Semiparametric multi-parameter regression survival modelling

Burke, Kevin; Eriksson, Frank; Pipper, C. B.

Semiparametric multi-parameter regression survival modelling

Kevin Burke* Frank Eriksson† C. B. Pipper‡

Abstract

We consider a log-linear model for survival data, where both the location and scale parameters depend on covariates and the baseline hazard function is completely unspecified. This model provides the flexibility needed to capture many interesting features of survival data at a relatively low cost in model complexity. Estimation procedures are developed and asymptotic properties of the resulting estimators are derived using empirical process theory. Finally, a resampling procedure is developed to estimate the limiting variances of the estimators. The finite sample properties of the estimators are investigated by way of a simulation study, and a practical application to lung cancer data is illustrated.

Keywords. Counting processes; Empirical processes; Log-linear failure time model; Multi-parameter regression; Semiparametric regression; Survival data.

*Department of Mathematics and Statistics, University of Limerick; kevin.burke@ul.ie
†Section of Biostatistics, University of Copenhagen, Denmark.
‡Section of Biostatistics, University of Copenhagen, Denmark.

1 Introduction

In the context of survival analysis, we often consider log-linear models of the form $\log T = \mu + \sigma e$, where μ and σ are location and scale parameters, respectively, and e is a random error with an assumed parametric distribution on the real numbers. The familiar accelerated failure time model then arises by setting $\mu = -\beta^T X$ where β and X are, respectively, vectors of regression coefficients and covariates (Kalbfleisch and Prentice (2002, chap. 3) and Lawless (2003, chap. 6). As discussed in Burke and MacKenzie (2017), taking a multi-parameter regression approach (i.e., allowing both μ and σ to depend on covariates simultaneously) offers an intuitive and simple way of modelling complicated phenomena. For instance, the phenomenon of crossing survival curves is directly linked to the concentration of events at a given location which is governed by the scale parameter σ.

A limitation of fully parametric approaches is that the assumed baseline hazard may not always be realistic in practice. Thus, we propose to further extend the log-linear multi-parameter regression model by allowing the baseline hazard to vary freely. This semiparametric model therefore brings together the flexibility of multi-parameter regression with additional robustness afforded by relaxing the assumption of a parametric error distribution. The proposed extension not only generalises multi-parameter regression to semiparametric status but also generalises the semiparametric accelerated failure time model to multi-parameter regression status; the latter fact is noteworthy given that the semiparametric accelerated failure time model has been considered by many authors over the years (Miller, 1976; Prentice, 1978; Buckley and James, 1979; Tsiatis, 1990; Ritov, 1990; Lai and Ying, 1991; Ying, 1993a; Lin et al., 1998; Jin et al., 2003); see Martinussen and Scheike (2007, chap. 8) for a summary of developments in this area.

Other examples of semiparametric multi-parameter regression models, which differ from the model developed in this paper, exist in the literature. Chen and Jewell (2001) considered a model which combined the semiparametric accelerated failure time model and Cox's (1972) proportional hazards model and, therefore, had two regression components; the model also contained the lesser-known accelerated hazards model (Chen and Wang, 2000) as a special case. Scheike and Zhang (2002a,b) developed a different hybrid model which incorporated the Cox model and the Aalen model (Aalen, 1980) leading to two regression components. Somewhat closer to our work is that of Zeng and Lin (2007b) who considered transformation models with a covariate-dependent scale parameter and, therefore, like us, had regression components corresponding to location and scale. However, whereas their transformation was unspecified with a parametric baseline distribution, we, conversely, focus on the log-transformation with an unspecified baseline distribution.

From a practical perspective, inference based on the semiparametric accelerated failure time model has historically been somewhat cumbersome. This is partly due to the non-smooth nature of the estimating equations involved, but, more importantly, the precision of the resulting estimators does not lend itself to direct (i.e., plug-in) estimation due to its intractability. However, with recent resampling techniques, this is no longer an obstacle, and, specifically, we adapt the method of Zeng and Lin (2008) to our setting to facilitate inference for the regression coefficients. Moreover, we expand Zeng and Lin's approach to obtain the variance of the cumulative hazard estimator, and combine this with modern empirical process theory which permits straightforward inference for any functional of interest without the need for resolving estimating equations.

2 Model

2.1 Specification and interpretation

In line with the classical formulation of the accelerated failure time model (cf. Kalbfleisch and Prentice (2002, chap. 3)) we specify a regression model for $\log T$ with T denoting the failure time. In particular, for the ith individual, $i = 1, \ldots, n$, we assume that

$$\log T_i = \mu_i + \sigma_i e_i,$$

where the location and scale parameters, μ_i and σ_i, are related to $p + q$ covariates, $(X_i^T, Z_i^T)^T$, via

$$\mu_i = -\beta^T X_i,$$
$$\sigma_i = \exp(-\gamma^T Z_i),$$

where β and γ are vectors of regression coefficients. The error terms, $e_1, \ldots, e_n$, are assumed to be independent and identically distributed with cumulative hazard function $A(\cdot)$ which will be unspecified in our work.

The conditional quantile function for this model is given by

$$Q_i(\pi) = \exp(\mu_i) Q_0(\pi)^{\sigma_i},$$

where $\pi \in [0, 1]$ and $\log\{Q_0(\pi)\} = A^{-1}\{-\log(1 - \pi)\}$ is the quantile function for the error distribution, i.e., $Q_0(\pi)$ is the quantile function for a baseline individual. Consider individuals i and j whose respective X and Z vectors are denoted by X_i, X_j, Z_i, Z_j. The ratio of their quantile functions is then

$$\frac{Q_j(\pi)}{Q_i(\pi)} = \exp\{-\beta^T (X_j - X_i)\} Q_0(\pi)^{\exp(-\gamma^T Z_j) - \exp(-\gamma^T Z_i)}.$$

This quantile ratio provides insight into the interpretation of the location and scale regression coefficients and, indeed, can be used in practical applications to quantify the overall effect of a given covariate on lifetime. We immediately see that when $\gamma^T (Z_j - Z_i) = 0$, the quantile ratio reduces to the usual accelerated failure time constant, $\exp\{-\beta^T (X_j - X_i)\}$, so that the effect of covariates is quantile-independent, i.e., it applies across the whole lifetime, and, for example, $\beta^T (X_j - X_i) > 0$ implies reduced lifetime. Hence, the proposed model directly extends the accelerated failure time model, providing a lack of fit test of accelerated failure time effects.

It is worth noting that, since $Q_0(\pi)$ is an increasing function on $[0, \infty[$, the quantile ratio decreases with π for $\gamma^T(Z_j - Z_i) > 0$, increases with π for $\gamma^T(Z_j - Z_i) < 0$, and, in both cases, equals one for some π value if $\lim_{\pi \to 1} Q_0(\pi) = \infty$. Therefore, when $\gamma^T(Z_j - Z_i) \neq 0$, the model implies crossing quantile functions and, hence, crossing survivor functions.

2.2 Derivation of estimation equations

We now reformulate the model in a counting process framework (Andersen et al., 1993) to adopt potential right censoring, and to derive estimation equations based on the resulting intensity processes. For this we denote by C_i the censoring time, $\tilde{T}_i = C_i \wedge T_i$ the observed event time, and $\Delta_i = I(T_i \leq C_i)$ the failure indicator. With these quantities in place, the counting and at risk processes are, respectively, defined as

$$N_i(t) = \Delta_i I(\log \tilde{T}_i \leq t),$$
$$Y_i(t) = I(\log \tilde{T}_i \geq t).$$

We note that the hazard rate of $\log T_i$ is given by

$$\alpha_i(t) = \alpha\left\{\sigma_i^{-1}(t - \mu_i)\right\}\sigma_i^{-1},$$

where α denotes the derivative of A. Consequently, with independent right censoring (Andersen et al., 1993), the intensity process of $N_i(t)$ is given by $Y_i(t)\alpha_i(t)$. Furthermore, for some monotone increasing function, g_i, the time-transformed counting process

$$N_i^*(t) = N_i\{g_i(t)\}$$

has intensity

$$\lambda_i^*(t) = Y_i^*(t) g_i'(t) \alpha_i\{g_i(t)\}$$

where $Y_i^*(t) = Y_i\{g_i(t)\}$. In particular, with $g_i(t) = \sigma_i t + \mu_i$, we have that $N_i^*(t)$ has intensity $Y_i^*(t)\alpha(t)$. This observation motivates a Nelson-Aalen type estimator of A, that is, for a given value of $\theta = (\beta^T, \gamma^T)^T$, we estimate A by

$$\hat{A}_n(t, \theta) = \sum_{i=1}^n \int_{-\infty}^t \frac{dN_i^*(s)}{\sum_{j=1}^n Y_j^*(s)}.$$

To estimate the regression parameters, θ, we propose the use of a likelihood based approach. For given A, the score function for θ is given by

$$\sum_{i=1}^n \int_{-\infty}^{\infty} D_\theta \log\{\alpha_i(s)\}\{dN_i(s) - Y_i(s)\alpha_i(s)ds\}.$$

where $D_\theta = (\partial/\partial\beta_1, \ldots, \partial/\partial\beta_p, \partial/\partial\gamma_1, \ldots, \partial/\partial\gamma_q,)^T$ is the gradient operator. By observing that

$$D_\theta \log\{\alpha_i(s)\} = \left(\frac{\alpha'\{g_i^{-1}(s)\}}{\alpha\{g_i^{-1}(s)\}} \sigma_i^{-1} X_i^T, \left[\frac{\alpha'\{g_i^{-1}(s)\}}{\alpha\{g_i^{-1}(s)\}} g_i^{-1}(s) + 1 \right] Z_i^T \right)^T$$

we rewrite the score function as

$$\sum_{i=1}^n \int_{-\infty}^{\infty} \left[\frac{\alpha'(u)}{\alpha(u)} \sigma_i^{-1} X_i^T, \left\{ \frac{\alpha'(u)}{\alpha(u)} u + 1 \right\} Z_i^T \right]^T \{dN_i^*(u) - Y_i^*(u)\alpha(u)du\}.$$

To arrive at an operational estimation procedure, we modify this score function as follows. Firstly, we substitute the quantities $\alpha'(u)/\alpha(u)$ and $\alpha'(u)u/\alpha(u) + 1$ with known deterministic functions which we denote by $\rho_\beta(u)$ and $\rho_\gamma(u)$, respectively. Secondly, we replace $\alpha(u)du$ by $d\hat{A}_n(u, \theta)$. Thirdly, we truncate integration at an upper limit τ, where there is still a positive probability of being at risk. Doing so, we arrive at the estimating equations

$$\Psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\tau} \left\{ \rho_\beta(u) \sigma_i^{-1} X_i^T, \rho_\gamma(u) Z_i^T \right\}^T \{dN_i^*(u) - Y_i^*(u) d\hat{A}_n(u, \theta)\},$$

which, for ease of exposition in the developments to follow, is rewritten as

$$\Psi_n\{\theta, \hat{\eta}_n(\cdot, \theta)\} = \frac{1}{n} \sum_{i=1}^n I(\varepsilon_{\theta i} \leq \tau) \rho(\varepsilon_{\theta i}) \{(\sigma_i^{-1} X_i^T, Z_i^T) - \hat{\eta}_n(\varepsilon_{\theta i}, \theta)\}^T \Delta_i,$$

where $\rho(u)$ is a $(p + q) \times (p + q)$ diagonal matrix with diagonal elements given by $\rho_\beta(u)$ repeated p times followed by $\rho_\gamma(u)$ repeated q times, and

$$\begin{aligned}
\varepsilon_{\theta i} &= \sigma_i^{-1}(\log \tilde{T}_i - \mu_i), \\
\hat{\eta}_n(u, \theta) &= \{\hat{\eta}_n^\beta(u, \theta), \hat{\eta}_n^\gamma(u, \theta)\}, \\
\hat{\eta}_n^\beta(u, \theta) &= \frac{\sum_{j=1}^n Y_j^*(u) \sigma_j^{-1} X_j^T}{\sum_{j=1}^n Y_j^*(u)}, \\
\hat{\eta}_n^\gamma(u, \theta) &= \frac{\sum_{j=1}^n Y_j^*(u) Z_j^T}{\sum_{j=1}^n Y_j^*(u)}.
\end{aligned}$$

The resulting estimate $\hat{\theta}_n$ of the true parameter value θ_0 is obtained as a minimizer of $\|\Psi_n\{\theta, \hat{\eta}_n(\cdot, \theta)\}\|$ which in turn enables estimation of the cumulative hazard $A_0(\cdot)$ by $\hat{A}_n(\hat{\theta}_n, \cdot)$.

2.3 Weight functions

From the above we see that the weight functions in the efficient score function obey the relationship $\rho_\gamma(u) = u\rho_\beta(u) + 1$. Accordingly, we suggest using using weights of the form $\rho_\beta(u) = \rho(u)$ and $\rho_\gamma(u) = \rho(u)u + 1$ in practice as this choice mimics the efficient structure. As for the specific choice of $\rho(u)$, throughout the literature various authors have found rank-based estimation procedures, and associated variance estimators, to be quite insensitive to the choice of weight function (Lin et al., 1998; Chen and Jewell, 2001; Jin et al., 2003) and, in particular, typically suggest the use of (i) the log-rank weight, $\rho(u) = 1$, which assigns equal weight to all observations and is efficient when $e \sim$ Extreme Value (i.e., $T \sim$ Weibull), or (ii) the Gehan weight, $\rho(u) = \sum_{j=1}^{n} Y_j^*(u)/n$, which is somewhat more data-driven in that it assigns less weight to observations for which there is less information (i.e., those corresponding to survival times in the tail of the distribution).

Alternatively, a theoretically semiparametrically efficient procedure could be based on adaptively estimating $\rho(u)$ directly from the data, perhaps using kernel smoothing (Tsiatis, 1990; Lai and Ying, 1991; Zeng and Lin, 2007a). However, this step introduces additional complexity beyond the use of a deterministic weight function, which can introduce some instability into the numerical estimation procedure, and, moreover, one must then consider the selection of an optimal bandwidth – for which there are no clear guidelines in this context, and to which the results (particularly the variance estimators) can be sensitive (Zeng and Lin, 2007a). Furthermore, the resulting efficiency gain is not large in practice (cf. Chen and Jewell (2001), Jin et al. (2003), and Zeng and Lin (2007a)). For these reasons we propose the use of rank-based procedures within our semiparametric multi-parameter regression setting, and investigate some choices of weight function in Section 4.1 and the Online Supporting Information.

