Harvard University
Harvard University Biostatistics Working Paper Series

Year 2003 *Paper* 7

Statistical Inference for Infinite Dimensional Parameters Via Asymptotically Pivotal Estimating Functions

Meredith A. Goldwasser* Lu Tian†
L. J. Wei‡

*Harvard School of Public Health, mgoldwas@hsph.harvard.edu
†Harvard School of Public Health, ltian@hsph.harvard.edu
‡Harvard School of Public Health, wei@hsph.harvard.edu

Statistical inference for infinite dimensional parameters via asymptotically pivotal estimating functions

By M. A. GOLDWASSER, L. TIAN and L.J. WEI
Department of Biostatistics
Harvard School of Public Health
655 Huntington Avenue
Boston, MA 02115
mgoldwas@hsph.harvard.edu ltian@hsph.harvard.edu wei@sdac.harvard.edu

Summary

Suppose that a consistent estimator for an infinite-dimensional parameter can be readily obtained via a set of estimating functions which has a 'good' local linear approximation around the true value of the parameter. However, it may be difficult to estimate the variance function of this estimator well. We show that if the set of estimating functions evaluated at the true parameter value is 'asymptotically pivotal', then the 'fiducial' distribution of the parameter can be used to approximate the distribution of this consistent estimator. We present three examples to illustrate that the corresponding inference for the parameter can be made via a simple simulation technique without involving complex, high-dimensional nonparametric density estimates.

Some key words: Confidence band; Estimating equation; Gaussian process; Pivotal quantity; Quantile regression; Survival analysis.

1. Introduction

Suppose that we are interested in making inferences about an infinite-dimensional parameter $\Theta = \{\theta(t), t \in \mathcal{I}\}$ based on a system of asymptotically unbiased estimating functions $\mathcal{S}_X(\Theta) = \{S_X(\Theta; t), t \in \mathcal{I}\}$, where $\mathcal{I}$ is an interval in R, $\theta(t)$ assumes values in R^p, and X is the observable random quantity. Often a consistent estimator $\widehat{\Theta}_X = \{\hat{\theta}_X(t), t \in \mathcal{I}\}$ for Θ_0, the true value of Θ, can be readily obtained by solving the equation

$$\mathcal{S}_X(\Theta) \simeq 0. \tag{1.1}$$

If $\mathcal{S}_X(\Theta)$ has a 'good' local linear approximation around Θ_0, the standardised $\widehat{\Theta}_X$ converges weakly to a Gaussian process (van der Vaart, 1995). However, the variance-covariance function of this limiting process can be prohibitively complex. It is rather difficult, if not impossible, to study the large sample properties of $\widehat{\Theta}_X$ analytically via this Gaussian process.

In a recent technical report from the University of Washington, J. A. Wellner and Y. Zhan provided a formal justification of Efron's bootstrap method (Efron & Tibshirani, 1993) for making inferences about Θ_0 via a specific type of estimating function, which is a sum of independent, identically distributed random quantities. In this article, we deal with a much wider class of estimating functions and derive a relatively simple resampling method for studying the properties of $\widehat{\Theta}_X$.

Suppose that $\{S_X(\Theta_0; t), t \in \mathcal{I}\}$ converges weakly to a Gaussian process $\mathcal{W} = \{W(t), t \in \mathcal{I}\}$, whose distribution may depend on Θ_0. The process $\mathcal{S}_X(\Theta_0)$ is 'asymptotically pivotal' if for any observed x of X one can generate, without knowing Θ_0, a random process $\mathcal{W}_x = \{W_x(t), t \in \mathcal{I}\}$ which converges weakly to the same limiting process $\mathcal{W}$. Let $\Theta^*_x = \{\theta^*_x(t), t \in \mathcal{I}\}$ be a random function which is a solution to the system of stochastic equations

$$\{S_x(\Theta; t) \simeq W_x(t), t \in \mathcal{I}\}. \tag{1.2}$$

Under a set of regularity conditions on $\mathcal{S}_X(\Theta)$ given in § 2, the distribution of $(\widehat{\Theta}_X - \Theta_0)$ can be approximated by the conditional distribution of $(\Theta_x^* - \widehat{\Theta}_x)$. It is interesting to note that, if Θ is a finite-dimensional parameter and the distribution of $\mathcal{W}$ is free of Θ_0, Θ_x^* generates the so-called fiducial distribution of Θ.

