Semiparametric inference for proportional hazards
frailty regression models

Michael R. Kosorok, Ph.D.,
Bee Leng Lee, Ph.D.,
Jason P. Fine, Ph.D.
Semiparametric Inference for Proportional Hazards Frailty Regression Models

Department of Biostatistics and Medical Informatics Technical Report 156

Michael R. Kosorok, Bee Leng Lee, and Jason P. Fine

ABSTRACT

We consider inference for a class of semiparametric regression models which are one parameter extensions of the Cox (1972) model for independent observations. The added flexibility can substantially improve the prediction accuracy of the proportional hazards model. Fully efficient estimators are developed. Uniform consistency and weak convergence are established for all parameters, including the baseline cumulative hazard. A novel Markov chain Monte Carlo technique is useful when profiling over the baseline hazard. It is proved that the bootstrap gives valid inferences for the hazard. Simulations show that the proposals perform well with moderate sample sizes. The procedures are illustrated with a non-Hodgkin's lymphoma dataset.

KEY WORDS: Empirical process; Identifiability; Laplace transform; Nonparametric maximum likelihood; Semiparametric information bound; Unobservable heterogeneity

1Michael R. Kosorok is Associate Professor and Jason P. Fine is Assistant Professor of Statistics and Biostatistics & Medical Informatics at University of Wisconsin at Madison. Bee Leng Lee is Assistant Professor of Statistics and Applied Probability at National University of Singapore, Republic of Singapore. The first author was supported by grant CA75142 from the National Cancer Institute. We thank Robert J. Gray at Harvard University for providing the Non-Hodgkins Lymphoma data. (E-mail: kosorok@biostat.wisc.edu).
1 INTRODUCTION

An objective of many medical studies is a predictive model for survival. Consider the Non-Hodgkin's Lymphoma Prognostic Factors Project (1993) which analyzed data from a collection of international trials. A system was developed to classify patients according to baseline characteristics. The scheme employs a Cox (1972) model with five influential covariates. The ordinal and continuous predictors are dichotomized for clinical interpretation. There are also important risk factors which are omitted, such as treatment center. Diagnostics show that the proportional hazards model fits poorly (Gray, 2000). Furthermore, the survival estimates are quite biased by the misspecification.

There are several alternatives to the Cox model which might improve the fit. These include additive hazards regression models (Aalen, 1978, 1980; Lin and Ying, 1994), accelerated failure time models (Wei, Ying, and Lin, 1990; Tsiatis, 1990), and time-varying coefficient models (Sargent, 1997). Additional models have been developed for covariate-dependent heteroscedasticity and other departures from proportionality (Bagdonavicius and Nikulin, 1999; Hsieh, 2001).

Frailty models are a comparatively parsimonious representation which extend the Cox model in a natural way. The misspecified and omitted covariates are described by an unobservable random variable \( \log(W) \) unique to the linear predictor of each patient. Let \( T \) be the failure time and \( Z = Z(t) \) a \( d \times 1 \) vector of possibly time-dependent covariates. Denote \( \lambda\{t; \tilde{Z}(t), W\} \) as the hazard function of \( T \) conditionally on \( \tilde{Z}(t) = \{Z(s), s \leq t\} \) and \( W \). The proportional hazards frailty regression model is

\[
\lambda\{t; \tilde{Z}(t), W\} = \lambda_0(t) \exp\{\log(W) + \beta^T Z(t)\}
\]

(1)

where \( \beta \) is a \( d \times 1 \) regression parameter and \( \lambda_0(t) \) is an unspecified base hazard function. Taking \( f(w; \gamma) \) to be the density of \( W \), where \( \gamma \) is an unknown scalar, yields a rich class of semiparametric models. Examples include the inverse Gaussian frailty (Hougaard, 1984), the positive stable frailty (Hougaard, 1986), the lognormal frailty (McGilchrist and Aisbett, 1991), the power variance frailty (Aalen, 1988), the uniform frailty (Lee and Klein, 1988) and the threshold frailty (Lindley and Singpurwalla, 1986).

It is popular to let \( W \) have a gamma distribution with mean 1 and variance \( \gamma \). With time-independent covariates, the model is equivalent to the odds-rate regression (Dabrowska and Doksum, 1988),

\[
h(T) = -\beta^T Z + \epsilon_y,
\]

(2)
where \( h(t) \) is an unspecified strictly monotone increasing function, and \( \exp(\epsilon_{r}) \) has a pareto(\( \gamma \)) distribution. Fixing \( \gamma = 0 \) gives proportional hazards, while \( \gamma = 1 \) gives proportional odds (Bennett, 1983). For other \( \gamma \), a difficulty is that the interpretation of \( \beta \) in the linear model (2) for \( T \) is on the transformed scale \( h \), not the original scale (Hinkley and Runger, 1984). The formulation (1) provides a helpful explanation. The coefficients \( \beta \) are log hazard ratios in the presence of individual heterogeneity. This intuition generalizes when the covariates are time-varying.

If \( f(w; \gamma) \) has zero variance, then (1) reduces to the Cox model and efficient estimation of \( \beta \) is straightforward with the partial likelihood (Andersen and Gill, 1982). Estimation for the special case (2) with \( \gamma \) known has been studied extensively (Pettitt, 1982, 1984; Cheng, Wei, Ying, 1995, 1997; Murphy, Rossini, Van der Vaart, 1997; Fine, Ying, Wei, 1998; Scharfstein, Tsiatis, Gilbert, 1998; Shen, 1998). When the parameter in the frailty distribution is unknown, these methods are not applicable. Asymptotic theory for maximum likelihood estimation of model (2) with unknown \( \gamma \) and clusters of size \( \geq 2 \) was derived by Parner (1998). See Murphy (1994, 1995) for related work. A unified approach to estimation in model (1) with correlated data is not available.

In this paper, the focus is the practically important case of independent observations. Bagdonavičius and Nikulin (1999) suggested ad hoc estimators for the parameters, but their large sample properties were not established rigorously. In section 2, a likelihood-based procedure is formally proposed. An issue is without clusters the model (1) may not be identifiable (Heckman and Taber, 1994). Conditions are given for \( f(w; \gamma) \) and \( Z \) which ensure the uniqueness of the model in this scenario. The uniform consistency and weak convergence of the maximizers of a nonparametric likelihood is proved. The estimators achieve the semiparametric variance bound (Sasieni, 1992; Bickel et al., 1993) and are fully efficient. Since martingale methods are not applicable in this setting, modern empirical process techniques play an essential role.

Because the parametric and nonparametric components in (1) are estimated simultaneously, inference is complicated. The EM algorithm has proven to be a powerful tool for estimating parameters in gamma (Klein, 1992; Nielsen, et al, 1992) and log-normal (McGilchrist and Aisbett, 1991) frailty models with cluster sizes \( \geq 2 \) when the frailty parameter \( \gamma \) is specified. The penalized partial likelihood of Therneau and Grambsch (2000) enables shortcuts in the EM computations in some settings. However, these shortcuts are often not applicable, and MCMC may be needed (Vaida and Xu, 2000). For the shared frailty model, an estimate of \( \gamma \) can then be obtained by maximizing the profile likelihood for \( \gamma \) in conjunction with the EM algorithm. Parner (1998) showed that the vari-
ance of these estimates can be consistently estimated by inverting a discrete observed information matrix. Because the number of rows in the matrix is of the same order as the number of observed failure times, the numerical stability of this approach can be problematic for moderate sample sizes.

To assess the variability of the estimator for \((\gamma, \beta)\), we introduce in section 3 a novel application of Markov chain Monte Carlo to the profile likelihood. This method automatically maximizes the profile likelihood and estimates its curvature without computing derivatives. This is in contrast to the EM approach in which extra steps are needed to estimate \(\gamma\) and the variance of the parameter estimates. Ordinarily, with parametric models, the posterior of the full likelihood may be evaluated via MCMC to approximate frequentist inference (Ghosal, Ghosh, and Samanta, 1995; Johnson, 1967, 1970; Johnson and Ladalla, 1979). Our algorithm permits semiparametric inference about \(f(w; \gamma)\) and the regression coefficients separately from \(\lambda_0\). Its validity involves an extension of Murphy and Van der Vaart (2000). To construct confidence statements for the survival probabilities of a patient with certain covariates, the bootstrap is used. It is shown to give asymptotically correct inferences.

In section 4, numerical studies demonstrate that the procedure performs well in moderately sized samples. The methods are illustrated on the lymphoma data in section 5. Remarks conclude in section 6.

2 ESTIMATION

2.1 The Data and Model Set-up

The data \(\{(X_i, \delta_i, Z_i), i = 1 \ldots n\}\) consist of \(n\) i.i.d. realizations of \((X, \delta, Z)\), where \(X = T \wedge C, \delta = I\{T \leq C\}\), \(x \wedge y\) denotes the minimum of \(x\) and \(y\), \(I\{B\}\) is the indicator of \(B\), and \(C\) is the right censoring time. The covariate \(Z(\cdot)\) is assumed to be a caglad (left-continuous with right-hand limits) process. The analysis is restricted to the interval \([0, \tau]\) with \(\tau < \infty\) such that \(P[C \geq \tau | Z] = P[C = \tau | Z] > 0\) almost surely.

The failure time \(T\) is assumed to have the hazard function (1) given \(Z\) and the frailty \(W\). After integrating over \(W\), the survival function for \(T\) given \(Z\) is

\[
S(t|Z) \equiv P[T > t | Z] = E \left[ \exp \left\{ -W \int_0^t e^{\beta Z(s)} dA(s) \right\} | Z \right] \\
= \Lambda_\gamma \left\{ \int_0^t e^{\beta Z(s)} dA(s) \right\},
\]

(3)
where $A(s) = \int_0^s \lambda_0(u)du$ and $\Lambda_\gamma(x) = \int_0^\infty e^{-ux} f(w; \gamma) dw$ is the Laplace transform of $W$. The following are instances of frailty transforms:

1. The gamma frailty has $\Lambda_\gamma(x) = (1 + \gamma x)^{-1/\gamma}$.

2. The inverse Gaussian frailty (Hougaard, 1984) has $\Lambda_\gamma(x) = \exp \{-\gamma^{-1} [(1 + 2\gamma x)^{1/2} - 1]\}$.

3. The lognormal frailty (McGilchrist and Aisbett, 1991) has

$$
\Lambda_\gamma(x) = \int_{\mathbb{R}} \exp \left\{ -xe^{\gamma^2x - \gamma^2/2} \right\} \phi(v) dv.
$$

4. The positive stable frailty (Hougaard, 1986) has $\Lambda_\gamma(x) = \exp \{ -x^\gamma \}$.

Let $S_\tau \equiv \{ t \in [0, \tau] : dA_0(t)/dt > 0 \}$, $t_0 \equiv \inf S_\tau$, and $F(t+) \equiv \lim_{s \to t} F(s)$. Some assumptions on the data and the true parameter values (denoted by subscript 0) are:

(A) $A_0$ is continuous on $[0, \tau]$, $A_0(t) = 0$ for $t \in [0, t_0]$, and $0 < A_0(\tau) < \infty$.

(B) $\gamma_0 \in [0, M_0)$, where $M_0 < \infty$, and $\beta_0 \neq 0$ is in the interior of a compact set $B_0 \subset \mathbb{R}^d$.

(C) Censoring is independent of $T$ given $Z$ and uninformative of $\gamma$, $\beta$, and $A$.

(D) The total variation of $Z(\cdot)$ on $[0, \tau]$ is $\leq M_1 < \infty$ and $\text{var}[Z(t_0+)]$ is positive definite.

Conditions (A) and (B) constrain the parameter space. Condition (C) is noninformative censoring. Condition (D) is bounded variation of the covariate. Additional conditions (E) through (G) contain the technical assumptions on $\Lambda_\gamma$ and are in appendix A.1. In the next section, we show that frailty transforms 1 through 3 above satisfy the assumptions.

### 2.2 Identifiability

Identifiability of the frailty model with independent observations is theoretically challenging. Previous work is reviewed by Heckman and Taber (1994). See Elbers and Ridder (1982) and Heckman and Singer (1984). A key assumption is $\beta_0 \neq 0$. It allows the variability of the covariates to distinguish $\gamma$ from $A$. Another restriction imposed in past research is that the support of $Z$ is connected. This enables nonparametric estimation of the frailty distribution (Horowitz, 1999) but implies that at least one component of $Z$ is semi-continuous. Our result permits all discrete covariates. The weaker condition is critical in applications, like the non-Hodgkin's lymphoma data, where the predictors are binary. In this scenario, nonparametric identification of the frailty distribution is impossible without stronger assumptions, even when the mean of $W$ is finite, as in Elbers and Ridder (1982).
Lemma 1 Under conditions (A) through (F), model (3) is identifiable.