3 Asymptotic properties

3.1 Key results

We show that $\hat{\theta}_n$ is consistent, that $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges to a zero mean Gaussian distribution, and that $n^{1/2}\{\hat{A}_n(\hat{\theta}_n, \cdot) - A_0(\cdot)\}$ converges to a tight zero mean Gaussian process. Regularity conditions and proofs, extending the arguements of Nan et al. (2009) to the multi-parameter regression setting, can be found in the Online Supporting Information.

First we turn to the consistency. For this purpose let $\Psi(\theta, \eta)$ denote the limit of $\Psi_n(\theta, \eta)$ and let $\eta_0(\cdot, \theta)$ denote the limit of $\hat{\eta}_n(\cdot, \theta)$. Then we have the following result.

Theorem 1 *Assume that $\theta_0 \in \Theta$ is the unique solution of $\Psi\{\theta, \eta_0(\cdot, \theta)\} = 0$. Then an approximate root $\hat{\theta}_n$ satisfying $\Psi_n\{\hat{\theta}_n, \hat{\eta}_n(\cdot, \hat{\theta}_n)\} = o_{P^*}(1)$ is consistent for θ_0.*

Next, for detailing the weak convergence of $n^{1/2}(\hat{\theta}_n - \theta_0)$, we adopt the following notation. Let $O_i = (\log \tilde{T}_i, \Delta_i, X_i, Z_i)$ denote what we observe on the ith individual, and

$$\varepsilon_\theta(O) = \exp(\gamma^T Z)(\log \tilde{T} + \beta^T X)$$

so that $\varepsilon_\theta(O_i) = \varepsilon_{\theta i}$. In line with this, we shall use the short notation ε_θ for $\varepsilon_\theta(O)$ and also denote ε_{θ_0} by ε_0. Moreover we define

$$\psi(O; \theta, \eta) = I(\epsilon_\theta \leq \tau)\rho(\varepsilon_\theta)[\{\exp(\gamma^T Z)X^T, Z^T\} - \eta(\varepsilon_\theta, \theta)]^T \Delta$$

so that $\Psi_n(\theta, \eta) = \frac{1}{n}\sum_{i=1}^n \psi(O_i; \theta, \eta)$, and

$$\begin{aligned} J(O; \theta, \eta, A) = \psi\{&O; \theta, \eta(\cdot, \theta)\} \\ &- \int_{-\infty}^{\tau} \rho(t) I(\varepsilon_\theta \geq t)[\{\exp(\gamma^T Z)X^T, Z^T\} - \eta(t, \theta)]^T dA(t). \end{aligned}$$

We also define the $(p+q) \times (p+q)$ matrices $\dot{\eta}_\theta(\varepsilon_\theta, \theta) = D_\theta \eta(\varepsilon_\theta, \theta)$ and $\dot{\Psi}_\theta\{\theta_0, \eta_0(\cdot, \theta_0)\} = D_\theta \Psi\{\theta_0, \eta_0(\cdot, \theta_0)\}$ where the θ subscript in $\dot{\eta}_\theta$ and $\dot{\Psi}_\theta$ serves as a reminder that these derivatives are taken with respect to θ.

Finally, we adopt the usual empirical process notation

$$\begin{aligned} Pf &= \int f(o) dP(o), \\ \mathbb{P}_n f &= \frac{1}{n}\sum_{i=1}^n f(O_i), \\ \mathbb{G}_n f &= n^{1/2}(\mathbb{P}_n f - Pf), \end{aligned}$$

where f denotes some bounded function on the sample space.

Theorem 2 *Let $\hat{\theta}_n$ be an approximate root satisfying $\Psi_n\{\hat{\theta}_n, \hat{\eta}(\cdot, \hat{\theta}_n)\} = o_{P^*}(n^{-1/2})$. Suppose that $\theta \mapsto \eta_0(\varepsilon_\theta, \theta)$ is differentiable with uniformly bounded and continuous derivative $\dot{\eta}_\theta(\varepsilon_\theta, \theta)$. Then if $\dot{\Psi}_\theta\{\theta_0, \eta_0(\cdot, \theta_0)\}$ is non-singular,*

$$n^{1/2}(\hat{\theta}_n - \theta_0) = -\dot{\Psi}_\theta^{-1}\{\theta_0, \eta_0(\cdot, \theta_0)\} \times \mathbb{G}_n J(\theta_0, \eta_0, A_0) + o_{P^*}(1).$$

Now we turn to the weak convergence of of $n^{1/2}\{\hat{A}_n(\hat{\theta}_n, \cdot) - A_0(\cdot)\}$. For this we define

$$\phi(t, \theta) = P\left\{I(\varepsilon_\theta \leq t)\Delta d^{(0)}(\varepsilon_\theta, \theta)^{-1}\right\},$$

where $d^{(0)}(t, \theta) = PI(\varepsilon_\theta \geq t)$, and, furthermore, the associated vector of derivatives $\dot{\phi}_\theta = D_\theta \phi$. We further define

$$H(O; t, \theta, D^{(0)}, A) = I(\varepsilon_\theta \leq t)\Delta D^{(0)}(\varepsilon_\theta, \theta)^{-1} - \int_{-\infty}^{t} I(\varepsilon_\theta \geq s) D^{(0)}(s, \theta)^{-1} dA(s).$$

With this notation we have the following result

Theorem 3 *Let $\hat{\theta}_n$ be an estimator of θ_0 such that $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges weakly to a zero mean normal distribution. Then $n^{1/2}\{\hat{A}_n(\hat{\theta}_n, \cdot) - A_0(\cdot)\}$ converges weakly to a tight zero mean Gaussian process on $]-\infty, \tau]$ and the following holds*

$$n^{1/2}\{\hat{A}_n(\hat{\theta}_n, t) - A(t)\} = \dot{\phi}_\theta(t, \theta_0) n^{1/2}(\hat{\theta}_n - \theta_0) + \mathbb{G}_n H(t, \theta_0, d^{(0)}, A_0) + o_{P^*}(1).$$

3.2 A resampling procedure for estimating asymptotic variances

We note that the limiting covariance matrix for $n^{1/2}(\hat{\theta}_n - \theta_0)$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{\dot{\Psi}}_{\theta}^{-1}\hat{J}_i)^{\bigotimes 2}$$

where $a^{\bigotimes 2} = aa^T$, and $\hat{\dot{\Psi}}_{\theta}$ and $\hat{J}_i = J(O_i; \hat{\theta}_n, \hat{\eta}_n, \hat{A}_n)$ are estimates of their theoretical counterparts. While $\hat{J}_i$ is easily computed, estimation of $\dot{\Psi}_{\theta}$ would require an estimate of the error hazard $\alpha(\cdot)$ which is difficult to obtain reliably. The classical solution to this problem is to produce a sample $\hat{\theta}^b$, $b = 1, \ldots m$, by solving perturbed estimating equations (often based on Parzen et al. (1994)), from which the limiting covariance matrix can be estimated directly; such procedures are, however, computationally intensive owing to solving estimating equations multiple times.

As the representation in Theorem 2 is of the form considered in Zeng and Lin (2008), we may apply their more modern resampling approach which requires re-evaluating (but not re-solving) estimating equations. Their approach is based on the fact that $n^{-1/2}\Psi_n(\hat{\theta}_n + n^{-1/2}G) = \dot{\Psi}^{-1}G + o_{P^*}(1)$, where G is a zero-mean Gaussian $(p+q)$-vector independent of the data, which motivates the least squares estimate $\hat{\dot{\Psi}}_{\theta} = (M^TM)^{-1}M^TU$ where M and U are matrices whose bth rows are, respectively, given by the zero-mean Gaussian vector G^b, and the vector $n^{-1/2}\Psi_n(\hat{\theta}_n + n^{-1/2}G^b)$, $b = 1, \ldots, m$.

In a similar manner to the regression coefficients, we can estimate the limiting variance of $n^{1/2}\{\hat{A}_n(\hat{\theta}_n, t) - A_0(t)\}$ using

$$\frac{1}{n}\sum_{i=1}^{n}\{-\hat{\dot{\phi}}_{\theta}(t)\hat{\dot{\Psi}}_{\theta}^{-1}\hat{J}_i + \hat{H}_i(t)\}^2$$

where $\hat{\dot{\Psi}}_{\theta}$ is estimated using least squares as described in the previous paragraph, and $\hat{H}_i(t) = H(O_i; t, \hat{\theta}_n, D_n^{(0)}, \hat{A}_n)$ where $D_n^{(0)}(t, \theta) = \mathbb{P}_n I(\varepsilon_{\theta} \geq t) = \frac{1}{n}\sum_{i=1}^{n} Y_i^*(t)$. However, an estimator $\hat{\dot{\phi}}_{\theta}(t)$ of $\dot{\phi}_{\theta}(\theta_0, t)$ is difficult to obtain directly. Instead we adapt the resampling idea of Zeng and Lin (2008) to obtain an estimator of $\dot{\phi}_{\theta}(\theta_0, t)$. In particular, according to the asymptotic representation of $n^{1/2}\{\hat{A}_n(\hat{\theta}_n, \cdot) - A_0(\cdot)\}$ in Theorem 3, we find that

$$\begin{aligned}
&n^{1/2}\{\hat{A}_n(\hat{\theta}_n + n^{-1/2}G, t) - \hat{A}_n(\hat{\theta}_n, t)\} \\
&= n^{1/2}\{\hat{A}_n(\hat{\theta}_n + n^{-1/2}G, t) - A_0(t)\} - n^{1/2}\{\hat{A}_n(\hat{\theta}_n, t) - A_0(t)\} \\
&= \dot{\phi}_{\theta}(t, \theta_0)G + o_{P^*}(1)
\end{aligned}$$

which motivates the least squares estimate $\hat{\dot{\phi}}_{\theta}(t) = \{(M^TM)^{-1}M^T\tilde{U}\}^T$ where M is as before, and $\tilde{U}$ is a matrix whose bth row is given by $n^{1/2}\{\hat{A}_n(\hat{\theta}_n + n^{-1/2}G^b, t) - \hat{A}_n(\hat{\theta}_n, t)\}$, $b = 1, \ldots, m$.

With the estimates of the limiting variances of $\hat{\theta}_n$ and $\hat{A}_n(\hat{\theta}_n, t)$, we can straightforwardly produce Wald-type confidence intervals for the parameters, and confidence bands for the error cumulative hazard. In the latter case, as is standard, it is preferable to produce confidence bands on the $\log A_0(t)$ scale first and back-transform to the $A_0(t)$ scale. For functionals of θ_0 and $A_0(t)$, one could apply the functional delta method (Andersen et al., 1993). However, in line with the resampling approaches discussed above, we suggest the use of the conditional multiplier method from empirical process theory (van der Vaart and Wellner, 1996) which is described in the Online Supporting Information.

4 Numerical studies

4.1 Simulation

We now investigate the performance of our procedure in finite samples by way of a simulation study. In particular, we generated survival times according to the following setup: $\mu = -(X_1\beta_1 + X_2\beta_2)$ and $\log\sigma = -X_1\gamma_1$ with covariates $X_1 \sim \text{Bernoulli}(0.5)$ and $X_2 \sim \text{Uniform}(0,1)$, parameter vector $\theta = (\beta_1, \beta_2, \gamma_1) = (1,1,1)$, and error distribution $e \sim N(0,1)$. Furthermore, we considered a sample size of 100 where survival times were randomly censored according to a log-normal distribution with unit-scale and location set to achieve censored proportions of approximately 20% or 50% respectively.

Since the estimation equations are non-smooth step-functions of the parameters, we applied the Nelder-Mead optimisation procedure as implemented in the `optim` function in the `R` programming language. It is also worth highlighting the fact that the estimation equations, as we have presented them, depend on a threshold, τ, which was required for our asymptotic derivations to ensure a non-empty risk set so that denominators are theoretically bounded away from zero. To investigate the sensitivity of our approach to the inclusion of τ, we consider thresholds of 2 and ∞. As for the choice of weight function we consider the log-rank weight, $\rho(u) = 1$, and the normal weight, $\rho(u) = f_e(u)/(1-F_e(u)) - u$ where f_e and F_e are the normal pdf and cdf functions, which is the true, efficient weight in our simulation study.

In total we present eight simulation scenarios comprising one sample size, two censored proportions, two threshold values, and two choices of weight function. Each of these scenarios was replicated 5000 times. Within each replicate we estimated the following quantities: (i) the parameter vector, θ, (ii) the cumulative error hazard at the median error, $A(0) = 0.6931$, (iii) the conditional survivor function for covariate profile $x^{(1)} = (x_1^{(1)}, x_2^{(1)})^T = (1,1)^T$ evaluated at the median time for this covariate profile, $S(t_{0.5}^{(1)} \mid x^{(1)})$ =0.5, and (iv) the ratio of the median for $x^{(1)} = (1,1)^T$ to the median for $x^{(2)} = (0,1)^T$, $r(x^{(1)}, x^{(2)}) = 0.3679$. For quantities (i) and (ii), Wald-type confidence intervals were produced where the limiting variances were estimated using least squares (as described in Section 3.2) with $m = 1000$, and for quantites (iii) and (iv) the conditional multiplier method was used (as described in the Online Supporting Information) with $m = 1000$.

