Sankhyā : The Indian Journal of Statistics
Special Issue on Quantile Regression and Related Methods
2005, Volume 67, Part 2, pp 153-186

Quantile Regression in Transformation Models

Dorota M. Dabrowska
University of California, Los Angeles, CA 90095-1772, USA

Abstract

Conditional quantiles provide a natural tool for reporting results from regression analyses based on semiparametric transformation models. We consider their estimation and construction of confidence sets in the presence of censoring.

AMS (2000) *subject classification.* 62G08, 62G20.
Keywords and phrases. Transformation models, censored data, quantile process.

---

1 Introduction

One-sided transformation models provide a popular tool for regression analysis of failure time data. These models assume that the conditional distribution of a failure time T given a vector of covariates Z has distribution function

$$\overline{F}(t|z) = F(\Gamma(t), \theta|z) \quad \mu \quad \text{a.s.} \quad z\ , \tag{1}$$

where μ is the marginal distribution of covariates, Γ is an unknown increasing function mapping the support of the marginal distribution of T onto the positive half-line, and $\mathcal{F} = \{F(x, \theta|z) : \theta \in \Theta, x > 0\}$ is a parametric family of conditional cdf's supported on R^+. The most common choice corresponds to the scale regression model

$$\overline{F}(t|z) = G(\Gamma(t)e^{\theta^T z}) \quad \mu \quad \text{a.s.} \quad z, \tag{2}$$

where G is a known distribution function. In particular, the proportional hazard model is of this form. In this case G represents exponential distribution and the unknown transformation Γ is the so-called baseline cumulative hazard function. Proportionality of hazards means that the conditional distribution of T given $Z = z$ has hazard rates $h(t|z)$ satisfying

$$\frac{e^{-\theta^T z_1}}{e^{-\theta^T z_2}} = \frac{h(t|z_2)}{h(t|z_1)}$$

for any two distinct covariate levels z_1 and z_2. This interpretation of parameters (Γ, θ) is lost in other transformation models of type (2) because the shape of the function Γ depends on the distribution G.

It is convenient to consider quantiles

$$Q(p|z) = \inf\{t : \overline{F}(t|z) \geq p\}$$

of the conditional distribution of T given $Z = z$ as an alternative parameter. In transformation models (2), we have

$$Q(p|z) = \Gamma^{-1}(e^{-\theta^T z} G^{-1}(p)) \tag{3}$$

for all $p \in (0, 1)$ and μ almost all z. Thus the conditional quantiles are monotone in each coordinate of the vector $z = (z_1, \ldots, z_d)$. In addition, the direction of monotonicity does not depend on p:

$$\text{sign}\left[\frac{d}{dz_k} Q(p|z)\right] = \text{sign}\,(-\theta_k) \qquad \text{for} \quad k = 1, \ldots, d\ .$$

Invariance of the model with respect to the group of increasing transformations implies also that for any $p_1 \neq p_2$ we have

$$\frac{\Gamma(Q(p_1|z))}{\Gamma(Q(p_2|z))} = \frac{G^{-1}(p_1)}{G^{-1}(p_2)} \quad \mu \quad \text{a.s.} \quad z \tag{4}$$

and for any $z_1 \neq z_2$

$$\frac{\Gamma(Q(p|z_1))}{\Gamma(Q(p|z_2))} = \frac{e^{-\theta^T z_1}}{e^{-\theta^T z_2}} \tag{5}$$

for all $p \in (0, 1)$. These three identities can be perhaps better understood by noting that (2) represents a linear regression model

$$\log \Gamma(T) = -\theta^T Z + \varepsilon\ ,$$

where Z and ε are independent and $\exp \varepsilon$ has distribution function G. In linear regression models assuming that the transformation Γ is known and equal to $\Gamma(t) = t$, the conditional quantiles are linear in z but the slope of the regression does not change with p. Likewise, the identities (4) and (5) have their additive analogue. However, if the transformation is unknown, then the model is much more difficult to interpret in terms of the parameters (θ, Γ).

Properties of quantile regression in the proportional hazard model are further discussed in Koenker and Geling (2001) and Portnoy (2003). In particular, Koenker and Geling (2001) proposed to measure the local effect of the regression coefficient on the conditional quantile p in terms of a parameter $b(p, EZ) = [b_k(p, EZ), k = 1, \dots, d]$, where

$$b_k(p, z) = \frac{d}{dz_k} Q(p|z) \ .$$

This parameter can be applied to any regression model. In (2) we have

$$b(p, EZ) = -\theta^T \frac{e^{-\theta^T EZ} G^{-1}(p)}{\gamma(Q(p|EZ))} \ ,$$

provided the unknown transformation has density γ with respect to Lebesgue measure in a neighbourhood of $Q(p|EZ)$. While $b(p, EZ)$ is proportional to the regression coefficient θ, the local effect of the regression coefficient is determined by the shape of the density γ. Portnoy (2003) considered direct modeling of the conditional quantiles under the assumption that Γ is the identity map. His model takes form

$$Q(p|z) = e^{\theta(p)^T z} \ ,$$

so that for fixed p the log-conditional quantiles are linear in z, but also the quantile regression coefficient changes with p. However, the choice of the identity map may be problematic. For other choices of the transformation, we have $Q(p|z) = \Gamma^{-1}(\exp \theta(p)^T z)$. Koenker and Geling's measure is given by

$$b(p, EZ) = \theta(p)^T \frac{e^{\theta(p)^T EZ}}{\gamma(Q(p|EZ))} \ .$$

It shows that the model is more flexible than the semiparametric transformation model (2), but it is not clear how to estimate the transformation function in this setting.

In many practical situations researchers may be also interested in the conditional distribution of T given $\varphi(Z)$, where φ is a known function. In particular, if $Z = (V, W)$ represents a high-dimensional covariate, then the choice $\varphi(Z) = V$ may correspond to a low-dimensional vector of "main" covariates. If V and W are dependent variables, then the conditional distribution of T given V follows the more flexible transformation model (1). For example, if (2) represents the proportional hazard model with parameters $\theta = (\theta_1, \theta_2)$ and the conditional distribution of $\exp[\theta_2^T W]$ given V is gamma with shape and scale equal to $\exp \xi(v)$ for a possibly nonlinear function ξ of v, then the marginal conditional distribution of T given V has distribution function of the form (1) with

$$F(x, \theta_1, \xi|v) = 1 - (1 + \exp[\theta_1 v + \xi(v)]x)^{-\exp[-\xi(v)]} .$$

The ratio of conditional hazards is

$$\frac{h(x|v_2)}{h(x|v_1)} = \frac{e^{-\theta_1 v_1}}{e^{-\theta_1 v_2}} \left[\frac{1 + e^{\theta_1 v_1 + \xi(v_1)}\Gamma(x)}{1 + e^{\theta_1 v_2 + \xi(v_2)}\Gamma(x)} \right]$$

For $x = 0$ the right-hand side is equal to $\exp[-\theta_1(v_1 - v_2)]$ and changes to $\exp[\xi(v_1) - \xi(v_2)]$ as $x \uparrow \infty$. It represents an increasing function if $\theta_1(v_1 - v_2) \geq \xi(v_2) - \xi(v_1)$ and a decreasing function, if the inequality is reversed. The conditional quantile function is equal to

$$Q(p|v) = \Gamma^{-1}(F^{-1}(p, \theta_1, \xi|v))$$

where

$$F^{-1}(p, \theta, \xi|v) = \exp[-\xi(v) - \theta_1 v][(1 - p)^{-\exp \xi(v)} - 1] .$$

If $\xi(v)$ is constant for almost all v, then we obtain the model (2). Otherwise the shape of the quantile function changes with p. The ratios of the transformed quantiles (4) and (5) are no longer constant in v and p, respectively.

In the general case, the conditional distribution of $e^{\theta_2^T W}$ given V will not have a simple analytical form, even if specified via a parametric model. However, quantile regression of the marginal conditional distributions of the failure time T can also be estimated by combining nonparametric regression with estimates of the parameters (θ, Γ).

In this paper we consider estimation of the conditional quantiles of T given $\varphi(Z)$, where φ is a function assuming a finite number of values. In particular, if $Z = (Z_1, \ldots, Z_d)$ has one or more discrete components, then results of this paper can be applied to estimation of quantiles of the marginal conditional distributions of T given any discrete component of Z. On the other hand in the case of continuous covariates estimation of the marginal conditional distribution and quantiles requires smoothing and may be difficult to accomplish in moderate or heavily censored samples. In such circumstances grouping observations into a small number of categories provides an alternative. For purposes of estimation of the parameters (θ, Γ) in transformation models (1) and (2), we use procedures proposed by Bogdanovicius and Nikulin (1999) and Dabrowska (2005). The approach allows for estimation of quantiles of the conditional distribution of T given $Z = z$ much in the same way as in the proportional hazard model, i.e. based on the substitution of estimates of (θ, Γ) into (3) (Dabrowska and Doksum, 1987, Burr and Doss, 1993). Here we derive asymptotic structure of the estimates of the conditional quantiles under the assumption that φ is a finite valued function, and consider construction of pointwise and simultaneous confidence sets. We also develop a Gaussian multiplier method for setting simultaneous confidence sets for the conditional quantile function. It extends the Gaussian multiplier method for setting confidence bands for the conditional survival function in the proportional hazard model (Lin, Fleming and Wei, 1994) to transformation models of type (1). In Section 3 we use data from a Vateran's Administration lung cancer clinical trial (Kalbfleisch and Prentice, 2000) to illustrate the results. Section 4 contains proofs.

2 Estimation

We assume that the vector (X, δ, Z) represents a nonnegative withdrawal time (X), a binary withdrawal indicator ($\delta = 1$ for failure and $\delta = 0$ for loss-to-follow-up) and covariate (Z). The triple (X, δ, Z) is defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and (X, δ) are given by $X = T \wedge \widetilde{T}$, $\delta = 1(X = T)$, where T and $\widetilde{T}$ represent failure and censoring times. The variables T and $\widetilde{T}$ are conditionally independent given Z and the conditional cumulative hazard function of T given Z is of the form

$$H(t|z) = A(\Gamma_0(t), \theta_0|z) \quad \mu \quad \text{a.s.} \quad z \ ,$$

where Γ_0 is an unbounded continuous increasing function, $\{A(x, \theta|z) : \theta \in \Theta\}$ is a parametric family of cumulative hazard functions with hazard rate $\alpha(u, \theta, z)$, and θ_0 is the "true" parameter. It is assumed throughout the paper that the parameters of the conditional distribution of the censoring times are non-informative on (Γ, θ).

Let $N(t) = 1(X \leq t, \delta = 1)$ and $Y(t) = 1(X \geq t)$ denote the counting and risk processes associated with the pair (X, δ). We also set

$$\tau_0 = \sup\{t : EY(t) > 0\}$$

and assume the following regularity conditions.

(i) The covariate Z has a nondegenerate distribution μ and is bounded: $\mu(|Z| \le C) = 1$ for some constant C.

(ii) The function $EY(t)$ has at most a finite number of atoms, and $EN(t)$ is continuous.

(iii) The point $\tau > 0$ satisfies $\inf\{t : E[N(t)|Z = z] > 0\} < \tau$ for μ a.s. z. In addition $\tau < \tau_0$ if τ_0 is a continuity point of the survival function $EY(t)$, and $\tau = \tau_0$, if τ_0 is an atom of this survival function.

(iv) The parameter set $\Theta \subset R^d$ is open, and the parameter θ is identifiable in the core model: $\theta \neq \theta'$ iff $A(\cdot, \theta|z) \not\equiv A(\cdot, \theta'|z)$ μ a.s. z.

(v) There exist constants $0 < m_1 < m_2 < \infty$ such that the hazard rate α satisfies

$$m_1 \le \alpha(x, \theta, z) \le m_2 \tag{6}$$

for μ a.s. z and all $\theta \in \Theta$, or (6) and (vi) hold for $\widetilde{\alpha}(x, \theta, z) = \alpha(\Phi(x), \theta, z)\Phi'(x)$, where Φ a strictly increasing unbounded twice continuously differentiable function Φ such that $\Phi(0) = 0$.

(vi) The function $\ell(x, \theta, z) = \log \alpha(x, \theta, z)$ is twice continuously differentiable with respect to both x and θ. The derivatives with respect to x (denoted by primes) satisfy

$$|\ell'(x, \theta, z)| \le \psi(x), \quad |\ell''(x, \theta, z)| \le \psi(x) ,$$

where ψ is a constant or a continuous bounded decreasing function. The derivatives with respect to θ (denoted by dots) satisfy

$$|\dot{\ell}(x, \theta, z)| \le \psi_1(x), \quad |\ddot{\ell}(x, \theta, z)| \le \psi_2(x)$$

and

$$|g(x, \theta, z) - g(x', \theta, z)| \le \psi_3(x)[|x - x'| + |\theta - \theta'|] ,$$

where $g = \ddot{\ell}, \dot{\ell}'$ and ℓ''. The functions $\psi_p, p = 1, 2, 3$ are continuous, bounded or strictly increasing and such that $\psi_p(0) < \infty$,

$$\int_0^\infty e^{-x}\psi_1^2(x)dx < \infty, \int_0^\infty e^{-x}\psi_2(x)dx < \infty, \int_0^\infty e^{-x}\psi_3(x)dx < \infty.$$

The assumption that the covariate Z is bounded is restrictive, but standard for analysis of semiparametric models assuming that the transformation

Γ is unknown. In the special case of the proportional hazard model, Andersen and Gill (1982) required only existence of moments $EZ^2 e^{\theta^T Z} 1(X \geq x)$, for $x \geq 0$ in a neighbourhoood $\Theta \subset R^d$ of the true parameter θ_0. However, setting $x = 0$, we see that this moment condition may lead to a constrained optimization problem which cannot be correctly stated, if the distribution Z is unspecified. For example, if Z is multivariate normal $N(0, \Sigma)$ and Σ is a known non-singular matrix, then the moment condition is satisfied for all $\theta \in R^d$ and the usual unrestricted partial likelihood approach towards fitting the regression coefficients applies. However, if Z is a univariate lognormal variable, $Z \sim \exp \mathcal{N}(0, 1)$, then the parameter θ must be estimated under the added side condition $\theta \leq 0$. Thus the boundedness assumption is restrictive, but allows for parameter estimation without additional assumptions on the marginal distribution of the covariate.