This is shown in appendix A.1. The monotonicity in $\gamma$ of $\bar{G}_\gamma(0^+)$, where $\bar{G}_\gamma(x) = \frac{\partial^2 \log \Lambda_\gamma(x)}{\partial x^2}$, as given in condition (F), is the key to establishing identifiability of the Laplace transform. Since $-\bar{G}_\gamma(0^+)$ is the variance of $W$ when $\gamma \geq 0$, this is the same as requiring that $\text{var}[W]$ be a monotone function of $\gamma$. The positive stable frailty model violates condition (F) and is not identifiable without clustered data. The gamma, inverse Gaussian, and lognormal frailties are identifiable.

Proposition 1 Conditions (E), (F) and (G) are satisfied by the gamma, inverse Gaussian, and lognormal frailty models.

This is shown in appendix A.1. It is the first time that identification of the inverse Gaussian and lognormal models has been demonstrated. Verification of the lognormal is hard technically since $\Lambda_\gamma$ does not have a closed form. Note that condition (G) is needed for the consistency proof in section 2.4 but not for lemma 1.

2.3 The Nonparametric Maximum Likelihood Estimator

The nonparametric log-likelihood for $\psi = (\gamma, \beta, A)$ has the form

$$L_n(\psi) = \mathbb{P}_n \left\{ \int_0^T \left[ \log \bar{G}_\gamma(H^\psi(t)) + \beta^T Z(s) + \log a(s) \right] dN(s) - G_\gamma(H^\psi(t)) \right\},$$

(4)

where $N(t) \equiv I\{X \leq t, \delta = 1\}$, $Y(t) \equiv I\{X \geq t\}$, $H^\psi(t) \equiv \int_0^t Y(s) e^{\beta^T Z(s)} dA(s)$, $a \equiv dA/dt$, and $\mathbb{P}_n$ is the empirical probability measure. That is, for a known function $f$ depending on $\psi$,

$$\mathbb{P}_n f(X, \delta, Z; \psi) = n^{-1} \sum_{i=1}^n f(X_i, \delta_i, Z_i; \psi).$$

As discussed by Murphy, Rossi, and van der Vaart (1997), the maximum likelihood estimator for $a$ does not exist. The issue is that any unrestricted maximizer of (4) puts mass only at observed failure times and is not a continuous hazard.

Instead, we compute the maximizer by profiling over $A$. This yields estimators for $\theta = (\gamma, \beta)$ and $A$, but not $a$. The profile likelihood is $pL_n(\theta) \equiv \sup_A L_n(\psi)$. Consider one-dimensional submodels for $A, t \mapsto A_t(\cdot) \equiv \int_0^t \{1 + th(s)\} dA(s)$, where $\cdot$ denotes an argument ranging over $[0, \tau]$ and $h$ is the class of functions $h(s) = I\{s \leq u\}, u \geq 0$. Since $A = A_{t=0}$, one may differentiate $L_n(\theta, A_t)$ with respect to $t$ and solve for $A$ with $t = 0$. In this manner, $pL_n(\theta) = L_n(\theta, \hat{A}_\theta)$, where $\hat{A}_\theta$ is the
solution of

$$\hat{A}_\theta(u) = \int_0^u \left( \mathbb{P}_n \left[ Y(s)e^{g^T Z(s)} \left( \hat{G}_\gamma \left\{ H_{\hat{\psi}_u}(X) \right\} - \delta \frac{\hat{G}_\gamma \left\{ H_{\hat{\psi}_u}(X) \right\}}{\hat{G}_\gamma \left\{ H_{\hat{\psi}_u}(X) \right\}} \right] \right) d\mathbb{P}_n(N(s)) \right)$$

(5)

and $\hat{\psi}_u \equiv (\hat{\theta}_u, \hat{A}_u)$. The same maximizer occurs with $\Delta A$ in place of $a$ in $L_n(\psi)$, where $\Delta A(s) = A(s - A(s -)$ and $A(0) = \lim_{t \to 0} A(t)$. That is, one maximizes $L_n(\psi)$ over all $A$ with jumps at the observed failure times. We denote by $\hat{L}_n(\psi)$ the log-likelihood expression with $\Delta A$ in place of $a$.

The nonparametric maximum likelihood estimator is $\hat{\psi}_n \equiv (\hat{\theta}_n, \hat{A}_{\hat{\psi}_n})$, where $\hat{\theta}_n \equiv \arg \max \hat{L}_n(\theta)$. Equivalently, $\hat{\psi}_n = \arg \max \hat{L}_n(\psi)$.

2.4 Consistency

The following theorem establishes the consistency of $\hat{\psi}_n$:

**Theorem 1** Under model (3) and conditions (A) through (G), $\hat{\psi}_n$ converges outer almost surely to $\psi_0$ in the uniform norm.

The proof is given in the appendix. Because the unknown “parameter” $A$ is a function, classical methods for euclidean parameters are not appropriate. The modern empirical process approach of Parner (1998) is adapted to achieve both consistency of $\hat{\theta}_n$ and uniform consistency of $\hat{A}_n$, i.e., $\sup_{t \in [0,1]} \left| \hat{A}_n(t) - A_0(t) \right| \to 0$, almost surely. The first step is to establish that $\hat{A}_n(t)$ is asymptotically bounded and equicontinuous almost surely. This permits the Helly selection theorem to be used to show that the only possible limiting values of $\hat{\psi}_n$ are in a compact subset of the parameter space. The next step is to show that the limiting Kullback-Leibler information is zero. This means that the only possible limiting value is $\psi_0$, by identifiability (lemma 1). Consequently, every limiting value of $\hat{\psi}_n$ is $\psi_0$, almost surely.

2.5 Asymptotic Normality and Efficiency

**Theorem 2** Under model (3) and conditions (A) through (G), $\sqrt{n}(\hat{\psi}_n - \psi_0)$ converges weakly to a tight, mean zero Gaussian element $Z_0$ with covariance $V_0$, and is asymptotically linear, regular, and efficient.

The proof is given in appendix A.2. Here, the idea of efficiency, as reflected in theorem 5.2.1 of Bickel et al. (1993) (hereafter BKRW), is stronger than the usual notion of semiparametric
efficiency (Sasieni, 1992). In the latter, the finite dimensional \( \hat{\theta}_n \) is efficient, while in the former both \( \hat{\theta}_n \) and \( \hat{A}_n \) are efficient. This is understood intuitively via a convolution formula. Theorem 2 states that \( \sqrt{n}(\hat{\psi}_n - \psi_0) \) converges weakly to a random element \( Z_0 \). The estimator \( \hat{\psi}_n \) is efficient if for any other regular estimate of \( \psi_0 \), say \( \hat{\psi}_n \), \( \sqrt{n}(\hat{\psi}_n - \psi_0) \) converges weakly to \( Z_0 + E_0 \), where \( E_0 \) is independent of \( Z_0 \). This means that any smooth functional of \( \hat{\psi}_n \) has smaller variance than that of \( \hat{\psi}_n \). Restricting to functionals of \( \hat{\theta}_n \) gives efficiency in the sense of Sasieni (1992). A key and difficult step is demonstrating:

**Lemma 2** The information operator \( \sigma \) is one-to-one.

The proof of lemma 2 draws heavily on the tangent set \( H_p \), as defined in the proof of theorem 2 in appendix A.2. The definition of \( H_p \) for the semiparametric model (1) generalizes that for the finite dimensional case. For a \( d \)-dimensional parametric model, the score function is a vector in \( \mathbb{R}^d \) and the tangent set consists of vectors in \( \mathbb{R}^d \). With an infinite dimensional parameter, the tangent set is essentially the space of directions for the "derivatives" of the log-likelihood. The issue is showing that for \( h \in H_p \), \( \sigma(h) = 0 \) implies \( h = 0 \). This gives that \( \sigma \) is continuously invertible and onto, and thus the influence function is contained in the closed linear span of the score operator. Efficiency then follows from theorem 5.2.1 of BKRW.

**Proof of lemma 2.** With any \( h \in H_p \) such that \( \sigma(h) = 0 \), define the regular parametric one-dimensional submodel \( \psi_0(h) \equiv \psi_0 + t\{h_1, h_2, \int_0^t h_3(s) dA_0(s)\} \). \( \sigma(h) = 0 \) implies

\[
P_0 \left\{ \frac{\partial^2}{(\partial t)^2} L_n(\psi_0) \bigg\vert_{t=0} \right\} = P_0 \{ U^\tau(\psi_0)(h) \}^2 = 0,
\]

where the score operator \( U^\tau \) is the derivative of the log-likelihood based on the data available at time \( \tau \) and is explicitly given by (18) in appendix A.2. But this implies \( P_0 \{ U^\tau(\psi_0)(h) \} \mathcal{G}(n, y, t) \}^2 = 0, \)

where the random set

\[
\mathcal{G}(n, y, t) \equiv \{ N, Y : N(s) = n(s), Y(s) = y(s), s \in \mathbb{R} \}
\]

has non-zero probability. This then implies that \( U^t(\psi_0)(h) = 0 \) almost surely for all \( t \in [0, \tau] \).

Assuming that \( \{(N(s), Y(s), Z(s)), s \geq 0\} \) is censored at \( X \in (t_0, \tau] \),

\[
0 = G_{(1)}(H^{\psi_0}(t))h_1 + G_{(2)}(H^{\psi_0}(t)) \int_0^t Y(s) e^{\int_0^s Z(s)} \{ h_2 Z(s) + h_3(s) \} dA_0(s).
\]

Taking the Radon-Nikodym derivative with respect to \( A_0(t) \) and letting \( t \downarrow t_0 \) yields \( h_2 Z(t_0+) + h_3(t_0+) = 0 \) since \( G_{(1)}(0+) = 0 \) and \( G_{(2)}(0+) = 1 \) by assumption (F). But this implies \( h_2 = 0 \) by
condition (D). Dividing (6) by \( \hat{G}_{\gamma_0}(H^{\psi_0}(t)) \), differentiating with respect to \( A_0(t) \), and taking \( h_2 = 0 \) yields

\[
0 = \left[ \hat{G}_{\gamma_0}^{(1)}(H^{\psi_0}(t)) \hat{G}_{\gamma_0}(H^{\psi_0}(t)) - G_{\gamma_0}^{(1)}(H^{\psi_0}(t)) \hat{G}_{\gamma_0}(H^{\psi_0}(t)) \right] h_1 + \left[ \hat{G}_{\gamma_0}(H^{\psi_0}(t)) \right]^2 h_3(t).
\]

Differentiating again with respect to \( A_0(t) \) and letting \( t \downarrow t_0 \) gives

\[
0 = \hat{G}_{\gamma_0}^{(1)}(0^+) e^{\psi_0(t_0^+)} h_1 + \hat{h}_3(t_0^+),
\]

where \( h_3 \equiv \frac{d\hat{h}_3}{dA_0} \). Now (B), (D), and (F) yield \( h_1 = 0 \). Thus (6) implies \( h_3(t) = 0 \) for all \( t \in [0, \tau] \).□

3 INFERENCE

The complexity of the limiting variance of \( \sqrt{n}(\hat{\psi}_n - \psi_0) \) makes inference through direct estimation of \( V_0 \) extremely complicated. The problem is that \( V_0 \) involves linear operators defined on functional spaces. Nonparametric functional estimation is required and may be unstable with realistic sample sizes. When inference on \( \theta \equiv (\gamma, \beta) \) only is of interest, both the maximizer and curvature of \( pL_n(\theta) \equiv \sup_\theta L_n(\theta) \) can be accurately estimated using Markov chain Monte Carlo (MCMC). This is described in section 3.1 below. When joint inference on \( \gamma, \beta, \) and \( A \) is desired, bootstrapping is a valid technique. This is discussed in section 3.2.

