beta_0) + \begin{pmatrix} \sum_{k=1}^{K_i}\int_{-\infty}^{\tau} Y_{ik}(u,\beta_0)\{a_{ik}^{(1)}(u,\beta_0) - \bar{a}^{(1)}(u,\beta_0)\}Z_{ik}^{\mathrm{T}} d\lambda_0(u)\,(\beta-\beta_0) \\ \vdots \\ \sum_{k=1}^{K_i}\int_{-\infty}^{\tau} Y_{ik}(u,\beta_0)\{a_{ik}^{(m)}(u,\beta_0) - \bar{a}^{(m)}(u,\beta_0)\}Z_{ik}^{\mathrm{T}} d\lambda_0(u)\,(\beta-\beta_0) \end{pmatrix}.$$

Furthermore, let $Q_n^*(\beta) = \{S_n^*(\beta)\}^{\mathrm{T}}\{\Sigma_n^*(\beta)\}^{-1}S_n^*(\beta)$, where $\Sigma_n^*(\beta) = n^{-1}\sum_{i=1}^n U_i^*(\beta)\{U_i^*(\beta)\}^{\mathrm{T}} - S_n^*(\beta)\{S_n^*(\beta)\}^{\mathrm{T}}$. We then have that

$$\sup_{\|\beta-\beta_0\|<\delta/\sqrt{n}} \left\| Q_n(\beta) - Q_n^*(\beta)\right\| = o_p(n^{-1}).$$

Since $\hat\beta$ minimizes $Q_n(\beta)$, and equivalently $Q_n^*(\beta)$, and by Assumption A7, we obtain that

$$\frac{\partial Q_n^*(\beta)}{\partial\beta} = 2\{G_n^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)G_n(\beta_0)\}(\beta-\beta_0) + 2G_n^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)S_n(\beta_0) + o_p(n^{-1/2}),$$

$$\frac{\partial^2 Q_n^*(\beta)}{\partial\beta\partial\beta^{\mathrm{T}}} = 2G_n^{\mathrm{T}}(\beta_0)\Sigma^{-1}(\beta_0)G_n(\beta_0) + o_p(1).$$

The second derivative matrix is asymptotically positive definite, which guarantees a unique minimum. Since $\hat\beta$ satisfies $\partial Q_n^*(\beta)/\partial\beta\|_{\hat\beta} = 0$ and $Q_n(\beta_0) = Q_n^*(\beta_0)$, and by the continuity of $\partial Q_n^*(\beta)/\partial\beta$ at β, $\hat\beta$ converges to β_0 in probability, as $n\to\infty$. □

Proof of Theorem 2. Recall that $a_{ik}^{(l)}(u,\beta_0) = \sum_{j=1}^{K_i} Z_{ij}w_{ijk}^{(l)}(u,\beta_0)$. We define $\mu_{ik}^{(l)}(u,\beta_0)$ the same as $a_{ik}^{(l)}(u,\beta_0)$ except for replacing $w_{ijk}^{(l)}(u,\beta_0)$ by its limit. Since $\phi(u,\beta)$ satisfies Condition 5 in Ying (1993) and the cluster size K_i is bounded, it follows from the Skorohod strong embedding theorem (Shorack & Wellner, 1986, § 2.5) and similar arguments in Lee et al. (1993) that $S_{(l)}(\beta_0)$ is asymptotically equivalent to

$$S_{(l)}^{\dagger}(\beta_0) = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\int_{-\infty}^{\tau}\{\mu_{ik}^{(l)}(u,\beta_0) - \bar\mu^{(l)}(u,\beta_0)\}\,dM_{ik}(u,\beta_0).$$

Therefore, by the multivariate central limit theorem for a sum of independent terms, we can show that $n^{1/2}S_{(l)}^{\dagger}(\beta_0)$, and thus $n^{1/2}S_{(l)}(\beta_0)$, converge to a zero-mean normal distribution with a variance-covariance matrix asymptotically equivalent to

$$n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K_i}\sum_{j=1}^{K_i}\zeta_{ik}^{(l)}(\beta_0)\{\zeta_{ij}^{(l)}(\beta_0)\}^{\mathrm{T}},$$

where

$$\zeta_{ik}^{(l)}(\beta_0) = \int_{-\infty}^{\tau}\{\mu_{ik}^{(l)}(u,\beta_0) - \bar\mu^{(l)}(u,\beta_0)\}\,dM_{ik}(u,\beta_0).$$