The results are summarised in Table 1. It is clear that the estimates are reasonably

Table 1: Results of simulation study

τ	Cens.	Parameter	Log-rank Bias	SE	SEE	Cov.	Normal (true, efficient) Bias	SE	SEE	Cov.
2	20%	β_1	<0.001	0.099	0.099	94.8	-0.001	0.097	0.096	94.4
		β_2	0.008	0.189	0.185	94.0	0.002	0.178	0.169	93.1
		γ_1	-0.001	0.174	0.179	94.1	0.004	0.176	0.157	91.8
		A	-0.009	0.168	0.156	94.5	-0.002	0.164	0.153	94.4
		S	<0.001	—	—	94.6	<0.001	—	—	94.3
		r	0.003	—	—	95.1	0.004	—	—	94.8
2	50%	β_1	<0.001	0.127	0.124	94.2	-0.006	0.120	0.121	94.8
		β_2	0.005	0.216	0.210	93.3	0.003	0.201	0.193	93.7
		γ_1	-0.015	0.208	0.230	94.6	0.004	0.215	0.206	93.2
		A	-0.010	0.219	0.203	95.3	-0.003	0.217	0.201	94.8
		S	0.008	—	—	95.3	0.005	—	—	95.5
		r	0.006	—	—	95.5	0.008	—	—	95.9
∞	20%	β_1	<0.001	0.100	0.099	94.3	-0.001	0.098	0.096	94.3
		β_2	0.003	0.189	0.183	93.6	-0.002	0.179	0.169	93.4
		γ_1	-0.007	0.175	0.177	93.4	0.003	0.171	0.156	93.2
		A	-0.004	0.165	0.156	94.3	-0.005	0.165	0.152	94.2
		S	0.001	—	—	94.5	0.004	—	—	94.4
		r	0.004	—	—	95.3	0.005	—	—	94.6
∞	50%	β_1	0.003	0.126	0.124	94.5	<0.001	0.124	0.121	94.1
		β_2	0.001	0.215	0.208	93.1	-0.001	0.207	0.191	92.6
		γ_1	-0.016	0.215	0.230	93.7	0.008	0.224	0.206	92.0
		A	-0.010	0.230	0.204	95.1	-0.008	0.223	0.201	94.5
		S	0.009	—	—	95.3	0.005	—	—	94.4
		r	0.005	—	—	95.4	0.007	—	—	95.3

Cens., censored proportion; Bias, median bias; SE, standard error of estimates; SEE, median of estimated standard error, Cov., empirical coverage percentage for 95% confidence interval; $A = A(0)$; $S = S(t_{0.5}^{(1)} \,|\, x^{(1)})$; $r = r(x^{(1)}, x^{(2)})$. Since the variances for the functionals S and r are not estimated directly within our scheme, SE and SEE are not shown in those cases.

unbiased in all cases and the associated 95% confidence intervals achieve a coverage percentage which is close to the desired nominal level for both choices of weight function, and, in either case, the results for $\tau = 2$ and $\tau = \infty$ are very similar. The estimated standard errors capture the true variations adequately, and, moreover, the efficiency based on the log-rank weights is very close to that of the true, efficient weights (which is in line with the findings of other authors in the simpler accelerated failure time model context). Additional simulation results are given in the Online Supporting Information which cover $n = 50$ and $n = 500$, and Gehan weights; the results are comparable with those shown here.

4.2 Lung cancer data

We now apply our model to data arising from a lung cancer study which was the subject of a 1995 Queen's University Belfast PhD thesis by P. Wilkinson (previously analysed in Burke and MacKenzie (2017)). This observational study pertains to 855 individuals who were diagnosed with lung cancer during the one-year period 1st October 1991 to 30th September 1992, and these individuals were followed up until 30th May 1993 (approximately 20% of survival times were right-censored). The primary interest was to investigate the differences between the following treatment groups: palliative care, surgery, chemotherapy, radiotherapy, and a combined treatment of chemotherapy and radiotherapy. While various other covariates were measured (see Burke and MacKenzie (2017)), the aim here is to illustrate our semiparametric multi-parameter regression methodology for the treatment model.

The results of the fitted model are given in Table 2 where the log-rank weights were used in the estimation procedure (see the Online Supporting Information for other choices of weights which yield numerically very similar results). Firstly note that the β coefficients are all negative, and statistically significant, suggesting an improvement in survival relative to palliative care group. The γ coefficients of radiotherapy and the combined treatment of both chemotherapy and radiotherapy differ statistically from zero, indicating that the quantile ratios are non-constant. Furthermore, note that the γ coefficients are positive and thus the quantile ratios decrease over the timeframe, i.e., the effectiveness of these treatments diminishes over time. Recall that, this being an observational study, the estimated effects are not "treatment effects" in the sense of a randomised trial, but, notwithstanding this, analyses of observational effects are still useful in their own right.

It is also of interest to test whether the overall effect of a given treatment is statistically significant, i.e., testing $\beta_j = \gamma_j = 0$ for the jth group. Asymptotic normality of the estimated parameter vector means that this can be achieved by comparing $(\hat{\beta}_j, \hat{\gamma}_j)\hat{\Sigma}^{-1}_{\beta_j,\gamma_j}(\hat{\beta}_j, \hat{\gamma}_j)^T$ to a χ^2_2 distribution where $\Sigma_{\hat{\beta}_j,\hat{\gamma}_j}$ is the 2×2 covariance matrix for the pair $(\hat{\beta}_j, \hat{\gamma}_j)$. The resulting p-values for this test are shown in the last column of Table 2.

Figure 1 compares the fitted survivor curves to the Kaplan-Meier curves. We can see that the model provides an excellent fit to the data. Furthermore, we see that,

Table 2: Regression coefficients for model fitted to lung cancer data

		Location			Scale			Joint
Treatment Group	Sample size	Est.	SE	P-val	Est.	SE	P-val	P-val
Palliative care	441	0.00	—	—	0.00	—	—	—
Surgery	79	-2.65	0.17	<0.01	0.31	0.20	0.12	<0.01
Chemotherapy	45	-0.54	0.27	0.04	0.04	0.11	0.73	0.06
Radiotherapy	256	-1.08	0.10	<0.01	0.30	0.07	<0.01	<0.01
Chemo. & radio	34	-1.87	0.12	<0.01	0.94	0.17	<0.01	<0.01

Est., estimated location (β) or scale (γ) coefficient; SE, standard error; P-val, p-value. The joint p-value corresponds to testing that $\beta_j = \gamma_j = 0$.

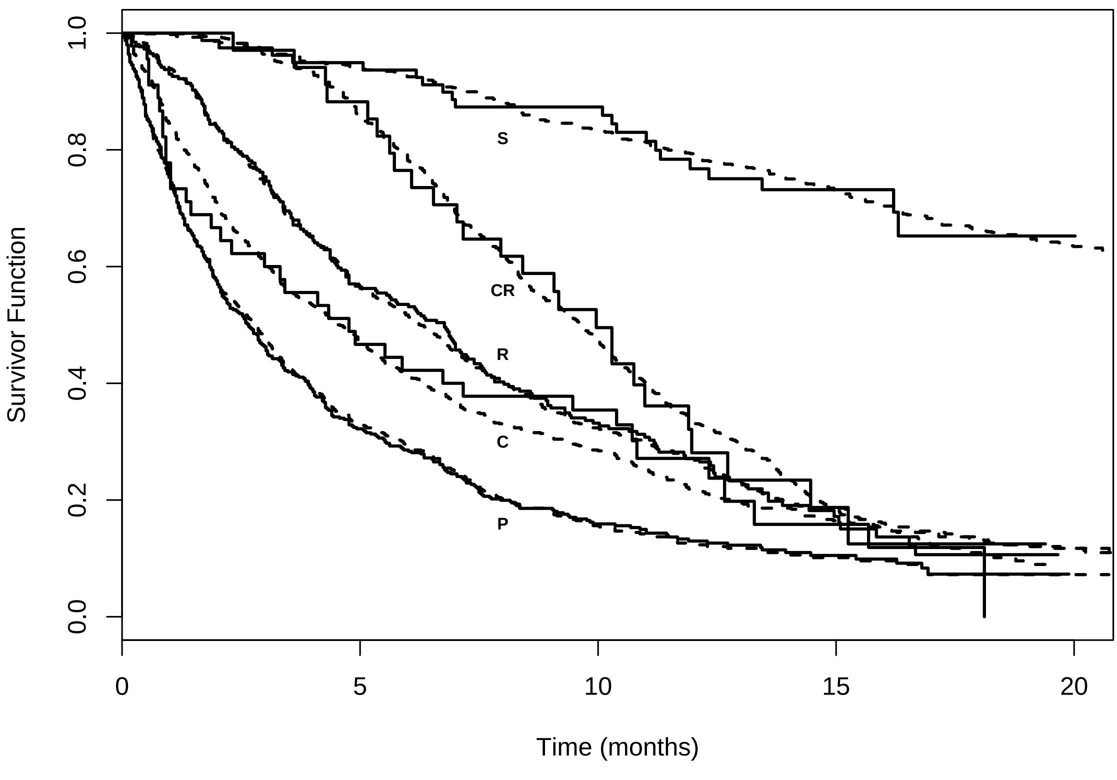

Figure 1: Kaplan-Meier (solid) curves with model-based curves (dash) overlayed where P = palliative, C = chemotherapy, R = radiotherapy, CR = chemotherapy & radiotherapy combined, and S = surgery, respectively.

relative to the palliative care curve, the various curves have different shapes, particularly those of radiotherapy and the combined treatment. Indeed, the curves converge at a rate which is indicative of a reduction in treatment effectiveness over time, and is something which cannot be handled by a basic accelerated failure time model. This highlights the flexibility of the multi-parameter regression extension wherein the scale parameter depends on covariates thereby facilitating survivor curves corresponding to non-constant quantile ratios.

While the results of Table 2 provide useful information on the nature of the treatment effects, and which are statistically significant, we now consider the quantile ratios which allow us to quantify the effect of each treatment on lifetime. From Figure 2, we can see that the population assigned to surgery has the highest survival, and the difference is sustained with time. Even though it may appear to diminish with time, it is clear that a horizontal line corresponding to a constant quantile ratio easily fits within the confidence bands. This is in line with the non-significant scale coefficient seen in Table 2. The combined treatment of chemotherapy and radiotherapy is particularly effective early on but drops sharply in effectiveness over time. Radiotherapy provides a more modest improvement in lifetime but has a similar performance to the combined treatment later in time. Finally, chemotherapy appears to have a relatively weak effect over the whole lifetime.

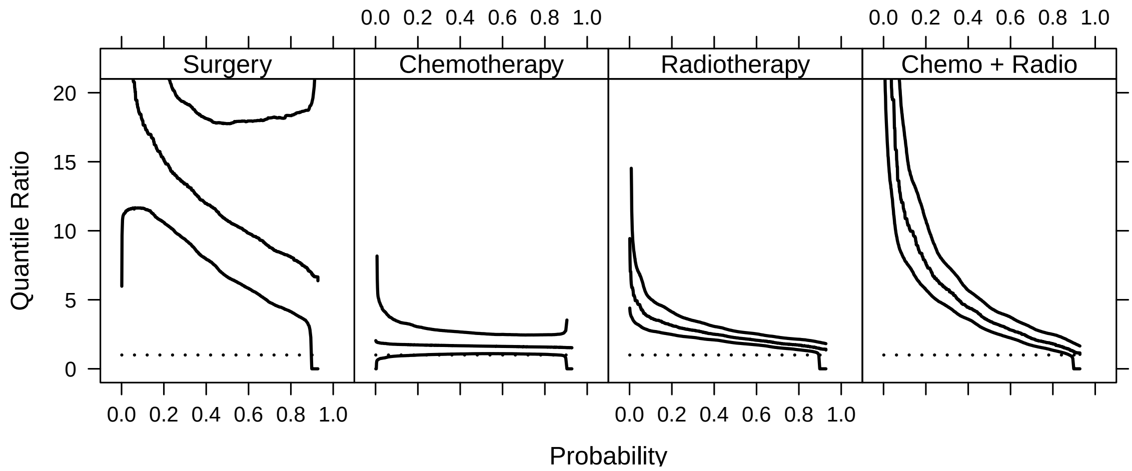

Figure 2: Quantile ratios (relative to palliative care) for each treatment group relative to palliative care along with pointwise 95% confidence intervals (solid). Reference line at unity (dot) also shown.

5 Discussion

In this paper we have extended the semiparametric accelerated failure time model to multi-parameter regression status by jointly modelling the location and scale parameters of its log-linear model representation. This brings together the structural flexibility of multi-parameter regression modelling and the robustness of an unspecified baseline hazard. The resulting model can be interpreted on the lifetime scale through its quantile ratio. The form of the quantile ratio directly generalizes that of the accelerated failure time model in that it depends on the quantile in question (rather than being constant over all quantiles) which allows for effects which change over the individual lifetime. Moreover, this modelling framework produces a new semiparametric test of the accelerated lifetime property for a given covariate which is adjusted for other covariates in the model.

Clearly, the combination of multi-parameter regression and semiparametric modelling is fruitful, and some approaches in this direction exist in current literature, e.g., the proportional-hazards-accelerated-failure-time hybrid of Chen and Jewell (2001) and the proportional-hazards-Aalen hybrid of Scheike and Zhang (2002a,b). However, in the aforementioned models, the two regression components essentially both correspond to distributional scale-type parameters, i.e., the components play similar roles. Furthermore, the models do not have a natural scale for interpretation which is related to the previous point. In contrast, the choice of jointly modelling location and dispersion of the error distribution (or scale and shape of the survival distribution) provides somewhat more "orthogonal" components for the inclusion of covariates – heteroscedastic linear models are familiar in other areas such as econometrics – and, as mentioned above, yields an interpretation on the lifetime scale. Note also that the Aalen model (Aalen, 1980), while not related to multi-parameter regression, is another flexible and robust survival regression model, being fully non-parametric, but which can be somewhat difficult to interpret (owing to the regression functions being on a cumulative hazard scale); Martinussen and Pipper (2013, 2014) considered its interpretation through the "odds of concordance".

On model interpretability, we could even go as far as criticizing hazard-based models in general, as did D.R. Cox when he stated that "accelerated life models are in many ways more appealing [than hazard modelling] because of their quite direct physical interpretation" (Reid, 1994), i.e., such models are interpretable on the lifetime scale as is the case for our model. While we might expect the basic accelerated failure time to be more popular on the basis of this "direct physical interpretation", inference in semiparametric accelerated failure time models has, historically, been comparatively more difficult than in the Cox model, which we discuss in the following paragraph.