3.1 Using Markov Chain Monte Carlo

The Metropolis-Hastings algorithm (Metropolis, et al., 1953; Hastings, 1970) can be used to generate a Markov chain \( \{\theta^{(1)}, \theta^{(2)}, \ldots\} \) with stationary distribution proportional to \( \exp\{pL_n(\theta)\} \). This is done as follows: Begin with an initial value for the chain \( \theta^{(1)} \). For each \( k > 1 \), obtain a proposal \( \hat{\theta}^{(k+1)} \) by random walk from \( \hat{\theta}^{(k)} \). Then compute \( A_{\hat{\theta}^{(k+1)}} \) and \( pL_n(\hat{\theta}^{(k+1)}) \), and decide whether to keep \( \hat{\theta}^{(k+1)} \) by evaluating \( \exp\{pL_n(\hat{\theta}^{(k+1)}) - pL_n(\theta^{(k)})\} \) multiplied by the Hastings's adjustment and applying an acceptance rule.

After generating a sufficiently long chain, one computes the mean or mode of the chain to estimate \( \hat{\theta}_n \) and the variance of the chain to estimate \( \text{var} \left[ \hat{\theta}_n \right] \). For an infinitely long chain, the mean is precisely \( \hat{\theta}_n \). The mean, on the other hand, is a Bayes estimate under squared error loss with an uninformative prior on \( \theta \) and a prior on \( A \) which after integration yields the profile likelihood. This proposed use of MCMC for semiparametric inference is novel.

The following result gives that the density proportional to \( \exp\{pL_n(\theta)\} \) is asymptotically equal to the density of a normal random variable with mean \( \hat{\theta}_n \) and variance \( A_0^{-1} \). Thus, the mean of the
Markov chain described above is $\sqrt{n}$-consistent for $\hat{\theta}_n$. The variance—after multiplying by $n$—is consistent for the inverse Fisher information corresponding to the $\theta$ terms in $V_0$ denoted $\tilde{I}_0^{-1}$.

**Theorem 3** For all compact $K \subset \mathbb{R}^{d+1}$ including zero as an interior point,

$$
\sup_{\theta \in \hat{\theta}_n + K/\sqrt{n}} \left| -\frac{1}{2} n (\theta - \hat{\theta}_n)^T \tilde{I}_0 (\theta - \hat{\theta}_n) - \left\{ pL_n(\theta) - pL(\hat{\theta}_n) \right\} \right| \rightarrow 0, \quad (7)
$$

in outer probability, as $n \rightarrow \infty$.

The proof is provided in appendix A.2. Note that the expansion in theorem 3 implies that the first two moments of the distribution proportional to $pL_n(\theta)$ converge to the desired Gaussian moments only if

$$
\lim_{\epsilon \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{\theta \in \hat{\theta}_n + K/\sqrt{n}} \int_{\hat{\theta}_n}^{\hat{\theta}_n + \epsilon} n(\theta - \hat{\theta})^2 \exp \left\{ pL_n(\theta) - pL(\hat{\theta}) \right\} d\theta = 0. \quad (8)
$$

This condition is hard to verify analytically. However, (8) is clearly satisfied if

$$
\lim_{n \rightarrow \infty} \sup_{\theta \in \hat{\theta}_n + K/\sqrt{n}} \left| \sqrt{n} (\theta - \hat{\theta}) \right|^{2+\epsilon} \exp \left\{ pL_n(\theta) - pL(\hat{\theta}) \right\} d\theta < \infty,
$$

for some $\epsilon > 0$. For simplicity, set $\epsilon = 1$. Observe that (8) can be evaluated empirically by checking whether $n^{3/2}$ times the centered outputs of the Markov chain are bounded. To be more precise, let $\theta^{(1)}, \ldots, \theta^{(N)}$ be $N$ steps of a Markov chain (in equilibrium) with stationary distribution proportional to $pL_n(\theta)$. Let $\bar{\theta}$ and $\bar{V}_\theta$ be the sample mean and covariance from this chain, and $\chi_j = (\theta^{(j)} - \bar{\theta})^T \bar{V}_\theta^{-1} (\theta^{(j)} - \bar{\theta})$, $j = 1, \ldots, N$. The quantiles of $\{\chi_1^{3/2}, \ldots, \chi_N^{3/2}\}$ should be the same order of magnitude as the quantiles of $N$ independent chi-squared variates with $d + 1$ degrees of freedom to the $3/2$th power.

### 3.2 Using the Bootstrap

The usual nonparametric bootstrap resamples with replacement from the observed data. A disadvantage is that ties can arise with censored survival data. We propose an alternative weighted bootstrap. In each bootstrap sample, one generates $n$ independent and identically distributed nonnegative weights $\zeta_1, \ldots, \zeta_n$ with mean and variance one, and with $\int_0^\infty \sqrt{P[\zeta_i > x]} dx < \infty$. Each weight is divided by the average weight (rejecting samples with all zeros) to obtain “standardized weights” $\zeta_1^*, \ldots, \zeta_n^*$ which sum to $n$. Distributions satisfying the moment conditions include the unit exponential and the Poisson with mean 1. For the nonparametric bootstrap, the weights $\zeta_1^*, \ldots, \zeta_n^*$ are generated from a multinomial distribution with $E \zeta_i^* = 1$, $i = 1 \ldots n$, and $\sum_{i=1}^n \zeta_i^* = n$. 

10
For a known function $f$, let $\mathbb{P}_n f(X_i, \delta; Z; \psi) \equiv n^{-1} \sum_{i=1}^n \zeta_i f(X_i, \delta_i, Z_i; \psi)$ define the weighted empirical measure $\mathbb{P}_n^\circ$. The weighted bootstrap estimate $\hat{\psi}_n^\circ$ is computed by substituting $\mathbb{P}_n^\circ$ for $\mathbb{P}_n$ in the expressions in section 2.3 and maximizing over $\psi$. $\mathbb{P}_n^\bullet$ is defined similarly to $\mathbb{P}_n^\circ$ with the weights $\zeta_1^\bullet, \ldots, \zeta_n^\bullet$ in place of $\zeta_1^\circ, \ldots, \zeta_n^\circ$. The nonparametric bootstrap estimate $\hat{\psi}_n^\bullet$ is computed by using $\mathbb{P}_n^\bullet$ in place of $\mathbb{P}_n$ in section 2.3.

The following establishes the validity of both the nonparametric and the weighted bootstraps. The proof is given in appendix A.2.

**Corollary 1.** The conditional bootstrap of $\hat{\psi}_n$, based either on $\hat{\psi}_n^\bullet$ or $\hat{\psi}_n^\circ$, is asymptotically consistent for the limiting process $Z_0$. That is, $\sqrt{n}(\hat{\psi}_n - \psi_n)$ and $\sqrt{n}(\hat{\psi}_n^\bullet - \psi_n)$ are asymptotically measurable,

$$
(i) \sup_{g \in BL_1} \left| E_{\cdot} g \left( \sqrt{n}(\hat{\psi}_n^\bullet - \psi_n) \right) - E_{\psi} g(Z_0) \right| \to 0 \text{ in outer probability, and}
$$

$$
(ii) \sup_{g \in BL_1} \left| E_{\cdot} g \left( \sqrt{n}(\hat{\psi}_n^\circ - \psi_n) \right) - E_{\psi} g(Z_0) \right| \to 0 \text{ in outer probability},
$$

where $BL_1$ is the space of functions mapping $\mathbb{R}^{d+1} \times \ell^\infty([0, r]) \to \mathbb{R}$ with Lipschitz norm $\leq 1$, $\ell^\infty(B)$ is the space of bounded functionals on $B$, and conditional on the data, $E_{\cdot}$ and $E_{\psi}$ are expectations over the multinomial and standardized weights, respectively.

While the choice of $\{\zeta_i\}$ in the weighted bootstrap has no effect asymptotically, the rate of convergence may be affected. Newton and Raftery (1994) discuss different choices in the context of parametric maximum likelihood. They demonstrate that unit exponential weights, which are Dirichlet after standardizing, perform well. Our own experience is that exponential weights also work well for semiparametric inference. A detailed analysis of the distribution of the weights is beyond the scope of this paper.

### 4 Numerical Studies

The data were generated from the odds-rate model with $\gamma = 0, 0.25, 0.5, 1$ or 2, and sample sizes of 100, 200 or 500. There were two covariates. The first was uniformly distributed on $(0, 1)$ and the second was a Bernoulli(0.5) variable. The regression coefficients, $\beta_1$ and $\beta_2$, were 1 and $A(t) = \exp(t) - 1$. The censoring time $C$ was $r$ with 10% probability and was uniformly distributed on $(0, r)$ with 90% probability. The value of $r$ was chosen to give approximately 15% censoring.

11
Two estimation procedures were applied: \( \gamma \) unspecified (Case 1) and \( \gamma \) fixed at \( \gamma_0 \) (Case 2). For each combination of sample size and model parameters, 200 datasets were simulated.

For each dataset, we constructed a Markov chain of length 20,000 using the Metropolis algorithm. The first 5000 were discarded as burn-in. The proposal distribution for each parameter was univariate normal, with standard deviation tuned to an acceptance rate of 25\%-50\%. In Case 1, the parameters were updated in two blocks, \( \gamma \) and \((\beta_1, \beta_2)\). The order of update was randomized. The support of \( \gamma \) was restricted to the interval \([-0.8, 5.0]\). Within the \((\beta_1, \beta_2)\)-block, each coefficient was updated separately in randomized order. In Case 2, only the \((\beta_1, \beta_2)\)-block was updated.

Let \( \hat{\gamma}, \hat{\beta}_1, \) and \( \hat{\beta}_2 \) denote the estimates for \( \gamma, \beta_1, \) and \( \beta_2 \), respectively, based on the empirical mean of the Markov chain in each simulation. Let \( \bar{\gamma}, \bar{\beta}_1, \) and \( \bar{\beta}_2 \) be the estimates based on the mode. For each simulated chain, the modes are computed by maximizing a kernel estimate of the density with a Gaussian window. The window width is chosen so that as the length of the chain goes to infinity, this estimate converges to the maximum likelihood estimate. When the mode did not converge, the value of the Markov chain element with the largest likelihood was selected. Typically, about 5\% of the kernel estimates were not convergent.

In Table 1, \( \bar{\gamma} \) is the empirical mean of \( \hat{\gamma} \) from Case 1 estimation, \( \bar{\sigma}_0 \) is the empirical mean of the corresponding standard error estimates, and \( \sigma_0^\gamma \) is the Monte Carlo standard error. The numbers in parentheses are the values from the mode estimators. The symbols \( \bar{\beta}_i, \bar{\sigma}_i, \) and \( \sigma_i^\gamma \) \((i = 1, 2)\) are similarly defined in Table 1 for Case 1 and in Table 3 for Case 2. In Table 2, we report the empirical coverage probabilities of the 95\% confidence intervals in Case 1 from the mean (left side) and mode (right side) estimators. The confidence intervals were calculated by adding and subtracting 1.96\( \bar{\sigma}_i \) \((i = 0, 1, 2)\). The Monte Carlo standard error for the coverage estimates is 0.015.

In Table 1, the estimates of the regression coefficients are nearly unbiased, except when \( \gamma = 2 \). In this case, the true values are underestimated. The estimate for the odds-rate parameter may be somewhat biased, which has previously been observed in the shared frailty model. However, this diminishes with increasing sample size, and the results for \( n = 500 \) are encouraging. In Table 2, the coverage probabilities for \( \beta_1 \) and \( \beta_2 \) are generally close to the nominal level. When \( \gamma = 2 \), the coverages for \( \gamma \) from the mode are considerably below 0.95 for all \( n \). The confidence intervals using the mean are more reliable.

In Table 3, there is less bias in the estimates of the regression coefficients. The Monte Carlo standard errors are uniformly smaller than those in Case 1. This is expected, in view of the extra
variability introduced by the estimation of \( \gamma \). The ratio of the Monte Carlo standard error of the estimate in Case 2 to that in Case 1 (R.E.) is also provided. When \( \gamma \) is unknown, the estimate of \( \beta_1 \) is fairly efficient, while the loss of precision in the estimate of \( \beta_2 \) is larger.

Our simulations were limited to 200 samples for each combination of \( n \) and \( \theta \). Due to the nature of MCMC, the inferences are computationally intensive and require careful monitoring. Unless the procedure is automated, it is infeasible to monitor the convergence of every chain (there were 1500 chains in Case 1 alone). We sampled roughly 10% of the chains in each configuration. Overall, the proposed method worked well. The poor performance of \( \hat{\gamma} \) for large \( \gamma \) is a well-known phenomenon and is not attributable to the MCMC implementation.