A consistent estimator for the covariance matrix can be obtained if we replace the true quantities in $\zeta_{ik}^{(l)}(\beta_0)$ with their empirical counterparts. Likewise, $n^{1/2}S_n(\beta_0)$ converges to a zero-mean normal distribution with the variance-covariance matrix $\Sigma(\beta_0)$; that is, in distribution, $n^{1/2}S_n(\beta_0)\to N\{0, \Sigma(\beta_0)\}$. Finally, noting Assumption A5 and by Slutsky's theorem, we have that, in distribution,

$$n^{1/2}(\hat\beta - \beta_0)\to N\left[0, \{\mathcal{G}^{\mathrm T}(\beta_0)\Sigma^{-1}(\beta_0)\mathcal{G}(\beta_0)\}^{-1}\right].$$

□

Proof of Theorem 3. In a small neighbourhood of β_0, $Q_n(\beta)$ and $Q_n^*(\beta)$ are asymptotically equivalent. It follows that

$$Q_n^*(\gamma_0, \tilde\theta) - Q_n^*(\hat\gamma, \hat\theta) = (\tilde\theta - \theta_0)^{\mathrm T} J_{\theta\theta}(\tilde\theta - \theta_0) - 2S_n^{\mathrm T}(\beta_0)\Sigma^{-1}(\beta_0)G_n(\beta_0)\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \tilde\theta\end{pmatrix}$$
$$-\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix}^{\mathrm T}\begin{pmatrix}J_{\gamma\gamma} & J_{\gamma\theta}\\ J_{\theta\gamma} & J_{\theta\theta}\end{pmatrix}\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix} + o_p(n^{-1}).$$

Furthermore, both $\partial Q_n^*(\hat\gamma, \hat\theta)/\partial\theta$ and $\partial Q_n^*(\gamma_0, \tilde\theta)/\partial\theta$ are approximately zero, since $Q_n(\gamma, \theta)$ and $Q_n(\gamma_0, \theta)$ take their minimum values at $(\hat\gamma, \hat\theta)$ and $(\gamma_0, \tilde\theta)$, respectively. Thus, the second item on the right-hand side of the above equation is equivalent to

$$2\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix}^{\mathrm T}\begin{pmatrix}J_{\gamma\gamma} & J_{\gamma\theta}\\ J_{\theta\gamma} & J_{\theta\theta}\end{pmatrix}\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix} - 2(\tilde\theta - \theta_0)^{\mathrm T} J_{\theta\theta}(\tilde\theta - \theta_0) + o_p(n^{-1}).$$

Therefore,

$$Q_n^*(\gamma_0, \tilde\theta) - Q_n^*(\hat\gamma, \hat\theta) = \begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix}^{\mathrm T}\begin{pmatrix}J_{\gamma\gamma} & J_{\gamma\theta}\\ J_{\theta\gamma} & J_{\theta\theta}\end{pmatrix}\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix}$$
$$-(\tilde\theta - \theta_0)^{\mathrm T} J_{\theta\theta}(\tilde\theta - \theta_0) + o_p(n^{-1}).$$

From the Taylor-series expansion of $\partial Q_n^*(\gamma_0, \tilde\theta)/\partial\theta$ and $\partial Q_n^*(\hat\gamma, \hat\theta)/\partial\theta$ around θ_0 and (γ_0, θ_0), respectively, and after some matrix algebra, we have that $\tilde\theta - \theta_0 = J_{\theta\theta}^{-1}J_{\theta\gamma}(\hat\gamma - \gamma_0) + (\hat\theta - \theta_0) + o_p(n^{-1/2})$. Therefore, $n\{Q_n^*(\gamma_0, \tilde\theta) - Q_n^*(\hat\gamma, \hat\theta)\}$ is asymptotically equivalent to $n(\hat\gamma - \gamma_0)^{\mathrm T}(J_{\gamma\gamma} - J_{\gamma\theta}J_{\theta\theta}^{-1}J_{\theta\gamma})(\hat\gamma - \gamma_0)$. By the conclusion of Theorem 2, we have that, in distribution,