Given an iid sample $(N_i, Y_i, Z_i), i = 1, \ldots, n$ of the (N, Y, Z) processes, we set $N_.(t) = n^{-1} N_i(t)$,

$$S(x, \theta, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t) \alpha_i(x, \theta) \ .$$

and $\alpha_i(x, \theta) = \alpha(x, \theta | Z_i)$. Following Bogdanovicius and Nikulin (1999), define

$$\Gamma_{n\theta}(t) = \int_0^t \frac{N_.(du)}{S(\Gamma_{n\theta}(u-), \theta, u)} \ , \quad \Gamma_{n\theta}(0-) = 0$$

for any $\theta \in \Theta$. The process $\{\Gamma_{n\theta} : \theta \in \Theta\}$ is here thought as the sample analogue of the Volterra integral equation

$$\Gamma_\theta(t) = \int_0^t \frac{EN(du)}{s(\Gamma_\theta(u-), \theta, u)} \ , \quad \Gamma_\theta(0-) = 0, \quad \theta \in \Theta \ , \tag{7}$$

where $s(x, \theta, u) = EY_i(u) \alpha_i(x, \theta)$. The condition (iv) was used in Dabrowska (2005) to verify that this equation has a unique locally bounded solution, and such that $\Gamma_\theta(\tau_0) < \infty$ if τ_0 is an atom of the survival function $EY(t)$, and $\lim_{t \uparrow \tau_0} \Gamma_\theta(t) \uparrow \infty$, if τ_0 is a continuity point of $EY(t)$. In particular, the latter applies to uncensored data. Therein we show that in the case of scale transformation models (2), the condition (v) is satisfied by half-logistic, half-normal and half-t distributions, proportional odds ratio distribution, frailty models with decreasing heterogeneity with fixed frailty parameter and polynomial hazards with nonnegative constant coefficients. These models have smooth differentiable hazards with respect to both x and θ and integrability condition (vi) imply also that Fisher information is finite. Affine independence of covariates is sufficient for the condition (iv) to hold. In the case of transformation models (1), the regularity conditions are satisfied in the gamma frailty model with frailty parameter representing a function of covariates dependent on a Euclidean parameter. They are also satisfied in regular polynomial hazard regression models with nonnegative coefficients representing parametric functions of covariates. In these models, the conditional hazard rates are twice differentiable with respect to x, while the condition (vi) imposes a second order differentiability assumption on the functions of covariates. Such differentiability conditions are in general not needed in regular parametric models. However, here we use semiparametric models and estimation of the parameter θ will be based on a conditional rank statistics score equation. We do not know at present time, how to relax these differentiability conditions to allow for estimation based on ranks.

For any τ satisfying conditions (i)-(vi), the function $\{\Gamma_\theta(t) : t \in [0, \tau], \theta \in \Theta\}$ is Fréchet differentiable with respect to θ and the derivative satisfies the linear Volterra equation

$$\dot{\Gamma}_\theta(t) = -\int_0^t \dot{s}(\Gamma_\theta(u-), \theta, u) C_\theta(du) - \int_0^t \dot{\Gamma}_\theta(u-) s'(\Gamma_\theta(u-), \theta, u) C_\theta(du) ,$$

where $\dot{s}(\Gamma_\theta(u-), \theta, u) = EY_i(u)\dot{\alpha}_i(\Gamma_\theta(u-), \theta)$, $s'(\Gamma_\theta(u-), \theta, u) = EY_i(u)\alpha_i'$ $(\Gamma_\theta(u-), \theta)$ and

$$C_\theta(t) = \int_0^t \frac{EN(du)}{s^2(\Gamma_\theta(u-), \theta, u)} .$$

In the case of the proportional hazard model, the function s' is identically equal to 0. Otherwise, the solution to this Volterra equation is given by

$$\begin{aligned} \dot{\Gamma}_\theta(t) &= -\int_0^t \dot{s}(\Gamma_\theta(u-), \theta, u) C_\theta(du) \mathcal{P}_\theta(u, t) , \\ \mathcal{P}_\theta(u, t) &= \pi_{(u,t]}(1 - s'(\Gamma_\theta(w-), \theta, w) C_\theta(dw)) . \end{aligned}$$

Here for any function b of bounded variation, $\pi_{(u,t]}(1+b(du))$ is the product integral, i.e.

$$\pi_{(u,t]}(1 + b(dw)) = \prod_{u<w\le t} (1 + b(\Delta w)) \exp[b_c(t)]$$

where b_c is the continuous part of b and the product is taken over its atoms. To make the definition complete, in the case of the proportional hazard model we set $\mathcal{P}_\theta(u, t) \equiv 1$. With this choice, the form of the function $\dot{\Gamma}_\theta$ is the same for all models of type (1) considered in this paper.

Let $\alpha_i(x,\theta) = \alpha(x,\theta,Z_i)$ and $\ell_i(x,\theta) = \log\alpha(x,\theta,Z_i)$. We shall apply the same convention to derivatives of the functions α_i and ℓ_i with respect to θ and x. Define functions

$$
\begin{aligned}
\overline{v}(u,\theta) &= \frac{EY_i(u)[\dot{\ell}_i^{\otimes 2}\alpha_i](\Gamma_\theta(u),\theta)}{s(\Gamma_\theta(u),\theta,u)} - \left(\frac{\dot{s}}{s}\right)^{\otimes 2}(\Gamma_\theta(u),\theta,u)\\
v(u,\theta) &= \frac{EY_i(u)[\ell_i'^2\alpha_i](\Gamma_\theta(u),\theta)}{s(\Gamma_\theta(u),\theta,u)} - \left(\frac{s'}{s}\right)^{2}(\Gamma_\theta(u),\theta,u)\\
\rho(u,\theta) &= \frac{EY_i(u)[\dot{\ell}_i\ell_i'\alpha_i](\Gamma_\theta(u),\theta)}{s(\Gamma_\theta(u),\theta,u)} - \left(\frac{\dot{s}}{s}\right)\left(\frac{s'}{s}\right)(\Gamma_\theta(u),\theta,u)
\end{aligned}
$$

and

$$
\begin{aligned}
K_\theta(t,t') &= \int_0^{t\wedge t'} C_\theta(du)\mathcal{P}_\theta(u,t)\mathcal{P}_\theta(u,t')\\
B_\theta(t) &= \int_0^t v(u,\theta)EN(du)\ .
\end{aligned}
$$

Suppose that $v(u,\theta) \not\equiv 0$ a.e.–EN and let $\varphi_\theta = \int_0^{\cdot} g_\theta d\Gamma_\theta$ be a vector valued function with d components and square integrable with respect to B_θ.

Define matrices

$$
\begin{aligned}
\Sigma_1(\theta) &= \int_0^\tau v_\varphi(t,\theta)EN(du)\\
\Sigma_2(\theta) &= \int_0^\tau\int_0^\tau K_\theta(t,u)\rho_\varphi(t,\theta)\rho_\varphi(u,\theta)^T EN(du)EN(dt)\\
\Sigma(\theta) &= \Sigma_1(\theta)+\Sigma_2(\theta)
\end{aligned}
$$

where

$$
\begin{aligned}
v_\varphi(t,\theta) &= \overline{v}(t,\theta) + v(t,\theta)\varphi_\theta^{\otimes 2}(t) - \rho(t,\theta)\varphi_\theta^T(t) - \varphi_\theta(t)\rho(t,\theta)^T\\
\rho_\varphi(t,\theta) &= \rho(t,\theta) - v(t,\theta)\varphi_\theta(t)\ .
\end{aligned}
$$

In the following we choose φ_θ as solution to the Fredholm equation

$$
\varphi_\theta(t)+\int_0^\tau K_\theta(t,u)v(u,\theta)\varphi_\theta(u)EN(du) = -\dot{\Gamma}_\theta(t)+\int_0^\tau K_\theta(t,u)\rho(u,\theta)EN(du), \tag{8}
$$

or equivalently

$$
\begin{aligned}
\varphi_\theta(t)+\dot{\Gamma}_\theta(t) &= \int_0^\tau K_\theta(t,u)\rho_\varphi(u,\theta)EN(du)\\
&= \int_0^\tau K_\theta(t,u)\rho_{-\dot{\Gamma}}(u,\theta)EN(du) - \int_0^\tau K_\theta(t,u)[\varphi_\theta+\dot{\Gamma}_\theta](u)B_\theta(du).
\end{aligned}
$$

This equation has a unique solution, square integrable with respect to B_θ. We define it as $\varphi_\theta = -\dot{\Gamma}_\theta$ if $\rho_{-\dot{\Gamma}}(u,\theta) \equiv 0$. In this case we have $\Sigma_2(\theta) = 0$. Finally, if $v(t,\theta) \equiv 0$ a.e. EN, then $\rho(t,\theta) \equiv 0$ as well. For the sake of completeness we, set in this case $\varphi_\theta = -\dot{\Gamma}_\theta$. We also have $\Sigma_2(\theta) = 0$, and $\Sigma_1(\theta)$ simplifies to $\Sigma_1(\theta) = \int \overline{v}(u,\theta)EN(du)$. This last choice corresponds to the proportional hazard model, and the scale regression models with regression coefficient $\theta = 0$. (Note that if $v(u,\theta) \equiv 0$, then the φ_θ function does not enter into the score equation below).

To estimate the parameter θ, we use a solution to the score equation $U_n(\theta) = 0$, where

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau [b_{1i}(\Gamma_{n\theta}(t),t,\theta) - b_{2i}(\Gamma_{n\theta}(t),t,\theta)\varphi_{n\theta}(t)]N_i(dt) \ , \tag{9}$$

$\varphi_{n\theta}$ is an estimator of φ_θ, and

$$b_{1i}(x,t,\theta) = \dot{\ell}_i(x,\theta) - \frac{\dot{S}(x,\theta,t)}{S(x,\theta,t)} \ , \quad b_{2i}(x,t,\theta) = \ell_i'(x,\theta) - \frac{S'(x,\theta,t)}{S(x,\theta,t)} \ .$$

If Γ_0 is a known function, e.g. $\Gamma_0(t) = t$, then under the assumption of conditional independence of failure and censoring times, the MLE score equation for estimation of the parameter θ is given by $\widetilde{U}_n(\theta) = 0$, where

$$\widetilde{U}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \dot{\ell}_i(\Gamma_0(t),\theta)N_i(dt) - \int_0^\tau \dot{S}(\Gamma_0(t),\theta,t)\Gamma_0(dt)$$

and $\dot{S}(x,\theta,t) = n^{-1}\sum_{i=1}^n Y_i(t)\dot{\alpha}_i(x,\theta)$. In addition, the assumption of conditional independence of failure and censoring times implies that the function (7) satisfies $\Gamma_{\theta_0}(t) = \Gamma_0(t)$ at the true value θ_0 of the parameter θ. This last identity remains to hold also when the transformation Γ_0 is unknown. Therefore a natural approach to estimation of the parameter θ is to consider solving the score equation $\widehat{U}_n(\theta) = 0$, where

$$\widehat{U}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau b_{1i}(\Gamma_{n\theta}(t),t,\theta)N_i(dt) \ .$$

In particular, this is the usual score equation for estimation of the parameter θ in the proportional hazard model. In general transformation models (1), this choice leads to an asymptotically inefficient estimate of the parameter θ. It may also lead to estimates of poor performance in moderate sample sizes. This also applies to score processes of the form (9), where $\varphi_{n\theta}$ is an estimate of some square integrable function φ_θ with respect to B_θ. For example, Bogdanovicius and Nikulin (1999) considered the choice of $-\dot{\Gamma}_\theta$, corresponding to the score equation derived from a modified partial likelihood function. Under mild regularity conditions on the estimator of the function φ_θ, the solution to the score equation (9) exists with probability tending to 1 and is unique in local neighbourhoods of the true parameter θ_0. However, its asymptotic variance assumes the usual "sandwich" form because the process $\Gamma_{n\theta}$ has a non-trivial contribution to both asymptotic variance of the score process and the negative derivative of it with respect to θ. The choice of the φ_θ function corresponding to the solution of to the Fredholm equation (8) leads to an M estimator whose asymptotic variance is of non-sandwich form and equal to the inverse of the asymptotic variance of the score function. The form of the solution to this equation can be found in Dabrowska (2005). The resulting estimator can also be shown to be asymptotically efficient under the assumption that the point $\tau_0 = \sup\{t : EY(t) > 0\}$ forms an atom of the survival function $EY(t)$. The following proposition summarizes some properties of the estimates of (θ, Γ).