We use a simple example to illustrate the concept of an asymptotic pivot. Let X be a random sample $\{Y_1, \cdots, Y_n\}$ from a distribution function $\{\theta_0(t), -\infty < t < \infty\}$, let $\hat{\theta}_X(\cdot)$ be the empirical distribution function, and let $S_X(\Theta; t) = n^{1/2}\{\hat{\theta}_X(t) - \theta(t)\} = n^{-1/2}\sum_{i=1}^n\{I(Y_i \leq t) - \theta(t)\}$, where $I(\cdot)$ is the indicator function. Then $\{S_X(\Theta_0; t), t \in \mathcal{I}\}$ converges weakly to a zero-mean Gaussian process $\mathcal{W}$. Note that the covariance function of this limiting process still depends on Θ_0. Let $W_x(t) = n^{-1/2}\sum_{i=1}^n\{I(y_i \leq t) - \hat{\theta}_x(t)\}G_i$, where y is the observed value of Y, and $\{G_i, i = 1, \cdots, n\}$ is a random sample from the standard normal distribution, which is independent of the data. By the uniform consistency of $\widehat{\Theta}_x$ and the conditional multiplier central limit theorem (van der Vaart & Wellner, 1996, Lemma 2.9.5), $\mathcal{W}_x$ converges weakly to $\mathcal{W}$. The distribution of $\mathcal{W}_x$ is free of any unknown parameters; this implies that $\{S_X(\Theta_0; t)\}$ is asymptotically pivotal. It follows from (1.2) that $\theta_x^*(t) = \hat{\theta}_x(t) - n^{-1}\sum_{i=1}^n\{I(y_i \leq t) - \hat{\theta}_x(t)\}G_i$. Since the processes $(\widehat{\Theta}_X - \Theta_0)$ and $(\Theta_x^* - \widehat{\Theta}_x)$ are tight and asymptotically have the same mean and covariance function, it follows that they have the same limiting distribution.

In practice, one may generate a large number, M say, of realisations from $\mathcal{W}_x$, and for each realisation obtain a Θ_x^* via (1.2). If it is relatively easy to solve equation (1.2), inferences about Θ_0 can be made based on the empirical distribution constructed from these realisations.

In § 2, we demonstrate that the conditional distribution of $(\Theta_x^* - \widehat{\Theta}_x)$ can be used to approximate the distribution of $(\widehat{\Theta}_X - \Theta_0)$ under certain regularity conditions. In § 3, three extensive examples are given to illustrate the new proposal. For the first example, we revisit the one-sample, nonparametric estimation of the cumulative hazard function with censored data, and show empirically that the confidence intervals and bands constructed using the above approach are practically identical to those derived analytically from standard martingale theory (Fleming & Harrington, 1991, Ch. 6). In the second example, we deal with a brand new problem in predicting the percentiles of the subject-specific survival function based on the Cox regression model (Cox, 1972; Lin et al., 1994). For the third problem, we derive confidence bands for the regression coefficients under a quantile regression setting without making any parametric assumptions about the error distribution or involving nonparametric density estimates, in contrast to the existing inference procedures for quantile regression in the literature (Koenker & Machado, 1999).

For the case of a finite-dimensional parameter Θ_0, various resampling methods that perturb an estimating function or its equivalent have been proposed and justified by Arcones & Gine (1992), Parzen et al. (1994), Hu & Kalbfleisch (2000), Jin et al. (2001) and the aforementioned technical report by Wellner and Zhan.

2. Conditional distribution of Θ_x^*

Note that Θ and $\mathcal{S}_X(\Theta)$ are mappings from $\mathcal{I}$ to R^p. For a generic function $\mathcal{A}(\Theta) = \{A(\Theta; t), t \in \mathcal{I}\}$, whose value is a mapping from $\mathcal{I}$ to R^p, let $||\mathcal{A}(\Theta)||$ be $\sup_{\{t\in\mathcal{I}\}}||A(\Theta;t)||_{R^p}$, where $||\cdot||_{R^p}$ is the standard Euclidean norm for R^p. Assume that the distribution of $\mathcal{W}_x$ is generated by a random element $\mathcal{G}$ whose distribution is free of x. This implies that the unconditional distribution of $\mathcal{W}_X$ is defined under a product probability measure P of two measures P_X and $P_\mathcal{G}$, where P_X and $P_\mathcal{G}$ are generated by X and $\mathcal{G}$, respectively. For the simple example of the empirical distribution function mentioned in § 1, $\mathcal{G}$ is the random sample $\{G_i, i = 1, \cdots, n\}$ from the standard normal distribution. In this section, we also assume that both $\widehat{\Theta}_X$ and Θ_X^* are uniformly consistent estimators for Θ_0. For the three cases discussed in § 3, we show that it is relatively easy to justify the consistency of these estimators. To study the large sample properties of $\widehat{\Theta}_X$ and Θ_X^*, we further impose the following conditions.