Estimation of parameters and baseline cumulative hazard function for our model is based on the creation of a time-transformed counting process (see Section 2.2), yielding a set of estimation equations which directly generalize those of the basic accelerated failure time model. It is noteworthy that our counting process formulation means that the estimating equations extend immediately to more general settings such as multiple events and time-varying covariates. Unlike those of the Cox model, however, the estimation equations for accelerated failure time models (and, hence, for our model) are such that the resulting covariance matrix for the estimators has a non-analytic form. However, we have overcome this by making use of modern empirical process theory in combination with a modification of a resampling procedure due to Zeng and Lin (2008) (as described in Section 3 and the Online Supplementary Information). In particular, our proposal does not require resolving of estimation equations, and permits straightforward inference for any (conditional) survival functional of interest.

In summary, the semiparametric multi-parameter regression model of this paper achieves flexibility through its basic model structure, robustness with respect to distributional assumptions as its baseline distribution is unspecified, and is interpretable on the lifetime scale similar to that of the accelerated failure time model which it directly extends. Our inferential framework combines modern approaches in a novel way which is useful for rank-based estimation in general (beyond our setting and the accelerated failure time model which we generalize), for example, those used in the estimation of transformation models.

Acknowledgement

This research was supported by the Irish Research Council (New Foundations Award) and the Royal Irish Academy (Charlemont Award).

References

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory*, pp. 1–25. Springer.

Andersen, P., Ø. Borgan, R. Gill, and N. Keiding (1993). *Statistical models based on counting processes*. New York: Springer.

Buckley, J. and I. James (1979). Linear regression with censored data. *Biometrika 66*, 429–436.

Burke, K. and G. MacKenzie (2017). Multi-parameter regression survival modelling: An alternative to proportional hazards. *Biometrics 73*, 678–686.

Chen, Y. and N. Jewell (2001). On a general class of semiparametric hazards regression models. *Biometrika 88*, 687–702.

Chen, Y. and M. Wang (2000). Analysis of accelerated hazards models. *J. Am. Statist. Ass. 95*, 608–618.

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

Jin, Z., D. Lin, L. Wei, and Z. Ying (2003). Rank-based inference for the accelerated failure time model. *Biometrika 90*, 341–353.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data* (Second ed.). Wiley.

Kim, E. and D. Zeng (2013). Semiparametic ROC analysis using accelerated regression models. *Stat. Sin. 23*, 829–851.

Lai, T. L. and Z. Ying (1991). Large sample theory of a modified buckley-james estimator for regression analysis with censored data. *Ann. Statist. 19*, 1370–1402.

Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data* (Second ed.). Wiley.

Lin, D., L. Wei, and Z. Ying (1998). Accelerated failure time models for counting processes. *Biometrika 85*, 605–618.

Martinussen, T. and C. B. Pipper (2013). Estimation of odds of concordance based on the Aalen additive model. *Lifetime Data Analysis 19*, 100–116.

Martinussen, T. and C. B. Pipper (2014). Estimation of causal odds of concordance using the Aalen additive model. *Scand. J. Statist. 41*, 141–151.

Martinussen, T. and T. Scheike (2007). *Dynamic regression models for survival data.* Springer.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika 63*, 449–464.

Nan, B., J. Kalbfleisch, and M. Yu (2009). Asymptotic theory for the semiparametric accelerated failure time model with missing data. *Ann. Statist. 37*, 2351–2376.

Parzen, M., L. Wei, and Z. Ying (1994). A resampling method based on pivotal estimating functions. *Biometrika 81*, 341–350.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika 65*, 167–179.

Reid, N. (1994). A conversation with Sir David Cox. *Statist. Sci. 9*, 439–455.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist. 18*, 303–328.

Scheike, T. H. and M. J. Zhang (2002a). An additive-multiplicative Cox-Aalen model. *Scand. J. Statist. 28*, 75–88.

Scheike, T. H. and M. J. Zhang (2002b). Extensions and applications of the Cox-Aalen survival model. *Biometrics 59*, 1033–1045.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist. 18*, 354–372.

van der Vaart, A. (1998). *Asymptotic statistics.* Cambridge: Cambridge university press.

van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes : with applications to statistics.* New York: Springer.

Ying, Z. (1993a). A large sample study of rank estimation for censored regression data. *Ann. Statist. 21*, 76–99.

Ying, Z. (1993b). A large sample study of rank estimation for censored regression data. *Ann. Statist. 21*, 76–99.

Zeng, D. and D. Lin (2007a). Efficient estimation for the accelerated failure time model. *J. Amer. Statist. Assoc. 102*, 1387–1396.

Zeng, D. and D. Lin (2007b). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Statist. Soc.* B *69*, 507–564.

Zeng, D. and D. Lin (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostatistics 9*, 355–363.

Appendix

A Assumptions

Assumption 1 *The parameter* $\theta_0 = (\beta_0^T, \gamma_0^T)^T$ *lies in the interior of a compact set* $\mathcal{B}\times\Gamma = \Theta \subset \mathbb{R}^{p+q}$.

Assumption 2 *There exists constants* $\tau < \infty$ *and* ϵ*, such that* $pr(\varepsilon_\theta \geq \tau) \geq \epsilon > 0$ *for all* X*,* Z *and* $\theta \in \Theta$.

Assumption 3 *The covariates* X *and* Z *are uniformly bounded with probability one.*

Assumption 4 *The error density* f *and its derivative* f' *are bounded and* $\int \{f'(t)/f(t)\}^2 f(t)dt < \infty$.

Assumption 5 $\log(C)$ *has uniformly bounded densities.*

Assumption 6 *The diagonal of* $\rho(\varepsilon_\theta)$ *is differentiable in* θ *with bounded continuous derivative* $\dot{\rho}_\theta(\varepsilon_\theta)$ *and* $\{\rho(\varepsilon_\theta) : \theta \in \Theta\}$ *is a bounded Donsker class.*

Assumption 7 ε_0 *has finite second order moment.*

Remark 1 *Assumptions 4 and 5 are those assumed in Nan et al. (2009) directly adapted from Ying (1993b).*

B Asymptotic results

The asymptotic properties are established by extending the proofs of Nan et al. (2009) to the multi parameter regression setting. We show asymptotic linearity of $\Psi_n(\theta, \hat{\eta}_n)$ in θ in a neighborhood of the true value θ_0. This will rely heavily on the fact that the function classes $\mathcal{F}_0$ and $\mathcal{F}_1$ defined below have bracketing numbers of polynomial order.

$$\mathcal{F}_0 = \{I(\varepsilon_\theta \geq t) : t \in]-\infty, \tau], \theta \in \Theta\}$$

and

$$\mathcal{F}_1 = \left\{ I(\varepsilon_\theta \geq t)(e^{\gamma^T Z} X^T, Z^T) : t \in]-\infty, \tau], \theta \in \Theta \right\}.$$

With $N_{[]}\{\epsilon, \mathcal{F}_0, L_2(P)\}$ and $N_{[]}\{\epsilon, \mathcal{F}_1, L_2(P)\}$ denoting the bracketing numbers of $\mathcal{F}_0$ and $\mathcal{F}_1$, respectively, we have the following result

Lemma 1 *There exist* $K_1 > 0$*,* $K_2 > 0$ *so that for all* $\epsilon < 1$*,*

$$N_{[]}\{\epsilon, \mathcal{F}_0, L_2(P)\} \leq K_1 \epsilon^{-3(p+q+3/2)}, \tag{B.1}$$
$$N_{[]}\{\epsilon, \mathcal{F}_1, L_2(P)\} \leq K_2 \epsilon^{-(6p+7q+9)}.$$

Proof. First note that due to the compactness of Θ and Assumption 3 there exist $m \in L_2(P)$ so that $|\varepsilon_{\theta_1}(o) - \varepsilon_{\theta_2}(o)| \leq m(o)\|\theta_1 - \theta_2\|$. It follows as in van der Vaart (1998) Example 19.7 that there exist $\tilde{K}_1 > 0$ and $\tilde{\delta} > 0$ so that the class $\mathcal{G} = \{\varepsilon_\theta : \theta \in \Theta\}$ can be covered by I brackets of the form $[f_i - \epsilon m, f_i + \epsilon m]$, $f_i \in \mathcal{G}$ where $I \leq \tilde{K}_1 \epsilon^{-(p+q)}$ for $\epsilon < \tilde{\delta}$. Now let $-\epsilon^{-1/2} = t_1 < \ldots < t_J = \tau$ be a partition of the interval $[-\epsilon^{-1/2}, \tau]$ so that $|t_j - t_{j+1}| \leq \epsilon$ and so that $J \leq 1 + \tau\epsilon^{-1} + \epsilon^{-3/2}$. From this partition define

$$
\begin{aligned}
l_{i,j} &= f_i - \epsilon m - t_j \text{ for } j = 1, \ldots, J, i = 1, \ldots, I \\
u_{i,j} &= f_i + \epsilon m - t_{j-1} \text{ for } j = 2, \ldots, J, i = 1, \ldots, I \\
u_{i,1} &= \infty
\end{aligned}
$$

and note that the brackets $\{I(l_{i,j} \geq 0), I(u_{i,j} \geq 0)\}$ cover $\mathcal{F}_0$. Moreover note that from Assumptions 3 and 7 the class $\mathcal{G}$ has an $L_2(P)$ envelope which we shall term F. It now follows that

$$
\begin{aligned}
\|I(l_{i,1} \geq 0) - I(u_{i,1} \geq 0)\|_2^2 &= \int \{1 - I(l_{i,1} \geq 0)\}^2 dP \\
&= \int I(f_i - \epsilon m + \epsilon^{-1/2} < 0) dP \\
&\leq pr(-F - \tilde{\delta} m < -\epsilon^{-1/2}) \\
&\leq \|F + \tilde{\delta} m\|_2^2 \epsilon
\end{aligned}
$$

for all $\epsilon \leq \tilde{\delta}$, where the last inequality follows from direct application of Markov's generalized inequality. Similarly for $j > 1$ and for all $\tilde{C} > 0$

$$
\begin{aligned}
\|I(l_{i,j} \geq 0) - I(u_{i,j} \geq 0)\|_2^2 &= \int I(-\epsilon m + t_{j-1} \leq f_i \leq \epsilon m + t_j) dP \\
&\leq \int I\{-\epsilon(m+1) \leq f_i - t_j \leq \epsilon(m+1)\} dP \\
&= \int_{\{m \geq \tilde{C}\}} I\{-\epsilon(m+1) \leq f_i - t_j \leq \epsilon(m+1)\} dP \\
&\quad + \int_{\{m < \tilde{C}\}} I\{-\epsilon(m+1) \leq f_i - t_j \leq \epsilon(m+1)\} dP \\
&\leq pr(m \geq \tilde{C}) + pr\{|f_i - t_j| \leq \epsilon(\tilde{C} + 1)\}.
\end{aligned}
$$

Using Markov's inequality we get $pr(m \geq \tilde{C}) \leq \|m\|_2^2 \tilde{C}^{-2}$. Furthermore from Assumptions 4 and 5 there exists a constant $\tilde{K}_2$ such that for all ϵ and $\tilde{C}$

$$
pr\{|f_i - t_j| \leq \epsilon(\tilde{C} + 1)\} \leq \tilde{K}_2(\tilde{C} + 1)\epsilon.
$$

Consequently, by choosing $\tilde{C} = \epsilon^{-1/3}$, we obtain the following bound for $j > 1$

$$
\|I(l_{i,j} \geq 0) - I(u_{i,j} \geq 0)\|_2^2 \leq \left(\|m\|_2^2 + \tilde{K}_2\right) \epsilon^{2/3} + \tilde{K}_2 \epsilon
$$

Combining all the bounds we conclude that there exists $K > 0$ and $\delta > 0$ such that for all $\epsilon \leq \delta$, $N_{[]}\{K\epsilon^{1/3}, \mathcal{F}_0, L_2(P)\} \leq IJ$. A rescaling then proves (B.1).

For the second part of the lemma note that $\mathcal{F}_0$ is bounded as is $\mathcal{G}_1 = \{e^{\gamma^T Z} X^T : \gamma \in \Gamma\}$ and $\mathcal{G}_2 = \{Z^T\}$. Finally according to van der Vaart (1998, Example 19.7), the bracketing number of $\mathcal{G}_1$ is less than of the order q. Adding up we see that the bracketing number of $\mathcal{F}_1$ is less than of the order $q + 3(p + q + 3/2) + 3(p + q + 3/2) = 6p + 7q + 9$.

■

Theorem 1 *Assume that $\theta_0 \in \Theta$ is the unique solution of $\Psi(\theta, \eta_0(\cdot, \theta)) = 0$. Then an approximate root $\hat{\theta}_n$ satisfying $\Psi_n\{\hat{\theta}_n, \hat{\eta}_n(\cdot, \hat{\theta}_n)\} = o_{P^*}(1)$ is consistent for θ_0.*

(i).

Let $\|\cdot\|$ denote the supremum norm. Since θ_0 is the unique solution to $\Psi\{\theta, \eta_0(\cdot, \theta)\} = 0$ and Θ is compact it follows that for any fixed $\epsilon > 0$, there exists a $\delta > 0$ such that $pr\left(\|\hat{\theta}_n - \theta_0\| > \epsilon\right) \leq pr\left[\|\Psi\{\hat{\theta}_n, \eta_0(\cdot, \hat{\theta}_n)\}\| > \delta\right]$. If we can show that

$$\|\Psi\{\hat{\theta}_n, \eta_0(\cdot, \hat{\theta}_n)\}\| = o_{P^*}(1), \tag{B.2}$$

then the consistency of $\hat{\theta}_n$ follows.

We first show that $\|\hat{\eta}_n(t, \theta) - \eta_0(t, \theta)\| = o_{P^*}(1)$. Define

$$\begin{aligned} D_n^{(0)}(t, \theta) &= \mathbb{P}_n \{I(\varepsilon_\theta \geq t)\}, \\ D_n^{(1)}(t, \theta) &= \mathbb{P}_n \left\{I(\varepsilon_\theta \geq t)(e^{\gamma^T Z} X^T, Z^T)\right\}, \\ d^{(0)}(t, \theta) &= P \{I(\varepsilon_\theta \geq t)\}, \\ d^{(1)}(t, \theta) &= P \left\{I(\varepsilon_\theta \geq t)(e^{\gamma^T Z} X^T, Z^T)\right\}. \end{aligned}$$

Thus, $\hat{\eta}_n(t, \theta) = D_n^{(1)}(t, \theta)/D_n^{(0)}(t, \theta)$ and $\eta_0(t, \theta) = d^{(1)}(t, \theta)/d^{(0)}(t, \theta)$.