PROPOSITION 2.1 *Suppose that the conditions (i)-(vi) are satisfied. Let $\Sigma_1(\theta_0)$ be non-singular, and let $\varphi_{n\theta}$ be an estimator of this function such that $\|\varphi_{n\theta_0} - \varphi_{\theta_0}\|_\infty \to_P 0$, $\limsup_n \|\varphi_{n\theta_0}\|_v = O_P(1)$, $\varphi_{n\theta} - \varphi_{n\theta'} = (\theta - \theta')\psi_{n\theta,\theta'}$, where*

$$\sup\{\limsup_n \|\psi_{n\theta,\theta'}\|_v : \theta \in B(\theta_0, \varepsilon_n)\} = O_P(1)$$

and $B(\theta_0, \varepsilon_n) = \{\theta : \|\theta - \theta_0\| \le \varepsilon_n\}$ for some sequence $\varepsilon_n \downarrow 0, \sqrt{n}\varepsilon_n \to \infty$. Then, with probability tending to 1, the score equation $U_n(\theta) = 0$ has a unique solution $\widehat{\theta}$ in $B(\theta_0, \varepsilon_n)$. Moreover, $[\widehat{T}, \widehat{W_0}], \widehat{T} = \sqrt{n}(\widehat{\theta} - \theta_0)$, $\widehat{W_0} = \sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_{\theta_0} - (\widehat{\theta} - \theta_0)\dot{\Gamma}_{\widehat{\theta}}]$ converges weakly in $R^p \times \ell^\infty([0, \tau])$ to a mean zero Gaussian process $[T, W_0]$ with covariance

$$\begin{aligned} &\mathrm{cov}\ T = \Sigma^{-1}(\theta_0) \quad \mathrm{cov}\ (W_0(t), T) = -\Sigma^{-1}(\theta_0)[\varphi_{\theta_0} + \dot{\Gamma}_{\theta_0}](t) \\ &\mathrm{cov}\ (W_0(t), W_0(t')) = K_{\theta_0}(t, t')\ . \end{aligned}$$

An example of an estimator of the function φ_θ is given in Section 3. The asymptotic covariances can be estimated using substitution method.

Let us assume now that $\mathcal{D} = \{D_j : j = 1, \ldots, k\}$ is a finite partition of the covariate space such that

$$\pi(D) = P(Z \in D) > 0, \quad D \in \mathcal{D}\ . \tag{10}$$

We denote by $F_D(t) = P(T \in t|Z \in D)$ the cdf of the conditional distribution of T given $Z \in D, D \in \mathcal{D}$. Under the assumption of the transformation model, this function is of the form

$$F_D(t) = \frac{1}{\pi(D)} E1[Z \in D] F(\Gamma_0(t), \theta_0|Z) .$$

In practice, the partition D will be chosen based on the observations. For example, if $Z = (Z_1, \ldots, Z_d)$ is a multivariate covariate, whose first component is continuous, then a natural partition of the covariate space may correspond to selection of $k = 4$ intervals determined by the sample quartiles of Z_1. If subjects are ranked according to values of the exponential factors $e^{\beta^T Z}$ than a natural partition may correspond to several groups determined by the distribution of $e^{\beta^T Z}$. Any selection of such a partition requires some form of estimation of parameters of the marginal distribution of the covariates. Here we consider a naive situation in which the cell probabilities can be estimated nonparametrically by means of sample proportions. This choice arises in analyses of models with possibly high-dimensional discrete or mixed discrete-continuous covariates, whenever interest is only in analyses of marginal conditional distributions corresponding to discrete variables representing treatment types, patients' gender etc. In the data example given in section 3, a many valued discrete variable representing a quantitative measurement patient's performance status, admits a natural partition into three groups corresponding to a more intuitive qualitative description of health condition at the time of entry into the clinical trial.

As an estimate $\widehat{F}_D(t)$ of the function $F_D(t)$ we take

$$\begin{aligned} \widehat{F}_D(t) &= \frac{1}{\widehat{\pi}(D)} \frac{1}{n} \sum_{i=1}^n 1(Z_i \in D) F(\Gamma_{n\widehat{\theta}}(t), \widehat{\theta}|Z_i) , \\ \widehat{\pi}(D) &= \frac{1}{n} \sum_{i=1}^n 1(Z_i \in D) . \end{aligned}$$

We also define scalar and vector valued functions

$$\begin{aligned} \widehat{\psi}_1(t, D) &= \frac{1}{n} \frac{1}{\widehat{\pi}(D)} \sum_{i=1}^n 1(Z_i \in D) f(\Gamma_{n\widehat{\theta}}(t), \widehat{\theta}|Z) , \\ \widehat{\psi}_2(t, D) &= \widehat{\psi}_1(t, D) \dot{\Gamma}_{n\widehat{\theta}}(t) + \frac{1}{n} \frac{1}{\widehat{\pi}(D)} \sum_{i=1}^n 1(Z_i \in D) \dot{F}(\Gamma_{n\widehat{\theta}}(t), \widehat{\theta}|Z_i) , \end{aligned}$$

where $\dot{F}(x, \theta|z)$ is the derivative of $F(x, \theta|z)$ with respect to θ.

Finally, we denote by $\|\cdot\|$ the supremum norm on $\mathcal{T} = [0,\tau]\times\mathcal{D}$ and let $\ell^\infty(\mathcal{T})$ be the space of bounded functions on $\mathcal{T}$ endowed with the supremum norm.

PROPOSITION 2.2 *Suppose that the conditions of Proposition 2.1 are satisfied and (10) holds.*

(i) *We have* $\|\widehat{F}-F\| \to_P 0$ *and* $\widehat{W} = \{\widehat{W}(t,D) = \sqrt{n}[\widehat{F}(t,D) - F(t,D)] : (t,D)\in\mathcal{T}\}$ *converges weakly in* $\ell^\infty(\mathcal{T})$ *to* W*, a mean zero Gaussian processes. Its covariance function is given in Section 4.*

(ii) *Let* $V_i = (V_{1i}, V_{2i}), i = 1,2,\ldots,n$ *and* $V_3 = (V_{31},\ldots,V_{3d})$ *be mutually independent* $\mathcal{N}(0,1)$ *variables, independent of the observations* $(X_i,\delta_i,Z_i), i = 1,\ldots,n$. *Define*

$$\widehat{W}_1^\#(t,D) = \frac{1}{\sqrt{n}}\frac{1}{\widehat{\pi}(D)}\sum_{i=1}^n V_{1n}1(Z_i\in D)[F(\widehat{\Gamma}_{\widehat{\theta}}(t),\widehat{\theta}|Z_i) - \widehat{F}_D(t)],$$
$$\widehat{W}_2^\#(t,D) = \widehat{W}_0^\#(t)\widehat{\psi}_1(t,D) + \int_0^\tau \widehat{W}_0^\#(s)\widehat{\rho}_{\widehat{\varphi}_n}(s,\widehat{\theta})N_.(ds)\Sigma_n^{-1}(\widehat{\theta})\widehat{\psi}_2(t,D),$$
$$\widehat{W}_3^\#(t,D) = V_3\Sigma_{1n}^{1/2}(\widehat{\theta})\Sigma_n^{-1}(\widehat{\theta})\widehat{\psi}_1(t,D),$$

where

$$\widehat{W}_0^\#(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{2i}\frac{1[X_i\le t,\delta_i = 1]}{S(\Gamma_{n\widehat{\theta}}(X_i-),\widehat{\theta},X_i)}\mathcal{P}_{n\widehat{\theta}}(X_i,t)$$

and $\Sigma_{1n}(\widehat{\theta})$, $\Sigma_n(\widehat{\theta})$, $\widehat{\varphi}_n = \varphi_{n\widehat{\theta}}$, $\widehat{\rho}_{\widehat{\varphi}_n}(u,\widehat{\theta})$ *and,* $\mathcal{P}_{n\widehat{\theta}}(u,t)$ *are estimates of* $\Sigma_1(\theta_0)$, $\Sigma(\theta_0)$, φ_{θ_0}, $\rho_{\varphi_{\theta_0}}(u,\theta_0)$, $\mathcal{P}_{\theta_0}(u,t)$ *obtained using substitution method. The process* $\widehat{W}^\# = \{\widehat{W}^\#(t,D) = \sum_{j=1}^3 \widehat{W}_j^\#(t,D) : (t,D)\in\mathcal{T}\}$ *converges weakly (unconditionally) in* $\ell^\infty(\mathcal{T})$ *to a Gaussian process* $W^\#$ *with the same covariance function as the process* W *of part (i) and independent of it. Conditionally, the process* $\widetilde{W}$ *converges weakly to* W *in probability.*

The proof is given in Section 4. In the first part of the proposition, the observations $R_i = (X_i,\delta_i,Z_i), i = 1,\ldots,n,\ldots$ are defined as coordinate projections on the product probability space $(\Omega^\infty,\mathcal{F}^\infty,P^\infty)$. In the second part, we use the product probability space $(\Omega^\infty\times\mathcal{V}\times\mathcal{V}',\mathcal{F}^\infty\times\mathcal{B}\times\mathcal{B}',P^\infty\times Q\times Q')$. The variables $R_i = (X_i,\delta_i,Z_i), i = 1,\ldots,n,\ldots$, $V_i, i = 1,\ldots,n\ldots$

and V_3 are defined as first, second and last projections. Conditional weak convergence in probability means

$$\sup_{f \in BL_1} |E^*_V f(W^{\#}) - Ef(W)| \to 0$$

in (outer) probability, where BL_1 is the set of all real functions on $\ell^\infty(\mathcal{T})$ with a Lipschitz norm bounded by 1 (van der Vaart and Wellner, 1996, Ch. 2.9).

We proceed to the discussion of the properties of the quantile regression. For $p \in (0,1)$ and (fixed) $D \in \mathcal{D}$ let

$$\ell_D(p) = \inf\{t : F_D(t) \geq p\} , \quad u_D(p) = \sup\{t : F_D(t) \leq p\} .$$

Then $\ell_D(p) \leq u_D(p)$ and the p-th quantiles of the conditional distribution of T given $Z \in D$ are defined as the set of numbers in the closed interval $[\ell_D(p), u_D(p)]$. We denote by $\widehat{\ell}_D(p)$ and $\widehat{u}_D(p)$ the sample counterparts of these points, i.e.

$$\widehat{\ell}_D(p) = \inf\{t : \widehat{F}_D(t) \geq p\}, \quad \widehat{u}_D(p) = \sup\{t : \widehat{F}_D(t) \leq p\} .$$

If $u_D(p) < \tau$, then under assumptions of Proposition 2.2, we have

$$\ell_D(p) \leq \liminf_n \widehat{\ell}_D(p) \leq \limsup_n \widehat{u}_D(p) \leq u_D(p) \tag{11}$$

with probability tending to 1. Indeed, let $\varepsilon = \varepsilon(D) > 0$ be arbitrary but small enough so that $u_D(p) + \varepsilon < \tau$. Then

$$F_D(\ell_D(p) - \varepsilon) < p, \quad F_D(u_D(p) + \varepsilon) > p$$

and uniform consistency of the estimate $\widehat{F}_D(\cdot)$ implies that with probability tending to 1, we also have

$$\widehat{F}_D(\widehat{\ell}_D(p) - \varepsilon) \leq p, \quad \widehat{F}_D(\widehat{u}_D(p) + \varepsilon) \geq p .$$

This in turn implies (11).

