Incorporating Correlation for Multivariate Failure Time Data When Cluster Size Is Large

L. Xue,∗∗∗ L. Wang,∗ and A. Qu∗∗∗

1Department of Statistics, Oregon State University, Corvallis, Oregon 97330, U.S.A.
2Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820, U.S.A.

∗email: xnel@stat.oregonstate.edu
∗∗email: wangli@stat.oregonstate.edu
∗∗∗email: anniequ@illinois.edu

SUMMARY. We propose a new estimation method for multivariate failure time data using the quadratic inference function (QIF) approach. The proposed method efficiently incorporates within-cluster correlations. Therefore, it is more efficient than those that ignore within-cluster correlation. Furthermore, the proposed method is easy to implement. Unlike the weighted estimating equations in Cai and Prentice (1995, *Biometrika* 82, 151–164), it is not necessary to explicitly estimate the correlation parameters. This simplification is particularly useful in analyzing data with large cluster size where it is difficult to estimate intracluster correlation. Under certain regularity conditions, we show the consistency and asymptotic normality of the proposed QIF estimators. A chi-squared test is also developed for hypothesis testing. We conduct extensive Monte Carlo simulation studies to assess the finite sample performance of the proposed methods. We also illustrate the proposed methods by analyzing primary biliary cirrhosis (PBC) data.

KEY WORDS: Chi-squared test; Correlated failure times; Cox’s model; Generalized estimating equation; Marginal hazard rate; Quadratic inference function.

1. Introduction

Multivariate survival data arise frequently in clinical trials and epidemiological studies where patients are monitored for multiple biological endpoints throughout a follow-up period, or a single biological endpoint is followed for members within the same family. In both cases, survival times of the same cluster are likely to be correlated. Incorporating intracluster correlation is essential to achieving correct inference on the effects of treatments or covariates on the survival times. The Cox proportional hazard model (Cox, 1972) has been studied extensively to analyze univariate survival time data. However, it is still challenging to apply the Cox model in multivariate survival data since it is not straightforward to estimate and take into account intracluster correlation without specific model assumptions. Two major classes of models have been proposed for multivariate survival data: frailty models (Clayton and Cuzick, 1985; Nielsen et al., 1992; Glidden and Self, 1999; Hougaard, 2000; Therneau and Grambsch, 2000; Gorfine, Zucker, and Hsu, 2006) and marginal hazard models (Wei, Lin, and Weissfeld, 1989; Lee, Wei, and Amato, 1992; Cai and Prentice, 1995, 1997; Prentice and Hsu, 1997; Spiekerman and Lin, 1998). In addition, extensions beyond the Cox proportional hazard model can also be found in Cai, Wei, and Wilcox (2000), and Cai, Cheng, and Wei (2002).

In this article, we are especially interested in marginal hazard model approaches, where the marginal hazard of failure for individuals within a cluster is specified by the Cox proportional hazard model while the intracluster correlation is left unspecified. The marginal hazard models are appropriate for finding population-averaged covariate effects when the intracluster correlation is not of interest. Wei et al. (1989), Lee et al. (1992), and Spiekerman and Lin (1998) provide consistent estimators of the hazard ratio parameters by maximizing the quasi-partial likelihood. However, they ignore intracluster dependence in the estimation procedures, and only adjust for correlation in the inference step by applying a robust sandwich variance estimator. Although their approaches are easy to be carried out, these estimators are not efficient since the correlation information is not incorporated in the estimation.

To improve the efficiency of estimation for multivariate failure time data, Cai and Prentice (1995, 1997) propose an estimation procedure by incorporating correlation explicitly into the partial likelihood score equation. Their approach is motivated by applying the generalized estimating equation (GEE) of Liang and Zeger (1986) into survival data settings, and adding a weighting matrix based on the inverse correlation matrix of the marginal martingales. The corresponding estimator is consistent and asymptotically normal. Their approach performs well numerically and provides a more efficient estimator than using independent structure when the cluster size is small, such as 2. But when the cluster size is large, it is still challenging to implement their method since the estimation of the weighting correlation matrix in general is rather difficult and extremely computationally intensive. In order to estimate the correlation matrix, it is necessary to either impose strong parametric model assumptions or estimate high-dimensional parameters when the cluster size is large.

© 2009, The International Biometric Society
However, in many studies the subjects are measured over a long period of time and the event of interest can occur a large number of times. For such data, it is infeasible to estimate the weighting matrix, therefore it is difficult to implement Cai and Prentice’s (1995, 1997) weighted estimating equation approach. This motivates us to develop Cai and Prentice’s (1995, 1997) weighted estimating equation in a rather different approach by applying the method of quadratic inference functions (QIFs) for parameter estimation and statistical inference. The QIF was proposed by Qu, Lindsay, and Li (2000) as an extension of GEEs to improve the efficiency of the regression estimator in longitudinal data analysis. Qu et al. (2000) showed that the QIF is more efficient than the GEE under the same misspecified working correlation. The last section provides further discussion. Proofs and technical lemmas are found in the Appendix.

In this article, we will develop a counterpart of the QIF method for correlated multivariate failure time data. The major advantage of our approach is that it improves estimation efficiency without requiring the specification of the correlation formula or estimating the correlation parameters. This advantage becomes more important when the cluster size is large and the specification of the correlation formula might be impossible to obtain when the true information is unknown. Another advantage of the QIF method is that the inference can be easily conducted on the hazard ratio parameters, since it naturally provides a chi-squared inference function for hypothesis testing. Furthermore, a standard error formula is proposed, and shown to be consistent with the empirical standard errors in the simulation studies.

We organize the paper as follows. In Section 2, we introduce basic settings and generalized estimation equations for multivariate survival data. In Section 3, we propose the QIF for multivariate survival data and establish the consistency and asymptotic normality property of the proposed estimators. In Section 4, a standard error formula is given and a chi-squared test is developed for hypothesis testing. Section 5 presents simulation results and application to an empirical example. The last section provides further discussion. Proofs and technical lemmas are found in the Appendix.

2. Generalized Estimating Equations for Marginal Hazard Models

For multivariate survival data, let $T_{ki}$ be the $i$th ($i = 1, \ldots, n_k$) type of failure time of the $k$th cluster ($k = 1, \ldots, K$), and $C_{ki}$ be the potential censoring time for $T_{ki}$. The failure times $T_{ki}$ usually are not completely observable. Instead, one observes bivariate vectors $(X_{ki}, \Delta_{ki})$, where $X_{ki} = \min(T_{ki}, C_{ki})$ and $\Delta_{ki} = I(T_{ki} \leq C_{ki})$, with $I(\cdot)$ being an indicator function.

Let $Z_{ki}(t) = (Z_{ki}^{1}(t), \ldots, Z_{ki}^{p}(t))^T$ be a $p \times 1$ vector of, possibly time varying, covariates for the $k$th cluster with respect to the $i$th type of failure. We assume that failure times from the same cluster might be correlated, while observations from different clusters are assumed to be independent. Furthermore, the failure times $T_{ki} = (T_{k1i}, \ldots, T_{kni})^T$ and the censoring times $C_{ki} = (C_{k1i}, \ldots, C_{kni})^T$ are assumed to be independent, conditional on $Z_{ki}(t) = (Z_{ki}^{1}(t), \ldots, Z_{ki}^{p}(t))^T$. For the $i$th type of failure in the $k$th cluster, Cox’s (1972) marginal proportional hazard model assumes that the hazard function $\lambda_{ki}(t)$ takes the following form:

$$
\lambda_{ki}(t) = Y_{ki}(t)\lambda_0(t)e^{Z_{ki}^T\beta}, \quad k = 1, \ldots, K, \ i = 1, \ldots, n_k,
$$

where $Y_{ki}(t) = I(X_{ki} \geq t)$ is an at-risk indicator process for the $i$th component of the $k$th response vector, $\lambda_0(\cdot)$ is an unspecified baseline hazard that may vary according to different types of failure within each cluster, and $\beta$ is a $p \times 1$ vector of the hazard ratio parameters to be estimated. Without loss of generality, assume $n_1 \geq n_2 \geq \cdots \geq n_K$. Let $n = \max\{n_1, \ldots, n_K\} = n_k$, $K_i$ be the number of individuals for the $i$th failure type for $i = 1, \ldots, n$.