Condition C.1. $||\mathcal{S}_X(\widehat{\Theta}_X)|| = o_{P_X}(1)$.

Condition C.2. There exists a continuously invertible, bounded linear operator D, which is a mapping from the set of bounded functions into itself, such that, for $l = 1, 2$,

$$\sup_{||\Theta_l - \Theta_0|| \leq \epsilon_n} \frac{\|\mathcal{S}_X(\Theta_2) - \mathcal{S}_X(\Theta_1) - D\{n^{1/2}(\Theta_2 - \Theta_1)\}\|}{1 + n^{1/2}\|\Theta_2 - \Theta_1\|} = o_{P_X}(1),$$

where Θ_l is a non-stochastic element, n is the sample size and $\{\epsilon_n\}$ is any sequence of constants which converges to 0.

Condition C.3. $\|\mathcal{S}_X(\Theta_X^*) - \mathcal{W}_X\| = o_P(1)$.

Conditions C.1 and C.3 provide a formal interpretation of the roots of equations (1.1) and (1.2). An estimator satisfying C.1 is called an asymptotic, generalised functional M-estimator by Bickel et al. (1993). Furthermore, Condition C.2 indicates that the set of estimating functions has a good linear approximation near the true parameter value.

With the consistency of $\widehat{\Theta}_X$, C.1 and C.2, it follows from van der Vaart's generalised version of Huber's theorem (van der Vaart, 1995) that

$$n^{1/2}(\widehat{\Theta}_X - \Theta_0) = -D^{-1}\{\mathcal{S}_X(\Theta_0)\} + o_{P_X}(1 + n^{1/2}\|\widehat{\Theta}_X - \Theta_0\|).$$

By the continuous mapping theorem, the right-hand side of the above equation converges weakly to $-D^{-1}(\mathcal{W})$. Furthermore, for $\delta > 0$ and a sequence of constants $\{\epsilon_n\}$ converging to 0, such that $n^{1/2}\epsilon_n \to \infty$, under C.1

$$\text{pr}\left(\frac{\|\mathcal{S}_X(\Theta_X^*) - D\{n^{1/2}(\Theta_X^* - \hat{\Theta}_X)\}\|}{1 + n^{1/2}\|\Theta_X^* - \hat{\Theta}_X\|} > \delta\right) \le \text{pr}\left(\|\hat{\Theta}_X - \Theta_0\| > \epsilon_n, \|\Theta_X^* - \Theta_0\| > \epsilon_n\right) +$$

$$\text{pr}\left(\sup_{\|\Theta_l - \Theta_0\| \le \epsilon_n,\, l=1,2} \frac{\|\mathcal{S}_X(\Theta_2) - \mathcal{S}_X(\Theta_1) - D\{n^{1/2}(\Theta_2 - \Theta_1)\}\|}{1 + n^{1/2}\|\Theta_2 - \Theta_1\|} > \delta\right).$$

Since $\hat{\Theta}_X$ and Θ_X^* are consistent, under C.2 the right hand side of the above inequality goes to 0, as $n \to \infty$. This implies that

$$n^{1/2}(\widehat{\Theta}_X - \Theta_X^*) = -D^{-1}\{\mathcal{S}_X(\Theta_X^*)\} + o_P(1 + n^{1/2}\|\widehat{\Theta}_X - \Theta_X^*\|) = -D^{-1}(\mathcal{W}_X) + o_P(1 + n^{1/2}\|\widehat{\Theta}_X - \Theta_X^*\|).$$

The right-hand side of the above equation converges weakly to $-D^{-1}(\mathcal{W})$ in probability with respect to the product probability measure P. It follows that

$$\sup_h \left| E\left[h\{n^{1/2}(\Theta_X^* - \widehat{\Theta}_X)\} \mid X\right] - E\left[h\{n^{1/2}(\widehat{\Theta}_X - \Theta_0)\}\right]\right| = o_{P_X}(1),$$