In the following we shall assume that the transformation function Γ_0 has density γ with respect to the Lebesgue measure, and the function γ is uniformly continuous and bounded away from 0 on an interval $[\tau_1 - \varepsilon, \tau_2 + \varepsilon], 0 < \tau_1 - \varepsilon, \tau_2 + \varepsilon \leq \tau \leq \tau_0$ and such that

$$\tau_1 = \min\{\ell_D(p_1) : D \in \mathcal{D}\}, \quad \tau_2 = \max\{u_D(p_2) : D \in \mathcal{D}\} . \tag{12}$$

Let $I = [p_1, p_2]$ and set $\mathcal{I} = I \times \mathcal{D}$. In this case the conditional distribution of T given $Z \in D$ has a unique p-th quantile $Q_D(p)$ for any $p \in I$ and we define its sample analogue by setting

$$\widehat{Q}_D(p) = \widehat{\ell}_D(p) = \inf\{t : \widehat{F}_D(t) \geq p\} .$$

Then (11) implies that $\widehat{Q}_D(p) \to_P Q_D(p)$ pointwise in $(p, D) \in \mathcal{I}$. Using finiteness of the class $\mathcal{D}$, monotonicity of $F_D(t)$ and $\widehat{F}_D(t)$, and an argument similar to the classical Glivenko-Cantelli theorem, we also have

$$\sup\{|\widehat{Q}_D(p) - Q_D(p)| : (p, D) \in \mathcal{I}\} \to_P 0 .$$

PROPOSITION 2.3 *Suppose that the conditions of Proposition 2.2 hold, and Γ_0 has density γ with respect to the Lebesgue measure such that γ is uniformly continuous and bounded away from 0 on an interval $[\tau_1 - \varepsilon, \tau_2 + \varepsilon], 0 < \tau_1 - \varepsilon, \tau_2 + \varepsilon \leq \tau$ satisfying (12). The normalized quantile process $\widehat{V} = \{\widehat{V}(p, D) : (p, D) \in \mathcal{I}\}$ given by*

$$\widehat{V}(p, D) = \sqrt{n}[\widehat{Q}_D - Q_D](p) ,$$

converges weakly in $\ell^\infty(\mathcal{I})$ to $V = \{V(p, D) = -h(p, D)W(Q_D(p), C) : (p, D) \in \mathcal{I}\}$, where

$$h(p, D) = [f_D(Q_D(p))\gamma(Q_D(p))]^{-1} .$$

PROOF. We have $\widehat{V}(p, D) = \widehat{h}(p, D)\widehat{R}(p, D)$, where

$$\begin{aligned} \widehat{h}(p, D) &= \left(\frac{\widehat{Q}_D - Q_D}{F_D \circ \widehat{Q}_D - F_D \circ Q_D} \right)(p) , \\ \widehat{R}(p, C) &= \sqrt{n}[F_D \circ \widehat{Q}_D - F_D \circ Q_D](p) . \end{aligned}$$

Since the function γ is positive and uniformly continuous on $[\tau_1 - \varepsilon, \tau_2 + \varepsilon]$, uniform consistency of the sample quantile function implies

$$\sup\{|\widehat{h} - h|(p, D) : (p, D) \in \mathcal{I}\} \to_P 0 .$$

The process $\widehat{R}(p, D)$ is on the other hand given by $\widehat{R}(p, D) = \sum_{j=1}^{3} \widehat{R}_j(p, D)$, where

$$\begin{aligned} \widehat{R}_1(p, D) &= -(\widehat{W}_D \circ Q_D)(p) , \\ \widehat{R}_2(p, D) &= -(\widehat{W}_D \circ \widehat{Q}_D - \widehat{W}_D \circ Q_D)(p) , \\ \widehat{R}_3(p, D) &= \sqrt{n}[\widehat{F}_D \circ Q_D(p) - p] . \end{aligned}$$

We have $\sup\{|\widehat{R}(p,D)| : (p,D) \in \mathcal{I}\} \leq \sup\{|\widehat{W}_D(u) - \widehat{W}_D(u-)| : u \in [\tau_1 - \varepsilon, \tau_2+\varepsilon], D \in \mathcal{D}\} = O_p(n^{-1/2})$ because the function $\widehat{F}_D(x)$ has jumps of order $O_p(n^{-1})$. Application of the Skorohod-Dudley-Wichura construction implies also that $\sup\{|\widehat{R}_2(p,D)| : (p,D) \in \mathcal{I}\} \to_P 0$, while the process $\{\widehat{R}_1(p,D) : (p,D) \in \mathcal{I}\}$ converges weakly in $\ell^\infty(\mathcal{I})$ to $\{-W_D \circ Q_D(p) : (p,D) \in \mathcal{I}\}$. □

We shall apply now this result to construct pointwise confidence intervals for the p-th quantile. Let $v_D(t)$ be the asymptotic variance function of the process $\{W(t,D) : (t,D) \in \mathcal{T}\}$. It is derived in Section 4. Here we shall use only that this function is positive and continuous on the interval $[\tau_1 - \varepsilon, \tau_2 + \varepsilon]$, and its plug-in analogue $\widehat{v}_D(t)$ is uniformly consistent on the set $[\tau_1 - \varepsilon, \tau_2 + \varepsilon] \times \mathcal{D}$.

For $p \in (0,1)$ and $D \in \mathcal{D}$, let

$$p_n^{\pm} = p \pm \frac{1}{\sqrt{n}} \widehat{v}_D(\widehat{Q}_D(p)) z(\alpha) \ ,$$

where $z(\alpha)$ is the upper $\alpha/2$ percentile of $\mathcal{N}(0,1)$ distribution. Proposition 2.3 and the inequalities

$$\begin{aligned} \widehat{Q}_D(p) \geq s \quad &\text{iff} \quad p \geq \widehat{F}_D(s) \ , \\ Q_D(p) \geq s \quad &\text{iff} \quad p \geq F_D(s) \ , \end{aligned}$$

imply that $[\widehat{Q}_D(p_n^-), \widehat{Q}_D(p_n^+)]$ is a $100\% \times (1-\alpha)$ asymptotic pointwise confidence interval for the conditional quantile $Q_D(p)$.

Unfortunately, in practice the points $p_n^{\pm}$ may fall outside the range $[0,1]$. To circumvent this problem, we follow the approach of Bie *et al.* (1987) and consider confidence intervals based on transformations. Let g be a strictly monotone cdf with density g' supported on the whole real line. Set

$$p_{nD}^{\pm} = g^{-1}(p) \pm \frac{1}{\sqrt{n}} \frac{\widehat{v}_D(\widehat{Q}_D(p))}{g'(g^{-1}(p))} z(\alpha) \ .$$

With probability tending to 1, the inequalities

$$\widehat{Q}_D(g(p_{nD}^-)) \leq Q_D(p) \leq \widehat{Q}_D(g(p_{nD}^+))$$

are equivalent to

$$-z(\alpha) \leq g'(g^{-1}(p)) \sqrt{n} \frac{g^{-1}(\widehat{F}_D(Q_D(p))) - g^{-1}(p)}{\widehat{v}_D(\widehat{Q}_D(p))} \leq z(\alpha)$$

and application of delta method implies that $[\widehat{Q}_D(p^-_{nD}), \widehat{Q}_D(p^+_{nD})]$ is a $100\% \times (1-\alpha)$ asymptotic confidence interval for the conditional quantile $Q_D(p)$.

Construction of simultaneous confidence sets for the function $\{Q_D(p) : (p, D) \in \mathcal{I}\}$ is more difficult because the process W appearing in Propositions 2.2 and 2.3 forms a sum of independent Gaussian processes with correlated increments. Therefore, following Burr and Doss (1993) and Lin, Fleming and Wei (1994), we propose the use of simulated confidence sets.

Define

$$\begin{aligned} U &= \sup\left\{\tfrac{|W(Q_D(p),C)|}{v_D(Q_D(p))} : (p,D) \in \mathcal{I}\right\} \\ &= \sup\left\{\tfrac{|W(t,D)|}{v_D(t)} : t \in [Q_D(p_1), Q_D(p_2)], D \in \mathcal{D}\right\} \end{aligned}$$

and let $u(\alpha)$ be the upper $100\%(1-\alpha)$ percentile of its distribution. To obtain an approximation to the critical level $u(\alpha)$, we generate mutually independent standard normal vectors V defined as in Proposition 2.3, and form

$$U^{\#} = \sup\left\{\frac{|\widehat{W}^{\#}(t,D)|}{\widehat{v}_D(t)} : t \in [\widehat{Q}_D(p_1), \widehat{Q}_D(p_2)], D \in \mathcal{D}\right\}$$

The procedure is repeated independently m times, for some large m, to obtain m iid copies $U_1^{\#}, \ldots, U_m^{\#}$. The estimate $u^{\#}(\alpha)$ of the critical point $u(\alpha)$ is taken as the empirical $(1-\alpha)$ quantile of $U_1^{\#}, \ldots, U_m^{\#}$. The corresponding simulated confidence set for $\{Q_D(p) : (p, D) \in \mathcal{I}\}$ is chosen as

$$\{[\widehat{Q}_D(\widehat{p}^-_{nD}), \widehat{Q}_D(\widehat{p}^+_{nD})] : D \in \mathcal{D}\}\ ,$$

where

$$\widehat{p}^{\pm}_{nD} = g^{-1}(p) \pm \frac{1}{\sqrt{n}} \frac{\widehat{v}_D(\widehat{Q}_D(p))}{g'(g^{-1}(p))} u^{\#}(\alpha)\ .$$

Application of Propositions 2.2 and 2.3 implies that $u^{\#}(\alpha)$, the upper α–quantile of this (conditional) distribution satisfies $u^{\#}(\alpha) \to u(\alpha)$ in probability.

An alternative approach to construction of simultaneous confidence sets may be based on bootstrap. Lin, Fleming and Wei (1994) argued that in the case of Cox regression with external time dependent covariates, it is not clear how to implement bootstrap to construct simultaneous confidence bands for the conditional survival function, or other functionals related to it. In our setting covariates are time independent, and confidence sets can be based on "obvious" bootstrap. We can draw $R_n^* = [(X_i^*, \delta_i^*, Z_i^*) : i = 1, \ldots, n]$ by sampling with replacement from the empirical distribution function of the $[(X_i, \delta_i, Z_i) : i = 1, \ldots, n]$ observations. For each sequence $R^*_{nj} : j = 1, \ldots, m$ we can compute bootstrap estimates $\{Q^*_D(p), (p, D) \in \mathcal{I}\}$ and next use them to approximate the distribution of the quantile process. Although it is possible to show consistency of this procedure, its drawback lies in the computational burden needed to construct estimates $(\theta^*_n, \Gamma^*_{n\theta^*})$ for each of the m simulated data sets. In the case of the proportional hazard model, Hjort (1985) proposed the use of "model based" bootstrap. Burr and Doss (1993) applied it to the construction of simultaneous confidence bands for the conditional median. In this approach, the distribution of the quantile process is approximated based on artificial observations $(X^*_i, \delta^*_i), i = 1, \ldots, n$ defined as $X^*_i = T^*_i \wedge \widetilde{T}^*_i, \delta^*_i = 1(T^*_i \leq \widetilde{T}^*_i)$, where T^*_i is sampled from the distribution $F(\widehat{\Gamma}_{n\widehat{\theta}}(t), \widehat{\theta}|Z_i)$ and $\widetilde{T}^*_i$ is sampled from $\widehat{G}(t) = 1-$ Kaplan-Meier estimate of the censoring distribution. This approach uses the assumption that censoring time is independent of covariates, which need not be satisfied in many practical situations. It is in principle possible to relax it by choosing a parametric or a semi-parametric model for the conditional distribution of censoring times, however, selection of such a model is often quite difficult, and its misspecifaction may affect the performance of confidence procedures.

3 Example

For illustrative purposes we consider now data from the Veteran's Administration lung cancer trial (Kalbfleisch and Prentice, 2000). In this trial males with inoperative lung cancer were randomized to either a standard or an experimental chemotherapy treatment and subsequently followed until death or withdrawal from the study. We shall look at the subgroup of 97 patients, who received no prior therapy, and use two covariates corresponding to performance status at the time of entry into the clinical trial and histopathological type of tumor (squamous, small cell, adeno and large cell).

Several authors (e.g. Bennett, 1983, Pettit, 1984, Cheng et al., 1995 and Murphy, Rossini and van der Vaart, 1996) proposed the use of the proportional odds ratio for analysis of this dataset. Our estimates are easy to compute in this case because the hazard rate of the i-th subject satisfies

$$\alpha_i(x, \theta) = e^{\theta^T Z_i}(1 + e^{\theta^T Z_i} x)^{-1}, \ \ell'_i(x, \theta) = -\alpha_i(x, \theta), \ \dot{\ell}_i(x, \theta) = Z_i e^{-\theta^T Z_i} \alpha_i(x, \theta). \tag{13}$$

For fixed θ, the estimate $\Gamma_{n\theta}$ is computed based on the recurrent formula given by Bogdanovicius and Nikulin (1999):

$$\Gamma_{n\theta}(t) = \Gamma_{n\theta}(t-) + \frac{N_{\cdot}(\Delta t)}{S(\Gamma_{n\theta}(t-), \theta, t)}$$

with the initial condition $\Gamma_{n\theta}(0-) = 0$. The sample version of the function $\dot\Gamma_\theta$ can be evaluated as

$$\dot\Gamma_{n\theta}(t)=\dot\Gamma_{n\theta}(t-)-[\dot S(\Gamma_{n\theta}(t-),\theta,t)+S'(\Gamma_{n\theta}(t-)\theta,t)\dot\Gamma_{n\theta}(t-)]\frac{N_{.}(\Delta t)}{S^2(\Gamma_{n\theta}(t-),\theta,t)}$$

and $\dot\Gamma_{n\theta}(0-) = 0$. The solution to the Fredholm equation can be obtained as follows. Let $X_{(1)} < \ldots < X_{(m)}, m \leq n$ be the distinct uncensored observations in the sample. Dropping dependence on the parameter θ, let B_n, C_n be the plug-in sample analogues of the functions B_θ and C_θ. These are step functions with jumps at points $X_{(i)}$ and we arrange their jumps into $m \times m$ diagonal matrices $\mathbf{B_n}(\mathbf{\Delta X}) = \text{diag}\,\{B_n(\Delta X_{(i)}) : i = 1, \ldots, m\}$, and $\mathbf{C_n}(\mathbf{\Delta X}) = \text{diag}\,\{C_n(\Delta X_{(i)}) : i = 1, \ldots, m\}$. let $\rho_\mathbf{n}(\mathbf{X})$ be an $m \times d$ matrix of the sample analogues of the conditional covariances $\rho_{-\dot\Gamma}(u, \theta)$ at points $X_{(i)}, i = 1, \ldots, m$. (Here d is dimension of the parameter θ). The matrix $\mathbf{C_n}(\mathbf{\Delta X})$ has positive entries, the matrix $\mathbf{B_n}(\mathbf{\Delta X})$ nonnegative. If $\mathbf{B_n}(\mathbf{\Delta X}) \equiv 0$ then also $\rho_\mathbf{n}(\mathbf{X}) \equiv 0$. Setting $\psi_{n\theta} = \varphi_{n\theta} + \dot\Gamma_{n\theta}$, the discrete version of the Fredholm equation corresponds to