2.1 Generalized Estimating Equations

Cai and Prentice (1995) propose to estimate the hazard ratio parameters $\beta$ in (1) by solving a weighted estimation equation (WEE). Let $N_{ki}(t) = \Delta_{ki}I(X_{ki} \leq t)$. For $i = 1, \ldots, n$, define

$$
\tilde{\Lambda}_0(t) = \int_0^t \left\{ \sum_{i=1}^{K_i} Y_{ki}(u)\lambda_0(u)\right\}^{-1} \sum_{i=1}^{K_i} N_{ki}(du),
$$

which is an estimator of the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ for any given $\beta$. Let $\tilde{M}_k(t) = (\tilde{M}_{ki}(t), \ldots, \tilde{M}_{kni}(t))^T$, where $\tilde{M}_{ki}(t) = N_{ki}(t) - \int_0^t Y_{ki}(u)\lambda_0(u)\lambda_0(u)\tilde{\Lambda}_0(t)du$ is an estimator of the marginal martingale $M_k(t) = N_{ki}(t) - \int_0^t Y_{ki}(u)\lambda_0(u)\lambda_0(u)\tilde{\Lambda}_0(t)du$. Then Cai and Prentice (1995) estimate $\beta$ in (1) by solving the weighted partial-likelihood score equation,

$$
\sum_{k=1}^{K} \int_0^\infty Z_k^T(u)W_k(\beta, u)\tilde{M}_k(du) = 0,
$$

where $W_k(\beta, u) = W_k(\beta) = \text{corr}^{-1}\{M_k(X_k)\}$. The weighting matrix $W_k(\beta)$ is crucial in incorporating intrachannel correlation and improving estimation efficiency. However, the estimation of the weighting matrix is often rather difficult. Parametric methods require us to impose additional assumptions on the joint distributions for each pair of elements $(T_{k1}, T_{k2})$. On the other hand, nonparametric methods avoid additional parametric distributional assumptions, but involve nonparametric estimation of the joint survival function of $(T_{1k}, \ldots, T_{nk})$. This might only work well when the cluster size is small because of the large number of parameter estimation involved.

We propose a new estimation method, which does not require us to estimate the correlation parameters explicitly. This approach is motivated by using the QIF method, which is proposed in Qu et al. (2000). Define $F_{k1} = \sigma\{N_{ki}(u), Y_{ki}(u), Z_{ki}(u) : 0 \leq u < t\}$, and $F_T = \{F_{T1}, \ldots, F_{Tn}, \ldots, F_{T1-K}, \ldots, F_{T1-Kn}\}$. Let $D_k(\beta, u) = (\frac{\partial M_{k1}(u)}{\partial \beta}, \ldots, \frac{\partial M_{kni}(u)}{\partial \beta})^T$, and $V_k(\beta, u) = \text{diag}\{\text{var}\{dM_{k1}(u)|F_T\}, \ldots, \text{var}\{dM_{kni}(u)|F_T\}\}$. Consider the following GEE for multivariate survival data:

$$
\sum_{k=1}^{K} \int_0^\infty D_k^T(\beta, u)V_k^{-1/2}(\beta, u)R_k^{-1}(\alpha)V_k^{1/2}(\beta, u)\tilde{M}_k(du) = 0,
$$

where $R_k(\alpha)$ are working correlation matrices whose common structure is fully specified by a vector of nuisance parameters $\alpha$. Note that $D_k(\beta, u) \approx -A_k(\beta, u)Z_k(u)du$, and $V_k(\beta, u) \approx A_k(\beta, u)Z_k(u)du$.
\( A_k(\beta, u)du, \) where \( A_k(\beta, u) = \text{diag}\{\lambda_{k,1}(u), \ldots, \lambda_{k,n_k}(u)\}. \)

Therefore, (4) is simplified as

\[
\sum_{k=1}^{K} \int_0^{\infty} Z_i^T(u) A_i^{-1}(\beta, u) R_i^{-1}(\alpha) A_i^{-1}(\beta, u) \tilde{M}_k(du) = 0. \tag{5}
\]

The GEE (5) is a natural extension of the GEE proposed by Liang and Zeger (1986) for longitudinal data analysis. It is derived by incorporating the covariance of the martingale residual \( dM_i \) of the clustered failure times. Notice that if \( R_i(\alpha) = I_k \), (5) coincides with (3) with an independent working correlation. In general, (3) and (5) are different. The structure of the working correlation matrix \( R_k(\alpha) \) depends on the correlation form of \( dM_i(\beta, u) \); it could be unspecified, or be of some simple form such as identity, equal-correlated, or AR(1).

The diagonal matrix \( A_i(\beta, u) \) in (5) involves the baseline hazard rates \( \{\lambda_{ki}(u)\}_{i=1}^{n_k} \) which need to be estimated in implementation. Following Klein and Moeschberger (1997, p. 153) and Andersen et al. (1993, p. 230), we use a simple kernel smoothed estimator based on \( \Delta \hat{\lambda}_0(t) = \hat{\lambda}_0(t) - \hat{\lambda}_0(t-1) \), where \( \hat{\lambda}_0(t) \) is given in (2). Define

\[
\hat{\lambda}_0(u) = h^{-1} \sum_{k=1}^{K} H \left( \frac{u - X_{ki}}{h_i} \right) \Delta \hat{\lambda}_0(X_{ki}), \tag{6}
\]

where \( H(\cdot) \) is a kernel function with integral 1, and \( h_i \) is the bandwidth. In the implementation, we have used an Epanechnikov kernel and the rule-of-thumb bandwidth, i.e., \( h_i = 1.06\sigma(X_i)K^{-1/5} \), where \( \sigma(X_i) \) is the sample standard deviation of \( X_i = (X_{i1}, \ldots, X_{iK})' \). Since the Epanechnikov kernel is of bounded support, \( \hat{\lambda}_0(u) \) may be 0 at some locations. As a result, \( A_i(\beta, u) \) is not invertible at such locations. If this occurs, we use \( \hat{\lambda}_0(u) = \hat{\lambda}_0(t_{\max})/K_i \), where \( t_{\max} \) is the maximum observed failure time.

