1,

$$IC(Y \mid Q, G, D_h(\cdot \mid \beta, \Lambda_0)) = IC_0(Y \mid G, D_h(\cdot \mid \beta, \Lambda_0)) - \int Q(u, \mathcal{F}(u)) dM_G(u),$$

where

$$Q = E(IC_0(Y) \mid C = u, \bar{X}(u)) - E(IC_0(Y) \mid C > u, \bar{X}(u)).$$

Note that $Q(u, \mathcal{F}(u))$ is the difference between the regression $E(IC_0(Y) \mid \bar{A}(u), \bar{X}(\min(C, u))$ evaluated at $\bar{A}(u) = I(C = u)$ and $\bar{A}(u) = 0$, where we recall $A(u) = I(C \leq u)$. The latter suggests a simple estimation procedure for $Q(u, \mathcal{F}(u))$. If $IC_0 = IC_{01}$, then

$$Q(u, \mathcal{F}(u)) = -E(IC_{01}(Y \mid G, D_h) \mid C > u, \bar{X}(u)).$$

If $IC_0 = IC_{02}$, then it is straightforward to show that

$$Q(u, \mathcal{F}(u)) = E\left(\int_u^\infty \{h(t, \bar{Z}(t)) - g(h)(t)\} \frac{I(C > t)}{\bar{G}(t \mid X)} dM(t) \mid \bar{X}(u), C > u\right).$$

Given an initial consistent estimator β_n^0 and estimates G_n, $\Lambda_{0n}(\cdot \mid \beta_n^0)$, Q_n (the latter is discussed in Section 3.5), one can now define the one-step estimator

$$\beta_n^1 = \beta_n^0 + \frac{1}{n} \sum_{i=1}^n IC(Y_i \mid Q_n, G_n, D_{h_n^*}(\cdot \mid \beta_n^0, \rho_n)).$$

Even when $G(\cdot \mid X) = G(\cdot \mid Z)$, this estimator will generally improve on β_n^0. If we implement the c_{nu} extension (Chapter 2, Section 2.3), then we can guarantee that this estimator improves on β_n^0. Finally, if h_n converges to h_{opt} (see Section 2.6 or Chapter 2), then it is efficient, but for computational ease we would usually recommend our simple choice h_n^*. Robins (1993a), however, shows h_{opt} is the solution to a Fredholm type 2 integral equation and describes how to construct a locally efficient estimator of β by numerically solving that integral equation to obtain an estimate h_n of h_{opt}.

Remark.

Robins and Rotnitzky (1992) show that if we replace CAR by the weaker, also non-identifiable, assumption that the cause specific hazard of C at t given $\overline{X}(t-)$ and $\{N(u); u \geq t\}$ does not depend on $\{N(u); u \geq t\}$ for each t, then the model for the observed data Y is identical to the model $M(CAR)$. Thus both the theory and methodology developed under the assumption of CAR actually hold under this weaker assumption.

3.3.3 Extension to proportional rate models.

As the full data model, we now consider a proportional rate model

$$E(dN(t) \mid \bar{Z}^*(t-)) = Y(t)\lambda_0(t)\exp(\beta W(t)), \tag{3.13}$$

where $\bar{Z}^*(t-)$ is a covariate process not including the past $\bar{N}(t)$ of the counting process itself and $Y(t)$,$W(t)$ are functions of $\bar{Z}^*(t-)$. Proportional rate models avoid the need to model the effect of $\bar{N}(t-)$ on $dN(t)$. These models have been considered by Pepe and Cai (1993), Lawless (1995), Lawless and Nadeau (1995), Lawless, Nadeau, and Cook (1997), and Lin, Wei, Yang and Ying (2000), .

These authors propose to use the analog of the Andersen–Gill partial likelihood estimating functions to obtain parameter estimates for the proportional rate model. The estimates obtained are only consistent and asymptotically normally distributed under the assumption that censoring only depends on the covariates entering the proportional rate model; that is, $\lambda_C(t \mid \bar{X}(t)) = \lambda_C(t \mid \bar{Z}^*(t))$. This assumption becomes more questionable as the conditioning set $\bar{Z}^*(t)$ decreases, which is what the use of proportional rate models encourages. In particular, in recurrent event applications the past of the counting process is often a predictor of censoring; for example, the number of asthma attacks or hospital admissions might predict the censoring time (e.g., dropout time by change of treatment) of a subject. In addition, these estimates are inefficient, in general, even if the full data structure is observed. The reason for this is that partial likelihood is not the correct likelihood in the proportional rate model.

Our methods described in the previous subsection are readily applicable to the proportional rate model as well. As the class of full data estimating functions, one can use $D_h = \int h(t, \bar{Z}^*(t-)) dM_r(t)$, where $dM_r(t) \equiv dN(t) - E(dN(t) \mid \bar{Z}^*(t-))$ and h is arbitrary. As in the full data intensity models, the orthogonal complement of the nuisance tangent space in the full data model is a subset of these estimating functions. We will not aim to specify this precise subset but instead just accept using full data estimating functions that are not necessarily orthogonal to Λ_0. We can map these full data estimating functions into a class of observed data estimating functions with the same mappings $IC(Y \mid Q, G, D_h)$, presented above. In particular, our proposed choices for the index h of the full data estimating function can still be applied. This yields simple-to-implement estimators that are at least as efficient as the "partial-likelihood"-based estimating functions used in Lin, Wei, Yang, and Ying (2000) and remain consistent if $\lambda_C(t \mid \bar{X}(t)) \neq \lambda_C(t \mid \bar{Z}^*(t))$. If we do not enforce D_h to be orthogonal to the nuisance tangent space of the baseline hazard in the full data model, then our confidence intervals based on the observed data estimating function itself are not necessarily conservative anymore. Therefore, obtaining a meaningful estimate of variance requires either calculating the projection onto the nuisance tangent space of the baseline hazard (in