where $h(\cdot)$ is any uniformly bounded and Lipschitz continuous mapping and the first expectation E inside the above absolute value is conditional on the data X. Although this type of approximation is weaker than the almost sure version, it has been used frequently to justify the validity of resampling methods in the literature (Hall, 1988). For the present case, this means that, for large n, there is a reasonably high probability with respect to P_X that the distribution of $(\Theta_X^* - \widehat{\Theta}_X)$, conditional on X, provides a good approximation to the unconditional distribution of $(\widehat{\Theta}_X - \Theta_0)$. More generally, if $g(\cdot)$ is a continuously differentiable function, the distribution of $\{v_X(t)\big[g\{\hat{\theta}_X(t)\} - g\{\theta_0(t)\}\big], t \in \mathcal{I}\}$ can be approximated by the conditional distribution of $\{v_x(t)\big[g\{\theta_x^*(t)\} - g\{\hat{\theta}_x(t)\}\big], t \in \mathcal{I}\}$, where $v_X(\cdot)$ is a known, positive weight function which converges uniformly to a deterministic function, as $n \to \infty$. Inferences about $g\{\theta_0(t)\}$ can then be made based on the empirical distribution of $\{v_x(t)\big[g\{\theta_x^*(t)\} - g\{\hat{\theta}_x(t)\}\big], t \in \mathcal{I}\}$.

3. Examples

3.1. *Estimation of the cumulative hazard function*

Let $T^{(0)}$ be a failure time with the cumulative hazard function $\Theta_0 = \{\theta_0(t), t \in \mathcal{I} \subset [0, \infty)\}$. One can only observe $T = \min(T^{(0)}, C)$ and $\Delta = I(T^{(0)} \leq C)$, where C is the censoring variable. The data consist of n independent, identical copies of (T, Δ); that is, $X = \{(T_i, \Delta_i), i = 1, \cdots, n\}$. To estimate $\{\theta_0(t), t \in \mathcal{I} = [t_1, t_2]\}$, where t_1 and t_2 are predetermined constants such that both $\mathrm{pr}(T < t_1)$ and $\mathrm{pr}(T > t_2)$ are positive, consider the system of estimating functions

$$S_X(\Theta; t) = n^{-1/2} \sum_{i=1}^{n} \{N_i(t) - \int_0^t I(T_i \geq s) d\theta(s)\},$$

where $N_i(t) = \Delta_i I(T_i \leq t)$. Then

$$\hat{\theta}_X(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} I(T_j \geq s)},$$

which is the well-known Nelson-Aalen estimator. By the martingale central limit theorem (Fleming & Harrington, 1991, Ch. 5; Andersen et al., 1993), $\mathcal{S}_X(\Theta_0)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}$. Note that the distribution of $\mathcal{W}$ depends on the unknown cumulative hazard function. Let $\mathcal{W}_X(t) = n^{-1/2} \sum_{i=1}^{n} N_i(t) G_i$, where $\{G_i, i = 1, \cdots, n\}$ is a random sample from the standard normal distribution, which is independent of the data. Note that the variance of $W_x(t)$ is $n^{-1} \sum_{i=1}^{n} N_i(t)$, which converges to the variance of $W(t)$. Furthermore, it follows from Lin et al. (1993) that, conditional on the data x, $\mathcal{W}_x$ converges weakly to the same limiting process $\mathcal{W}$, and can be generated without knowing Θ_0. This implies that $\mathcal{S}_X(\Theta_0)$ is asymptotically pivotal.

The root of equation (1.2) is

$$\theta^*_X(t) = \sum_{i=1}^{n} \frac{\Delta_i I(T_i \leq t)(1 - G_i)}{\sum_{j=1}^{n} I(T_j \geq T_i)}.$$

Using the martingale central limit theorem again, one can easily show that both $\widehat{\Theta}_X$ and Θ^*_X are uniformly consistent for $\{\theta_0(t), t \in \mathcal{I}\}$. Since $S_x(\widehat{\Theta}_x; t)$ and $S_x(\Theta^*_x; t) - W_x(t)$ are exactly zero for $t \in \mathcal{I}$, Conditions C.1 and C.3 are satisfied. Furthermore, it follows from Example 1 of § 5 in the technical report by Wellner and Zhan that Condition C.2 is satisfied. Therefore, the conditional distribution of $\{g(\Theta_x^*) - g(\widehat{\Theta}_x)\}$ is a good approximation to the distribution of $\{g(\widehat{\Theta}_X) - g(\Theta_0)\}$.

The above approximation can be used to derive confidence bands for $g(\Theta_0)$. A $(1-\alpha)$ confidence band for $g\{\theta_0(t)\}$ is

$$g\{\hat{\theta}_X(t)\} \pm c_\alpha v_X^{-1}(t), \tag{3.1}$$

where c_α is a cut-off point such that

$$\text{pr}_X\left(\sup_{t\in\mathcal{I}}\left[v_X(t)\left|g\{\hat{\theta}_X(t)\} - g\{\theta_0(t)\}\right|\right] \le c_\alpha\right) = 1-\alpha. \tag{3.2}$$