As the classes $\mathcal{F}_0$ and $\mathcal{F}_1$ are Donsker by Lemma 1 it follows that $\|D_n^{(0)} - d^{(0)}\| = o_{P^*}(1)$ and $\|D_n^{(1)} - d^{(1)}\| = o_{P^*}(1)$ and $n^{1/2}\{D_n^{(k)}(t, \theta) - d^{(k)}(t, \theta)\}$ converge to zero mean Gaussian processes on $]-\infty, \tau] \times \Theta$. Since $D_n^{(0)}$ (almost surely) and $d^{(0)}$ are bounded away from zero,

$$\|\hat{\eta}_n - \eta_0\| = o_{P^*}(1). \tag{B.3}$$

The random functions $D_n^{(0)}(\varepsilon_\theta, \theta) = n^{-1}\sum_{i=1}^n I(\varepsilon_{\theta i} \geq \varepsilon_\theta)$ and $D_n^{(1)}(\varepsilon_\theta, \theta) = n^{-1}\sum_{i=1}^n I(\varepsilon_{\theta i} \geq \varepsilon_\theta)(e^{\gamma^T Z} X^T, Z^T)$ can be expressed as the limit of convex combinations of elements of the Donsker classes $\{I(s \geq \varepsilon_\theta) : s \in]-\infty, \tau], \theta \in \Theta\}$ and $\{I(s \geq \varepsilon_\theta)(e^{\gamma^T Z} X^T, Z^T) : s \in]-\infty, \tau], \theta \in \Theta\}$ and are bounded. Thus, they belong to the closed convex hull of those classes which is Donsker by van der Vaart and Wellner (1996, Theorem 2.10.3). By Assumption 2, $D_n^{(0)}$ is bounded away from zero almost surely, so that $\{\hat{\eta}_n(\varepsilon_\theta, \theta) : \theta \in \Theta\}$ is

Donsker by van der Vaart and Wellner (1996, Example 2.10.9). See Kim and Zeng (2013, pp. 844-845).

The class of bounded functions

$$\left\{\psi(O;\theta,\hat{\eta}_n) = I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_\theta,\theta)\}^T\Delta : \theta \in \Theta\right\}$$

is a Glivenko-Cantelli class. By adding and subtracting the same term, and by the triangle inequality, we then have that

$$\begin{aligned}\|\Psi_n(\theta,\hat{\eta}_n) - \Psi(\theta,\eta_0)\| &= \|\mathbb{P}_n\psi(O;\theta,\hat{\eta}_n) - P\psi(O;\theta,\eta_0)\| \\ &\leq \|(\mathbb{P}_n - P)\psi(O;\theta,\hat{\eta}_n)\}\| \\ &\quad + \|PI(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\{\hat{\eta}_n(\varepsilon_\theta,\theta) - \eta_0(\varepsilon_\theta,\theta)\}\Delta\|\end{aligned}$$

The first term on the right-hand side converges to zero in outer probability by the Glivenko-Cantelli property. Further,

$$\|PI(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)(\hat{\eta}_n - \eta_0)\Delta\| \leq \|\hat{\eta}_n - \eta_0\|\|PI(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\Delta\| = o_{P^*}(1),$$

by (B.3). Thus, $\|\Psi_n(\theta,\hat{\eta}_n) - \Psi(\theta,\eta_0)\| = o_{P^*}(1)$,which establishes (B.2) as

$$\begin{aligned}\|\Psi\{\hat{\theta}_n,\eta_0(\cdot,\hat{\theta}_n)\}\| &\leq \|\Psi\{\hat{\theta}_n,\hat{\eta}_n(\cdot,\hat{\theta}_n)\}\| + \|\Psi\{\hat{\theta}_n,\hat{\eta}_n(\cdot,\hat{\theta}_n)\} - \Psi\{\hat{\theta}_n,\eta_0(\cdot,\hat{\theta}_n)\}\| \\ &= o_{P^*}(1) + o_{P^*}(1) = o_{P^*}(1).\end{aligned}$$

■

Lemma 2 *Let $\hat{\theta}_n$ be an approximate root satisfying $\Psi_n\{\hat{\theta}_n,\hat{\eta}_n(\cdot,\hat{\theta}_n)\} = o_{P^*}(n^{-1/2})$. Suppose $\Psi\{\theta_0,\eta_0(\cdot,\theta_0)\}$ is differentiable with bounded continuous derivative $\dot{\Psi}_\theta\{\theta_0,\eta_0(\cdot,\theta_0)\}$, and $\dot{\Psi}_\theta\{\theta_0,\eta_0(\cdot,\theta_0)\}$ is non-singular. Then, $\|\hat{\eta}_n - \eta_0\| = O_{P^*}(n^{-1/2})$ and $\|\hat{\theta}_n - \theta_0\| = O_{P^*}(n^{-1/2})$.*

Proof. First consider the asymptotic representation of $\hat{\eta}_n$,

$$\begin{aligned}&n^{1/2}\{\hat{\eta}_n(t,\theta) - \eta_0(t,\theta)\} \\ &\quad= n^{1/2}\left[\frac{1}{d^{(0)}(t,\theta)}\left\{D_n^{(1)}(t,\theta) - d^{(1)}(t,\theta)\right\} - \frac{D_n^{(1)}(t,\theta)}{D_n^{(0)}(t,\theta)d^{(0)}(t,\theta)}\{D_n^{(0)}(t,\theta) - d^{(0)}(t,\theta)\}\right] \\ &\quad= n^{1/2}\left[\frac{1}{d^{(0)}(t,\theta)}\left\{D_n^{(1)}(t,\theta) - d^{(1)}(t,\theta)\right\} - \frac{d^{(1)}(t,\theta)}{d^{(0)}(t,\theta)^2}\{D_n^{(0)}(t,\theta) - d^{(0)}(t,\theta)\}\right] + o_{P^*}(1) \\ &\quad= d^{(0)}(t,\theta)^{-1}n^{1/2}\left[\left\{D_n^{(1)}(t,\theta) - D_n^{(0)}(t,\theta)\eta_0(t,\theta)\right\} - \{d^{(1)}(t,\theta) - d^{(0)}(t,\theta)\eta_0(t,\theta)\}\right] + o_{P^*}(1) \\ &\quad= d^{(0)}(t,\theta)^{-1}\mathbb{G}_n I(\varepsilon_\theta \geq t)\left\{(e^{\gamma^T Z}X^T, Z^T) - \eta_0(t,\theta)\right\} + o_{P^*}(1). \qquad \text{(B.4)}\end{aligned}$$

For the second equality we used that

$$\|n^{1/2}\{D_n^{(k)}(t,\theta) - d^{(k)}(t,\theta)\}\| = O_{P^*}(1),\ k = 0, 1. \qquad \text{(B.5)}$$

This follows from Lemma 1 which enables the use of van der Vaart and Wellner (1996, Theorem 2.14.9) from which exponentially decaying tail bounds are obtained.

The classes of functions $\{I(\varepsilon_\theta \geq t) : t \in]-\infty, \tau], \theta \in \Theta\}$, $\{\exp(\gamma^T Z)X^T : \theta \in \Theta\}$ and $\{Z\}$ are Donsker and η_0 is a bounded deterministic function. Thus, $\{I(\varepsilon_\theta \geq t)[\{\exp(\gamma^T Z)X^T, Z^T\} - \eta_0(t,\theta)] : t \in]-\infty, \tau], \theta \in \Theta\}$ is Donsker. Because $d^{(0)}(t,\theta)^{-1}$ is bounded, $n^{1/2}\|\hat{\eta}_n - \eta_0\| = O_{P^*}(1)$.

Then,

$$\begin{aligned}
&\|n^{1/2}[\Psi_n\{\theta, \hat{\eta}_n(\cdot,\theta)\} - \Psi\{\theta, \eta_0(\cdot,\theta)\}]\| \\
&\quad = \left\|\mathbb{G}_n\psi\{\theta, \hat{\eta}_n(\varepsilon_\theta,\theta)\} + n^{1/2}P[\psi\{\theta, \hat{\eta}_n(\varepsilon_\theta,\theta)\} - \psi\{\theta, \eta_0(\varepsilon_\theta,\theta)\}]\right\| \\
&\quad \leq \|\mathbb{G}_n\psi\{\theta, \hat{\eta}_n(\varepsilon_\theta,\theta)\}\| + n^{1/2}\|\hat{\eta}_n - \eta_0\|\|PI(\varepsilon_\theta \leq \tau)\rho(\epsilon_\theta)\Delta\| \\
&\quad = O_{P^*}(1)
\end{aligned}$$

Using that $\Psi_n\{\hat{\theta}, \hat{\eta}_n(\cdot,\hat{\theta})\} = o_{P^*}(n^{-1/2})$ and $\Psi\{\theta_0, \eta_0(\cdot,\theta_0)\} = 0$,

$$\begin{aligned}
O_{P^*}(1) &= -n^{1/2}\left[\Psi_n\{\hat{\theta}, \hat{\eta}_n(\cdot,\hat{\theta})\} - \Psi\{\hat{\theta}, \eta_0(\cdot,\hat{\theta})\}\right] \\
&= o_{P^*}(1) + n^{1/2}\Psi\{\hat{\theta}, \eta_0(\cdot,\hat{\theta})\} - n^{1/2}\Psi\{\theta_0, \eta_0(\cdot,\theta_0)\} \\
&= o_{P^*}(1) + \left[\dot{\Psi}_\theta\{\theta_0, \eta_0(\cdot,\theta_0)\} + o_{P^*}(1)\right] n^{1/2}\left(\hat{\theta}_n - \theta_0\right).
\end{aligned}$$

The invertibilty of $\dot{\Psi}_\theta$ gives $n^{1/2}\left(\hat{\theta}_n - \theta_0\right) = O_{P^*}(1)$.

■

Theorem 2 *Let $\hat{\theta}_n$ be an approximate root satisfying $\Psi_n\{\hat{\theta}_n, \hat{\eta}(\cdot,\hat{\theta}_n)\} = o_{P^*}(n^{-1/2})$. Suppose that $\theta \mapsto \eta_0(\varepsilon_\theta,\theta)$ is differentiable with uniformly bounded and continuous derivative $\dot{\eta}_\theta(\varepsilon_\theta,\theta)$. Then if $\dot{\Psi}_\theta\{\theta_0, \eta_0(\cdot,\theta_0)\}$ is non-singular,*

$$n^{1/2}(\hat{\theta}_n - \theta_0) = -\dot{\Psi}_\theta^{-1}\{\theta_0, \eta_0(\cdot,\theta_0)\}\mathbb{G}_n J(\theta_0, \eta_0, A_0) + o_{P^*}(1).$$

Proof. Lemma 2 shows that there exists a $K < \infty$, such that $\|\theta - \theta_0\| \leq Kn^{-1/2}$. Then,

$$
\begin{aligned}
& n^{1/2}\left[\Psi_n\{\theta, \hat{\eta}_n(\cdot, \theta)\} - \Psi_n\{\theta_0, \hat{\eta}_n(\cdot, \theta_0)\}\right] \\
&\quad = n^{1/2}\Bigg[\mathbb{P}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_\theta, \theta)\right\}^T \Delta \\
&\qquad\quad - \mathbb{P}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta\Bigg] \\
&\qquad + n^{1/2}\Bigg[\mathbb{P}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\qquad\quad - \mathbb{P}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_0)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)^T\right\}^T \Delta\Bigg] \\
&\qquad + n^{1/2}\Bigg[\mathbb{P}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_0)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\qquad\quad - \mathbb{P}_n I(\varepsilon_0 \leq \tau)\rho(\varepsilon_0)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta\Bigg] \qquad \text{(B.6)} \\
&\qquad + n^{1/2}\Bigg[\mathbb{P}_n I(\varepsilon_0 \leq \tau)\rho(\varepsilon_0)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\qquad\quad - \mathbb{P}_n I(\varepsilon_0 \leq \tau)\rho(\varepsilon_0)\left\{(e^{\gamma_0^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta\Bigg].
\end{aligned}
$$

Consider the first difference on the right-hand side above,

$$
\begin{aligned}
& - n^{1/2}\mathbb{P}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\{\hat{\eta}_n(\varepsilon_\theta, \theta) - \hat{\eta}_n(\varepsilon_0, \theta_0)\}^T \Delta \\
&\quad = -\mathbb{G}_n I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_\theta, \theta) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \qquad \text{(B.7)} \\
&\qquad - n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_\theta, \theta) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta.
\end{aligned}
$$

The class $\{I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\hat{\eta}_n(\varepsilon_\theta, \theta)^T\Delta : \theta \in \Theta\}$ is Donsker by the arguments used in the proof of Theorem 1, and $I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_\theta, \theta) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta$ converges to zero in $L_2(P)$. Thus, the first term on the right-hand side above is $o_{P^*}(1)$.