## 5 NON-HODGKIN’S LYMPHOMA DATA

The data is a subset of 1385 patients with aggressive non-Hodgkin's lymphoma (NHL) from 16 institutions and cooperative groups in North America and Europe. These patients were treated with a particular chemotherapy regimen. Survival was documented from start of treatment until either death or loss to follow-up. The censoring rate was 54.7%. Information on the following pre-treatment covariates is complete for all patients in the subset: age at the diagnosis of NHL (\( \leq 60 \) or \( > 60 \) years); performance status (ambulatory or non-ambulatory); serum lactate dehydrogenase level (below normal or above normal); number of extranodal disease sites (\( \leq 1 \) or \( > 1 \)); and Ann Arbor classification of tumor stage (stage I or II [localized disease], or stage III or IV [advanced disease]). Each characteristic is coded 0 for the first group in the parentheses and 1 for the second. These predictors are the basis for the original model (Non-Hodgkin’s Lymphoma Prognostic Factors Project, 1993).

To estimate the parameters in the odds-rate regression (2), we generated a Markov chain of length 120,000. The first 20,000 were discarded as burn-in. The initial value for the odds-rate parameter, \( \gamma \), was 0.5, and those for the regression parameters were 0.1. The mean of the chain and 95% confidence intervals are given in Table 4. Results from the Cox model are also provided. The odds-rate parameter is significantly greater than 1. This suggests that neither the proportional hazards model (\( \gamma = 0 \)), nor the proportional odds model (\( \gamma = 1 \)), is a good fit to the data. Note that the magnitude of each coefficient in the odds-rate model is roughly twice that in the proportional hazards model. This exemplifies the well-known attenuation of the covariate effects in a misspecified
Cox model.

In Fig. 1, we plot the Kaplan-Meier estimates of the marginal survival distributions for the performance status groups and the tumor stage groups. The estimates from the odds-rate model and the \( \gamma = 0 \) model are also displayed. The survival estimate in a group (for example, patients with status = 0) from the odds-rate model is \( \exp\{-H(t)\} \), where

\[
\hat{H}(t) = \int_0^t \frac{\sum_i Y_i(s) \exp\{\hat{\beta}^T Z_i(s)\} \left(1 + \gamma_n \int_0^s \exp\{\hat{\beta}^T Z_i(u)\} d\hat{A}_n(u)\}^{-1}}{\sum_i Y_i(s)} \ d\hat{A}_n(s).
\]

The estimate based on the Cox model uses the partial likelihood estimator, Breslow’s estimator, and 0 in the place of \( \hat{\beta}_n \), \( \hat{A}_n \), and \( \hat{\nu}_n \), respectively. The summations are over all patients in the group. That is, \( \hat{H}(t) \) is a model-based estimate of the cumulative hazard which averages over the observed covariate distribution in that group. The frailty model estimates are almost indistinguishable from the Kaplan-Meier curves. The proportional hazards model is a poor fit. The plots for other covariates are similar.

The diagnostic plots for the Markov chain in \( \beta_1 \) are given in Fig. 2. The scatter plot and rapid decay in the autocorrelation plot suggest the chain is in equilibrium. The histogram shows the (marginal) distributions of the chain are unimodal, fairly symmetric, and approximately Gaussian. The moment diagnostic plot proposed in section 3.5 was satisfactory. The characteristics of these diagnostics are similar for other parameters.

Survival predictions for two covariate values, representing an elderly high risk patient (\( Z = (1, 1, 1, 1)^T \)) and an elderly low risk patient (\( Z = (1, 0, 0, 0)^T \)), are in Fig. 3. Also shown are 95% pointwise confidence intervals for the odds-rate prediction using 500 multiplier bootstrap samples with Dirichlet weights. For each bootstrap, a maximum likelihood estimate \( (\hat{\beta}^k, \hat{A}^k) \) was obtained from the average of a Markov chain of length 1000 based on the profile likelihood (\( k = 1 \ldots 500 \)). The proportional hazards prediction for the high risk patient significantly underestimates the long-term survival probability relative to the odds-rate model. The corresponding prediction for the low risk patient overestimates survival at most time points.

6 DISCUSSION

Proportional hazards frailty regression models are useful one parameter extensions to the Cox model which may lead to improvements in survival predictions. The reason is that the frailty acknowledges
missing and mismeasured covariates. The choice of \( f(u; \gamma) \) is an open question. It would be worthwhile to develop methods for selecting the distribution. This is a topic for future research.

We demonstrated that maximum likelihood estimation can be implemented by profiling over the baseline hazard and then using MCMC. This computational technique may be widely applicable to semiparametric maximum likelihood inference. It should be particularly advantageous when the infinite dimensional nuisance parameter is not estimable at the \( \sqrt{n} \) rate, as happens, for example, with interval censoring. The theoretical issues related to this approach are currently being explored.

\section{Appendix}

\subsection{A.1 Regularity Conditions and Related Results}

We need the following conditions on \( \Lambda \gamma \), where \( \Lambda \gamma = \partial \Lambda \gamma (u) / (\partial u) \), \( \hat{\Lambda} \gamma = \partial \hat{\Lambda} \gamma (u) / (\partial u) \), \( G \gamma = -\log \Lambda \gamma \), \( \hat{G} \gamma = \partial G \gamma (x) / (\partial x) \), \( \hat{G} \gamma = \partial \hat{G} \gamma (x) / (\partial x) \), \( G \gamma^{(1)} = \partial G \gamma (x) / (\partial \gamma) \), \( \hat{G} \gamma^{(1)} = \partial \hat{G} \gamma (x) / (\partial \gamma) \), and \( \hat{G} \gamma^{(1)} = \partial \hat{G} \gamma^{(1)} (x) / (\partial x) \):

(E) For each nonnegative \( x < \infty \), there exists an extension of \( \Lambda \gamma (\cdot) : [0, M_0] \times [0, x] \mapsto [0, 1] \), depending on \( x \), with domain \([-\epsilon_0(x), M_0] \times [0, x] \), where \( 0 < \epsilon_0(x) < 3/4 \), \( \lim_{x \to \infty} \epsilon_0(x) = 0 \). In addition, the following is true for all \( \gamma \in [-\epsilon_0(x), M_0] \): \( \Lambda \gamma : [0, x] \mapsto [0, 1] \) is monotone decreasing, with \( \Lambda \gamma (0+) = 1 \) and \( \Lambda \gamma (x) \leq 0 \), and both \( 0 < -\hat{\Lambda} \gamma (u) < \infty \) and \( 0 \leq \hat{\Lambda} \gamma (u) < \infty \) for all \( u \in [0, x] \).

(F) For each nonnegative \( x < \infty \) and all \( \gamma \in [-\epsilon_0(x), M_0] \), \( \partial^2 \hat{G} \gamma (u) / (\partial u)^2 \), \( \partial \hat{G} \gamma^{(1)} (u) / (\partial u) \), \( \partial^2 \hat{G} \gamma (u) / (\partial \gamma) \), and \( \partial^2 \hat{G} \gamma^{(1)} (u) / (\partial \gamma)^2 \) exist and are bounded for all \( u \in [0, x] \), \( \hat{G} \gamma (0+) = 1 \) and \( \hat{G} \gamma (0+) \) is bounded and strictly monotone in \( \gamma \). Also, for all \( \gamma \in [0, M_0] \) and all \( u : 0 \leq u < \infty \), \( \hat{G} \gamma (u) \leq 0 \) and \( \left| \hat{G} \gamma^{(1)} (0+) \right| > 0 \).

(G) For \( \{ \gamma_k \} \in [0, M_0] \) with \( \gamma_k \to \gamma > 0 \), \( \lim_{k \to \infty} \sup_{u \geq 0} u^{c_1(\gamma)} \Lambda \gamma_k (u) < \infty \) and \( \lim_{k \to \infty} \sup_{u \geq 0} \left| u^{1+c_1(\gamma)} \hat{\Lambda} \gamma_k (u) \right| < \infty \), for some \( c_1(\gamma) > 0 \). For all sequences \( x_k \to \infty \) and \( \{ \gamma_k \} \in (-\epsilon(x_k), M_0) \) with \( \gamma_k \to \gamma \geq 0 \), \( \lim_{k \to \infty} \inf_{u \in [0, x_k]} x_k \hat{G} \gamma_k (u) \geq c_2(\gamma) \), where \( c_2(\gamma) = \infty \) if \( \gamma = 0 \) and \( c_2(\gamma) > 0 \) if \( \gamma > 0 \).

The condition (E) ensures that \( \gamma_0 = 0 \) is an interior point. Parts of (F) and (G) are conditions on the moments of \( W \). For (F), this follows since \( \hat{G} \gamma (0+) = E[W] \) and \( \hat{G} \gamma (0+) = -\text{var}[W] \). The
first sentence in (G) is satisfied if for each \( \gamma > 0 \), there exists a continuous function \( c_1(\gamma) > 0 \) such that \( \mathbb{E} \left[ W^{-c_1(\gamma)} \right] < \infty \).

**Proof of lemma 1.** Define \( \psi \equiv (\gamma, \beta, A) \), \( Y(t) \equiv I \{ X \geq t \} \), and \( H^\psi(t) = \int_0^t Y(s)e^{\beta T Z(s)}dA(s) \), for \( t \in [0, \tau] \). We need to establish

\[
G_\gamma(H^\psi(t)) = G_{\gamma_0}(H^{\psi_0}(t))
\]  

(9)

for all \( t \in [0, \tau] \) implies \( \psi = \psi_0 \) almost surely. Taking the Radon-Nikodym derivative of both sides of (9) as functions of \( t \) with respect to \( A_0(t) \) yields

\[
\dot{G}_\gamma(H^\psi(t))Y(t)e^{\beta T Z(t)}b(t) = \dot{G}_{\gamma_0}(H^{\psi_0}(t))Y(t)e^{\beta_0 T Z(t)},
\]  

(10)

where \( b \equiv dA/dA_0 \). Letting \( t \downarrow t_0 \) in (10) gives \( b(t_0+)e^{\beta T Z(t_0+)} = e^{\beta_0 T Z(t_0+)} \) by condition (F). This implies \( \beta = \beta_0 \) since \( \text{var}[Z(t_0+)] \) is positive definite. Hence \( b(t_0+) = 1 \) by conditions (B) and (D).

Setting \( \beta = \beta_0 \) and dividing both sides of (10) by \( Y(t)e^{\beta T Z(t)} \), differentiating with respect to \( A_0(t) \), and letting \( t \downarrow t_0 \), gives \( \dot{G}_\gamma(0+)e^{\beta T Z(t_0+)} + \dot{b}(0+) = \dot{G}_{\gamma_0}(0+)e^{\beta_0 T Z(t_0+)} \), where \( \dot{b} = da/dA_0 \). This proves \( \gamma = \gamma_0 \) since \( \dot{G}_\gamma(0+) \) is monotone and bounded in \( \gamma \). Now \( A = A_0 \) follows trivially. \( \Box \)

**Proof of proposition 1.** For the gamma frailty, the conditions hold with \( \epsilon_0(x) = (2/3)(1 \vee x)^{-1} \) and \( c_1(\gamma) = 1/\gamma \). For the inverse Gaussian frailty, the conditions hold with \( \epsilon_0(x) = (2/3)(1 \vee x)^{-1} \) and \( c_1(\gamma) = 1 \). Establishing these results is straightforward. For the lognormal, we now show that the conditions hold with \( \epsilon_0(x) = (1 \vee x)^{-4}/8 \) and \( c_1(\gamma) = 1 \). Complex analysis is involved since \( \sqrt{\gamma} \) is imaginary for \( \gamma < 0 \). However, the imaginary components of \( \Lambda_\gamma, G_\gamma \) and their derivatives are all zero. Moreover, \( \Lambda_\gamma(x) \) and its first two derivatives in \( x \) have the following form, with \( \xi \equiv \sqrt{\gamma} \) and \( u \equiv xe^{\xi^2/2} \):

\[
(-1)^k e^{k\xi^2/2} \int_\mathbb{R} e^{-u \cos \xi v} \cos(u \sin \xi v - k\xi v) \phi(v)dv,
\]  