### 3. Quadratic Inference Functions for Marginal Hazard Model

In this section, we develop the QIF method for multivariate survival data analysis. The key idea of the QIF is to approximate the inverse of the working correlation matrix in (5) by a linear combination of independent basis matrices. Let \( R(\alpha) \) be the \( n \times n \) working correlation matrix. Then \( R^{-1}(\alpha) = a_1 B_1 + \cdots + a_m B_m \), where \( B_1 = I \) is the identity matrix and \( \{B_m\}_{m=2}^{\infty} \) are symmetric matrices. The advantage of this approach is that it does not require us to estimate the linear coefficients \( a_i \)'s, which can be viewed as nuisance parameters associated with the correlation. The basis matrices \( \{B_m, m = 1, \ldots, m\} \) are chosen to be rich enough to approximate the inverse of the true correlation structure of \( dM_i \). For example, if the working correlation is exchangeable, two basis matrices are needed, \( B_2 \) being 0 on the diagonal and 1 off-diagonal. If the working correlation is AR(1) instead, then three basis matrices are needed with \( B_2 \) being 1 on the sub-diagonals and 0 elsewhere, and \( B_3 \) being 1 on (1, 1) and \( (n, n) \) components and 0 elsewhere. That is, for AR(1) working correlation,

\[
B_2 = \begin{bmatrix}
0 & 1 & \cdots & 0 \\
1 & \ddots & \ddots & \\
\vdots & \ddots & \ddots & 1 \\
0 & \cdots & 0 & 1
\end{bmatrix}_{n \times n}, \quad B_3 = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 0 & \cdots & \vdots \\
\vdots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & 1
\end{bmatrix}_{n \times n}
\]

If there is no prior information on working correlation, Qu and Lindsay (2003) provide an adaptive estimation equation approach to approximate the true correlation empirically, and their approach does not require the inversion of a large dimensional unrestructured correlation matrix. See also Qu and Li (2006) for more details on choice of basis matrices.

If the clustered data are unbalanced, define \( B_ki = \Xi_k'B_i\Xi_k \), where \( k = 1, \ldots, K, i = 1, \ldots, m \), and \( \Xi_k \) is a transformation matrix with dimension \( n \times n_k \) and its \( s, s' \)th components being 1 for \( s = 1, \ldots, n_k \), and 0 elsewhere. The transformation guarantees that the basis matrices for the inverse of working correlation still have the same dimension even though the cluster size could be different for different clusters. The GEE (5) can be viewed as a linear combination of elements of the vector

\[
G_K(\beta) = \frac{1}{K} \sum_{k=1}^{K} g_k(\beta)
\]

where

\[
G_K(\beta) = \left( \int_0^{\infty} Z_i^T(u) A_i^{-1}(\beta, u) R_i^{-1}(\alpha) A_i^{-1}(\beta, u) \tilde{M}_k(du) \right)
\]

\[
= \frac{1}{k} \sum_{k=1}^{K} \left( \int_0^{\infty} Z_i^T(u) A_i^{-1}(\beta, u) R_i^{-1}(\alpha) A_i^{-1}(\beta, u) \tilde{M}_k(du) \right)
\]

\[
\cdot \left( \int_0^{\infty} Z_i^T(u) A_i^{-1}(\beta, u) R_i^{-1}(\alpha) A_i^{-1}(\beta, u) \tilde{M}_k(du) \right)^{-1}
\]

Since there are more equations than unknown parameters, we estimate \( \beta \) by setting \( G_K(\beta) \) as close to zero as possible in the sense of minimizing the QIF,

\[
\hat{\beta}_{QIF} = \arg\min_{\beta} Q_K(\beta) = \arg\min_{\beta} G_K^T(\beta) W_K^{-1}(\beta) G_K(\beta), \tag{8}
\]

where \( W_K = \frac{1}{n^2} \sum_{i,j} g_k(\beta) g_l(\beta) \) and \( KW_K \) is a consistent estimator of \( \text{var}\{g_k(\beta)\} \) for any fixed \( \beta \). The function \( Q_K(\beta) = \hat{G}_K(\beta) W_K^{-1}(\beta) \hat{G}_K(\beta) \) is called the QIF because it also provides an inference function for testing \( \beta \).

To solve for \( \hat{\beta}_{QIF} \) in (8), one can use the Newton–Raphson algorithm. Suppose \( \hat{\beta}^{(i)} \) is an initial value, and \( \hat{\beta}^{(i)} \) is the estimator of \( \beta \) in the \( i \)th step. In the implementation, we have used the estimation equation in (5) with independent working correlation as the initial values. Then,

\[
\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} - \hat{Q}_K^{-1}(\hat{\beta}^{(i)}) \hat{Q}_K(\hat{\beta}^{(i)}), \tag{9}
\]

where \( \hat{Q}_K \) and \( \hat{Q}_K \) are the first- and second-order derivatives of \( Q_K \), respectively. As shown in Qu et al. (2000), asymptotically they can be approximated by \( \hat{Q}_K(\beta) \approx 2\hat{G}_K W_K^T \hat{G}_K \), and \( \hat{Q}_K(\beta) \approx 2\hat{G}_K W_K^T \hat{G}_K \), where \( \hat{G}_K \) is the mp \( \times \) mp matrix \( \{\partial G_K / \partial \beta\} \). Then, \( \hat{\beta}_{QIF} \) is obtained by iterating (9) until
convergence. In the implementation, we have used \(|\hat{\beta}^{(i+1)} - \hat{\beta}^{(i)}| < 10^{-6}\) for convergence, where \(|\cdot|\) is a vector L2 norm.

Now, we present the asymptotic properties of the proposed QIF estimator. Under mild regularity conditions, we have the following theorems.

**Theorem 1.** Let \(\beta_0\) be the true parameters. Under assumptions (A1)–(A6) in the Appendix, the parameter estimator \(\hat{\beta}_{QIF}\) defined in (8) exists and is a consistent estimator of \(\beta_0\).

**Theorem 2.** Under assumptions (A1)–(A6) in the Appendix, the parameter estimator \(\tilde{\beta}_{QIF}\) in (8) is asymptotically normal. That is, \(\sqrt{K}(\beta_{QIF} - \beta_0) \sim N(0, (J_0^{-1} W_0^{-1} J_0)^{-1})\), where \(J_0 = J(\beta_0)\) with \(J(\cdot)\) as given in equation (A.4), and \(W_0 = W(\beta_0)\) as defined in equation (A.1).

4. Inference on Hazard Ratio Parameters

In this section, we discuss a standard error formula and a chi-squared test method for making inferences on the hazard ratio parameters in (1).

The standard errors for the proposed parameter estimates in (8) can be obtained directly using the sandwich formula,

\[
\hat{\text{cov}}(\beta_{QIF}) = \{\hat{Q}_K(\hat{\beta}_{QIF})\}^{-1} \hat{\text{cov}}(\hat{Q}_K(\hat{\beta}_{QIF})) \{\hat{Q}_K(\hat{\beta}_{QIF})\}^{-1}
\]

This can be shown to be a consistent estimator of the asymptotic covariance matrix \(\text{cov}(\beta_{QIF})\) as given in Theorem 2. In the simulation study, we will illustrate the accuracy of this standard error formula for a finite sample.

The proposed QIF also provides a natural way to make inferences about the hazard ratio parameters \(\beta\) in model (1). Consider the partitioning of \(\beta\) such that \(\beta^T = (\gamma^T, \delta^T)\), where \(\beta\) is a \(p \times 1\) vector and \(\gamma\) is a \(q \times 1\) vector of the first \(r\) components of \(\beta\) \((1 \leq r \leq p)\). We are interested in testing \(H_0: \gamma = \gamma_0\) versus \(H_1: \gamma \neq \gamma_0\). Let \(\delta = \text{argmin}_\delta Q_K(\gamma_0, \delta)\), and \((\hat{\gamma}, \hat{\delta}) = \text{argmin}_{\gamma, \delta} Q_K(\gamma, \delta)\). Similar to the log-likelihood function, \(Q_K(\gamma_0, \delta)\) and \(Q_K(\hat{\gamma}, \hat{\delta})\) measure how well the model fits the data under \(H_0\) and \(H_1\), respectively. If \(H_1\) is true, then \(Q_K(\gamma_0, \delta)\) should be systematically larger than \(Q_K(\hat{\gamma}, \hat{\delta})\). Therefore, we consider the following test statistic:

\[
T = Q_K(\gamma_0, \delta) - Q_K(\hat{\gamma}, \hat{\delta}).
\]

**Theorem 3.** Suppose assumptions (A1)–(A6) in the Appendix are satisfied. Under \(H_0\), the test statistic \(T\) asymptotically follows \(\chi_r^2\).