For the second term on the right-hand side of (B.7),

$$
\begin{aligned}
& n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_\theta, \theta) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\quad = n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_\theta, \theta) - \eta_0(\varepsilon_\theta, \theta)\right\}^T \Delta \\
&\qquad - n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta \qquad \text{(B.8)} \\
&\qquad + n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\eta_0(\varepsilon_\theta, \theta) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta.
\end{aligned}
$$

We now argue that the first two terms on the right-hand side of (B.8) are asymptotically negligible. Similar to (B.4),

$$
\begin{aligned}
& n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left\{\hat{\eta}_n(\varepsilon_\theta,\theta)-\eta_0(\varepsilon_\theta,\theta)\right\}^T \Delta \\
&\quad = n^{1/2} P I(\varepsilon_\theta \leq \tau)\rho(\varepsilon_\theta)\left[\frac{1}{d^{(0)}(\varepsilon_\theta,\theta)}\left\{D_n^{(1)}(\varepsilon_\theta,\theta_0)-d^{(1)}(\varepsilon_\theta,\theta)\right\}\right. \\
&\qquad \left. -\frac{D_n^{(1)}(\varepsilon_\theta,\theta)}{D_n^{(0)}(\varepsilon_\theta,\theta)d^{(0)}(\varepsilon_\theta,\theta)}\left\{D_n^{(0)}(\varepsilon_\theta,\theta)-d^{(0)}(\varepsilon_\theta,\theta)\right\}\right]^T \Delta \\
&\quad = n^{1/2}\int I(t' \leq \tau)\rho(t')\left[\frac{1}{d^{(0)}(t',\theta)}\left\{D_n^{(1)}(t',\theta_0)-d^{(1)}(t',\theta)\right\}\right. \qquad \text{(B.9)} \\
&\qquad \left. -\frac{D_n^{(1)}(t',\theta)}{d^{(0)}(t',\theta)d^{(0)}(t',\theta)}\left\{D_n^{(0)}(t',\theta)-d^{(0)}(t',\theta)\right\}\right]^T \delta dP_{\varepsilon_0,\Delta,X,Z}(t,\delta,x,z)
\end{aligned}
$$

where $t' = t'(\theta,x,z) = \exp\{(\gamma-\gamma_0)^T z\}t + \exp\{\gamma^T z\}(\beta-\beta_0)^T x$ and $P_{\varepsilon_0,\Delta,X,Z}$ is the joint probability law of $(\varepsilon_0,\Delta,X,Z)$.

Now, as in Nan et al. (2009)[p. 2368],

$$
\begin{aligned}
& \left\| n^{1/2}\int I(t' \leq \tau)\rho(t')\left[\frac{1}{d^{(0)}(t',\theta)}\left\{D_n^{(1)}(t',\theta)-d^{(1)}(t',\theta)\right\}\right.\right. \\
&\qquad \left. -\frac{D_n^{(1)}(t',\theta)}{D_n^{(0)}(t',\theta)d^{(0)}(t',\theta)}\left\{D_n^{(0)}(t',\theta)-d^{(0)}(t',\theta)\right\}\right]^T \delta dP_{\varepsilon_0,\Delta,X,Z}(t,\delta,x,z) \\
&\quad - n^{1/2}\int I(t' \leq \tau)\rho(t')\left[\frac{1}{d^{(0)}(t',\theta)}\left\{D_n^{(1)}(t',\theta)-d^{(1)}(t',\theta)\right\}\right. \\
&\qquad \left.\left. -\frac{d^{(1)}(t',\theta)}{d^{(0)}(t',\theta)^2}\left\{D_n^{(0)}(t',\theta)-d^{(0)}(t',\theta)\right\}\right]^T \delta dP_{\varepsilon_0,\Delta,X,Z}(t,\delta,x,z)\right\| \\
&\quad = \left\| n^{1/2}\int I(t' \leq \tau)\rho(t')\left\{\frac{d^{(1)}(t',\theta)}{d^{(0)}(t',\theta)^2}-\frac{D_n^{(1)}(t',\theta)}{D_n^{(0)}(t',\theta)d^{(0)}(t',\theta)}\right\}^T\right. \\
&\qquad \left. \times\left\{D_n^{(0)}(t',\theta)-d^{(0)}(t',\theta)\right\}\delta dP_{\varepsilon_0,\Delta,X,Z}(t,\delta,x,z)\right\| \\
&\quad \leq \|\rho(t)\| \left\|\frac{d^{(1)}(t,\theta)}{d^{(0)}(t,\theta)^2}-\frac{D_n^{(1)}(t,\theta)}{D_n^{(0)}(t,\theta)d^{(0)}(t,\theta)}\right\| \left\| n^{1/2}\left\{D_n^{(0)}(t,\theta)-d^{(0)}(t,\theta)\right\}\right\| \\
&\quad = O_{P^*}(1)o_{P^*}(1)O_{P^*}(1) = o_{P^*}(1).
\end{aligned}
$$

Thus, similar to (B.4), (B.9) is

$$
\begin{aligned}
& \int I(t' \leq \tau)\rho(t')\delta d^{(0)}(t',\theta)^{-1} \\
&\qquad \times \mathbb{G}_n I(\varepsilon_\theta \geq t')\left[\left\{\exp(\gamma^T Z)X^T, Z^T\right\}-\eta_0(t',\theta)\right]^T dP_{\varepsilon_0,\Delta,X,Z}(t,\delta,x,z) \\
&\quad = \int \mathbb{G}_n I(t' \leq \tau)\rho(t')\ell(t',\theta,X,Z,\varepsilon_\theta) dP_{\varepsilon_0,\Delta,X,Z}(t,1,x,z)
\end{aligned}
$$

where $\ell(t', \theta, X, Z, \varepsilon_\theta) = d^{(0)}(t', \theta)^{-1} I(\varepsilon_\theta \geq t') \left[\left\{\exp(\gamma^T Z) X^T, Z^T\right\} - \eta_0(t', \theta)\right]^T$.

Similarly,

$$
\begin{aligned}
& n^{1/2} P I(\varepsilon_\theta \leq \tau) \rho(\varepsilon_\theta) \left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta \\
& \qquad = \int \mathbb{G}_n I(t' \leq \tau) \rho(t') \ell(t, \theta_0, X, Z, \varepsilon_0) dP_{\varepsilon_0, \Delta, X, Z}(t, 1, x, z) + o_{P^*}(1).
\end{aligned}
$$

Thus, the first two terms on the right-hand side of (B.8) equate to

$$
\int \mathbb{G}_n I(t' \leq \tau) \rho(t') \left\{\ell(t', \theta, X, Z, \varepsilon_\theta) - \ell(t, \theta_0, X, Z, \varepsilon_0)\right\} dP_{\varepsilon_0, \Delta, X, Z}(t, 1, x, z)
$$

which is $o_{P^*}(1)$ as $\{\ell(t, \theta, X, Z, \varepsilon_\theta) : t \in]-\infty, \tau], \theta \in \Theta\}$ is a Donsker class, and $\ell(t', \theta, X, Z, \varepsilon_\theta) - \ell(t, \theta_0, X, Z, \varepsilon_0)$ converges to zero in $L_2(P)$.

Thus, (B.8) is

$$
\begin{aligned}
& n^{1/2} P I(\varepsilon_\theta \leq \tau) \rho(\varepsilon_\theta) \left\{\eta_0(\varepsilon_\theta, \theta) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta + o_{P^*}(1) \\
& \quad = n^{1/2} \left\{P I(\varepsilon_\theta \leq \tau) \rho(\varepsilon_\theta) \dot{\eta}_\theta(\varepsilon_0, \theta_0)^T + o_{P^*}(1)\right\} \Delta(\theta - \theta_0) + o_{P^*}(1) \qquad \text{(B.10)} \\
& \quad = \left\{P I(\varepsilon_0 \leq \tau) \rho(\varepsilon_0) \dot{\eta}_\theta(\varepsilon_0, \theta_0)^T \Delta\right\} n^{1/2}(\theta - \theta_0) + o_{P^*}(1).
\end{aligned}
$$

The first equality in (B.10) follows from using the assumption of bounded density functions for failure and censoring times, that is, assumptions 4 and 5, together with the dominated convergence theorem.

The second difference on the right-hand side of (B.6) is

$$
\begin{aligned}
& n^{1/2} \mathbb{P}_n I(\varepsilon_\theta \leq \tau) \left\{\rho(\varepsilon_\theta) - \rho(\varepsilon_0)\right\} \left\{(e^{\gamma^T Z} X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
& \quad = \mathbb{P}_n I(\varepsilon_\theta \leq \tau) diag \left\{(e^{\gamma^T Z} X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\} \left\{\dot{\rho}_\theta(\varepsilon_0) + o_{P^*}(1)\right\} \Delta n^{1/2}(\theta - \theta_0) \qquad \text{(B.11)} \\
& \quad = \left[P I(\varepsilon_0 \leq \tau) diag \left\{(e^{\gamma_0^T Z} X^T, Z^T) - \eta_0(\varepsilon_0, \theta_0)\right\} \dot{\rho}_\theta(\varepsilon_0) \Delta\right] n^{1/2}(\theta - \theta_0) + o_{P^*}(1).
\end{aligned}
$$

For the third difference on the right-hand side of (B.6) let $h(t, \theta, X, Z) = pr(\varepsilon_\theta \leq t,\ \Delta = 1 | X, Z)$. Then, according to assumptions 2, 3 and 4, h has bounded continuous derivative $\dot{h}_\theta(t, \theta, X, Z)$ with respect to θ. We will establish

$$
\begin{aligned}
& n^{1/2} \mathbb{P}_n \left\{I(\varepsilon_\theta \leq \tau) - I(\varepsilon_0 \leq \tau)\right\} \rho(\varepsilon_0) \left\{(e^{\gamma^T Z} X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \qquad \text{(B.12)} \\
& \quad = \left[P \rho(\tau) \left\{(e^{\gamma_0^T Z} X^T, Z^T) - \eta_0(\tau, \theta_0)^T\right\} \dot{h}_\theta(\tau, \theta_0, X, Z)\right] n^{1/2}(\theta - \theta_0) + o_{P^*}(1).
\end{aligned}
$$

To see this, first note that

$$
\begin{aligned}
& n^{1/2} \mathbb{P}_n \left\{I(\varepsilon_\theta \leq \tau) - I(\varepsilon_0 \leq \tau)\right\} \rho(\varepsilon_0) \left\{(e^{\gamma^T Z} X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
& \quad = \mathbb{G}_n \left\{I(\varepsilon_\theta \leq \tau) - I(\varepsilon_0 \leq \tau)\right\} \rho(\varepsilon_0) \left\{(e^{\gamma^T Z} X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
& \qquad + n^{1/2} P \left\{I(\varepsilon_\theta \leq \tau) - I(\varepsilon_0 \leq \tau)\right\} \rho(\varepsilon_0) \left\{(e^{\gamma^T Z} X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta.
\end{aligned}
$$

By similar arguments as above the first term on the right rand side is $o_{P^*}(1)$. For the second term on the right hand side we have

$$
\begin{aligned}
&n^{1/2}P\left\{I(\varepsilon_\theta \le \tau) - I(\varepsilon_0 \le \tau)\right\}\rho(\varepsilon_0)\left\{(e^{\gamma^T Z}X^T, Z^T) - \hat{\eta}_n(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\quad = n^{1/2}P\left\{I(\varepsilon_\theta \le \tau) - I(\varepsilon_0 \le \tau)\right\}\rho(\varepsilon_0)\left\{(e^{\gamma_0^T Z}X^T, Z^T) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\qquad + n^{1/2}P\left\{I(\varepsilon_\theta \le \tau) - I(\varepsilon_0 \le \tau)\right\}\rho(\varepsilon_0)\left\{(e^{\gamma^T Z}X^T, Z^T) - (e^{\gamma_0^T Z}X^T, Z^T)\right\}^T \Delta \\
&\qquad - n^{1/2}P\left\{I(\varepsilon_\theta \le \tau) - I(\varepsilon_0 \le \tau)\right\}\rho(\varepsilon_0)\left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta.
\end{aligned}
$$

The second term on the right hand side is clearly $o_{P^*}(1)$. For the third term note that, similar to (B.9), we have

$$
\begin{aligned}
&n^{1/2}P\left\{I(\varepsilon_\theta \le \tau) - I(\varepsilon_0 \le \tau)\right\}\rho(\varepsilon_0)\left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \\
&\quad = \int \mathbb{G}_n\{I(t' \le \tau) - I(t \le \tau)\}\rho(t)\ell(t, \theta_0, X, Z, \varepsilon_0)dP_{\varepsilon_0, \Delta, X, Z}(t, 1, x, z) + o_{P^*}(1).
\end{aligned}
$$

Accordingly this term is also $o_{P^*}(1)$ since $\{I(t' \le \tau) - I(t \le \tau)\}\rho(t)\ell(t, \theta_0, X, Z, \varepsilon_0)$ converges to zero in $L_2(P)$. Finally, for the first term we have

$$
\begin{aligned}
&n^{1/2}P\left\{I(\varepsilon_\theta \le \tau) - I(\varepsilon_0 \le \tau)\right\}\rho(\varepsilon_0)\left\{(e^{\gamma_0^T Z}X^T, Z^T) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\quad = \left[P\rho(\tau)\left\{(e^{\gamma_0^T Z}X^T, Z^T) - \eta_0(\tau, \theta_0)^T\right\}\dot{h}_\theta(\tau, \theta_0, X, Z)\right]n^{1/2}(\theta - \theta_0) + o_{P^*}(1),
\end{aligned}
$$

where we have used that for a continuous density p on the real line, continuous function f with $\int_{-\infty}^{\infty}|f(t)|p(t)dt < \infty$, and continuously differentiable function $g(\theta)$ we have:

$$
\begin{aligned}
\int_{-\infty}^{g(\theta)} f(s)p(s)ds - \int_{-\infty}^{g(\theta_0)} f(s)p(s)ds &= f\{g(\theta_0)\}p\{g(\theta_0)\}\{g(\theta) - g(\theta_0)\} + o(|g(\theta) - g(\theta_0|) \\
&= f\{g(\theta_0)\}p\{g(\theta_0)\}\dot{g}_\theta(\theta_0)(\theta - \theta_0) + o(\|\theta - \theta_0\|).
\end{aligned}
$$

Now, consider the last difference on the right-hand side of (B.6):

$$
\begin{aligned}
&n^{1/2}\mathbb{P}_n I(\varepsilon_0 \le \tau)\rho(\varepsilon_0)\begin{bmatrix} \{\exp(\gamma^T Z) - \exp(\gamma_0^T Z)\}X \\ 0_{q\times 1} \end{bmatrix}\Delta \\
&\quad = \mathbb{P}_n I(\varepsilon_0 \le \tau)\rho(\varepsilon_0)\begin{Bmatrix} 0_{p\times q} & XZ^T\exp(\gamma_0^T Z) \\ 0_{q\times q} & 0_{q\times p} \end{Bmatrix}\Delta n^{1/2}(\theta - \theta_0) + o_{P^*}(1) \\
&\quad = \left[PI(\varepsilon_0 \le \tau)\rho(\varepsilon_0)\begin{Bmatrix} 0_{p\times q} & XZ^T\exp(\gamma_0^T Z) \\ 0_{q\times q} & 0_{q\times p} \end{Bmatrix}\Delta\right]n^{1/2}(\theta - \theta_0) + o_{P^*}(1). \qquad \text{(B.13)}
\end{aligned}
$$