(11)

for \( k = 0, 1, 2 \), respectively. If we establish that (11), for \( k = 2 \), is \( > 0 \) over the correct range, then (E) follows and showing (F) is easy. If there exists a \( v_0 \geq 2 \) and \( \xi_0 \in [0, \pi/(2v_0)] \) such that

\[
u \sin \xi_0 v_0 + 2\xi_0 v_0 = \pi/4
\]  

(12)

and such that the part of the integral over \( |v| > v_0 \) is completely dominated by the part over \( |v| \leq v_0 \), then (11) will be \( > 0 \) for all \( u' \in [0, u] \) and all \( \xi \in [0, \xi_0] \). Assume that \( v_0 \geq 2 \) and \( \xi_0 \) satisfies (12). Then

\[
\int_{-v_0}^{v_0} e^{-u \cos \xi_0 v} \cos(u \sin \xi_0 v - 2\xi_0 v) \phi(v)dv \geq \frac{0.95}{\sqrt{2}} e^{-u}
\]
and

\[
\int_{|v| > v_0} e^{-u \cos \xi_0 v} \cos (u \sin \xi_0 v - 2 \xi_0 v) \phi(v) dv \leq \frac{1}{\sqrt{2\pi}} e^{-u \nu_0^2/2}.
\]

Thus the total integral is positive if

\[
\frac{1}{\sqrt{2\pi}} e^{-u \nu_0^2/2} \left( \frac{0.95}{\sqrt{2}} e^{-u} \right)^{-1} \leq \frac{3}{4}.
\]

This is satisfied for \( v_0 = 2\sqrt{u} \). Choosing \( \xi_0 \leq \pi(1 \vee u)^{-3/2}/24 \), since \( u = e^{\xi^2/2} \), \( \xi_0 = \frac{(1/8)}{(1 \vee x)^{-5/2}} \) is sufficient and \( e_0(x) = (1 \vee x)^{-3/8} \) works. However, to satisfy (G), we reduce the rate to \((1 \vee x)^{-4/7}\). Lemma 3 below gives that this rate is sufficient. \( \square \)

**Lemma 3** Condition (G) is satisfied by the lognormal frailty model with \( e_0(x) = (1 \vee x)^{-4/8} \) and \( c_1(\gamma) = 1 \).

**Proof.** For the first part of (G), let \( W_k \equiv e_{\gamma_{\nu_k}}^{1/2} z_{-\nu_k/2, \gamma} \), where \( Z \sim N(0, 1) \). Now,

\[
\sup_{u \geq 0} u \Lambda_{\gamma_{\nu_k}}(u) = \sup_{u \geq 0} E \left[ u \exp (-u W_k) \right] \leq E \left[ \sup_{u \geq 0} u e^{-u W_k} \right] \leq E \left[ W_k^{-1} \right],
\]

\[
\sup_{u \geq 0} u^2 \Lambda_{\gamma_{\nu_k}}(u) = \sup_{u \geq 0} E \left[ u^2 W_k \exp (-u W_k) \right] \leq E \left[ \sup_{u \geq 0} u^2 W_k e^{-u W_k} \right] \leq E \left[ W_k^{-1} \right],
\]

and it follows since \( E \left[ W_k^{-1} \right] = e^{\gamma} \). For the second part, if \( x_k \to \infty \) and \( \gamma_k \to 0 \), then \( \gamma_k \) consists of one or both of two subsequences, one \( \leq 0 \) and one \( \geq 0 \). Without loss of generality, assume \( \gamma_k \to 0 \) from above or below but not both. We begin with a sequence approaching from below. Let \( \xi \equiv \sqrt{(\gamma_k)} \),

\[
\xi_k \equiv \sqrt{|\gamma_k|}, \; u \equiv x e^{\xi^2/2}, \; u_k \equiv x_k e^{\xi^2/2}, \; \text{and reparameterize } x \inf_{w \in [0, x]} \mathcal{G}_\gamma(w), \text{ for } \gamma < 0, \text{ as }
\]

\[
x \inf_{w \in [0, x]} \mathcal{G}_\gamma(w) = u \inf_{w \in [0, u]} \frac{\int_{\Re} e^{-w \cos \xi v} \cos (w \sin \xi v - \xi v) \phi(v) dv}{\int_{\Re} e^{-w \cos \xi v} \cos (w \sin \xi v) \phi(v) dv} = u \inf_{w \in [0, x]} g_k(w).
\]

For \( v > v_k \equiv u_k^{-3/3} \), the \( u^2 \) term in \( \phi(v) \) completely dominates \( u_k \) since \( v_k^2/\nu_k \to \infty \). Since \( u_k \xi_k v_k \leq u_k^{-1/3} \to 0 \),

\[
\inf_{w \in [0, u_k]} g_k(w) = \Theta(1) + \inf_{w \in [0, u_k]} \frac{\int_{\Re} e^{-w (1 - \cos \xi_k v)} \cos (w \sin \xi_k v - \xi_k v) \phi(v) dv}{\int_{\Re} e^{-w (1 - \cos \xi_k v)} \cos (w \sin \xi_k v) \phi(v) dv} \to 1.
\]

Hence \( u_k \inf_{w \in [0, u_k]} g_k(w) \to \infty \).

Now assume that \( \gamma_k \to 0 \) from above,

\[
\frac{-x \Lambda_{\gamma}(x)}{\Lambda_{\gamma}(x)} = \frac{x e^{\sqrt{\gamma} - \gamma/2} \exp \{ -x e^{\sqrt{\gamma} - \gamma/2} \} \phi(v) dv}{\int_{\Re} \exp \{ -x e^{\sqrt{\gamma} - \gamma/2} \} \phi(v) dv} = \left[ \frac{\int_{\Re} e^{-w \xi_k} \phi_{\gamma_k}(w) dw}{\int_{\Re} \xi_k(w) \phi_{\gamma_k}(w) dw} \right]^{-1}.
\]

17
where \( u \equiv x e^{-\gamma/2} \), \( \zeta_u(w) \equiv uw \exp\{-uw\} \), and \( \phi_\gamma(w) \equiv \gamma^{-1/2} \phi(\gamma^{-1/2} w) \). Since \( e^{-w} \) is a decreasing function, for any \( w_k \to \infty \),

\[
\frac{\int_{-w_k}^{w} -w \phi_\gamma(w)dw}{\int_{-w_k}^{w} \phi_\gamma(w)dw} \leq \frac{\int_{-w_k}^{w} -w \phi_\gamma(w)dw}{\int_{-w_k}^{w} \phi_\gamma(w)dw} = \frac{\int_{-w_k}^{w} -w \phi_\gamma(w)dw}{\int_{-w_k}^{w} \phi_\gamma(w)dw}.
\]

(13)

Now denote \( u_k \equiv x_k e^{-\gamma_k/2} \) and let \( w_k = \log(1 + \gamma_k u_k + u_k^{1/2}) \). Since

\[
\gamma_k^{-1} \left( w_k + u_k e^{-w_k} \right)^{-1} = w_k \left( \gamma_k u_k + \frac{\gamma_k u_k}{1 + \gamma_k u_k + u_k^{1/2}} \right)^{-1} \to \infty,
\]

for all \( w \geq w_k \), the \( \phi_\gamma(w) \) term dominates the expectation in (13) with \( u = u_k \) and \( \gamma = \gamma_k \). Hence, with this substitution, (13)\( = 1 + \gamma_k u_k + u_k^{1/2} + O(1) \), and \( x_k \hat{G}_{\gamma_k}(x_k) \to \infty \). The same arguments work when \( \gamma_k \to \gamma > 0 \), except that \( \liminf_{k \to \infty} \hat{G}_{\gamma_k}(x_k) \geq 1/\gamma \). Condition (G) follows since (F) implies for \( \gamma \geq 0 \) that \( \inf_{w \in [0,1]} \hat{G}_\gamma(w) = \hat{G}_\gamma(x) \).

A.2 Additional Results and Proofs

Proof of theorem 1. Let \( P_\psi \) be the distribution of a single observation from the model with parameter \( \psi \), \( dP_\psi \) be the density, and \( P_0 \equiv P_{\psi_0} \). In consequence of condition (D), let \( K_0 < \infty \) be the maximum possible value of both \( e^{\beta Z(t)} \) and \( e^{-\beta Z(t)} \) over \( t \in [0, \tau] \). Under condition (E), the parameter space for \( \gamma \) may be enlarged to \( [-\epsilon_0(K_0 A(\tau)), M_0] \), where \( \epsilon_0(\cdot) \) is defined in appendix A.1. This interval has 0 as an interior point. Lemma 4 below gives that, as \( n \to \infty \), \( \hat{A}_n(\tau) \) is asymptotically bounded and equicontinuous outer almost surely. This means the Helly selection theorem can be used to show that for any subsequence of \( \{n\} \), there exists a further subsequence \( \{n_k\} \) such that \( \hat{\psi}_{n_k} \) converges uniformly to some \( \psi \) with \( \beta \in B_0 \), \( A \) continuous, \( A(\tau) < \infty \), and \( \gamma \in [-\epsilon_0(K_0 A(\tau)), M_0] \). Define \( d\hat{P}_\psi \) to be \( dP_\psi \) with \( \alpha \) replaced by \( \Delta A \). By proposition 2 below,

\[
0 \leq \int \log \frac{d\hat{P}_{\psi_{n_k}}}{d\hat{P}_{\psi_{n_k}}} dP_{n_k} = \hat{L}_{n_k}(\psi_{n_k}) - \hat{L}_{n_k}(\psi_{n_k}) \to \int \log \frac{dP_{\psi}}{dP_0} dP_0,
\]

(14)

where \( \psi_n \equiv (\gamma_0, \beta_0, A_n) \) and

\[
A_n(t) \equiv \int_0^t \left( \mathbb{P}_n \left[ Y(s) e^{\beta Z(s)} \left( \hat{G}_{\gamma_0} \{ H^{\psi_0}(X) \}^\beta + \delta \frac{\hat{G}_{\gamma_0} \{ H^{\psi_0}(X) \}}{\hat{G}_{\gamma_0} \{ H^{\psi_0}(X) \}} \right) \right] \right) d\mathbb{P}_n \{ N(s) \}.
\]

Hence \( \int \log \frac{dP_{\psi_n}}{dP_0} dP_0 \leq 0 \). The nonnegativity of the Kullback-Leibler information implies \( \int \log \frac{dP_{\psi_n}}{dP_0} dP_0 = 0 \). Hence \( P_{\psi} = P_0 \) on the \( \sigma \)-algebra from a single observation, and \( \psi = \psi_0 \) by lemma 1. Since this is true for every convergent subsequence, the proof is complete.\( \square \)
Lemma 4 \( \dot{A}_n : [0, \tau] \to [0, \infty) \) is asymptotically bounded and equicontinuous.

Proof. Assume that \( \limsup_{n \to \infty} \dot{A}_n(\tau) = \infty \). Now suppose \( \dot{\gamma}_n \) has an accumulation point at \( \gamma < 0 \). But, this is impossible by (E). Thus \( \dot{\gamma}_n \) has no accumulation points < 0. Suppose \( \ddot{\gamma}_n \) has 0 as an accumulation point. Take a subsequence \( \{ n_k \} \) such that \( \dot{\gamma}_{n_k} \to 0 \) and \( \dot{A}_{n_k}(\tau) \to \infty \). Since by (E), \( G_{\gamma}(x) - \dot{G}_{\gamma}(x) \left\{ G_{\gamma}(x) \right\}^{-1} = -\dot{\Lambda}_{\gamma}(x) \left\{ \Lambda_{\gamma}(x) \right\}^{-1} \geq 0 \), we have that

\[
\dot{A}_{n_k}(\tau) = \int_0^\tau \left( \mathbb{P}_n \left[ Y(t) e^{\nabla_{\dot{\phi}_n} s(t)} \left( \dot{G}_{\gamma_{n_k}} H_{\dot{\phi}_n}(x) - \delta \frac{\dot{G}_{\gamma_{n_k}} H_{\dot{\phi}_n}(x)}{\dot{G}_{\gamma_{n_k}} H_{\dot{\phi}_n}(x)} \right) \right] \right) d\mathbb{P}_n\{N(t)\}
\leq O(1) \inf_{u \in [0, \ddot{\gamma}_n(\tau)]} \dot{G}_{\gamma_{n_k}}(u)^{-1}.
\]

Thus, by (G), \( 1 \leq O(1) \inf_{u \in [0, \ddot{\gamma}_n(\tau)]} \dot{G}_{\gamma_{n_k}}(u)^{-1} \to 0 \), which is a contradiction. Hence for any subsequence with \( \dot{A}_{n_k}(\tau) \to \infty \), the accumulation points of \( \ddot{\gamma}_{n_k} \) are > 0. Now let \( \{ n_k \} \) be a subsequence with \( \dot{A}_{n_k}(\tau) \to \infty \), \( \dot{\gamma}_{n_k} \to \gamma > 0 \), and \( P_{n_k} N \to \mu_0 = p_0 N \) uniformly on \( t \in [0, \tau] \).