$$[\mathbf{I} + \mathbf{K_n}(\mathbf{X})\mathbf{B_n}(\mathbf{\Delta X})]\psi_\mathbf{n}(\mathbf{X}) = \mathbf{K_n}(\mathbf{X})\rho_\mathbf{n}(\mathbf{X})\ ,$$

where $\psi_\mathbf{n}(\mathbf{X}) = [\psi_n(X_{(i)}) : i = 1, \ldots, m]^T$ is an $m \times d$ matrix of unknowns, $\mathbf{K_n}(\mathbf{X})$ is an $m \times m$ matrix with entries $\mathbf{K_n}(\mathbf{X}) = [K_n(X_{(i)}, X_{(j)})]$ and $\mathbf{I}$ represents an $m \times m$ identity. If $\mathbf{B_n}(\mathbf{\Delta X}) \equiv \mathbf{0}$ or $\rho_\mathbf{n}(\mathbf{X}) \equiv \mathbf{0}$ then the solution is $\psi_\mathbf{n}(\mathbf{X}) \equiv \mathbf{0}$. Otherwise, $\psi_\mathbf{n}(\mathbf{X}) = \mathbf{P_n^T}(\mathbf{X})\mathbf{g_n^{-1}}(\mathbf{X})\mathbf{P_n}(\mathbf{X})\rho_\mathbf{n}(\mathbf{X})$, where $\mathbf{g_n}(\mathbf{X}) = [g_{ij}]$ is a tridiagonal symmetric matrix with entries $g_{ii} = c_i + c_{i+1} + b_i, g_{i,i+1} = -c_{i+1} = g_{i+1,i}, i = 1, \ldots, m-1$ and $g_{mm} = c_m + b_m$, where $b_i = \mathcal{P}_{n\theta}(0, X_{(i)})^2 B_n(\Delta X_{(i)}),\ c_i = \mathcal{P}_{n\theta}(0, X_{(i)})^2 C_n(\Delta X_{(i)})^{-1}$ and $\mathbf{P_n}(\mathbf{X}) = \text{diag}\,[\exp(-\int_{[0,X_{(i)}]} S'(\Gamma_{n\theta}(u-), \theta, u)C_{n\theta}(du)) : i = 1, \ldots, m] \sim \text{diag}\,[\mathcal{P}_{n\theta}(0, X_{(i)}), i = 1, \ldots, m]$. (Dabrowska, 2005). After obtaining the solution, $\psi_{n\theta}$ we set $\varphi_{n\theta} = \psi_{n\theta} - \dot\Gamma_{n\theta}$. The estimate $\widehat\theta$ can be obtained using Fisher scoring algorithm. The algorithm can be started by setting $\widehat\theta^{(0)}$ obtained by solving the same score equation, but function $\varphi_{n\theta}$ set to 0 or $-\dot\Gamma_{n\theta}$.

The estimate $\Gamma_{n\widehat\theta}$ is a cadlag step function with jumps at uncensored observations, and so is the estimate $\widehat F_D(t)$ of the conditional distribution function of T given $Z \in D$. Thus the graph of the quantile function can be obtained by inverting graphically the plot of this function. The estimate $\widehat v_D(t)$ of the asymptotic variance of the $\sqrt{n}[\widehat F_D - F_D](t)$ and the process $\widehat{W}^\#(t, D)$ can be easily computed based on expressions given in Sections 2 and 4.

Table 1 provides regression coefficients and their standard errors for the Veteran's Administration lung cancer data. In this data set the performance score (PS) has range between 10 and 99, with lower values indicating poorer performance status at the time of entry into the trial. This covariate was used in the regression model after standardizing it to have average zero and standard deviation 1. The negative sign of the regression coefficient indicates that patients with higher performance score have lower odds on death and thereby a better survival experience. Patients with squamous tumor have a slightly lower odds on death than large cell tumor patients, however, the difference is not significant. Patients with adeno or small cell tumor have higher odds on death than patients with squamous or large cell types.

Table 1. Regression estimates and standard errors in the proportional odds ratio model.

covariate	theta	sd error	p-value
PS	-1.049	0.045	$< 10^{-5}$
SQUAMOUS	-0.246	0.428	0.71
SMALL	1.345	0.304	0.01
ADENO	1.275	0.342	0.02
LARGE	NA	NA	NA

We shall consider now two partitions $\mathcal{D}$ of the covariate space. In both cases, we shall consider quantile regression estimates in the range $p \in (.25, .75)$. Simultaneous confidence sets are based on the transformation $g^{-1}(p) = \log(-\log(1-p))$ and we used 1000 Monte Carlo simulations of the V vectors (section 2) to obtain the critical points.

The first partition corresponds to the four histopathological types of tumor. Figure 1 shows the corresponding quantile regression and confidence set for the conditional quantiles. The plots support results of Table 1 and show that patients with squamous or large cell tumor perform better than patients with adeno or small tumor cells. However, within each pair of tumor types, the confidence sets are nearly the same so that the differences are small.

Next we partition the covariate space according to the performance status at the time of entry into the trial. We consider patients, who are completely hospitalized (PS < 40), partially confined (PS $\in [40, 70)$) and who are not able to care (PS ≥ 70). In Figure 2, the confidence sets for the hospitalized and partially confined patients nearly overlap, suggesting similar survival experience after treatment. This experience is much worse than for patients who are not able to care. For example, the estimated median time till death for hospitalized, partially confined and unable to care patients is 25, 29 and 110 days, respectively. The corresponding confidence bounds are $(22, 35)$,

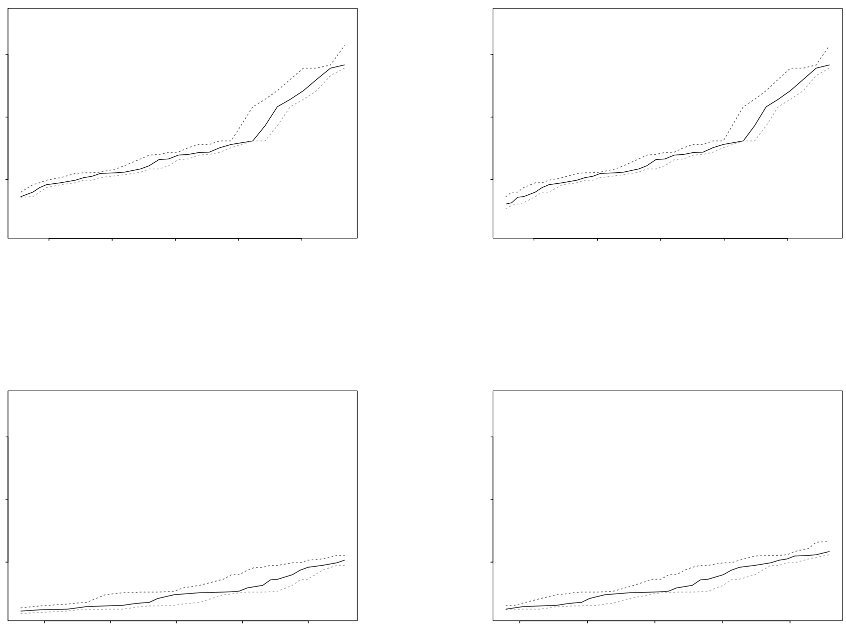

Figure 1. Quantile regression and 95% simultaneous confidence bands. Covariate space partitioned according to four tumor types.

$(24, 36)$ and $(103, 112)$ days. Figure 2 suggests also that effect of the PS score is not linear, and a regression model using a binary covariate: $Z = 1(0)$ if PS score $\geq (<)70$ may be more appropriate.

We have also considered the choice of the proportional hazard model and generalized inverse Gaussian frailty model. In each of these models the regression coefficients had the same sign, however, neither of the transformation models could be fully justified. In Figure 3 we show nonparametric plots of the Aalen-Nelson estimator, odds ratio function and Kaplan-Meier estimator of the survival function for the four tumor cell types : squamous c (solid line), large (dotted line), small (short dash) and adeno (log dash). The plots of the cumulative hazard function of the large and squamous cell type cross at around 150 days. Patients with squamous cell type are initially at a higher risk for death but at around 150 days after treatment the role of the two groups is reversed. The corresponding plots of the odds ratio function suggest that the choice of proportional hazard model may not be

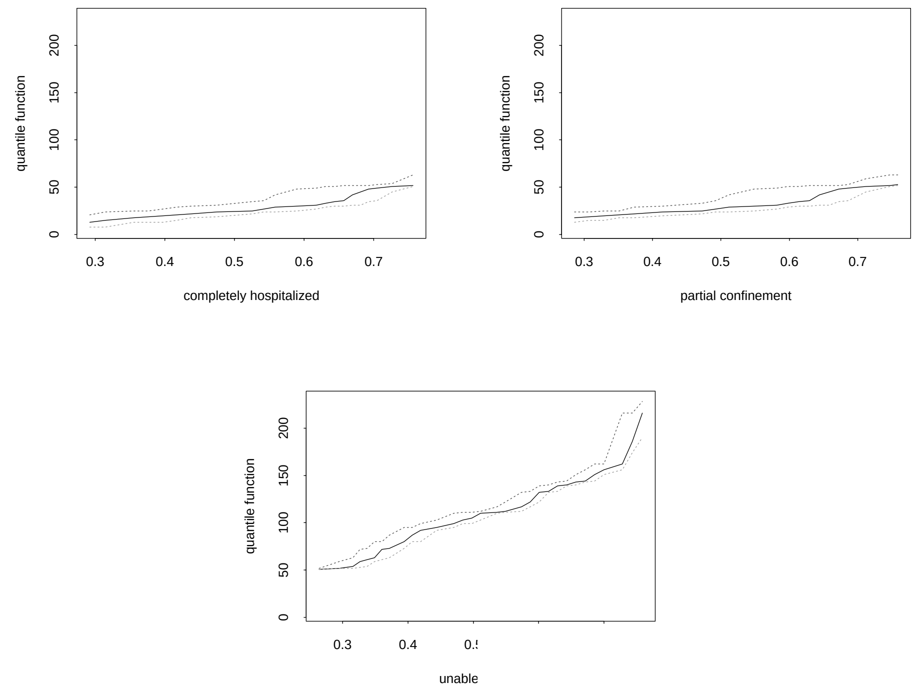

Figure 2. Quantile regression and 95% simultaneous confidence bands. Covariate space partitioned into three groups according to the of performance status (Karnofsky) score.

appropriate and that odds ratio functions are close for the two groups. In the case of the adeno and small cell tumor cell type groups, the graphs of both cumulative hazard and odds ratio functions cross only at the upper tail, however, the two groups can be only compared during the initial 180 days.

These graphs illustrate typical difficulty arising in regression analyses based on transformation models of type (1) or (2). The transformation models assume that the conditional distributions of the failure time T given $Z = z$ have the same support as the marginal distribution of T for μ- almost all z. This assumption fails to be satisfied in the fully nonparametric setting, not assuming any restrictions on the support or shape of the conditional distribution of T given $Z = z$. If $\overline{F}(t|z)$ represents the conditional distribution function of T given $Z = z$ and $\overline{G}$ is the corresponding marginal distribution function of T, then setting

$$\tau_1(z) = \inf\{t : \overline{F}(t|z) > 0\} \qquad \tau_2(z) = \sup\{t : \overline{F}(t|z) < 1\}$$
$$\tau_1 = \inf\{t : \overline{G}(t) > 0\} \qquad \tau_2 = \sup\{t : \overline{G}(t) < 1\}$$

we have $\tau_1 \le \tau_1(z) \le \tau_2(z) \le \tau_2$ for μ-almost all z, i.e. the marginal distribution of T has longer support than the conditional distributions. For different covariate levels z_1 and z_2, the intervals $[\tau_1(z_1), \tau_2(z_1)]$ and $[\tau_1(z_2), \tau_2(z_2)]$ may be very different.

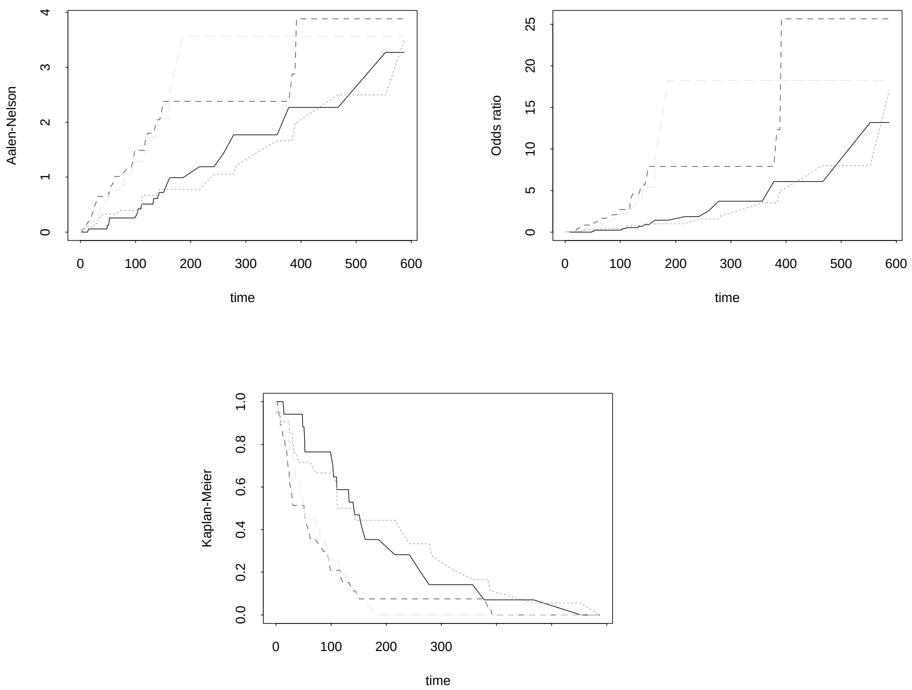

Figure 3. Aalen-Nelson, odds ratio function and Kaplan-Meier estimators for the four tumor cell types: squamous (solid line), large (dotted line), small (long dash) and adeno (short dash).