5. Examples

We conduct two simulation studies to assess the finite sample performance of the proposed QIF estimation and inference methods, and demonstrate the proposed methods with an analysis of primary biliary cirrhosis (PBC) data.

\[
\text{Corr} \{M_k(X_{ki}), M_j(X_{kj})\} = \frac{\text{Cov} \{M_k(X_{ki}), M_j(X_{kj})\}}{\sqrt{\text{Var} \{M_k(X_{ki})\} \text{Var} \{M_j(X_{kj})\}}}.
\]

5.1 Simulation Study 1

In this simulation study, 500 data sets were generated, each consisting of \(K = 100\) clusters with \(n = 5\) or 10 observations in each cluster. Failure times were generated from a multivariate extension of the model of Clayton and Cuzick (1985) using an algorithm described in Cai and Shen (2000), in which the joint survival function for \(n\) correlated samples \(T_k = (T_{ki1}, \ldots, T_{kn})^T\) in the \(k\)th cluster is

\[
S(t_1, \ldots, t_n) = Pr(T_{ki1} > t_1, \ldots, T_{kn} > t_n) = \left[\sum_{t_i = 1}^n \exp(-\theta t_i) - (n-1)\right]^{-\theta},
\]

where the parameter \(\theta\) characterizes the degree of dependence of \(T_{ki1}\) and \(T_{kj}\) \((i, j = 1, \ldots, n, i \neq j)\). When \(\theta > 0\), a positive dependence is implied and a decreasing value of \(\theta\) implies an increasing level of positive dependence. This Clayton model is well studied and various methods are available for estimating \(\beta_0\) and \(\theta\) in the literature, such as those proposed in Nielsen et al. (1992); Hsu and Prentice (1996); Prentice and Hsu (1997); Glidden and Self (1999); Glidden (2000); and Gorfine et al. (2006). Here, we focus on estimation of the marginal parameter \(\beta_0\) using the proposed QIF method that does not make use of the complete parametric survival function given in (12).

The relationship between \(\theta\) and Kendall’s \(\tau\) is \(\tau = 1/(2\theta + 1)\). We set \(\theta\) to be 0.25, 0.5, 0.8, 3, corresponding to Kendall’s \(\tau\) of 0.67, 0.38, 0.14 and representing varying degrees of correlation in survival times. A single binary covariate was included in the model with \(Z_{ki}\) being independent and taking value one with probability 0.5. The true scalar parameter was set to be \(\beta_0 = 0\). Censoring times \(C_k = (C_{k1}, \ldots, C_{kn})^T\) were selected independently of each other and of \(T_k, Z_k\), as the smaller of an exponentially distributed variate and a maximal follow-up time. Various choices of exponential censoring parameters were considered to achieve 10%, 50%, and 90% censorship.

We estimated the unknown parameter \(\beta_0\) using the proposed QIF method. The working correlation was taken to be exchangeable. The baseline hazard rates in \(A_k\) were estimated using the kernel smoothing method given in (6). Our simulation studies that are not reported indicate that the proposed baseline hazard rate estimator works reasonably well. Note that the true correlation structure for the Clayton model is exchangeable. To illustrate how different working correlations could affect the QIF estimator, we also considered a QIF using the misspecified AR(1) working correlation. The resulting estimators were denoted as \(\hat{\beta}_{QIF(EC)}\) and \(\hat{\beta}_{QIF(AR(1))}\), respectively. For comparison, we considered an estimating equation with independent working correlation in (3), which ignores within-cluster correlation. The resulting estimator was denoted as \(\hat{\beta}_U\).

We also conducted estimation using the weighted estimating equation proposed in Cai and Prentice (1995). Following Cai and Prentice (1995), the weighting matrix in (3) was set to be the inverse of the true correlation matrix of \(M_k = (M_{k1}(X_{k1}), \ldots, M_{kn}(X_{kn}))^T\), and

\[
\text{Cov} \{M_{ki}(X_{ki}), M_{kj}(X_{kj})\} = \frac{|1 - \exp(-C_k \exp(Z_{ki} \beta))|[1 - \exp(-C_k \exp(Z_{kj} \beta))]^{\theta/2}}{(1 - \exp(-C_k \exp(Z_{ki} \beta)))[1 - \exp(-C_k \exp(Z_{kj} \beta))]^{\theta/2}}.
\]
with
\[
\text{Cov}\{M_{ki}(X_{ki}), M_{kj}(X_{kj})\} \\
= S_{ij}(C_{ki}, C_{kj}, \theta - 1) + \int_0^{C_{ki}} S_{ij}(s, C_{kj}, \theta) \exp(Z_{ki} \beta) ds \\
+ \int_0^{C_{kj}} S_{ij}(C_{ki}, s, \theta) \exp(Z_{kj} \beta) ds \\
+ \int_0^{C_{ki}} \int_0^{C_{kj}} S_{ij}(s_1, s_2, \theta) \exp(Z_{ki} \beta) \exp(Z_{kj} \beta) ds_1 ds_2,
\]
where \(S_{ij}(s_1, s_2, \theta) = S(s, \mathbf{J}_i + s_2 \mathbf{J}_j)\), where \(S(\cdot)\) is defined in (12) and \(\mathbf{J}_i\) is a vector of length \(n\) with the \(i\)th component being 1 and 0 elsewhere, and \(\mathbf{J}_j\) is defined similarly. When implementing this parametric correlation matrix formula, one needs to specify the values for the parameters \(\beta\) and \(\theta\). We used the estimated values by the two-stage method proposed in Gliedden (2000) and implemented by function \texttt{two.stage()} in \texttt{R} package \texttt{timereg}. The resulting estimator was denoted as \(\hat{\beta}_{\text{WEE}}\).

Table 1 summarizes the empirical biases in absolute value and compares the empirical standard errors of \(\hat{\beta}_{\text{QIF(EC)}}, \hat{\beta}_{\text{QIF(AR(1))}}\), and \(\hat{\beta}_{\text{WEE}}\) with \(\beta_0\). To test accuracy of the standard error formula proposed in (10), the average of the estimated standard errors from 500 simulations were reported for two QIF methods. To assess the degree of correlation in the generated failure times after censoring, we calculated the averaged sample Kendall’s \(\hat{\tau} = \frac{1}{\min n^2 - n(n-1)} \sum_{i=1}^{n-1} \sum_{j=1}^{n} \hat{\tau}_ij\), where \(\hat{\tau}_{ij}\) is the sample Kendall’s \(\hat{\tau}\) between \(X_i\) and \(X_j\) with \(X_i = (X_{i1}, \ldots, X_{ik})^T\) in the \(i\)th replication. Table 1 shows that the proposed QIF method performs well in finite sample sizes. The parameter estimates are virtually unbiased. Both QIF methods are more efficient than \(\hat{\beta}_0\) for all values of \(\theta\) when the censoring probability is low or moderate (\(P = 0.1\) or 0.5). A noticeable efficiency gain is observed when \(\theta \geq 0.25\), which corresponds to very high dependency among correlated survival times and is uncommon in real correlated failure time data. But this efficiency gain decreases as \(\theta\) or the censoring probability \(P\) increases due to the reduction in the correlation among marginal martingales. When the censoring probability is high (\(P = 0.9\)), the QIF methods are less efficient than \(\hat{\beta}_0\) in several cases. Similar results are also observed for \(\hat{\beta}_{\text{WEE}}\). The column of \(\hat{\tau}\) in Table 1 shows that the degree of correlation decreases dramatically as censoring probability increases. When the censoring probability is high, the correlation in censored failure times is very low for all values of \(\theta\). All three weighting methods can be less efficient than \(\hat{\beta}_0\) due to extra variability cumulated from incorporating correlation of marginal martingales into estimation.