From (B.10), (B.11), (B.12), and (B.13), (B.6) is

$$
n^{1/2}\left[\Psi_n\{\theta, \hat{\eta}_n(\cdot, \theta)\} - \Psi_n\{\theta_0, \hat{\eta}_n(\cdot, \theta_0)\}\right] = \dot{\Psi}_\theta\{\theta_0, \eta_0(\cdot, \theta_0)\}n^{1/2}(\theta - \theta_0) + o_{P^*}(1).
$$

On the other hand, inserting $\hat{\theta}_n$ into the left-hand side above,

$$
\begin{aligned}
&n^{1/2}\Psi_n\{\hat{\theta}_n, \hat{\eta}_n(\cdot, \hat{\theta}_n)\} - n^{1/2}\Psi_n\{\theta_0, \hat{\eta}_n(\cdot, \theta_0)\} \\
&\quad = o_{P^*}(1) - n^{1/2}\Psi_n\{\theta_0, \hat{\eta}_n(\cdot, \theta_0)\} \\
&\quad = o_{P^*}(1) - \mathbb{G}_n\psi\{O; \theta_0, \eta_0(\cdot, \theta_0)\} + \mathbb{G}_n I(\varepsilon_0 \le \tau)\rho(\varepsilon_0)\left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\qquad + n^{1/2} P I(\varepsilon_0 \le \tau)\rho(\varepsilon_0)\left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta.
\end{aligned}
$$

The third term on the right-hand side above is $o_{P^*}(1)$. For the last term above,

$$
\begin{aligned}
&n^{1/2} P I(\varepsilon_0 \le \tau)\rho(\varepsilon_0)\left\{\hat{\eta}_n(\varepsilon_0, \theta_0) - \eta_0(\varepsilon_0, \theta_0)\right\}^T \Delta \\
&\quad = n^{1/2}\int_{-\infty}^{\tau} \rho(t) d^{(0)}(t, \theta)^{-1}\left[\left\{D_n^{(1)}(t, \theta_0) - D_n^{(0)}(t, \theta_0)\eta_0(t, \theta_0)\right\}\right. \\
&\qquad \left. - \left\{d^{(1)}(t, \theta_0) - d^{(0)}(t, \theta_0)\eta_0(t, \theta_0)\right\}\right]^T dP_{\varepsilon_0, \Delta}(t, 1) + o_{P^*}(1) \\
&\quad = \int_{-\infty}^{\tau} \rho(t) d^{(0)}(t, \theta_0)^{-1}\mathbb{G}_n I(\varepsilon_0 \ge t)\left[\left\{\exp(\gamma_0^T Z)X^T, Z^T\right\} - \eta_0(t, \theta_0)\right]^T dP_{\varepsilon_0, \Delta}(t, 1) + o_{P^*}(1) \\
&\quad = \mathbb{G}_n \int_{-\infty}^{\tau} \rho(t) I(\varepsilon_0 \ge t)\left[\left\{\exp(\gamma_0^T Z)X^T, Z^T\right\} - \eta_0(t, \theta_0)\right]^T dA_0(t) + o_{P^*}(1),
\end{aligned}
$$

where A_0 is the cumulative hazard of the error term $e = e^{\gamma_0^T Z}(\log T + \beta_0^T X)$.

Thus, combining the three displays above

$$
\begin{aligned}
&n^{1/2}(\hat{\theta}_n - \theta_0) \\
&\quad = -\dot{\Psi}_\theta^{-1}\{\theta_0, \eta_0(\cdot, \theta_0)\} \\
&\qquad \times \mathbb{G}_n\left(\psi\{O; \theta_0, \eta_0(\cdot, \theta_0)\} - \int_{-\infty}^{\tau} \rho(t) I(\varepsilon_0 \ge t)\{(\exp(\gamma_0^T Z)X^T, Z^T) - \eta(t, \theta_0)\}^T dA(t)\right) \\
&\qquad + o_{P^*}(1).
\end{aligned}
$$

■

Theorem 3 *Let $\hat{\theta}_n$ be an estimator of θ_0 such that $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges weakly to a zero mean normal distribution. Then $n^{1/2}\{\hat{A}_n(\hat{\theta}_n, \cdot) - A_0(\cdot)\}$ converges weakly to a tight zero mean Gaussian process on $]-\infty, \tau]$ and the following holds*

$$
n^{1/2}\{\hat{A}_n(\hat{\theta}_n, t) - A_0(t)\} = \dot{\phi}_\theta(t, \theta_0) n^{1/2}(\hat{\theta}_n - \theta_0) + \mathbb{G}_n H(t, \theta_0, d^{(0)}, A_0) + o_{P^*}(1).
$$

Proof. Let $\|\theta - \theta_0\| < K n^{-1/2}$ for some $K < \infty$. Then note that

$$
n^{1/2}\{\hat{A}(t, \theta) - \hat{A}(t, \theta_0)\} = n^{1/2}\mathbb{P}_n\left\{\frac{I(\varepsilon_\theta \le t)\Delta}{D_n^{(0)}(\varepsilon_\theta, \theta)} - \frac{I(\varepsilon_0 \le t)\Delta}{D_n^{(0)}(\varepsilon_0, \theta_0)}\right\}.
$$

As in (B.8) in the proof of Theorem 2 one may show that

$$n^{1/2}\mathbb{P}_n\left[I(\varepsilon_\theta \le t)\Delta\left\{\frac{1}{D_n^{(0)}(\varepsilon_\theta,\theta)}-\frac{1}{D_n^{(0)}(\varepsilon_0,\theta_0)}\right\}\right]$$
$$=-n^{1/2}P\left[I(\varepsilon_0\le t)\Delta\frac{\dot{d}_\theta^{(0)}(\varepsilon_0,\theta_0)}{d^{(0)}(\varepsilon_0,\theta_0)^2}\right](\theta-\theta_0)+o_{P^*}(1).$$

As in (B.12) in the proof of Theorem 2 one may also show that

$$n^{1/2}\mathbb{P}_n\left[\frac{\Delta\{I(\varepsilon_\theta\le t)-I(\varepsilon_0\le t)\}}{D_n^{(0)}(\varepsilon_0,\theta_0)}\right]=n^{1/2}P\left\{\frac{\dot{h}_\theta(t,\theta_0,X,Z)}{d^{(0)}(t,\theta_0)}\right\}(\theta-\theta_0)+o_{P^*}(1).$$

Combining the displays above, we get

$$n^{1/2}\{\hat{A}(t,\hat{\theta}_n)-\hat{A}(t,\theta_0)\}=\dot{\phi}_\theta(t,\theta_0)n^{1/2}(\hat{\theta}_n-\theta_0)+o_{P^*}(1). \tag{B.14}$$

Secondly note that

$$n^{1/2}\{\hat{A}(t,\theta_0)-A_0(t)\}$$
$$=\mathbb{G}_n\left[I(\varepsilon_0\le t)\Delta\left\{\frac{1}{D_n^{(0)}(\varepsilon_0,\theta_0)}-\frac{1}{d^{(0)}(\varepsilon_0,\theta_0)}\right\}\right]$$
$$+n^{1/2}P\left[I(\varepsilon_0\le t)\Delta\left\{\frac{1}{D_n^{(0)}(\varepsilon_0,\theta_0)}-\frac{1}{d^{(0)}(\varepsilon_0,\theta_0)}\right\}\right]+\mathbb{G}_n\left\{\frac{I(\varepsilon_0\le t)\Delta}{d^{(0)}(\varepsilon_0,\theta_0)}\right\}. \tag{B.15}$$

By similar arguments as in the proof of Theorem 2 one may show that the first term on the right hand side of (B.15) converges in probability to 0. For the second term note that

$$n^{1/2}P\left[I(\varepsilon_0\le t)\Delta\left\{\frac{1}{D_n^{(0)}(\varepsilon_0,\theta_0)}-\frac{1}{d^{(0)}(\varepsilon_0,\theta_0)}\right\}\right]$$
$$=n^{1/2}\int_{-\infty}^{t}\left\{\frac{1}{D_n^{(0)}(s,\theta_0)}-\frac{1}{d^{(0)}(s,\theta_0)}\right\}dP_{\varepsilon_0,\Delta}(s,1)$$
$$=-n^{1/2}\int_{-\infty}^{t}\frac{1}{D_n^{(0)}(s,\theta_0)d^{(0)}(s,\theta_0)}\left\{D^{(0)}(s,\theta_0)-d^{(0)}(s,\theta_0)\right\}dP_{\varepsilon_0,\Delta}(s,1)$$
$$=-n^{1/2}\int_{-\infty}^{t}\frac{1}{D_n^{(0)}(s,\theta_0)}\left\{D_n^{(0)}(s,\theta_0)-d^{(0)}(s,\theta_0)\right\}dA(s)$$
$$=-\int_{-\infty}^{t}\frac{1}{d^{(0)}(s,\theta_0)}\mathbb{G}_nI(\varepsilon_0\ge s)dA_0(s)+o_{P^*}(1)$$
$$=-\mathbb{G}_n\int_{-\infty}^{t}\frac{1}{d^{(0)}(s,\theta_0)}I(\varepsilon_0\ge s)dA_0(s)+o_{P^*}(1),$$

from which it follows by Lemma 1 that the second term converges weakly to a tight zero mean Gaussian process. It also follows from Lemma 1 that the third term in (B.15) converges weakly to a tight zero mean Gaussian process. Combining (B.14) and (B.15) we have that $n^{1/2}\{\hat{A}(\hat{\theta}_n,\cdot)-A(\cdot)\}$ converges weakly to a tight zero mean Gaussian process on $]-\infty,\tau]$ and that

$$n^{1/2}\{\hat{A}_n(\hat{\theta}_n,t)-A_0(t)\} = \\ \dot{\phi}_\theta(t,\theta_0)n^{1/2}(\hat{\theta}_n-\theta_0)+\mathbb{G}_n\left\{\frac{I(\varepsilon_0\le t)}{\Delta d^{(0)}(\varepsilon_0,\theta_0)}-\int_{-\infty}^{t}\frac{I(\varepsilon_0\ge s)}{d^{(0)}(s,\theta)}dA_0(s)\right\}+o_{P^*}(1).$$

■

C Conditional multiplier method

While it is straightforward to compute confidence intervals for θ_0 and $A_0(t)$ (once we can estimate the limiting variances as discussed in the main paper), we now discuss how confidence intervals for functions of θ_0 and $A_0(t)$ can be produced. For this, we rely on empirical process theory via the conditional multiplier method (cf. van der Vaart and Wellner (1996)).

From empirical process theory we have that for a Gaussian vector $G=(G_1,\dots,G_n)$, the limiting distribution of

$$-\dot{\Psi}_\theta^{-1}n^{-1/2}\sum_{i=1}^{n}J(O_i)G_i$$

is the same as that of $n^{1/2}(\hat{\theta}_n-\theta_0)$. Defining G^b to be one such randomly generated Gaussian vector ($b=1,\dots,m$), we may compute

$$\hat{\theta}^b=\hat{\theta}_n-\hat{\dot{\Psi}}_\theta^{-1}n^{-1}\sum_{i=1}^{n}\hat{J}_iG_i^b. \tag{C.16}$$

Hence, generating a sample of G^b vectors produces a sample of $\hat{\theta}^b$ vectors. Note that the quantiles of the sample $\{\hat{\theta}_j^1,\dots,\hat{\theta}_j^m\}$ may be used to form confidence intervals for θ_j.

We now turn to $A_0(t)$ where we have that

$$n^{-1/2}\sum_{i=1}^{n}\left[-\dot{\phi}_\theta(t)\dot{\Psi}_\theta^{-1}J(O_i)+H(O_i;t)\right]G_i$$

has the same limiting distribution as $n^{1/2}\{\hat{A}_n(\hat{\theta}_n,t)-A_0(t)\}$. However, it is well known that transforming to the unrestricted $\log A_0(t)$ scale (and subsequently back-transforming) is preferable. Thus, we consider

$$n^{-1/2}\frac{1}{\hat{A}_n(\hat{\theta}_n,t)}\sum_{i=1}^{n}\left[-\dot{\phi}_\theta(t)\dot{\Psi}_\theta^{-1}J(O_i)+H(O_i;t)\right]G_i$$

which has the same limiting distribution as $n^{1/2}\{\log \hat{A}_n(\hat{\theta}_n, t) - \log A_0(t)\}$. Hence, computing

$$\hat{A}^b(t) = \hat{A}_n(\hat{\theta}_n, t) \exp\left\{n^{-1}\frac{1}{\hat{A}_n(\hat{\theta}_n, t)}\sum_{i=1}^{n}\left[-\hat{\dot{\phi}}_\theta(t)\hat{\dot{\Psi}}_\theta^{-1}\hat{J}_i + \hat{H}_i(t)\right] G_i^b\right\} \tag{C.17}$$

for $b = 1, \ldots, m$ creates a sample whose quantiles may be used to produce confidence bands for $A_0(t)$.