In the present example, large and squamous cell type patient groups have longer support interval than the groups of squamous and adeno cell types. Apparently, patients for whom treatment is beneficial live longer. The choice of the proportional odds ratio model appears to be more appropriate than the proportional hazards model, however, it does not accommodate variable support intervals of conditional distributions of different subgroups. The problem applies to all transformation models of type (1) and (2). The plots of Kaplan-Meier estimators corresponding to the four groups are proper survival functions in this data example because data are lightly censored (Kalbfleisch and Prentice, 2000). In moderately or heavily censored samples, the grouped data Kaplan-Meier estimator will often form an improper survival function. In such circumstances, variable supports of Kaplan-Meier estimator may indicate also presence of informative censoring. The difficulties in handling variable supports of conditional distributions apply also to other common parametric and semiparametric regression models in survival analysis and are very common in practical applications.

4 Proofs

In this section, we denote by $M_i(t)$ the process

$$M_i(t) = 1(X_i \leq t) - \int_0^t Y_i(u)\alpha_i(\Gamma_{\theta_0}(u), \theta_0)\Gamma_{\theta_0}(du) \ ,$$

where $\Gamma_0 = \Gamma_{\theta_0}$ is the "true" transformation. Then M_i are independent mean zero martingales, with respect to natural filtration generated by $\mathcal{F}_t = \sigma\{(N_i(s), Y_i(s+), Z_i) : s \leq t, i = 1, \ldots, n\}$. For any measurable functions $g_q(u, z), q = 1, 2$ such that

$$E \int Y_i(u) g_q^2(u, Z_i)\alpha_i(\Gamma_0(u), \theta_0)\Gamma_0(du) < \infty$$

we have

$$\begin{aligned} \mathrm{cov}\left(\int g_1(u, Z_i)M_i(du), \int g_2(u, Z_i)M_i(du)\right) = \\ E \int Y_i(u) g_1(u, Z_i) g_2(u, Z_i)\alpha_i(\Gamma_0(u), \theta_0)\Gamma_0(du) \end{aligned}$$

Lemma 4.1 *Suppose that the conditions of Propositions 2.1 and 2.2 are satisfied.*

(i) The estimate $\widehat{\theta}$ *satisfies* $\sqrt{n}[\widehat{\theta} - \theta_0] = \Sigma(\theta_0)^{-1}\sqrt{n}\widetilde{U}_n(\theta_0) + o_P(1)$, *where* $\Sigma(\theta) = \Sigma_1(\theta) + \Sigma_2(\theta)$ *and* $\widetilde{U}_n(\theta_0) = n^{-1}\sum_{i=1}^n [U_{1i}(\theta_0) + U_{2i}(\theta_0)]$ *is given by*

$$\begin{aligned} U_{1i}(\theta_0) &= \int_0^\tau b_i(\Gamma_{\theta_0}(u), \theta_0, u)M_i(du) \ , \\ U_{2i}(\theta_0) &= -\int_0^\tau W_{0i}(t)\rho_{\varphi_{\theta_0}}(t, \theta_0)EN(dt) \ , \end{aligned}$$

$$W_{0i}(t) \;=\; \int_0^t \frac{M_i(du)}{s(\Gamma_0(u-),\theta_0,u)}\mathcal{P}_{\theta_0}(u,t)$$

and

$$\begin{aligned} b_i(\Gamma_{\theta_0},\theta_0,u) \;&=\; \dot\ell_i(\Gamma_{\theta_0}(u),\theta_0) - \ell_i'(\Gamma_{\theta_0}(u),\theta_0)\varphi_\theta(u) \\ &-\; \frac{\dot s}{s}(\Gamma_{\theta_0}(u),\theta_0,u) + \frac{s'}{s}(\Gamma_{\theta_0}(u),\theta_0,u)\varphi_{\theta_0}(u)\ . \end{aligned}$$

The sums $n^{-1/2}\sum_{i=1}^n U_{1i}(\theta_0)$ and $n^{-1/2}\sum_{i=1}^n U_{2i}(\theta_0)$ are uncorrelated and converge weakly to independent mean zero normal vectors with covariances $\Sigma_1(\theta_0)$ and $\Sigma_2(\theta_0)$. Moreover,

$$\sqrt{n}[\Gamma_{n\widehat\theta} - \Gamma_{\theta_0} - [\widehat\theta-\theta_0]\dot\Gamma_{\theta_0}](t) = \frac{1}{\sqrt n}\sum_{i=1}^n W_{0i}(t) + o_P(1)$$

uniformly in $t \in [0,\tau]$.

(ii) We have $\Sigma_{qn}(\widehat\theta) \to_p \Sigma_q(\theta_0)$ for $q = 1, 2$,

$$\|\Gamma_{n\widehat\theta} - \Gamma_{n\theta_0}\| \to_P 0, \qquad \left\|\dot\Gamma_{n\widehat\theta} - \dot\Gamma_{n\theta_0}\right\| \to_P 0,$$

$$\left\|\int_0^{\cdot} \widehat\rho_{\widehat\varphi}(u,\widehat\theta)N_{.}(du) - \int_0^{\cdot} \rho_{\varphi_{\theta_0}}(u,\theta_0)EN(du)\right\| \to_P 0,$$

$$\left\|\int_0^{\cdot} \frac{\dot S}{S^2}(\Gamma_{n\widehat\theta}(u-),\widehat\theta,u)N_{.}(du) - \int_0^{\cdot} \frac{\dot s}{s^2}(\Gamma_{\theta_0}(u-),\theta_0,u)EN(du)\right\| \to_P 0,$$

$$\left\|\int_0^{\cdot} \frac{S'}{S^2}(\Gamma_{n\widehat\theta}(u-),\widehat\theta,u)N_{.}(du) - \int_0^{\cdot} \frac{s'}{s^2}(\Gamma_{\theta_0}(u-),\theta_0,u)EN(du)\right\| \to_P 0,$$

$$\limsup_n \exp\int_0^\tau \frac{|S'|}{S^2}(\Gamma_{n\widehat\theta}(u-),\widehat\theta,u)N_{.}(du) = O_P(1)$$

and

$\widehat{\mathcal{P}}_{\widehat\theta}(u,t) \to_P \mathcal{P}_{\theta_0}(u,t)$ uniformly in $0 < u < t \le \tau$.

(iii) Let

$$\begin{aligned} \psi_1(t,D) \;&=\; \pi(D)^{-1}E1(Z_i \in D)f(\Gamma_0(t),\theta_0|Z_i)\ , \\ \psi_2(t,D) \;&=\; \psi_1(t,X)\dot\Gamma_0(t) + \pi(D)^{-1}E1(Z_i\in D)\dot F(\Gamma_0(t),\theta_0|Z_i) \end{aligned}$$

and let $\widehat\psi_p, p = 1,2$ be the estimate of this function obtained by replacing the pair (θ_0,Γ_0) and the function $\pi(D)$ by $(\widehat\theta,\Gamma_{n\widehat\theta})$ and $\widehat\pi(D)$. Then $\|\widehat\psi_q - \psi_q\| \to_P 0, q = 1,2$.

(iv) Part (ii) and (iii) remains to hold if the estimates $(\widehat{\theta}, \Gamma_{n\widehat{\theta}})$ are replaced by (θ^*, Γ_n^*) such that $\theta^* \to_P \theta_0$ and $\|\Gamma_n^* - \Gamma_{\theta_0}\| \to_P 0$.

We omit the proof of this lemma. Part (i)-(ii) and (iv) can be found in Dabrowska (2005), while part (iii) is a straightforward consequence of part (i)-(ii).

Proof of Proposition 2.2. We have

$$\widehat{W}(t, D) = \frac{\pi(D)}{\widehat{\pi}(D)} \sum_{j=1}^{4} \widehat{W}_j(t, D) \ ,$$

where

$$\begin{aligned}
\widehat{W}_1(t, D) &= \frac{1}{\sqrt{n}} \frac{1}{\pi(D)} \sum_{i=1}^{n} 1(Z_i \in D)[F(\Gamma_0(t), \theta_0 | Z_i) - F_D(t)] \ , \\
\widehat{W}_2(t, D) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{0i}(t)\psi_1(t, D) - \psi_2(t, D)^T \Sigma^{-1}(\theta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{2i}(\theta_0) \ , \\
\widehat{W}_3(t, D) &= \psi_2(t, D)^T \Sigma^{-1}(\theta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{1i}(\theta_0) \\
\widehat{W}_4(t, D) &= \frac{1}{\sqrt{n}\pi(D)} \sum_{i=1}^{n} 1(Z_i \in D)[F(\widehat{\Gamma}_{\widehat{\theta}}(t, \widehat{\theta} | Z_i) - F(\Gamma_0(t), \theta_0 | Z_i)] \\
&- \widehat{W}_2(t, D) - \widehat{W}_3(t, D) \ .
\end{aligned}$$

Here $\widehat{W}_j(t, D), j = 1, 2, 3$ represent uncorrelated sums of mean zero iid processes with finite variance and covariance

$$\begin{aligned}
\text{cov}\ (\widehat{W}_1(t_1, D_1), \widehat{W}_1(t_2, D_2)) =& \pi(D_1)\pi(D_2)^{-1} E 1(Z_i \in D_1 \cap D_2) \\
& F(t_1|Z)F(t_2|Z), -F_{D_1}(t_1)F_{D_2}(t_2) \\
\text{cov}\ (\widehat{W}_2(t_1, D_1), \widehat{W}_2(t_2, D_2)) =& \text{cov}\ (W_0(t), W_0(t'))\psi_1(t_1, D_1)\psi_1(t_2, D_2) \\
& + \psi_2(t_1, D_1)^T \text{cov}\ (W_0(t_1), T)\psi_1(t_2, D_2) \\
& + [\psi_2(t_1, D_1)^T \text{cov}\ (W_0(t_1), T)\psi_1(t_2, D_2)]^T \\
& + \psi_2(t_1, D_1)^T VarT \psi_2(t_2, D_2) \\
& - \text{cov}\ (\widehat{W}_3(t_1, D_1), \widehat{W}_3(t_2, D_2)), \qquad (14) \\
\text{cov}\ (\widehat{W}_3(t_1, D_1), \widehat{W}_3(t_2, D_2)) =& \psi_2(t_1, D_1)^T \Sigma^{-1}(\theta_0)\Sigma_1(\theta_0)\Sigma^{-1}(\theta_0)\psi_2(t_2, D_2)
\end{aligned}$$

and, from section 2,

$$\operatorname{cov} T = \Sigma^{-1}(\theta_0), \quad \operatorname{cov}(T, W_0(t)) = -\Sigma^{-1}(\theta_0)[\varphi_{\theta_0} + \dot\Gamma_{\theta_0}](t) ,$$
$$\operatorname{cov}(W_0(t), W_0(t')) = K_{\theta_0}(t,t') .$$

We also have

$$[\varphi_{\theta_0} + \dot\Gamma_{\theta_0}](t) = \int_0^\tau K_{\theta_0}(t,u)\rho_\varphi(u,\theta_0)EN.(du) .$$

By central limit theorem, finite dimensional distributions of the processes $\{\widehat{W}_j, j = 1,2,3\}$ converge weakly to a multivariate vector with covariance matrix given by (14).

For each $j = 1,2,3$, the process $\{\widehat{W}_j(t,D) : (t,D) \in \mathcal{T}\}$ can be represented as $n^{-1/2}\sum_{i=1}^n h_{t,D}^{(j)}(X_i,\delta_i,Z_i)$, with $h^{(j)}$ varying over a Euclidean class of functions $\mathcal{H}_j = \{h_{t,D}^{(j)} : (t,D) \in \mathcal{T}\}$ for a square integrable envelope (Nolan and Pollard, 1987). This can be verified, by noting that $\mathcal{D}$ is a finite collection of sets, and for each $D \in \mathcal{D}$, the relevant functions $h_{t,D}^{(j)} \in \mathcal{H}_j$ can be represented as finite linear combination of functions of bounded variation with respect to t. We also have $Eh_{t,D}^{(j)}(X_i,\delta_i,Z_i) = 0$ for each $h_{t,D}^{(j)} \in \mathcal{H}_j$. Hence the process $\widehat{W}_j = G_{n,j} = \{\sqrt{n}[P_n - P](h_{t,D}^{(j)}) : h_{t,D}^{(j)} \in \mathcal{H}_j\}$ is equicontinuous and $\mathcal{H}_j$ is totally bounded with respect to the variance semi-metric ρ_j. Set $\rho = \max \rho_j, j = 1,2,3$. Then $\mathcal{T}$ is totally bounded with respect to ρ, and $\{\widehat{W}_j : j = 1,2,3\}$ is asymptotically tight in $\ell^\infty(\mathcal{T})$ and converges weakly to a Gaussian process $\{W_j : j = 1,2,3\}$. Its components are independent, and W_j have covariance function given by the right–hand side of (14).