Let \( \nu_n \equiv (\ddot{\gamma}_n, \dot{\beta}_n, \mathbb{F}_n N) \). Then

\[
0 \leq \dot{L}_n(\dot{\psi}_n) - \dot{L}_n(\psi_n)
= \mathbb{P}_n \left\{ \int_0^\tau \log \left( \frac{\dot{G}_{\gamma}(H_{\dot{\psi}_n}(s))}{G_{\gamma}(H_{\dot{\psi}_n}(s))} \right) + \log(\nabla \dot{A}_n(s)) \right\} dN(s) + \log \left( \frac{\Lambda_{\gamma}(H_{\dot{\psi}_n}(\tau))}{\Lambda_{\gamma}(H_{\dot{\psi}_n}(\tau))} \right) \]
\leq O(1) + \mathbb{P}_n \left\{ \int_0^\tau \log(\nabla \dot{A}_n(s)) dN(s) - \delta (1 + c_1(\gamma) \log H_{\dot{\psi}_n}(\tau) - c_1(\gamma) \log \dot{A}_n(X) \right\}
\leq O(1) + \mathbb{P}_n \left\{ \int_0^\tau \log(\nabla \dot{A}_n(s)) dN(s) - (\delta + c_1(\gamma)) \log \dot{A}_n(X) \right\},
\]

since (G) implies \( \log \dot{G}_{\gamma}(H_{\dot{\psi}_n}(s)) \leq -(1 + c_1(\gamma)) \log H_{\dot{\psi}_n}(s) + O(1) \) and \( G_{\gamma}(H_{\dot{\psi}_n}(\tau)) \geq c_1(\gamma) \log \dot{A}_n(X) + O(1) \). For any sequence \( 0 = x_0 < x_1 < x_2 < \cdots < x_K = \tau \), let \( N^k(s) \equiv N(s) I\{X \in [x_{k-1}, x_k]\}, \]
\( k = 1, \ldots, K \). By Jensen's inequality,

\[
\int_0^\tau \log(\nabla \dot{A}_n(s)) d\mathbb{P}_n\{N^k(s)\} \leq \mathbb{P}_n N^k(\tau) \log \left( \int_0^{x_k} n \Delta \dot{A}_n(s) d\mathbb{P}_n N^k(s) / \mathbb{P}_n N^k(\tau) \right)
\leq O(1) + \log(\dot{A}_n(x_k)) \mathbb{P}_n(\delta I\{X \in [x_{k-1}, x_k]\})
\]

Thus (15) is dominated by

\[
O(1) + \sum_{k=1}^{K-1} \log(\dot{A}_n(x_k)) \mathbb{P}_n[\delta I\{X \in [x_{k-1}, x_k]\} - (c_1(\gamma) + \delta) I\{X \in [x_k, x_{k+1}]\}].
\]
Choose $\epsilon : 0 < \epsilon < P_0\{X = \tau\}$ and the sequence $\{x_k\}$ for finite $K$ such that $\epsilon/(c_1(\gamma) + 1)/\mu_0(x_{k+1}) - \mu_0(x_k) < \epsilon$ for $k = 0 \ldots K - 2$ and $0 < \mu_0(\tau) - \mu_0(K - 1) < c_1(\gamma)\epsilon/(c_1(\gamma) + 1)$. Since $\mathbb{P}_n N \to \mu_0$ uniformly, (16) goes to $-\infty$. This is a contradiction. Thus, $\limsup \hat{A}_n(\tau) < \infty$.

By previous arguments and (5), for $s, t \in [0, \tau]$, $|\hat{A}_n(s) - \hat{A}_n(t)| \leq O(1)\mathbb{P}_n|N(s) - N(t)|$, giving asymptotic equicontinuity of $\hat{A}_n$. □

**Proposition 2** Expression (14) is satisfied.

**Proof.** Assuming (E) and (F), the only difficult part is to establish that

$$
\int_0^T \log \left( \Delta \hat{A}_{nk}(t)/\Delta \hat{A}_{nk}(t) \right) d\mathbb{P}_n\{N(t)\} \to \int_0^T \log(dA(t)/dA_0(t))d\mu_0(t),
$$

where $\mu_0 \equiv P_0N$ as in lemma 4. However, for $t$ on the set of observed failure times,

$$
\frac{d\hat{A}_{nk}(t)}{dA_{nk}(t)} = \frac{\mathbb{P}_n \left[ Y(t)e^{\mathcal{T}_k \mathcal{Z}(t)} \left( \hat{G}_{\gamma_{nk}} \left\{ H^\psi_{\gamma_{nk}}(X) \right\} - \delta \frac{\hat{G}_{\gamma_{nk}} \left\{ H^\psi_{\gamma_{nk}}(X) \right\}}{\hat{G}_{\gamma_{nk}} \left\{ H^\psi_{\gamma_{nk}}(X) \right\}} \right] \right]}{\mathbb{P}_n \left[ Y(t)e^{\mathcal{T}_k \mathcal{Z}(t)} \left( \hat{G}_{\gamma_{nk}} \left\{ H^\psi(\mathcal{W}) \right\} - \delta \frac{\hat{G}_{\gamma_{nk}} \left\{ H^\psi(\mathcal{W}) \right\}}{\hat{G}_{\gamma_{nk}} \left\{ H^\psi(\mathcal{W}) \right\}} \right] \right]}
$$

uniformly, by conditions (E) and (F), where (17) is $dA(t)/dA_0(t)$. Since $dA(t)/dA_0(t)$ is bounded above and below by (F), we have that $\int_0^T \log \left\{ dA(t)/dA_0(t) \right\} dN(t)$ is $P_0$-Glivenko-Cantelli, and the desired result follows. □

**Proof of theorem 2.** We use Hoffmann-Jörgensen weak convergence as described in van der Vaart and Wellner (1996) (hereafter abbreviated VW). Define the one-dimensional submodels $t \mapsto \psi_1 \equiv \psi + t \{ h_1, h_2, \int_0^t h_3(s)dA(s) \}$, where $(h_1, h_2, h_3) \in H_p$ for some $p < \infty$, and where $H_p$ is the space of elements $h = (h_1, h_2, h_3)$ such that $h_1 \in \mathbb{R}$, $h_2 \in \mathbb{R}^d$, $h_3$ is a cadlag (right-continuous with left-hand limits) function, and $|h_1| + \sqrt{h_2^T h_2} + \| h_3 \|_v \leq \rho$, with $\| \cdot \|_v$ being the total variation norm. In $H_p$, restrict $h_3$ to have variation only on $\mathcal{S}_r$. $\psi$ can be represented as a functional on $H_p$ of the form $\psi(h) = h_1\gamma + h_2^T \beta + \int_0^t h_3(s)dA(s)$, and the parameter space $\Psi$ is then a subset of $\ell^\infty(H_p)$ with

20
norm $\|\psi\| = \sup_{h \in H_{\rho}} |\psi(h)|$. In the set $\Psi$, restrict $A$ to have variation only on $S_{\tau}$. Let

$$U_{n}^{\tau}(\psi)(h) = \frac{\partial}{\partial t} L_{n}(\psi_{t}) \bigg|_{t=0}$$

$$= \mathbb{P}_{n} \left\{ \left[ \frac{\delta \hat{G}_{\gamma}^{(1)}(H_{\psi}(\tau))}{G_{\gamma}(H_{\psi}(\tau))} - \frac{\hat{G}_{\gamma}^{(1)}(H_{\psi}(\tau))}{G_{\gamma}(H_{\psi}(\tau))} \right] h_{1} + \int_{0}^{\tau} \left[ h_{2}^{T} Z(s) + h_{3}(s) \right] dN(s) \right. \right.$$

$$+ \left. \left[ \frac{\delta \hat{G}_{\gamma}(H_{\psi}(\tau))}{G_{\gamma}(H_{\psi}(\tau))} - \frac{\hat{G}_{\gamma}(H_{\psi}(\tau))}{G_{\gamma}(H_{\psi}(\tau))} \right] \int_{0}^{\tau} Y(s) e^{\int_{0}^{s} Z(s)} \left[ h_{2}^{T} Z(s) + h_{3}(s) \right] dA(s) \right\}$$

$$\equiv \mathbb{P}_{n} U_{\tau}^{\gamma}(\psi)(h).$$

The score operator $U_{\tau}^{\gamma}(\psi)$ has expectation $U_{\rho}^{\tau}(\psi) \equiv P_{0} U_{\tau}^{\gamma}(\psi)$.

By (F), the Fréchet derivative of $U_{\rho}^{\tau}(\psi)(h)$ at $\psi_{0}$ exists and is obtained by differentiating the score operator for the submodels $t \mapsto \psi_{0} + t \psi$. This derivative is

$$-\dot{U}_{\psi_{0}}(\psi)(h) \equiv -\frac{\partial}{\partial t} U_{\rho}^{\tau}(\psi_{t})(h) \bigg|_{t=0} = \psi(\sigma(h)),$$

where the operator $\sigma : H_{\infty} \mapsto H_{\infty}$ is

$$\sigma(h) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} h_{1} \\ h_{2} \\ h_{3} \end{pmatrix}.$$  \hspace{1cm} (19)

The operators $\sigma_{jk}$, for $1 \leq j, k \leq 3$, are well defined and bounded. More specifically, $\sigma_{13}(h_{3}) = \int_{0}^{\tau} f_{1}(s) h_{3}(s) dA_{0}(s)$, $\sigma_{23}(h_{3}) = \int_{0}^{\tau} f_{2}(s) h_{3}(s) dA_{0}(s)$, and $\sigma_{33}(h_{3}) = g_{1}(\cdot) \int_{0}^{\tau} f_{3}(s) h_{3}(s) dA_{0}(s) + g_{2}(\cdot) h_{3}(\cdot)$, where $f_{1}, f_{2}, f_{3}, g_{1} : \mathbb{R} \mapsto \mathbb{R}$ and $f_{2} : \mathbb{R} \mapsto \mathbb{R}^{3}$ are bounded, and where

$$g_{2}(s) \equiv P_{0} \left\{ Y(s) e^{\int_{0}^{\tau} Z(s)} \left[ \hat{G}_{\gamma_{\psi}}(H_{\psi}(\tau)) - \frac{\delta \hat{G}_{\gamma_{\psi}}(H_{\psi}(\tau))}{\hat{G}_{\gamma_{\psi}}(H_{\psi}(\tau))} \right] \right\}.$$

From the proof of lemma 4, $0 < g_{2}(s) < \infty$ for all $s \in [0, \tau]$. Thus $\sigma = \sigma^{(1)} + \sigma^{(2)}$, where

$$\sigma^{(1)}(h) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & g_{2}(\cdot) \end{pmatrix} \begin{pmatrix} h_{1} \\ h_{2} \\ h_{3} \end{pmatrix}$$

is continuously invertible, $\sigma^{(2)}$ = $\sigma - \sigma^{(1)}$ is compact, and I denotes the identity. Since $\sigma$ is one-to-one by lemma 2, $\sigma$ is continuously invertible, with inverse $\sigma^{-1}$. This implies that for each $p > 0$, there is a $q > 0$ with $\sigma^{-1}(H_{q}) \subset H_{p}$.