When applying (C.16) and (C.17) above, we must maintain the same set of Gaussian vectors, $\{G^1, \ldots, G^m\}$, i.e., both $\hat{\theta}^b$ and $\hat{A}^b(t)$ are generated from the same Gaussian vector G^b (for $b = 1, \ldots, m$). This has the effect of respecting the dependence structure between the estimators $\hat{\theta}_n$ and $\hat{A}_n(\hat{\theta}_n, t)$ which propagates into any functions of these estimates. Because the limiting distribution of $\{\hat{\theta}^b, \hat{A}^b(t)\}$ is the same as that of $\{\hat{\theta}_n, \hat{A}_n(\hat{\theta}_n, t)\}$, we have, from the continuous mapping theorem, that the limiting distribution of $w\{\hat{\theta}^b, \hat{A}^b(t)\}$ is the same as that of $w\{\hat{\theta}_n, \hat{A}_n(\hat{\theta}_n, t)\}$ where $w(\cdot, \cdot)$ is a continuous function of the parameters and error cumulative hazard function. Hence, from the simulated sample $\{\hat{\theta}^b, \hat{A}^b(t)\}$, $b = 1, \ldots, m$, we may produce confidence bands for any functional of interest. As an example, consider the conditional survivor function for our proposed model which is given by

$$S(t \mid x_i, z_i) = \exp\left\{-A_0\left(\frac{\log t - \mu_{i0}}{\sigma_{i0}}\right)\right\}$$

where $\mu_{i0} = -x_i^T\beta_0$ and $\sigma_{i0} = \exp(-z_i^T\gamma_0)$. Hence, we can compute

$$\hat{S}^b(t \mid x_i, z_i) = \exp\left\{-\hat{A}^b\left(\frac{\log t - \hat{\mu}_i^b}{\hat{\sigma}_i^b}\right)\right\}$$

where $\hat{\mu}_i^b = -x_i^T\hat{\beta}^b$ and $\hat{\sigma}_i^b = \exp(-z_i^T\hat{\gamma}^b)$, $b = 1, \ldots, m$, from which confidence bands can be produced.

D Additional simulation results

In Section 4.1 of the main paper, we presented a subset of a larger simulation study, the results of which are contained here. The details of the full simulation study are as described in the main paper with the addition of the sample sizes $n = 50$ and $n = 500$, and, furthermore, the Gehan weight, $\rho(u) = \sum_{j=1}^{n} Y_j^*(u)/n$, was also considered.

Tables 3 - 5 below display the bias and coverage percentages for each of the three weight function types, while Tables 6 - 8 show the empirical and estimated standard errors. In all cases the bias is low, the coverage is close to the nominal level, and our proposed variance estimators are adequately capturing the true variations in estimation (and, indeed, the efficiency is similar across the three weight function choices).

Table 3: Log-rank bias and coverage

			$n = 50$		$n = 100$		$n = 500$	
τ	Cens.	Parameter	Bias	Cov.	Bias	Cov.	Bias	Cov.
2	20%	β_1	0.004	94.6	0.000	94.8	0.000	95.0
		β_2	0.017	94.3	0.008	94.0	-0.002	94.8
		γ_1	-0.033	93.8	-0.001	94.1	0.000	94.6
		A	-0.016	94.9	-0.009	94.5	-0.002	95.1
		S	0.000	96.6	0.000	94.6	0.000	94.8
		r	0.007	94.8	0.003	95.1	0.001	95.0
2	50%	β_1	-0.004	95.6	0.000	94.2	0.000	94.8
		β_2	0.029	95.2	0.005	93.3	-0.001	94.1
		γ_1	-0.064	93.1	-0.015	94.6	-0.004	94.3
		A	-0.018	93.5	-0.010	95.3	-0.003	94.5
		S	0.004	97.1	0.008	95.3	0.001	94.5
		r	0.016	95.0	0.006	95.5	0.000	94.8
∞	20%	β_1	0.000	95.1	0.000	94.3	-0.001	95.4
		β_2	0.017	94.8	0.003	93.6	0.000	94.6
		γ_1	-0.036	93.9	-0.007	93.4	0.002	94.3
		A	-0.012	94.7	-0.004	94.3	-0.001	94.9
		S	-0.002	96.2	0.001	94.5	0.000	95.2
		r	0.006	95.1	0.004	95.3	0.001	95.2
∞	50%	β_1	-0.005	95.0	0.003	94.5	-0.001	94.9
		β_2	0.031	95.2	0.001	93.1	0.000	94.5
		γ_1	-0.066	92.8	-0.016	93.7	0.000	94.4
		A	-0.016	94.5	-0.010	95.1	-0.002	94.8
		S	0.002	97.3	0.009	95.3	0.002	94.6
		r	0.018	94.9	0.005	95.4	0.001	94.7

Cens., censored proportion; Bias, median bias; Cov., empirical coverage percentage for 95% confidence interval; $A = A(0)$; $S = S(t_{0.5}^{(1)} \mid x^{(1)})$; $r = r(x^{(1)}, x^{(2)})$.

Table 4: Gehan bias and coverage

			$n = 50$		$n = 100$		$n = 500$	
τ	Cens.	Parameter	Bias	Cov.	Bias	Cov.	Bias	Cov.
2	20%	β_1	0.000	95.5	0.003	95.6	0.000	94.7
		β_2	0.015	95.2	0.002	94.7	0.000	94.7
		γ_1	-0.003	95.2	0.009	94.4	0.001	95.0
		A	-0.006	94.9	-0.008	94.7	-0.002	94.3
		S	0.003	96.2	-0.001	95.5	-0.001	95.1
		r	0.010	95.7	0.001	95.5	0.001	94.4
2	50%	β_1	-0.010	95.7	0.000	94.6	0.001	95.4
		β_2	0.031	95.2	0.007	93.6	-0.001	95.1
		γ_1	-0.018	95.1	0.012	94.4	0.007	94.8
		A	-0.003	94.4	-0.005	94.4	-0.002	94.8
		S	0.000	97.6	0.003	95.4	0.000	95.0
		r	0.022	95.4	0.007	95.1	0.000	95.1
∞	20%	β_1	-0.004	95.1	0.000	94.5	0.001	94.9
		β_2	0.013	95.6	0.002	93.8	-0.002	95.0
		γ_1	-0.008	95.0	0.008	94.3	0.004	94.6
		A	0.001	94.8	0.003	95.1	-0.001	95.0
		S	0.005	96.3	0.003	94.7	0.001	95.1
		r	0.014	95.5	0.007	95.4	0.000	95.1
∞	50%	β_1	-0.007	95.6	-0.005	94.3	-0.001	94.7
		β_2	0.028	95.4	0.009	93.2	0.002	94.6
		γ_1	-0.027	94.7	0.012	92.9	0.003	94.8
		A	-0.009	94.6	-0.001	94.7	-0.003	94.7
		S	-0.002	97.1	0.001	95.2	0.000	95.4
		r	0.017	95.5	0.010	95.1	0.002	94.7

Cens., censored proportion; Bias, median bias; Cov., empirical coverage percentage for 95% confidence interval; $A = A(0)$; $S = S(t_{0.5}^{(1)} \mid x^{(1)})$; $r = r(x^{(1)}, x^{(2)})$.

Table 5: Normal (true efficient) bias and coverage

			$n=50$		$n=100$		$n=500$	
τ	Cens.	Parameter	Bias	Cov.	Bias	Cov.	Bias	Cov.
2	20%	β_1	-0.005	94.2	-0.001	94.4	-0.001	94.7
		β_2	0.009	93.9	0.002	93.1	0.001	94.6
		γ_1	-0.011	91.8	0.004	91.8	0.002	94.9
		A	-0.009	94.1	-0.002	94.4	0.000	94.3
		S	0.000	95.3	0.000	94.3	-0.001	94.5
		r	0.012	95.2	0.004	94.8	0.001	94.6
2	50%	β_1	-0.013	95.2	-0.006	94.8	0.000	95.3
		β_2	0.037	93.8	0.003	93.7	-0.001	95.0
		γ_1	-0.030	92.2	0.004	93.2	-0.001	93.9
		A	-0.015	94.7	-0.003	94.8	-0.001	95.0
		S	-0.002	97.3	0.005	95.5	0.001	95.5
		r	0.020	96.3	0.008	95.9	0.001	94.8
∞	20%	β_1	-0.006	94.1	-0.001	94.3	-0.001	94.5
		β_2	0.017	93.1	-0.002	93.4	0.000	95.1
		γ_1	0.000	91.5	0.003	93.2	0.005	94.2
		A	-0.008	94.4	-0.005	94.2	0.000	95.2
		S	0.000	95.2	0.004	94.4	0.000	94.6
		r	0.010	95.2	0.005	94.6	0.001	94.9
∞	50%	β_1	-0.012	95.2	0.000	94.1	0.000	95.0
		β_2	0.027	94.2	-0.001	92.6	-0.002	94.6
		γ_1	-0.022	92.1	0.008	92.0	0.000	94.0
		A	-0.011	94.1	-0.008	94.5	-0.003	94.7
		S	0.001	96.9	0.005	94.4	0.002	95.0
		r	0.021	95.8	0.007	95.3	0.001	95.1

Cens., censored proportion; Bias, median bias; Cov., empirical coverage percentage for 95% confidence interval; $A = A(0)$; $S = S(t_{0.5}^{(1)} \mid x^{(1)})$; $r = r(x^{(1)}, x^{(2)})$.

Table 6: Log-rank standard errors

τ	Cens.	Parameter	$n = 50$ SE	$n = 50$ SEE	$n = 100$ SE	$n = 100$ SEE	$n = 500$ SE	$n = 500$ SEE
2	20%	β_1	0.148	0.142	0.099	0.099	0.045	0.044
		β_2	0.239	0.261	0.189	0.185	0.087	0.085
		γ_1	0.237	0.256	0.174	0.179	0.081	0.082
		A	0.252	0.223	0.168	0.156	0.071	0.071
2	50%	β_1	0.186	0.182	0.127	0.124	0.056	0.055
		β_2	0.264	0.301	0.216	0.210	0.099	0.097
		γ_1	0.296	0.327	0.208	0.230	0.107	0.105
		A	0.355	0.290	0.219	0.203	0.095	0.092
∞	20%	β_1	0.149	0.142	0.100	0.099	0.044	0.044
		β_2	0.240	0.256	0.189	0.183	0.085	0.085
		γ_1	0.236	0.252	0.175	0.177	0.082	0.081
		A	0.245	0.221	0.165	0.156	0.072	0.071
∞	50%	β_1	0.200	0.183	0.126	0.124	0.056	0.055
		β_2	0.268	0.298	0.215	0.208	0.099	0.097
		γ_1	0.306	0.328	0.215	0.230	0.106	0.105
		A	0.361	0.293	0.230	0.204	0.094	0.093

SE, standard error of estimates; SEE, median of estimated standard error.

Table 7: Gehan standard errors

			$n = 50$		$n = 100$		$n = 500$	
τ	Cens.	Parameter	SE	SEE	SE	SEE	SE	SEE
2	20%	β_1	0.137	0.142	0.096	0.099	0.044	0.044
		β_2	0.229	0.253	0.177	0.178	0.080	0.080
		γ_1	0.243	0.249	0.178	0.174	0.077	0.077
		A	0.244	0.223	0.166	0.156	0.072	0.070
2	50%	β_1	0.174	0.182	0.123	0.123	0.054	0.054
		β_2	0.250	0.289	0.207	0.201	0.092	0.091
		γ_1	0.294	0.318	0.229	0.225	0.102	0.100
		A	0.337	0.289	0.229	0.201	0.093	0.091
∞	20%	β_1	0.139	0.144	0.098	0.099	0.044	0.044
		β_2	0.229	0.255	0.180	0.178	0.079	0.080
		γ_1	0.242	0.247	0.176	0.174	0.078	0.077
		A	0.245	0.223	0.161	0.156	0.071	0.070
∞	50%	β_1	0.177	0.181	0.127	0.123	0.055	0.054
		β_2	0.252	0.289	0.207	0.202	0.092	0.092
		γ_1	0.301	0.320	0.234	0.223	0.101	0.100
		A	0.335	0.286	0.226	0.201	0.093	0.091

SE, standard error of estimates; SEE, median of estimated standard error.

Table 8: Normal (true, efficient) standard errors

			$n = 50$		$n = 100$		$n = 500$	
τ	Cens.	Parameter	SE	SEE	SE	SEE	SE	SEE
2	20%	β_1	0.141	0.137	0.097	0.096	0.043	0.043
		β_2	0.240	0.236	0.178	0.169	0.079	0.078
		γ_1	0.244	0.217	0.176	0.157	0.074	0.073
		A	0.242	0.216	0.164	0.153	0.071	0.069
2	50%	β_1	0.181	0.179	0.120	0.121	0.054	0.054
		β_2	0.285	0.274	0.201	0.193	0.089	0.088
		γ_1	0.319	0.292	0.215	0.206	0.100	0.096
		A	0.343	0.285	0.217	0.201	0.092	0.090
∞	20%	β_1	0.142	0.136	0.098	0.096	0.043	0.043
		β_2	0.247	0.235	0.179	0.169	0.078	0.077
		γ_1	0.249	0.216	0.171	0.156	0.074	0.073
		A	0.246	0.213	0.165	0.152	0.069	0.069
∞	50%	β_1	0.177	0.179	0.124	0.121	0.053	0.054
		β_2	0.279	0.274	0.207	0.191	0.089	0.088
		γ_1	0.305	0.291	0.224	0.206	0.098	0.096
		A	0.336	0.289	0.223	0.201	0.093	0.091

SE, standard error of estimates; SEE, median of estimated standard error.

E Lung cancer analysis

Below is the table of estimated coefficients and standard errors using the log-rank, Gehan, and normal weight functions for the lung cancer data presented in the main paper; the estimates and standard errors are very similar in all cases.

Table 9: Lung Cancer

		Log-rank		Gehan		Normal	
	Group	Est.	SE	Est.	SE	Est.	SE
Scale	Palliative	0.000	—	0.000	—	0.000	—
	Surgery	-2.645	0.174	-2.644	0.174	-2.623	0.173
	Chemotherapy	-0.537	0.267	-0.487	0.272	-0.536	0.219
	Radiotherapy	-1.075	0.104	-1.076	0.103	-1.010	0.111
	Chemo&Radio	-1.868	0.118	-1.867	0.120	-1.866	0.113
Shape	Palliative	0.000	—	0.000	—	0.000	—
	Surgery	0.308	0.195	0.296	0.194	0.298	0.168
	Chemotherapy	0.040	0.115	-0.006	0.100	0.053	0.093
	Radiotherapy	0.296	0.072	0.271	0.069	0.216	0.073
	Chemo&Radio	0.943	0.173	0.944	0.157	0.944	0.148