Using Taylor expansion, we also have $\widehat{W}_4(t,D) = \widehat{W}_{41}(t,D) + \widehat{W}_{42}(t,D)$, where

$$\begin{aligned}
\widehat{W}_{41}(t,D) &= \sqrt{n}(\Gamma_{n\theta}(t) - \Gamma_{\theta_0}(t) - (\widehat{\theta} - \theta_0)^T\dot\Gamma_0(t))\widehat{\psi}_1^*(t,D) \\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^n W_{0i}(t)\psi_1(t,D) , \\
\widehat{W}_{42}(t,D) &= \psi_2^*(t,D)^T\sqrt{n}(\widehat{\theta} - \theta_0) - \psi_2(t,D)^T\Sigma^{-1}(\theta_0)\sqrt{n}U_n(\theta_0)
\end{aligned}$$

and

$$\begin{aligned}
\psi_1^*(t,D) &= \frac{1}{n\pi(D)}\sum_{i=1}^n 1(Z_i \in D)f(\Gamma^*(t),\theta^*|Z_i) , \\
\psi_2^*(t,D) &= \psi_1^*(t,D)\dot\Gamma_0(t) + \frac{1}{n\pi(D)}\sum_{i=1}^n 1(Z_i \in D)\dot F(\Gamma^*(t),\theta^*|Z_i) .
\end{aligned}$$

Here θ^* is on a line segment between θ_0 and $\widehat{\theta}$, and $\|\Gamma^* - \Gamma_{\theta_0}\| \to_P 0$. By Lemma 4.1,

$$\sup\{|\widehat{W}_{4p}(t,D)| : (t,D) \in \mathcal{T}\} \to_P 0$$

for $p = 1, 2$. To complete the proof of part (i) of the Proposition 2.2, we note that $\widehat{\pi}(D) \to \pi(D)$ a.s. for $D \in \mathcal{D}$ so that $\widehat{W} = \big\{\frac{\pi(D)}{\widehat{\pi}(D)} \sum_{j=1}^4 \widehat{W}_j(t,D) : (t,D) \in \mathcal{T}\big\}$ converges weakly in $\ell^\infty(\mathcal{T})$ to $W = \{W(t,D) = \sum_{j=1}^3 W_j(t,D) : (t,D) \in \mathcal{T}\}$. Its variance function is given by $v_D(t) = \sum_{j=1}^3 \mathrm{var}\, W_j(t,D)$. For any D this is a continuous function with respect to t and positive on any interval $[\tau_1 - \varepsilon, \tau_2 + \varepsilon]$ on which Γ_{θ_0} forms a continuous strictly increasing function.

To show part (ii), first recall that $V_i = (V_{1i}, V_{2i}), i = 1, \ldots, n, \ldots$ and $V_3 = (V_{31}, \ldots, V_{3d})$, are mutually independent $\mathcal{N}(0,1)$ variables, independent of $R_i = (X_i, \delta_i, Z_i), i = 1, \ldots, n$. We let variables $R_i = i = 1, 2, \ldots$ be defined as coordinate projections on the "first" ∞ coordinates in the product probability space $(\Omega^\infty \times \mathcal{V} \times \mathcal{V}', \mathcal{F}^\infty \times \mathcal{B} \times B', P^\infty \times Q \times Q')$ and let $V_i, i = 1, \ldots, ..$ and V_3 be defined on the "last" two coordinates.

Set

$$\widetilde{W}_1(t,D) = \frac{1}{\sqrt{n}} \frac{1}{\pi(D)} \sum_{i=1}^n V_{1i} 1(Z_i \in D)[F(\Gamma_{\theta_0}(t), \theta_0 | Z_i) - F_D(t)] \, ,$$
$$\widetilde{W}_2(t,D) = \widetilde{W}_0(t)\psi_1(t,D) + \int_0^\tau \widetilde{W}_0(s)\rho_{\varphi_{\theta_0}}(s,\theta_0) EN_{\cdot}(ds)\Sigma^{-1}(\theta_0)\psi_2(t,D) \, ,$$
$$\widetilde{W}_3(t,D) = V_3 \Sigma_1^{1/2}(\theta_0)\Sigma^{-1}(\theta_0)\psi_1(t,D) \, ,$$

where

$$\widetilde{W}_0(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{2i} \frac{1[X_i \le t, \delta_i = 1]}{s(\Gamma_{\theta_0}(X_i-), \theta_0, X_i)} \mathcal{P}_{\theta_0}(X_i, t) \, .$$

For $j, k = 1, 2, 3, j \ne k$, we have

$$\begin{aligned}
&\mathrm{cov}\,(\widetilde{W}_j(t,D), \widetilde{W}_j(t',D')) = \mathrm{cov}\,(\widehat{W}_j(t,D), \widehat{W}_j(t',D')) \, , \\
&\mathrm{cov}\,(\widehat{W}_k(t,D), \widetilde{W}_j(t',D')) = \mathrm{cov}\,(\widetilde{W}_k(t,D), \widetilde{W}_j(t',D')) \\
&= \mathrm{cov}\,(\widehat{W}_k(t,D), \widehat{W}_j(t',D')) = 0 \, .
\end{aligned}$$

Also $\widetilde{W}_3$ does not involve n, the $R_i, i = 1, 2, \ldots$ or the $V_{ji}, j = 1, 2, i = 1, 2 \ldots$ sequences, and is independent of the processes $\widetilde{W}_j, j = 1, 2$ and $\widehat{W}_j, j = 1, 2, 3$.

Similarly to part (i), the processes $\{\widetilde{W}_j(t,D) : (t,D) \in \mathcal{T}, j = 1, 2\}$ are of the form $\widetilde{W}_j(t,D) = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_{ji} g^{(j)}(X_i, \delta_i, Z_i)$, where $g^{(j)}$ varies over

$\mathcal{G}_j = \{g^{(j)}_{t,D}(x,d,z) : (t,D) \in \mathcal{T}\}$, a Euclidean class of functions for a square integrable envelope and is totally bounded with respect to the semi-metric ρ. The class of products $\{vg^{(j)}_{t,D}(x,\delta,z) : (t,D) \in \mathcal{T}\}$ is also Euclidean. Therefore, unconditionally $[\widetilde{W}_j : j = 1,2]$ is asymptotically tight and converges to a Gaussian process $[W^{\#}_j : j = 1,2]$, whose components are independent and independent of $\widetilde{W}_3$ and $[W_1, W_2, W_3]$.

Alternatively, for $j = 1$, we have $g^{(1)}_{t,D} = h^{(1)}_{t,D}$ with $Ph^{(1)}_{t,D} = 0$ and

$$\widetilde{W}_1(t,D) = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{1i}(\delta_{R_i} - P)[g_{t,D}] = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{1i}\delta_{R_i}[g_{t,D}] \,.$$

For $j = 2$

$$\begin{aligned}\widetilde{W}_2(t,D) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{2i}(\delta_{R_i} - P)[g^{(2)}_{t,D}] + \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{2i}P[g^{(2)}_{t,D}] \\ &= \widetilde{W}_{21}(t,D) + \widetilde{W}_{22}(t,D)\end{aligned}$$

and the two components on the right-hand side are uncorrelated. Application of the unconditional multiplier central limit theorem in van der Vaart and Wellner (1996, Corollary 2.9.4, p.180) implies that the processes $[\widetilde{W}_1, \widetilde{W}_{21}, \widetilde{W}_{22}, \widetilde{W}_3]$ and $[\widehat{W}_1, \widehat{W}_2, \widehat{W}_3]$ converge jointly in $[\ell^\infty(\mathcal{T})]^4 \times [\ell^\infty(\mathcal{T})]^3$ to independent Gaussian processes, $[W^{\#}_1, W^{\#}_{21}, W^{\#}_{22}, W^{\#}_3 = \widetilde{W}_3]$ and $[W_1, W_2, W_3]$. By continuous mapping theorem, we also have unconditional weak convergence of $[\widehat{W} = \sum_{j=1}^3 \widehat{W}_j, \widetilde{W} = \sum_{j=1}^3 \widetilde{W}_j]$ in $\ell^\infty(\mathcal{T}) \times \ell^\infty(\mathcal{T})$ to a vector of independent Gaussian processes $[W, W^{\#}]$, with the same covariance function.

Conditionally on $R_1, R_2, \ldots, \ldots$ the processes $\widetilde{W}_1, \widetilde{W}_{21}$ and $\widetilde{W}_{22}$ have mean zero,

$$\begin{aligned}\text{cov}_V[\widetilde{W}_1(t_1,D_1), \widetilde{W}_1(t_2,D_2)] =& \frac{1}{n}\sum_{i=1}^n g^{(1)}_{t_1,D_1}(R_i) g^{(1)}_{t_2,D_2}(R_i)^T \\ &\to P(g^{(1)}_{t_1,D_1}[g^{(1)}_{t_2,D_2}]^T), \\ \text{cov}_V[\widetilde{W}_{21}(t_1,D_1), \widetilde{W}_{21}(t_2,D_2)] =& \frac{1}{n}\sum_{i=1}^n g^{(2)}_{t_1,D_1}(R_i) g^{(2)}_{t_2,D_2}(R_i)^T \\ &\quad - Pg^{(2)}_{t_1,D_1}[Pg^{(2)}_{t_2,D_2}]^T \\ &\to \text{cov}\,(g^{(2)}_{t_1,D_1}(R_1), g^{(2)}_{t_2,D_2}(R_1))\,, \\ \text{cov}_V[\widetilde{W}_{22}(t_1,D_1), \widetilde{W}_{22}(t_2,D_2)] =& Pg^{(2)}_{t_1,D_1}[Pg^{(2)}_{t_2,D_2}]^T\,,\end{aligned}$$

$$\begin{aligned}
\text{cov}\,_V[\widetilde{W}_{21}(t_1,D_1),\widetilde{W}_{22}(t_2,D_2)] =& \frac{1}{n}\sum_{i=1}^{n}[g^{(2)}_{t_1,D_1}(R_i) - Pg^{(2)}_{t_1,D_1}][Pg^{(2)}_{t_2,D_2}]^T \to 0,\\
\text{cov}\,_V[\widetilde{W}_1(t_1,D_1),\widetilde{W}_{2j}(t_2,D_2)] =& 0, \quad j=1,2\ ,\\
\text{cov}\,_V[\widetilde{W}_3(t_1,D_1),\widetilde{W}_{2j}(t_2,D_2)] =& 0, \quad j=1,2\ ,\\
\text{cov}\,_V[\widetilde{W}_3(t_1,D_1),\widetilde{W}_1(t_2,D_2) =& 0,
\end{aligned}$$

for almost all $R_1, R_2, \ldots$ (Actually, conditionally on $R_1, R_2, \ldots$, $\widetilde{W}_j$ processes are independent). By conditional multiplier CLT, we have that conditionally on $R_1, R_2, \ldots$, the finite dimensional distributions of $\widetilde{W}_1$ and $\widetilde{W}_2$ are asymptotically multivariate normal and independent, for almost all $R_1, R_2 \ldots$. The covariance function is the same as of finite dimensional distributions of W_1 and W_2. By continuous mapping theorem, we also have that conditionally on $R_1, R_2, \ldots$, the finite dimensional distributions of $\widetilde{W}$ converge weakly to a multivariate normal distribution for almost all $R_1, R_2, \ldots$. The covariance of the multivariate normal distributions is the same as the covariance of the corresponding finite dimensional distributions of W.