Furthermore, Table 1 shows that \(\hat{\beta}_{\text{QIF(EC)}}\) is more efficient than \(\hat{\beta}_{\text{QIF(AR(1))}}\) with misspecified working correlation as expected. Overall, the \(\hat{\beta}_{\text{WEE}}\) is more efficient than two QIF methods. This is not surprising since the estimation of \(\hat{\beta}_{\text{WEE}}\) uses the true formula to calculate the correlation matrix, which is often unknown in real data analysis. In the second simulation example, we show \(\hat{\beta}_{\text{WEE}}\) is sensitive to the specification of correlation matrix and it can fail when the working correlation is misspecified.

Finally, we illustrate the performance of the proposed chi-squared test method. We generated data from the same model as the above with \(n = 5, P = 0.1\), and \(\theta = 0.25\). We consider the hypotheses that \(H_0: \beta = 0, H_1: \beta \\neq 0\). We calculated \(\beta\) by minimizing (8) with an exchangeable or AR(1) type working correlation. Since the difference between the number of parameters under \(H_0\) and \(H_1\) is 1, the test statistics \(Q(\hat{\beta}) - Q(0)\) asymptotically follow \(\chi^2\). Figure 1 provides a quantile-quantile plot from 500 replications. It illustrates that under \(H_0\), the empirical quantiles of \(Q(\hat{\beta}) - Q(0)\) follow the theoretical chi-squared quantiles rather well for both exchangeable and AR(1) working correlation. We also examined the power of the proposed chi-squared test when \(\beta\) is different from 0. The powers were evaluated under a sequence of alternatives: \(H_1: \beta = \beta_1\), where \(\beta_1\) is taken to be a grid of equally spaced points in \([0, 0.25]\). When \(\beta_1 = 0\), it collapses into the null hypothesis. Based on 500 replications, Figure 2 plots the power functions with significance level \(\alpha = 0.05\) for both exchangeable and AR(1) working correlation. The powers at \(\beta_1 = 0.048\), and 0.053 for the exchangeable and AR(1) working correlation, respectively. It shows that the proposed chi-squared test gives the right level for testing. The power functions increase rapidly as \(\beta_1\) increases for both exchangeable and AR(1) working correlation, however the test power with correct exchangeable working correlation is uniformly higher than the one with the misspecified AR(1) working correlation.

5.2 Simulation Study 2
We use this simulation study to compare the performance of the proposed QIF method with the WEE when the working correlations in both methods are misspecified. A comparison to the WEE with correct correlation is not considered here since this has been addressed in subsection 5.1. Also it is difficult to derive the exact correlation matrix of the marginal martingales, since joint survival functions are not readily available for the two high-dimensional survival models considered in this subsection. Multivariate survival data are simulated from two multivariate positive stable frailty models. For both models, the cluster size is \(n = 10\), and the number of clusters is \(K = 100\) or 250. The number of replication is 500. We first generate positive stable frailty variates \(\{W_k = (W_{k1}, \ldots, W_{kn})^T\}_{k=1}^K\) independently from one of the following models. We use \(U \sim \text{PS}(\alpha)\) to denote that the random variable \(U\) follows the positive stable distribution with parameter \(\alpha\).

**• Model 1:** \(W_{ki} = (U_{ki})^{1/\alpha} U_{1, ki} \), where \(U_{ik} \sim \text{PS}(\alpha_1)\), \(U_{1, ki} = \cdots = U_{1, k5} = U_1^k\), and \(U_{1, k6} = \cdots = U_{1, kn} = U_2^k\) with \(U_1^k\) and \(U_2^k\) being i.i.d. PS(\(\alpha_2\)), and independent of \(U_{ik}\).

**• Model 2:** \(W_{ki} \sim \text{PS}(\alpha), W_{ki} = (1/2) W_{(k-1)} + \varepsilon_{ki}\), for \(i = 2, \ldots, n\), where \(\sigma = \{1 - (1/2)^n\}^{1/\alpha}\) and \(\{\varepsilon_{ki}\}_{i=2}^n\) are i.i.d. PS(\(\alpha\)).

We have used \(\alpha_1 = 3/4, \alpha_2 = 2/3\), and \(\alpha = 1/2\). As a result, \(W_{ki}\) marginally follows a positive stable distribution with parameter \(\alpha = 1/2\) in both models. But the vector \(\{W_i\}\) have different correlation structure. Failure times are then
Table 1
Summary statistics for estimates of hazard ratio $\beta$ when $K = 100$. |Bias| and SE are absolute value of empirical bias and standard error, and SEE are the means of the estimated standard errors from 500 simulations. RSE1, RSE2, and RSE3 represent the ratios of the empirical standard error of $\hat{\beta}_U$ to $\hat{\beta}_{QIF(\text{EC})}$, $\hat{\beta}_{QIF(\text{AR}(1))}$, and $\hat{\beta}_{WEE}$, respectively.