Continuous invertibility of $-\dot{U}_{\psi_{0}}$ on $\text{lin}\Psi$, where $\text{lin}$ denotes linear span, follows by Prop. A.1.7 of BKRW since

$$\inf_{\psi \in \text{lin}\Psi} \frac{\|\dot{U}_{\psi_{0}}(\psi)\|_{p}}{\|\psi\|_{p}} \geq \inf_{\psi \in \text{lin}\Psi} \frac{\sup_{h \in \sigma^{-1}(H_{q})} |\psi(\sigma(h))|}{\|\psi\|_{p}} = \frac{\|\psi\|_{q}}{\inf_{\psi \in \text{lin}\Psi} \|\psi\|_{p}} \geq \frac{q}{3p}.$$
This proposition also implies that \( \sigma \) is onto. By proposition 3 below,

\[
\sqrt{n} \left( U_n^+(\hat{\psi}_n)(h) - U_0^+(\hat{\psi}_n)(h) \right) - \sqrt{n} \left( U_n^-(\psi_0) - U_0^-(\psi_0) \right) = o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|) \tag{20}
\]

uniformly over \( h \in H_p \). Theorem 3.3.1 of VW now gives the desired weak convergence of \( \sqrt{n} (\hat{\psi}_n - \psi_0) \), provided \( U_0^+(\psi_0)(\cdot) = 0, U_n^+(\hat{\psi}_n)(\cdot) = 0 \) asymptotically, and \( U^+(\psi_0)(\cdot) \) is \( P_0 \)-Donsker. The first two conditions follow from lemma 1 and the fact that \( \hat{\psi}_n \) is asymptotically an interior point. Because products of bounded \( P_0 \)-Donsker classes are \( P_0 \)-Donsker, showing \( \delta h_3(X) \) and \( \int_0^T Y(s) e^{\beta T} Z(s) h_3(s) dA_0(s) \) are \( P_0 \)-Donsker is sufficient. This follows because \( Z \) and \( \delta A_0 \) are uniformly bounded in total variation.

Using arguments similar to Parmer (1998), the above results imply that \( \hat{\psi}_n \) is asymptotically linear with influence function \( \hat{\ell}(h) = U^+(\psi_0)(\sigma^{-1}(h)) \) contained in the closed linear span of the tangent space of the score operator \( U^+(\psi_0)(h) \) since \( \sigma \) is continuously invertible. Thus \( \hat{\psi}_n \) is regular by theorem 5.2.3 of BKRW and efficient by theorem 5.2.1 of BKRW. The covariance of \( \hat{\ell}(\cdot) \) has the form \( \psi_0^+(\sigma^{-1}(g)) \), where \( \psi_0^+ \equiv \{ h_1, h_2, \int_0^T h_3(s) dA_0(s) \} \) and \( g, h \in H_p \). Taking \( p \) large enough yields weak convergence in the uniform metric. \( \square \)

**PROPOSITION 3** Expression (20) holds.

**Proof.** Let \( \| \cdot \| \) denote the uniform metric. If for some \( \epsilon > 0, \{ U^+(\psi)(h) - U^+(\psi_0)(h) : \| \psi - \psi_0 \| < \epsilon, h \in H_p \} \) is \( P_0 \)-Donsker and \( \lim_{\psi \to \psi_0} \sup_{h \in H_p} P_0 \{ U^+(\psi)(h) - U^+(\psi_0)(h) \}^2 = 0 \), then (20) holds by lemma 3.3.5 of VW. The latter condition follows from (F). The Donsker condition requires more work. Let \( \Psi_\epsilon = \{ \psi : \| \psi - \psi_0 \| \leq \epsilon \} \). Take \( \epsilon \) small enough so that \( \Psi_\epsilon \subset \Psi \). Such an \( \epsilon \) always exists by (E). Because \( Z \) has bounded total variation, \( \exp(\cdot) \) is Lipschitz on compacts, and because \( \{ A : (A, \beta, A) \in \Psi_\epsilon \} \) is uniformly bounded in total variation, \( \int_0^T Y(s) e^{\beta T} Z(s) dA(s) \) is \( P_0 \)-Donsker as a process in \( L^\infty(\Psi_\epsilon) \). By (E) and (F), \( G_\gamma(x), \bar{G}_\gamma(x), G^{(1)}_\gamma(x), \bar{G}^{(1)}_\gamma(x), \) and \( \left[ G_\gamma(x) \right]^{-1} \) are Lipschitz in \( \gamma \) and \( x \) over the appropriate range. Thus

\[
\begin{bmatrix}
\delta G^{(1)}_\gamma(H^\psi(\tau)) - G^{(1)}_\gamma(H^\psi(\tau)) \\
\delta G_\gamma(H^\psi(\tau)) - G_\gamma(H^\psi(\tau))
\end{bmatrix}
\]

are \( P_0 \)-Donsker as processes in \( L^\infty(\Psi_\epsilon) \). Similar results and the fact that both sums of Donsker classes and products of bounded Donsker classes are Donsker gives the result. \( \square \)

**Proof of theorem 3.** If we establish the conditions of theorem 1 and corollary 1 of Murphy and van der Vaart (2000) (hereafter MV), the result is a consequence of the following argument. Let \( K_n(\hat{\theta}_n) \equiv \{ \theta : \sqrt{n}(\theta - \hat{\theta}_n) \in K \} \). Since \( K_n(\hat{\theta}_n) \) is compact, for each \( n \geq 1 \) there exists \( \hat{\theta}_n \in K_n(\hat{\theta}_n) \).
such that the supremum in (7) is attained at \( \hat{\theta}_n \). Since \( \hat{\theta}_n \) is consistent for \( \theta_0 \), by corollary 1 of MV,

\[
\sup_{\theta \in K_{\epsilon}(\hat{\theta}_n)} \left| -\frac{1}{2} n(\theta - \hat{\theta}_n)^T I_n(\theta) - \frac{1}{2} p L_n(\theta) - p L(\hat{\theta}_n) \right| = o_P \left( \sqrt{n} \| \hat{\theta}_n - \theta_0 \| + 1 \right)^2 ,
\]

and (7) follows.

We now check the conditions. From the proof of theorem 2, the least favorable direction for estimating \( \theta \) is the \( d + 1 \)-vector of functions \( g_0 : \mathbb{R} \to \mathbb{R}^{d+1} \) defined for \( h = (h_1, h_2) \), with \( h_1 \in \mathbb{R} \) and \( h_2 \in \mathbb{R}^d \), as \( h^T g_0 = \sigma^{-1}_{33} \{ \sigma_{31} h_1 + \sigma_{32} h_2 \} \). The components are as defined in (19) and \( \sigma^{-1}_{33} \) exists by the linearity and continuous invertibility of \( \sigma \). Let \( A_t(\theta, A) \equiv f_0^{(1)} \{ 1 + (\theta - t)^T g_0(s) \} dA(s) \), where \( t \in \mathbb{R}^{d+1} \). Now define \( I(t, \theta, A) \equiv \tilde{L}\{t, A_t(\theta, A)\} \), where \( \tilde{L}(\psi) \) is the log-likelihood (using \( \Delta A \) in place of \( a \)) of a single observation \((X, \delta, Z)\).

Let \( \hat{I}(t, \theta, A) \equiv \partial I(t, \theta, A)/(\partial t) \) and \( \tilde{I}(t, \theta, A) \equiv \partial \hat{I}(t, \theta, A)/(\partial t) \). For \( h \in \mathbb{R}^{d+1} \),

\[
h^T \hat{I}(t, \theta, A) = U^* \begin{pmatrix} t & h \\ A_t(\theta, A) \end{pmatrix} \begin{pmatrix} h^T g_0/[1 + (\theta - t)^T g_0] \end{pmatrix} .
\]

The variance \( \hat{I}_0 \) of the efficient score \( \hat{I}(\theta_0, \theta_0, A_0) \) is finite and invertible by the proof of theorem 2. Thus, by the consistency of \( \hat{\psi}_n \), corollary 1 of MV follows from their theorem 1. The arguments from our proof of theorem 2 yield a neighborhood \( D \) of \((\theta_0, \theta_0, A_0)\) such that \( \{ \hat{I}(t, \theta, A) : (t, \theta, A) \in D \} \) is \( P_0\)-Donsker with square integrable envelope and such that \( \{ \hat{I}(t, \theta, A) : (t, \theta, A) \in D \} \) is \( P_0\)-Glivenko-Cantelli and bounded in \( L_1(P_0) \).

The proof is complete if for any \( \hat{\theta}_n \to \theta_0 \) in outer probability, \( \hat{A}_{\hat{\theta}_n} \to A_0 \) in outer probability and

\[
P_0 I(\theta_0, \hat{\theta}_n, \hat{A}_{\hat{\theta}_n}) = o_P(\| \hat{\theta}_n - \theta_0 \| + n^{-1/2}) .
\]

This is implied by

\[
\| \hat{A}_{\hat{\theta}_n} - A_0 \| = O_P(\| \hat{\theta}_n - \theta_0 \| + n^{-1/2}) .
\]

(21)

Here, \( O_P \) represents a quantity bounded in outer probability. Note that \( \tilde{L}_n(\hat{\theta}_n, \hat{A}_{\hat{\theta}_n}) \geq \tilde{L}_n(\hat{\theta}_n, \hat{A}_{\hat{\theta}_n}) \) since \( \hat{\theta}_n \) and \( \hat{A}_{\hat{\theta}_n} \) are consistent, \( \hat{A}_{\hat{\theta}_n} \) is consistent by lemma 1. Equation (21) is satisfied by proposition 4 below. \( \square \)

**Proposition 4** Expression (21) holds.

**Proof.** Denote \( \hat{A}_n \equiv \hat{A}_{\hat{\theta}_n} \) and \( \hat{\psi}_n \equiv (\hat{\theta}_n, \hat{A}_n) \). For any cadlag function \( h \) bounded in total variation, let

\[
\hat{D}_n(A)(h) \equiv \mathbb{P}_n U^* \begin{pmatrix} \hat{\theta}_n \\ A \end{pmatrix} \begin{pmatrix} 0 \\ h \end{pmatrix} , \quad D_n(A)(h) \equiv \mathbb{P}_n U^* \begin{pmatrix} \theta_0 \\ A \end{pmatrix} \begin{pmatrix} 0 \\ h \end{pmatrix} , \quad \text{and}
\]

23
\[ D_0(A)(h) = P_0 U^\top \begin{pmatrix} \theta_0 \\ A \\ 0 \\ h \end{pmatrix}. \]

Note that \( \hat{D}_n(\hat{A}_n) = 0 \) and \( D_0(A_0) = 0 \). Using lemma 3.3.5 of VW, as in the proof of prop. 3, we obtain \( \sqrt{n}(\hat{D}_n - D_0)(\hat{A}_n) = \sqrt{n}(\hat{D}_n - D_0)(A_0) = o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|) \) and

\[
\sqrt{n}(\hat{D}_n - D_n)(A_0) = \sqrt{n}(\hat{D}_n - D_n)(A_0) - \sqrt{n}(D_n - D_0)(A_0) = o_P(1 + \sqrt{n}\|\hat{\theta}_n - \theta_0\|).
\]

Thus

\[
\sqrt{n}(D_0(\hat{A}_n) - D_0(A_0)) = \sqrt{n}(D_0(\hat{A}) - \hat{D}_n(\hat{A}_n)) = -\sqrt{n}(\hat{D}_n - D_0)(A_0) + o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|) = -\sqrt{n}(D_n - D_0)(A_0) + o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|) = O_P(1) + o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|) \tag{22}
\]

by weak convergence of \( \sqrt{n}(P_n - P_0) U^\top(\psi_0) \). Since \( \sigma \) is continuously invertible, for some \( c > 0 \), \( \|D_0(A) - D_0(A_0)\| \geq c\|A - A_0\| + o(\|A - A_0\|) \). Thus (22) implies \( \sqrt{n}\|\hat{A}_n - A_0\| \leq O_P(1) + o_P(1 + \sqrt{n}\|\hat{\psi}_n - \psi_0\|) \). Expression (21) follows. \( \square \)

**Proof of corollary 1.** We first prove (ii). Using arguments from the proof of theorem 2 and applying theorem 3.3.1 of VW gives that \( \sqrt{n}(\hat{\psi}_n^0 - \psi_0) = \sqrt{n}P_n U^\top(\psi_0)(\sigma^{-1}(\cdot)) + o_P(1) \) unconditionally, where \( o_P \) denotes a quantity \( \to 0 \) in outer probability. Since \( \sqrt{n}(\hat{\psi}_n^0 - \hat{\psi}_n) = \sqrt{n}(P_n - P_n)U^\top(\psi_0)(\sigma^{-1}(\cdot)) + o_P(1) \) unconditionally, (ii) follows by theorem 2.9.6 of VW. Similar arguments establish (i), but theorem 3.6.1 of VW is used in place of theorem 2.9.6. \( \square \)

**REFERENCES**


*Biometrics*, 47, 461–466.