Let BL_1 be the collection of functions f from $\ell^\infty(\mathcal{T})$ into $[0,1]$ that are Lipschitz continuous with Lipschitz continuity constant equal to 1. For fixed δ and $x \in \mathcal{T}$, let $\Pi_\delta(x)$ be the closest point to x in $\mathcal{T}$ in a partition of the set $\mathcal{T}$ with mesh-width δ (with respect to the semi-metric ρ). By triangular inequality

$$\begin{aligned}
&\sup_{f\in BL_1}|E_V f(\widetilde{W}) - Ef(W)| \le \sup_{f\in BL_1}|Ef(W\circ\Pi_\delta) - Ef(W)| + \\
&\sup_{f\in BL_1}|Ef(W\circ\Pi_\delta) - E_V f(\widetilde{W}\circ\Pi_\delta)| + \sup_{f\in BL_1}|E_V f(\widetilde{W}\circ\Pi_\delta) - E_V f(\widetilde{W})|\\
&= I_1 + I_2 + I_3\ .
\end{aligned}$$

As in van der Vaart and Wellner (1996, p. 182), the term I_1 converges to 0, because the process W has continuous paths with respect ρ and $W\circ\Pi_\delta \to W$ in almost surely as $\delta \downarrow 0$. For fixed $\delta > 0$, I_2 converges to 0 for almost all $R_1, R_2, \ldots$. This follows because conditionally on $R_1, R_2, \ldots$, the finite dimensional distributions of $\widetilde{W}$ converge in distribution to a multivariate normal vector, for almost all $R_1, R_2, \ldots$. Finally,

$$\begin{aligned}
I_3 &\le \sup_{f\in BL_1} E_V|f(\widetilde{W}\circ\Pi_\delta) - f(\widetilde{W})| \le E_{V_1}\|\widetilde{W}_1\circ\Pi_\delta - \widetilde{W}_1\|_{\mathcal{G}_{1\delta}} + \\
&+ E_{V_2}\|\widetilde{W}_2\circ\Pi_\delta - \widetilde{W}_2\|_{\mathcal{G}_{2\delta}} + E_{V_3}\|\widetilde{W}_3\circ\Pi_\delta - \widetilde{W}_3\|_{\mathcal{G}_{3\delta}}\\
&\le E_{V_1}\|\widetilde{W}_1\|_{\mathcal{G}_{1\delta}} + E_{V_2}\|\widetilde{W}_2\|_{\mathcal{G}_{2\delta}} + E_{V_3}\|\widetilde{W}_3\|_{\mathcal{G}_{3\delta}}\ ,
\end{aligned}$$

where $\mathcal{G}_{j\delta} = \{g - g' : g, g' \in \mathcal{G}_j : \rho(g - g') < \delta\}$, for $j = 1, 2, 3$. The first two expectation converge to 0 as $n \to \infty$ and $\delta \downarrow 0$, by Lemma 2.9.1 in van der Vaart and Wellner (1996, p 177). The last expected does not depend on n, and converges to 0 as $\delta \downarrow 0$.

It remains to consider the process $\widehat{W}^{\#}$ defined in Section 2. We show that unconditionally $\|\widehat{W}_j^{\#} - \widetilde{W}_j\| \to 0$ in probability. If this is the case, then for $\varepsilon > 0$, we have

$$\begin{aligned}
&\sup_{f \in BL_1} |E_V^* f(\widehat{W}^{\#}) - Ef(W)| \\
&\leq \sup_{f \in BL_1} |E_V f(\widetilde{W}) - Ef(W)| + \sup_{f \in BL_1} |E_V^* f(\widehat{W}^{\#}) - E_V f(\widetilde{W})| \\
&\leq \sup_{f \in BL_1} |E_V f(\widetilde{W}) - Ef(W)| + \varepsilon + 2P_V^*(\|\widehat{W}^{\#} - \widetilde{W}\| > \varepsilon).
\end{aligned}$$

The first term converges to 0 in probability. The last term converges to 0 in (outer) mean.

Clearly, for $j = 3$, we have $\widehat{\Sigma}_n(\widehat{\theta}) \to \Sigma(\theta_0)$, $\widehat{\Sigma}_{2n}(\widehat{\theta}) \to \Sigma_2(\theta_0)$ and $\|\widehat{\psi}_1 - \psi_1\|_\infty \to 0$ in probability so that $\|\widehat{W}_3^{\#} - \widetilde{W}_3\| \to_P 0$.

Next, for $j = 1, 2, 3$, define

$$\widetilde{H}_j(t, D) = \frac{1}{n} \sum_{i=1}^n V_{1i} 1(Z_i \in D) h_{jt}(Z_i) \ ,$$

where

$$\begin{aligned}
h_{jt}(Z) &= \pi(D)^{-1} & j = 1 \ , \\
&= \pi(D)^{-1} f(\Gamma_{\theta_0}(t), \theta_0 | Z) & j = 2 \ , \\
&= \pi(D)^{-1} \dot{F}(\Gamma_{\theta_0}(t), \theta_0 | Z) & j = 3 \ .
\end{aligned}$$

We have $E\widetilde{H}_j(t, D) = 0$ for $(t, D) \in \mathcal{T}$. Unconditionally, the strong law of large numbers, yields $\widetilde{H}_j(t, D) \to 0$ a.s. pointwise in $(t, D) \in \mathcal{T}$. The convergence is also uniform since for each D, the process $H_j(t, D)$ has paths of bounded variation. We also have $\widetilde{W}_1 - \widehat{W}_1^{\#} = \sum_{j=1}^4 \widetilde{W}_{1j}$, where

$$\begin{aligned}
\widetilde{W}_{11}(t, D) &= -\sqrt{n}[\widehat{F}_D - F_D](t) \widetilde{H}_1(t, D) \ , \\
\widetilde{W}_{12}(t, D) &= \sqrt{n}[\Gamma_{n\widehat{\theta}} - \Gamma_0 - (\widehat{\theta} - \theta_0)^T \dot{\Gamma}_{\theta_0}](t) \widetilde{H}_2(t, D) \ , \\
\widetilde{W}_{13}(t, D) &= \sqrt{n}[\widehat{\theta} - \theta_0]^T [\dot{\Gamma}_{\theta_0}(t) \widetilde{H}_2(t, D) + \widetilde{H}_3(t, D)] \ , \\
\widetilde{W}_{14}(t, D) &= O_P(1) \frac{1}{n} \sum_{i=1}^n |V_{1i}| O(\sqrt{n} \|\Gamma_{n\widehat{\theta}} - \Gamma_{\theta_0}\|^2 + \sqrt{n}(\widehat{\theta} - \theta_0)^2) \ .
\end{aligned}$$

These four terms satisfy $\|\widetilde{W}_{1j}\| \to 0$ in probability (unconditionally) and the same holds for the process $\widetilde{W}_1 - \widehat{W}_1^{\#}$.

Finally, define

$$\begin{aligned}
\widetilde{M}(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_{2i} 1(X_i \le t, \delta_i = 1) , \\
\widetilde{W}_4(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_{2i} \frac{1(X_i \le t, \delta_i = 1)}{s(\Gamma_0(X_i-), \theta_0, X_i)} = \int_0^t \frac{\widetilde{M}(du)}{s(\Gamma_{\theta_0}(u-), \theta_0, u)} , \\
\widehat{W}_4^{\#}(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_{2i} \frac{1(X_i \le t, \delta_i = 1)}{S(\Gamma_{n\widehat{\theta}}(X_i-), \widehat{\theta}, X_i)} = \int_0^t \frac{\widetilde{M}(du)}{S(\Gamma_{n\widehat{\theta}}(u-), \widehat{\theta}, u)} .
\end{aligned}$$

A similar argument as in analysis of the term $\widetilde{W}_2$ shows that $\widetilde{W}_4$ converges weakly (unconditionally) to a mean zero time transformed Brownian motion with variance function $C_{\theta_0}(t)$. Since EN is a continuous function, so is C_{θ_0}. We have

$$\widehat{W}_4^{\#}(t) - \widetilde{W}_4(t) = \int_0^t \left[\frac{s(\Gamma_{\theta_0}(u-), \theta_0, u)}{S(\Gamma_{n\widehat{\theta}}(u-), \widehat{\theta}, u)} - 1 \right] \widetilde{W}_4(du) .$$

Denote the term in the bracket by $a_n(u-)$. Then a_n is a process with left continuous and right-hand limits, $\|a_n\| \to_P 0$ and

$$\limsup_n \|a_n\|_v = O_P(1) ,$$

where $\|\cdot\|_v$ is the variation norm. For given $\delta > 0$, let $t_1 < t_2 < \ldots t_k$ be a partition of $[0, \tau]$, such that $C_{\theta_0}(t_i) - C_{\theta_0}(t_{i-1}) < \delta$. Define $\Pi_\delta(t) = t_{i-1}$ if $t \in [t_{i-1}, t_i)$. Then integration by parts, yields

$$\begin{aligned}
\widehat{W}_4^{\#}(t) &= \int_0^t a_n(u-)[\widetilde{W}_4 - \widetilde{W}_4 \circ \Pi_\delta](du) + \int_0^t a_n(u-)[\widetilde{W}_4 \circ \Pi_\delta](du) \\
&= [\widetilde{W}_4 - \widetilde{W}_4 \circ \Pi_\delta](t) a_n(t) + \int_0^t [\widetilde{W}_4 - \widetilde{W}_4 \circ \Pi_\delta](u) a_n(du) \\
&\quad + \int_0^t a_n(u-)[\widetilde{W}_4 \circ \Pi_\delta](du) .
\end{aligned}$$

The right-hand side converges then to 0 in probability uniformly in t, as $n \to \infty$, followed by $\delta \to 0$. We also have

$$\begin{aligned}
\widehat{W}_0^{\#}(t) &= \int_0^t \widehat{W}_4^{\#}(du) \mathcal{P}_{\widehat{\theta}}(u, t) , \\
\widetilde{W}_0(t) &= \int_0^t \widetilde{W}_4(du) \mathcal{P}_{\theta_0}(u, t) .
\end{aligned}$$

Then

$$\begin{aligned}\widehat{W}_0^{\#}(t) &= \widehat{W}_4^{\#}(t) - \int_0^t \widehat{W}_0^{\#}(u-)\frac{S'}{S^2}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)N_{\cdot}(du)\ ,\\ \widetilde{W}_0(t) &= \widetilde{W}_4(t) - \int_0^t \widetilde{W}_0(u-)\frac{s'}{s^2}(\Gamma_{\theta_0}(u-),\theta_0,u)EN(du)\ .\end{aligned}$$

We have

$$\begin{aligned}[\widehat{W}_0^{\#}(t) - \widetilde{W}_0(t)] =& \text{Rem}(t) - \int_0^t [\widehat{W}_0^{\#} - \widetilde{W}_0](u-)\frac{S'}{S^2}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)N_{\cdot}(du)\ ,\\ \text{Rem}(t) =& [\widehat{W}_4^{\#} - \widetilde{W}_4](t) - \int_0^t \widetilde{W}_0(u-)\left(\frac{S'}{S^2}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)N_{\cdot}(du)\right.\\ &\left.-\frac{s'}{s^2}(\Gamma_0(u-),\theta_0,u)EN(du)\right)\ .\end{aligned}$$

We have $\|\text{Rem}\| \to 0$ and $\|\text{Rem}^-\| \to 0$ in probability. Hence by Gronwall's inequality (Beesack, 1975)

$$\begin{aligned}&|\widehat{W}_0^{\#} - \widetilde{W}_0|(t) \le |\text{Rem}(t)|\\ &+ \int_0^t |\text{Rem}(u-)|\frac{|S'|}{S^2}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)N_{\cdot}(du)\exp\int_u^\tau \frac{|S'|}{S^2}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)N_{\cdot}(du)\\ &\le \max\sup_{t\le\tau}|\text{Rem}(t)|, |\text{Rem}(t-)|\limsup_n \exp\int_0^\tau \frac{|S'|}{S^2}(\Gamma_{n\widehat{\theta}}(u-),\widehat{\theta},u)N_{\cdot}(du).\end{aligned}$$

Application of Lemma 4.1 and integration by parts implies that this term converges to 0 in probability, and $\|\widehat{W}_0^{\#} - \widetilde{W}_0\| \to 0$ in probability. Similarly, we have $\|\widehat{W}_2^{\#} - \widetilde{W}_2\| \to 0$ in probability. □

Acknowledgement. I thank an anonymous reviewer and Roger Koenker for comments.

References

ANDERSEN, P.K. and GILL, R.D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.,* **10,** 1100-1120.

BENNETT. S. (1983). Analysis of the survival data by the proportional odds model. *Stat. Med.,* **2,** 273-277.

BEESACK, P.R. (1975). *Gronwall Inequalities.* Carlton Math. Lecture Notes **11,** Carlton University, Ottawa.

BIE, O., BORGAN, O. and LIESTOL, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small-sample properties. *Scand. J. Statist.,* **14,** 221-233.

Burr, D. and Doss, H. (1993). Confidence bands for the median survival time as a function of the covariates in the Cox model. *J. Amer. Statist. Assoc.,* **88,** 1330-1340.

Bogdanovicius, V. and Nikulin, M. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Anal.,* **5,** 329-350.

Cheng, S.C., Wei, L.J. and Ying, Z. (1995) Analysis of transformation models with censored data. *Biometrika,* **82,** 835-845.

Cox, D.R. (1972). Regression models in life tables. *J. Roy. Statist. Soc. Ser. B.,* **34,** 187-202.

Dabrowska, D.M. and Doksum, K.A. (1987). Estimates and confidence intervals for median and mean life in the proportional hazard model. *Biometrika,* **74,** 799-807.

Dabrowska, D.M. (2005). Estimation in a class of semiparametric transformation models. *J. Multivariate Anal.* (in revision).

Hjort, N. (1985). Bootstrapping Cox's regression model. Technical Report 241, Stanford University, Department of Statistics.

Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* New York, Wiley

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *J. Amer Statist. Assoc.,* **96,** 458-468.

Lin, D.Y., Fleming, T.R., Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika,* **81,** 73-81.

Murphy, S.A., Rossini, A.J. and van der Vaart, A.W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.,* **92,** 968-976.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.,* **15,** 780-799.

Pettitt, A.N. (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics,* **33,** 169-175.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.,* **98,** 1001–1013.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer Verlag, New York.

Dorota M. Dabrowska
Department of Biostatistics
University of California
Los Angeles, CA 90095-1772
USA
E-mail: dorota@brahms.ph.ucla.edu

Paper received: September 2004; revised March 2005.