<table>
<thead>
<tr>
<th>n</th>
<th>P</th>
<th>$\theta$</th>
<th>$\tau$</th>
<th>$\beta_{QIF(\text{EC})}$</th>
<th>$\beta_{QIF(\text{AR}(1))}$</th>
<th>$\beta_{WEE}$</th>
<th>$\hat{\beta}_U$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bias ($\times 10^2$)</td>
<td>SE ($\times 10^2$)</td>
<td>SEE ($\times 10^2$)</td>
<td>RSE1</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td>0.25</td>
<td>0.58</td>
<td>0.37</td>
<td>3.94</td>
<td>3.88</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.33</td>
<td>0.45</td>
<td>6.56</td>
<td>6.52</td>
<td>1.45</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.12</td>
<td>0.67</td>
<td>9.10</td>
<td>9.15</td>
<td>1.06</td>
<td>0.71</td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
<td>0.20</td>
<td>5.22</td>
<td>10.03</td>
<td>10.61</td>
<td>1.27</td>
<td>4.34</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.11</td>
<td>2.58</td>
<td>11.85</td>
<td>11.97</td>
<td>1.08</td>
<td>3.59</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.06</td>
<td>0.47</td>
<td>12.39</td>
<td>12.94</td>
<td>1.04</td>
<td>0.61</td>
</tr>
<tr>
<td>0.9</td>
<td>0.25</td>
<td>0.05</td>
<td>1.99</td>
<td>32.28</td>
<td>34.47</td>
<td>1.01</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.05</td>
<td>0.78</td>
<td>33.37</td>
<td>34.58</td>
<td>0.95</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.05</td>
<td>0.71</td>
<td>33.49</td>
<td>32.98</td>
<td>0.95</td>
<td>0.42</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
<td>0.25</td>
<td>0.59</td>
<td>1.04</td>
<td>2.54</td>
<td>2.55</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.34</td>
<td>0.21</td>
<td>4.29</td>
<td>4.31</td>
<td>1.56</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.13</td>
<td>3.73</td>
<td>5.95</td>
<td>6.06</td>
<td>1.10</td>
<td>0.61</td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
<td>0.21</td>
<td>0.57</td>
<td>6.44</td>
<td>6.25</td>
<td>1.39</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.11</td>
<td>1.14</td>
<td>7.81</td>
<td>7.64</td>
<td>1.16</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.06</td>
<td>0.89</td>
<td>8.72</td>
<td>8.88</td>
<td>1.02</td>
<td>0.85</td>
</tr>
<tr>
<td>0.9</td>
<td>0.25</td>
<td>0.05</td>
<td>0.44</td>
<td>18.89</td>
<td>18.62</td>
<td>1.07</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.05</td>
<td>0.65</td>
<td>19.79</td>
<td>20.52</td>
<td>1.01</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.05</td>
<td>1.77</td>
<td>21.57</td>
<td>21.11</td>
<td>0.97</td>
<td>2.07</td>
</tr>
</tbody>
</table>
randomly generated with the survival function \( S(t_{ki}|W_{ki}) = \exp\{-W_{ki}\exp(\beta_0 Z_{ki})\}^{t_{ki}} \), with \( \beta_0 = 1, \gamma = 2 \), and \( \{Z_{ki}\}_{k,i=1}^{K,n} \) being independent Uniform(0,1) random variables. Therefore, the failure times \( T_{ki} \) marginally follow a Weibull distribution with parameters \( \beta^* = \alpha \beta_0 \) and \( \gamma^* = \alpha \gamma \). The correlation structure of \( T_k = (T_{k1},...,T_{kn})^T \) in two models is different. In Model 1, the first five failure times in a cluster form a subgroup and the last five form the other subgroup. The failure times in the same subgroup have an exchangeable working correlation with Kendall’s \( \tau \) being 0.5. The failure times between subgroups have a weaker correlation with Kendall’s \( \tau \) about 0.25. In Model 2, the Kendall’s \( \tau \) of the failure times are of the form \( \tau(T_{ki},T_{kj}) = \rho^{(|i-j|)/2} \) with \( \rho \approx 0.78 \) for \( i \neq j \).

The censoring times \( C_k = (C_{k1},...,C_{kn})^T \) were selected independently of each other and of \( T_k, Z_k \), as the smaller of an exponentially distributed variate and a maximal follow-up time. The exponential censoring parameters were chosen to achieve 70% censorship for both models.

We consider estimation using the same four methods used in the simulation study 1: \( \hat{\beta}_{QIF(EC)} \), \( \hat{\beta}_{QIF(AR(1))} \), \( \hat{\beta}_{WEE} \), and \( \hat{\beta}_U \). For \( \hat{\beta}_{WEE} \), we use the same formula (13) of the Clayton model to compute the correlation matrix. This Clayton correlation matrix is misspecified for both multivariate frailty models. Table 2 summarizes the bias and compares standard errors of \( \hat{\beta}_{QIF(EC)}, \hat{\beta}_{QIF(AR(1))}, \hat{\beta}_{WEE} \) with \( \hat{\beta}_U \). It shows that both QIF methods are always more efficient than \( \hat{\beta}_U \), even when the working correlation is misspecified. However, \( \hat{\beta}_{WEE} \) with a misspecified working correlation are less efficient than \( \hat{\beta}_U \). It clearly shows that the WEE approach is sensitive to the specification of a correlation matrix. It can fail when the weighting matrix is misspecified. However, the proposed QIF method is more robust to the specification of weighting matrix.

### 5.3 An Empirical Example

In this section, we illustrate an application of the proposed estimation and testing methods in an analysis of PBC data (Therneau and Grambsch, 2000). PBC is a chronic cholestatic liver disease characterized by a progressive destruction of the bile ducts. A randomized double-blind trial was conducted at the Mayo Clinic from 1988 to 1992 to evaluate the efficacy of ursodeoxycholic acid (UDCA) treatment on PBC. Each of the 170 patients was monitored for several endpoints throughout the study period. This analysis focuses on the following four endpoints: liver transplant, histologic progression, appearance of esophageal varices, and death. At the beginning of the study, along with the treatment indicator (RX), three potential risk factors were observed for each patient: histologic stage (STAGE), bilirubin value (BILI), and Mayo PBC risk score (SCORE). A marginal Cox model (1) was used to analyze the effects of treatment and three potential risk factors on the times of the above four endpoints. See Therneau and Grambsch (2000) for more details of the study.

The hazard ratio coefficients were estimated using the proposed QIF method with both exchangeable and AR(1) type
Table 2
Summary statistics for estimates of hazard ratio $\beta$. Bias and SE are empirical bias and standard error from 500 simulations. RSE1, RSE2, and RSE3 represent the ratios of the empirical standard error of $\hat{\beta}_U$ to $\hat{\beta}_{QIF(EC)}$, $\hat{\beta}_{QIF(AR(1))}$, and $\hat{\beta}_{WEE}$, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>n</th>
<th>$\hat{\beta}_{QIF(EC)}$ Bias</th>
<th>RSE1</th>
<th>$\hat{\beta}_{QIF(AR(1))}$ Bias</th>
<th>RSE2</th>
<th>$\hat{\beta}_{WEE}$ Bias</th>
<th>RSE3</th>
<th>$\hat{\beta}_U$ Bias</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>100</td>
<td>0.0052</td>
<td>1.20</td>
<td>0.0054</td>
<td>1.12</td>
<td>0.0741</td>
<td>0.24</td>
<td>0.0023</td>
<td>0.0115</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-0.0006</td>
<td>1.25</td>
<td>0.0021</td>
<td>1.16</td>
<td>-0.0011</td>
<td>0.28</td>
<td>0.0007</td>
<td>0.0046</td>
</tr>
<tr>
<td>Model 2</td>
<td>100</td>
<td>-0.0050</td>
<td>1.05</td>
<td>-0.0043</td>
<td>1.07</td>
<td>-0.0338</td>
<td>0.18</td>
<td>-0.0071</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-0.0007</td>
<td>1.08</td>
<td>-0.0001</td>
<td>1.11</td>
<td>-0.0045</td>
<td>0.21</td>
<td>-0.0003</td>
<td>0.0045</td>
</tr>
</tbody>
</table>

Figure 2. Power function for QIF chi-squared test against $\beta_1$ for hypothesis testing $H_0: \beta = 0, H_1: \beta = \beta_1$. The solid line is power using exchangeable correlation, the dashed line is power using AR(1) correlation.

working correlation. The resulting estimation procedures were denoted as QIF1 and QIF2, respectively. We also fitted the data using the WEE approach, with the working correlation specified by the Clayton model formula (13) and an estimating equation with independent working correlation (IND). Table 3 summarizes the estimated coefficients with standard errors for each method. The standard error of IND is the robust standard error from the grouped jackknife method provided by command coxph in R. The standard error of WEE is estimated from 100 bootstrap samples obtained by resampling the patients. Table 3 shows that the estimates from both QIF methods have smaller standard errors than those from IND, and WEE gives larger standard errors than IND. Note that WEE relies on the Clayton model assumption, which is questionable for the PBC data. To test if treatment and the three risk factors have significant effects on the times of the four endpoints, we calculated p-values from a two-sided $t$-test. For the two QIF methods, the p-values from the proposed chi-squared test were also calculated. The results from all of the methods indicate that treatment (RX) is significant, and two risk factors (STAGE and BILI) are not significant at a significance level of 0.05. Both QIF and WEE found that SCORE has a significant effect at a level of 0.05. However, it is not significant at a level of 0.05 in the IND method. The proposed chi-squared test is similar to the likelihood-ratio test, since the QIF is an analog of minus twice the log likelihood. Therefore, it is asymptotically equivalent to the $t$-test. For small samples, the likelihood-ratio type of test performs better in general than the $t$-test, in the sense that it is more powerful.
6. Discussion