Non-Hodgkin’s Lymphoma: The International Non-Hodgkin’s Lymphoma Prognostic Factors 


Sargent, D. J. (1997), “A Flexible Approach to Time-Varying Coefficients in the Cox Regression 


Table 1: Results from simulation studies—estimation in the odds-rate model with the odds-rate parameter unspecified (Case 1). Corresponding results for the mode estimators are given in parentheses.

<table>
<thead>
<tr>
<th>n</th>
<th>γ</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\sigma}_0$</th>
<th>$\hat{\sigma}_0^*$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_1^*$</th>
<th>$\hat{\sigma}_1$</th>
<th>$\hat{\sigma}_1^*$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\sigma}_2$</th>
<th>$\hat{\sigma}_2^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.00</td>
<td>0.275</td>
<td>0.384</td>
<td>0.389</td>
<td>1.212</td>
<td>0.540</td>
<td>0.561</td>
<td>1.195</td>
<td>0.354</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.072)</td>
<td>(0.206)</td>
<td>(1.093)</td>
<td>(0.457)</td>
<td>(0.1087)</td>
<td>(0.283)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.419</td>
<td>0.468</td>
<td>0.375</td>
<td>1.126</td>
<td>0.570</td>
<td>0.550</td>
<td>1.132</td>
<td>0.374</td>
<td>0.344</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.181)</td>
<td>(0.279)</td>
<td>(1.012)</td>
<td>(0.488)</td>
<td>(1.023)</td>
<td>(0.292)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.611</td>
<td>0.576</td>
<td>0.378</td>
<td>1.032</td>
<td>0.613</td>
<td>0.553</td>
<td>1.032</td>
<td>0.396</td>
<td>0.340</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.332)</td>
<td>(0.451)</td>
<td>(0.912)</td>
<td>(0.507)</td>
<td>(0.910)</td>
<td>(0.331)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.876</td>
<td>0.730</td>
<td>0.466</td>
<td>0.948</td>
<td>0.673</td>
<td>0.605</td>
<td>0.927</td>
<td>0.425</td>
<td>0.404</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.447)</td>
<td>(0.678)</td>
<td>(0.802)</td>
<td>(0.561)</td>
<td>(0.796)</td>
<td>(0.416)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.261</td>
<td>0.927</td>
<td>0.914</td>
<td>0.831</td>
<td>0.749</td>
<td>0.774</td>
<td>0.858</td>
<td>0.476</td>
<td>0.406</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.695)</td>
<td>(1.410)</td>
<td>(0.674)</td>
<td>(0.761)</td>
<td>(0.699)</td>
<td>(0.467)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.00</td>
<td>0.121</td>
<td>0.238</td>
<td>0.220</td>
<td>1.105</td>
<td>0.342</td>
<td>0.353</td>
<td>1.092</td>
<td>0.231</td>
<td>0.239</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.153)</td>
<td>(1.035)</td>
<td>(0.312)</td>
<td>(1.027)</td>
<td>(0.211)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.340</td>
<td>0.333</td>
<td>0.290</td>
<td>1.108</td>
<td>0.390</td>
<td>0.380</td>
<td>1.038</td>
<td>0.260</td>
<td>0.230</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.185)</td>
<td>(0.253)</td>
<td>(1.031)</td>
<td>(0.346)</td>
<td>(0.965)</td>
<td>(0.217)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.546</td>
<td>0.427</td>
<td>0.360</td>
<td>1.019</td>
<td>0.423</td>
<td>0.397</td>
<td>0.990</td>
<td>0.280</td>
<td>0.276</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.342)</td>
<td>(0.359)</td>
<td>(0.933)</td>
<td>(0.383)</td>
<td>(0.909)</td>
<td>(0.281)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.877</td>
<td>0.592</td>
<td>0.429</td>
<td>0.862</td>
<td>0.469</td>
<td>0.425</td>
<td>0.940</td>
<td>0.313</td>
<td>0.286</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.596)</td>
<td>(0.555)</td>
<td>(0.778)</td>
<td>(0.438)</td>
<td>(0.851)</td>
<td>(0.306)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.262</td>
<td>0.796</td>
<td>0.912</td>
<td>0.838</td>
<td>0.528</td>
<td>0.542</td>
<td>0.816</td>
<td>0.345</td>
<td>0.380</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.811)</td>
<td>(1.311)</td>
<td>(0.718)</td>
<td>(0.570)</td>
<td>(0.688)</td>
<td>(0.495)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.00</td>
<td>0.032</td>
<td>0.137</td>
<td>0.123</td>
<td>1.024</td>
<td>0.203</td>
<td>0.200</td>
<td>1.031</td>
<td>0.143</td>
<td>0.143</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.021)</td>
<td>(0.117)</td>
<td>(0.991)</td>
<td>(0.196)</td>
<td>(0.999)</td>
<td>(0.138)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.270</td>
<td>0.208</td>
<td>0.200</td>
<td>1.066</td>
<td>0.237</td>
<td>0.238</td>
<td>0.997</td>
<td>0.165</td>
<td>0.148</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.196)</td>
<td>(0.197)</td>
<td>(0.970)</td>
<td>(0.234)</td>
<td>(0.961)</td>
<td>(0.153)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.531</td>
<td>0.286</td>
<td>0.266</td>
<td>0.977</td>
<td>0.264</td>
<td>0.245</td>
<td>0.999</td>
<td>0.186</td>
<td>0.176</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.429)</td>
<td>(0.256)</td>
<td>(0.938)</td>
<td>(0.247)</td>
<td>(0.959)</td>
<td>(0.180)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.952</td>
<td>0.427</td>
<td>0.382</td>
<td>0.973</td>
<td>0.311</td>
<td>0.289</td>
<td>0.973</td>
<td>0.211</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.794)</td>
<td>(0.411)</td>
<td>(0.925)</td>
<td>(0.289)</td>
<td>(0.927)</td>
<td>(0.219)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.444</td>
<td>0.646</td>
<td>0.757</td>
<td>0.890</td>
<td>0.352</td>
<td>0.359</td>
<td>0.858</td>
<td>0.240</td>
<td>0.279</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.138)</td>
<td>(1.027)</td>
<td>(0.826)</td>
<td>(0.377)</td>
<td>(0.842)</td>
<td>(0.318)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Results from simulation studies—empirical coverage probabilities of 95% confidence intervals (Case 1). The Monte Carlo standard error is 0.015.

<table>
<thead>
<tr>
<th>n</th>
<th>(\hat{\gamma})</th>
<th>(\hat{\gamma})</th>
<th>(\hat{\beta}_1)</th>
<th>(\hat{\beta}_2)</th>
<th>(\hat{\gamma})</th>
<th>(\hat{\beta}_1)</th>
<th>(\hat{\beta}_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.00</td>
<td>1.00</td>
<td>0.98</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.99</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.97</td>
<td>0.97</td>
<td>0.92</td>
<td>0.90</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.90</td>
<td>0.95</td>
<td>0.95</td>
<td>0.72</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>200</td>
<td>0.00</td>
<td>0.99</td>
<td>0.96</td>
<td>0.98</td>
<td>0.99</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
<td>0.95</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.97</td>
<td>0.98</td>
<td>0.95</td>
<td>0.92</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.86</td>
<td>0.91</td>
<td>0.86</td>
<td>0.65</td>
<td>0.91</td>
<td>0.81</td>
</tr>
<tr>
<td>500</td>
<td>0.00</td>
<td>0.98</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.96</td>
<td>0.94</td>
<td>0.98</td>
<td>0.89</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.97</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.93</td>
<td>0.96</td>
<td>0.94</td>
<td>0.88</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.85</td>
<td>0.93</td>
<td>0.89</td>
<td>0.72</td>
<td>0.91</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Table 3: Results from simulation studies—estimation in the odds-rate model with fixed odds-rate parameter (Case 2).

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma$</th>
<th>$\tilde{\beta}_1$</th>
<th>$\tilde{\sigma}_1$</th>
<th>$\sigma_1^*$</th>
<th>R.E.</th>
<th>$\tilde{\beta}_2$</th>
<th>$\tilde{\sigma}_2$</th>
<th>$\sigma_2^*$</th>
<th>R.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.00</td>
<td>1.055</td>
<td>0.408</td>
<td>0.407</td>
<td>0.53</td>
<td>1.042</td>
<td>0.245</td>
<td>0.239</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.072</td>
<td>0.474</td>
<td>0.496</td>
<td>0.81</td>
<td>1.068</td>
<td>0.283</td>
<td>0.274</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>1.017</td>
<td>0.532</td>
<td>0.525</td>
<td>0.90</td>
<td>1.006</td>
<td>0.313</td>
<td>0.300</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.006</td>
<td>0.637</td>
<td>0.629</td>
<td>1.08</td>
<td>0.984</td>
<td>0.372</td>
<td>0.372</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>0.986</td>
<td>0.809</td>
<td>0.797</td>
<td>1.06</td>
<td>1.061</td>
<td>0.473</td>
<td>0.404</td>
<td>0.99</td>
</tr>
<tr>
<td>200</td>
<td>0.00</td>
<td>1.032</td>
<td>0.280</td>
<td>0.304</td>
<td>0.74</td>
<td>1.023</td>
<td>0.170</td>
<td>0.176</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.073</td>
<td>0.329</td>
<td>0.338</td>
<td>0.79</td>
<td>1.003</td>
<td>0.195</td>
<td>0.183</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>1.014</td>
<td>0.369</td>
<td>0.370</td>
<td>0.87</td>
<td>0.982</td>
<td>0.218</td>
<td>0.206</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.992</td>
<td>0.444</td>
<td>0.418</td>
<td>0.97</td>
<td>1.000</td>
<td>0.260</td>
<td>0.247</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.024</td>
<td>0.567</td>
<td>0.573</td>
<td>1.12</td>
<td>0.995</td>
<td>0.330</td>
<td>0.339</td>
<td>0.80</td>
</tr>
<tr>
<td>500</td>
<td>0.00</td>
<td>1.008</td>
<td>0.175</td>
<td>0.174</td>
<td>0.76</td>
<td>1.014</td>
<td>0.106</td>
<td>0.117</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.998</td>
<td>0.206</td>
<td>0.208</td>
<td>0.76</td>
<td>0.995</td>
<td>0.122</td>
<td>0.108</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.977</td>
<td>0.232</td>
<td>0.232</td>
<td>0.90</td>
<td>0.997</td>
<td>0.137</td>
<td>0.142</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.998</td>
<td>0.278</td>
<td>0.257</td>
<td>0.79</td>
<td>1.000</td>
<td>0.164</td>
<td>0.157</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.025</td>
<td>0.355</td>
<td>0.357</td>
<td>0.99</td>
<td>0.988</td>
<td>0.207</td>
<td>0.208</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Table 4: Parameter estimates for the non-Hodgkin’s lymphoma data using the odds-rate model, with Cox model estimates given in italics below.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Parameter</th>
<th>Estimate</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(odds-rate)</td>
<td>$\gamma$</td>
<td>2.419</td>
<td>(1.482, 3.356)</td>
</tr>
<tr>
<td>age</td>
<td>$\beta_1$</td>
<td>1.091</td>
<td>(0.792, 1.390)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.671</td>
<td>(0.512, 0.830)</td>
</tr>
<tr>
<td>level</td>
<td>$\beta_2$</td>
<td>1.067</td>
<td>(0.743, 1.391)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.615</td>
<td>(0.440, 0.790)</td>
</tr>
<tr>
<td>status</td>
<td>$\beta_3$</td>
<td>1.355</td>
<td>(0.956, 1.754)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.584</td>
<td>(0.413, 0.756)</td>
</tr>
<tr>
<td>sites</td>
<td>$\beta_4$</td>
<td>0.710</td>
<td>(0.384, 1.036)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.395</td>
<td>(0.225, 0.566)</td>
</tr>
<tr>
<td>stage</td>
<td>$\beta_5$</td>
<td>0.614</td>
<td>(0.295, 0.933)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.384</td>
<td>(0.186, 0.581)</td>
</tr>
</tbody>
</table>
Figure 1: Estimated marginal survival distributions of non-Hodgkin's lymphoma patients for performance status groups (a) and tumor stage groups (b).
Figure 2: Diagnostic plots of the Markov chain in $\beta_1$ from the non-Hodgkin’s lymphoma data analysis.

(a) Plot of Markov chain against iteration number

(b) Plot of autocorrelations

(c) Histogram

Survival time in years

0 2 4 6 8 10