$$n^{1/2}\begin{pmatrix}\hat\gamma - \gamma_0\\ \hat\theta - \theta_0\end{pmatrix}\to N\left\{\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}J_{\gamma\gamma} & J_{\gamma\theta}\\ J_{\theta\gamma} & J_{\theta\theta}\end{pmatrix}^{-1}\right\}.$$

Thus, in distribution, $n^{1/2}(\hat\gamma - \gamma_0)\to N\{0, (J_{\gamma\gamma} - J_{\gamma\theta}J_{\theta\theta}^{-1}J_{\theta\gamma})^{-1}\}$, which implies the limiting chi-squared distribution of $n\{Q_n^*(\gamma_0, \tilde\theta) - Q_n^*(\hat\gamma, \hat\theta)\}$ under H_0. Under $H_1 : \gamma = \gamma_0 + n^{-1/2}\epsilon$, the limiting normal distribution of $n^{1/2}\{(\hat\gamma - \gamma_0)^{\mathrm T}, (\hat\theta - \theta_0)^{\mathrm T}\}^{\mathrm T}$ is not zero-mean, and this leads to a noncentral chi-squared distribution. □

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore, MD: Johns Hopkins University Press.

Buckley, J. & James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36.

Cai, J. & Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82**, 151–64.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Economet.* **34**, 305–34.

Cox, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

Diabetic Retinopathy Study Research Group (1985). Diabetic retinopathy study. *Investig. Ophthal. Visual Sci.* **21**, 149–226.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203–23.

Gray, R. J. (2003). Weighted estimating equations for linear regression analysis of clustered failure time data. *Lifetime Data Anal.* **9**, 123–38.

HALL, A. R. (2005). *Generalized Method of Moments*. Oxford: Oxford University Press.
HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–54.
HANSEN, L. P., HEATON, J. & YARON, A. (1996). Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Statist.* **14**, 262–80.
HUANG, Y. (2002). Calibration regression of censored lifetime medical cost. *J. Am. Statist. Assoc.* **97**, 318–27.
JIN, Z., LIN, D. Y., WEI, L. J. & YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.
JIN, Z., LIN, D. Y. & YING, Z. (2006). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scand. J. Statist.* **33**, 1–23.
JIN, Z., YING, Z. & WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381–90.
LAI, T. L. & SMALL, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method of moments approach. *J. R. Statist. Soc.* B **69**, 79–99.
LEE, E. W., WEI, L. J. & AMATO, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis*: *State of the Art*, Ed. J. P. KLEIN and P. K. GOEL, pp. 237–47. Dordrecht: Kluwer.
LEE, E. W., WEI, L. J. & YING, Z. (1993). Linear regression analysis for highly stratified failure time data. *J. Am. Statist. Assoc.* **88**, 557–65.
LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
LIN, J. S. & WEI, L. J. (1992). Linear regression analysis for multivariate failure time observations. *J. Am. Statist. Assoc.* **87**, 1091–7.
MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
NELDER, J. A. & MEAD, R. (1965). A simplex method for function minimization. *Comp. J.* **7**, 308–13.
NEWEY, W. K. (2004). Efficient semiparametric estimation via moment restrictions. *Econometrica* **72**, 1877–97.
PAKES, A. & POLLARD, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027–57.
PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.
PETO, R. & PETO, J. (1972). Asymptotically efficient rank invariant test procedures (with Discussion). *J. R. Statist. Soc.* A **135**, 185–206.
PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–80.
PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1989). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge University Press.
QU, A., LINDSAY, B. G. & LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–36.
SHORACK, G. R. & WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: John Wiley.
TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–72.
WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.
YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76–99.

[*Received November* 2006. *Revised September* 2008]