In the analysis of multivariate survival data, incorporation of the intracluster correlation is essential to improve estimation efficiency. However, modeling and estimation of such intracluster correlation for multivariate survival data in general is very challenging because of incomplete follow-up and the possibility of large cluster size. As a result, most existing estimation methods for marginal hazard models fail to accommodate the intracluster correlation. The proposed QIF method improves the efficiency of the hazard ratio estimators without explicitly estimating the intracluster correlation parameters. Instead, it approximates the inverse of the intracluster correlation by a linear combination of basis matrices, and treats the correlation parameters as nuisance parameters. This simplification is particularly useful when the cluster size is large and the intracluster correlation is difficult to estimate. Another advantage of the proposed method is that it provides a statistical inference function for hypothesis testing of the hazard ratio parameters. Furthermore, the test statistic follows a chi-squared distribution asymptotically whether or not the working correlation structure is correctly specified.

ACKNOWLEDGEMENTS

The authors are grateful to many constructive suggestions from two referees, the associate editor, and the co-editor. The work is supported in part by National Science Foundation grant DMS-03-48764.

REFERENCES


**APPENDIX**

**Notation and Conditions**

For notation simplicity, we assume $n_1 = \cdots = n_K = n$ throughout the Appendix. Proof with unequal class size follows similarly. For any $n \times n$ constant matrix $B$, denote $\Phi_i(\beta, u) = \Phi_i(\beta, u, B) = A_i^T(\beta, u)BA_i^{-1}(\beta, u)$. Write $A_i^T(\beta, u) = \text{diag}(a_{i1}(\beta, u), \ldots, a_{iK}(\beta, u))$, $B = (b_{ij})_{j=1}^n$, and $\Phi_i(\beta, u) = (\phi_{ijk}(\beta, u))_{j=1}^n$, with $\phi_{ijk}(\beta, u) = a_{ik}(\beta, u)b_{ij}A_{k-1}(\beta, u)$.

For $1 \leq j \leq n$, let $S^{(d)}(\beta, u) = K^{-1} \sum_{i=1}^n Y_{ij}(u)Z_{ij}(u)^d \exp\{Z_{ij}(u)\beta\}$ for $d = 0, 1, \ldots, K^{-1} \sum_{i=1}^n \sum_{j=1}^n Y_{ij}(u)Z_{ij}(u)^d \exp\{Z_{ij}(u)\beta\}$. Also define $E_j(\beta, u, B) = \frac{S^{(2)}(\beta, u, B)}{S^{(0)}(\beta, u)}$ and $V_j(\beta, u, B) = \frac{S^{(3)}(\beta, u, B)}{S^{(0)}(\beta, u)} - \frac{S^{(2)}(\beta, u, B)}{S^{(0)}(\beta, u)^2}$.

To establish the asymptotic properties of the QIF estimator as stated in Theorems 1 and 2, we need the following regularity conditions. Similar conditions are also given in Cai and Prentice (1995) and Andersen and Gill (1982).

(A1) The range of integrations in (7) are finite, say $[0, 1]$, and $\int_0^1 \lambda_{ij}(t) dt < \infty$, for $j = 1, \ldots, n$.

(A2) The parameter space $S$ is compact, and the true value $\beta_0$ is an interior point of $S$.

(A3) There exist scalar, vector, and matrix functions $s^{(d)}_j(\beta, u)(d = 0, 1)$, $s^{(d)}_{ij}(\beta, u, B)(d = 2, 3, 4)$, such that in probability $\sup_{\beta \in \Delta_1} \|s^{(d)}_j - s^{(d)}_0\| \to 0$, for any constant matrix $B$, and all $j = 1, \ldots, n$.

(A4) Let $s_j^{(d)}(\beta, u)(d = 0, 1), s^{(d)}_{ij}(\beta, u, B)(d = 2, 3, 4)$ be as in Condition (A3). Set $e_j(\beta, u, B) = s_j^{(2)}(\beta, u, B)/s_j^{(0)}(\beta, u)$, and $e_j(\beta, u, B) = s_j^{(3)}(\beta, u, B)/s_j^{(0)}(\beta, u)$.

Then, for all $\beta \in S, u \in [0, 1]$, and $j = 1, \ldots, n$, $s_j^{(1)}(\beta, u) = \partial s_j^{(0)}(\beta, u)/\partial \beta$, $s_j^{(3)}(\beta, u, B) = \partial s_j^{(2)}(\beta, u, B)/\partial \beta - s_j^{(4)}(\beta, u, B)$, and $s_j^{(0)}$ is bounded away from zero on $S \times [0, 1]$.

(A5) There exists a matrix function $w(\cdot, \cdot, \cdot)$, such that for any $n \times n$ constant matrices $B_1, B_2, K^{-1} \sum_{k=1}^K \sum_{i=1}^n \int_0^1 r_{ik}(\beta, u, B_1) r_{ik}(\beta, u, B_2)s_k^{(0)}(\beta, u) \lambda_{0j}(u) du \rightarrow W(\beta, B_1, B_2)$, in probability uniformly for $\beta \in S$, where $r_{ik}(\beta, u, B) = \sum_{j=1}^n Z_{ij}(u)\phi_{ijk}(\beta, u) - e_j(\beta, u, B)$. Furthermore, for any set of basis matrices $\{B_k\}_{k=1}^m$, the matrix $W_0 = W(\beta_0) = \left[ w(\beta_0, B_k, B_{mk}) \right]_{k=1}^m$ is positive definite.

(A6) Let $\Delta_1(B) = \left\{ \sum_{i=1}^n \int_0^1 r_{ij}(\beta, u, B) M_{ij}(du) \right\}$, where $\lambda_{ij}(t) = \lambda_{ij}(t) = \delta_0(t)$ for $t \in [0, 1]$ and $i = 1, \ldots, n$. According to Theorem VII.2.7 in Andersen et. al. (1993), the kernel smoothed baseline hazard

**Proof of Theorem 1**: In the proof, we consider the case where the baseline hazard rates $\{\lambda_{ij}(\cdot)\}_{ij=1}^n$ are known. The same asymptotic results hold when the estimated baseline hazards are consistent: that is, $\hat{\lambda}_{ij}(t) - \lambda_{ij}(t) \to \delta_0(t)$ for $t \in [0, 1]$ and $i = 1, \ldots, n$. According to Theorem VII.2.7 in Andersen et. al. (1993), the kernel smoothed baseline hazard
rates \(\hat{\lambda}_n(\cdot)\) in (6) are consistent under conditions (A1)–(A7) and with the rule-of-thumb bandwidth.

Based on a straightforward extension of Qu and Li (2006), one can show that \(\hat{\beta}_{QH}\) is consistent for \(\beta_0\) provided

(i) the parameter \(\beta\) is identifiable. That is, there is a unique \(\beta_0 \in S\), satisfying \(E[G_K(\beta_0)] = 0\).

(ii) \(E[G_K(\beta)]\) is continuous in \(\beta\), for \(\beta \in S\).

(iii) \(\partial G_K(\beta) / \partial \beta\) exists and is continuous, and it converges in probability to a fixed function \(J(\beta)\) uniformly for \(\beta \in S\).

(iv) \(K W_K(\beta) = \frac{1}{\hat{\gamma}} \sum_{k=1}^K g_k(\beta) g_k^T(\beta)\) converges in probability to a constant matrix \(W(\beta)\) uniformly for \(\beta \in S\).

(v) \(K W_K(\beta_0)\) is positive definite with probability going to one as \(K \to \infty\).

Denote \(h_K(\beta, B) = \frac{1}{K} \sum_{k=1}^K \int_0^1 Z_k(u) A_k^T(\beta, u) B A_k^T(\beta, u) d\hat{M}_k(u)\). Then, one can write \(G_K(\beta) = (h_K^T(\beta, B_1), \ldots, h_K^T(\beta, B_n))^T\). To show conditions (i)–(iii) hold, it is enough to show they hold for each component in \(G_K(\beta)\). Similarly as in Cai and Prentice (1995), one can write

\[
M_k = \frac{1}{K} \sum_{k=1}^K \int_0^1 Z_k(u) \phi_{kij}(\beta, u) - E_j(\beta, u, B) \right) M_{kij}(du).
\]

When \(\beta = \beta_0\), each \(M_{kij}\) is a mean-zero martingale, and each \(Z_k(u), \phi_{kij}(\beta_0, u)\), and \(E_j(\beta_0, u, B)\) are predictable processes. Therefore, \(E[h_K(\beta_0, B)] = 0\), for any constant matrix \(B\). As a result, \(E[\hat{G}_K(\beta_0)] = 0\), and (i) is satisfied. Clearly, condition (ii) is also satisfied, since \(h_K(\beta, B)\) is continuous in \(\beta\) for any fixed \(B\) based on (A.2). Now consider the first-order derivative of \(h_K(\beta, B)\) with respect to \(\beta\). One can write

\[
K \frac{\partial h_K(\beta, B)}{\partial \beta} = \frac{1}{K} \sum_{k=1}^K \int_0^1 H_{kij}(\beta, u, B) M_{kij}(du)
\]

where \(H_{kij}(\beta, u, B) = \sum_{j=1}^n Z_k(u) (\partial \phi_{kij}(\beta, u) / \partial \beta)^T\). Furthermore, as in Cai and Prentice (1995), one has \(\partial h_K(\beta, B) / \partial \beta \to -D(\beta, B)\) in probability, uniformly for \(\beta \in S\). Therefore, \(\partial G_K(\beta) / \partial \beta\) converges in probability to

\[
J(\beta) = \begin{pmatrix}
D(\beta, B_1) \\
\vdots \\
D(\beta, B_n)
\end{pmatrix},
\]

uniformly for \(\beta \in S\). For conditions (iv) and (v), we consider a typical component in \(K W_K\). Let

\[
I_K(\beta, B_1, B_2) = \frac{1}{K} \sum_{k=1}^K \left\{ \int_0^1 Z_k^T(u) A_k^T(\beta, u) B_1 A_k^T(\beta, u) d\hat{M}_k(u) \right\}
\]

\[
\times \left\{ \int_0^1 Z_k^T(u) A_k^T(\beta, u) B_2 A_k^T(\beta, u) d\hat{M}_k(u) \right\}^T.
\]

Let \(R_{kj}(\beta, u, B) = \sum_{j=1}^n Z_k(u) \phi_{kij}(\beta, u) - E_j(\beta, u, B)\) and

\[
I_{kjj'}(\beta, B_1, B_2) = \left\{ \int_0^1 R_{kj}(\beta, u, B_1) M_{kij}(du) \right\}
\]

\[
\times \left\{ \int_0^1 R_{kj}(\beta, u, B_2) M_{kij}(du) \right\}^T.
\]

One can show that

\[
I_K(\beta, B_1, B_2) = \frac{1}{K} \sum_{k=1}^K \left\{ \int_0^1 \sum_{j=1}^n R_{kj}(\beta, u, B_1) M_{kij}(du) \right\}
\]

\[
\times \left\{ \int_0^1 \sum_{j=1}^n R_{kj}(\beta, u, B_2) M_{kij}(du) \right\}^T.
\]

When \(j \neq j'\), the martingales \(M_{kij}\) and \(M_{kij'}\) are orthogonal. Therefore, \(\{I_{kjj'}\}_{j \neq j'}\) are also martingales. By the Lenglart inequality, one can show the second term on the right-hand side of the above equation converges to zero in probability, uniformly for \(\beta \in S\). When \(j = j'\), using the Lenglart inequality, one can show that the first term on the right-hand side of the above equation is asymptotically equivalent to

\[
\frac{1}{K} \sum_{k=1}^K \int_0^1 R_{kj}(\beta, u, B_1) R_{kj}(\beta, u, B_2) d(M_{kij})(u)
\]

\[
= \frac{1}{K} \sum_{k=1}^K \int_0^1 R_{kj}(\beta, u, B_1) R_{kj}(\beta, u, B_2) Y_k(u)
\]

\[
\times \exp \left( \int_0^1 Z_k^T(u) \lambda_{kij}(u) du \right) du.
\]

Therefore, by assumption (A5), \(I_K(\beta, B_1, B_2)\) converges in probability to \(W(\beta, B_1, B_2)\) and \(K W_K(\beta)\) converges in probability to \(W(\beta) = [\hat{w}(\beta, B_1, B_2)]_{k=1}^K\), uniformly for \(\beta \in S\). As a result, (iv) and (v) follow from assumption (A5).
Proof of Theorem 2: We denote \( \dot{Q} \) and \( \ddot{Q} \) to be the first- and second-order derivatives of the QIF \( Q \) with respect to \( \beta \). Similarly to Cai and Prentice (1995), one can use the multivariate central limit theory, e.g., Puri and Sen (1971), to show that,

\[
\sqrt{K} g_K (\beta_0) \rightarrow^d N(0, W_0), \tag{A.5}
\]

where \( W_0 = W(\beta_0) \) as defined in (A.1). Since \( \hat{\beta}_{QIF} \) is obtained by minimizing (8), \( \hat{\beta}_{QIF} \) satisfies \( K^{-1} \dot{Q}(\hat{\beta}_{QIF}) = 0 \). By Taylor expansion, \( 0 = K^{-1} \dot{Q}(\hat{\beta}_{QIF}) = K^{-1} \dot{Q}(\beta_0) + K^{-1} \ddot{Q}(\hat{\beta})(\hat{\beta}_{QIF} - \beta_0) \), where \( \hat{\beta} \) is between \( \beta_0 \) and \( \hat{\beta}_{QIF} \).

Therefore,

\[
\hat{\beta}_{QIF} - \beta_0 = -[K^{-1} \dot{Q}(\hat{\beta})]^{-1}[K^{-1} \dot{Q}(\beta_0)], \tag{A.6}
\]

in which \( K^{-1} \dot{Q}(\hat{\beta}) \) and \( K^{-1} \dot{Q}(\beta_0) \) converges to \( 2J_0^T W_0^{-1} J_0 \) and \( 2J_0^T W_0^{-1} g_K (\beta_0) \) in probability, respectively. Then, one has \( \sqrt{K}(\hat{\beta}_{QIF} - \beta_0) = -\sqrt{K}(J_0^T W_0^{-1} J_0)^{-1}\{J_0^T W_0^{-1} g_K (\beta_0)\} + o_p(1) \). Using (A.5), one has \( \sqrt{K}(\hat{\beta}_{QIF} - \beta_0) \rightarrow^d N(0, (J_0^T W_0^{-1} J_0)^{-1}) \).

Proof of Theorem 3: The proof of Theorem 3 is similar to the proof of Theorem 1 in Qu et al. (2000).