The c_α can be approximated by c_α^*, which is obtained from

$$\text{pr}_{\mathcal{G}}\left(\sup_{t\in\mathcal{I}}\left[v_x(t)\left|g\{\theta_x^*(t)\} - g\{\hat{\theta}_x(t)\}\right|\right] \le c_\alpha^*\right) = 1-\alpha. \tag{3.3}$$

The c_α^* can be estimated empirically from M realisations from $g(\Theta_x^*)$. To obtain the so-called equal-precision confidence band for $g\{\theta_0(t)\}$, we let $v_x^{-1}(t)$ be an estimate of the standard error of $g\{\hat{\theta}_X(t)\}$. A good candidate for this estimate is the sample standard error estimate based on those M independent realisations of $\{g\{\theta_x^*(t)\}, t \in \mathcal{I}\}$. Note that the pointwise confidence intervals for $g\{\theta_0(t)\}$ can also be obtained using this specific standard error estimate with c_α in (3.1) replaced by the $100(1-\alpha/2)$th percentile of the standard normal distribution.

We illustrate the above simultaneous confidence interval estimation procedure using the well-known Mayo Clinic dataset of patients with primary biliary cirrhosis disease. The data are given in Appendix D of Fleming & Harrington (1991). Although this was a placebo-controlled trial to evaluate the drug D-penicillamine, the drug was not found to have any significant benefit on patient survival. Therefore, in our analysis we use the failure times from all 418 patients who met the eligibility criteria of the study, as used in Fleming & Harrington (1991, Ch. 6). To construct interval estimates for the survival function of $T^{(0)}$, we let $g(x) = \exp(-x)$ in (3.1)-(3.3). In order to compare our results with those in the literature, we let the weight function $v_x^{-1}(t)$ be $\exp\{-\hat{\theta}_x(t)\}\{1 + n\hat{\sigma}_x^2(t)\}$, where $\hat{\sigma}_X(t)$ is an estimated standard error of $\hat{\theta}_X(t)$, which corresponds to the Hall-Wellner confidence band (Fleming & Harrington, 1991). Note that the Hall-Wellner band is one of very few confidence bands for the survival function which can be constructed analytically. On the other hand, our technique is applicable to the case with any weight function $v_X(\cdot)$. Figure 1 shows the estimate $\exp\{-\hat{\theta}_x(t)\}$ for the survival function, given by the solid line, the 0.95 pointwise confidence intervals, given by the dotted lines, and a 0.95 confidence band (3.1), given by the dashed lines. Our band is practically identical to the Hall-Wellner band displayed in Fig. 6.3.5a of Fleming & Harrington (1991). For this example, $\mathcal{I} = [t_1, t_2] = [0.1, 11.5]$, whose bounds correspond to the first and last observed death times in years. The intervals and band are constructed with $M = 1000$ realisations of Θ_x^*.

3.2. *Predicting the percentile process of the subject-specific failure time distribution*

In this section we consider a more complex case using censored failure time data. Here, for each failure time $T^{(0)}$, there is a covariate vector Z. Assume that the data X consist of n independent, identical copies $\{(T_i, \Delta_i, Z_i), i = 1, \cdots, n\}$ of (T, Δ, Z). Furthermore, assume that $T^{(0)}$ is related to Z via the Cox proportional hazards model, with the vector of regression coefficients β_0 (Cox, 1972) and a continuous and positive nuisance hazard function. Let $\hat{\beta}$ be the maximum partial likelihood estimator for β_0. Now, suppose that we are interested in predicting simultaneously the tth percentiles, $\theta_0(t), t \in \mathcal{I} = [t_1, t_2] \subset [0, 1]$, of the failure time distribution for subjects with a specific covariate vector z, where both $\text{pr}\{T < \theta_0(t_1)\}$ and $\text{pr}\{T > \theta_0(t_2)\}$ are positive. First, let $\Lambda(s)$ be the underlying cumulative hazard function for the proportional hazards model at time point s. The Breslow estimator for $\Lambda(s)$ is

$$\hat{\Lambda}(s) = \sum_{i=1}^{n} \frac{I(T_i \le s)\Delta_i}{\sum_{j=1}^{n} I(T_j \ge T_i) e^{\hat{\beta}' Z_j}} \quad \text{(Breslow, 1972)}.$$

To estimate $\Theta_0 = \{\theta_0(t), t \in \mathcal{I}\}$, consider the estimating function

$$S_X(\Theta; t) = n^{1/2}[\hat{\Lambda}\{\theta(t)\} - \Lambda\{\theta_0(t)\}] = n^{1/2}[\hat{\Lambda}\{\theta(t)\} + \log(1-t)].$$

Since the Breslow estimate is a step function, the estimating function $\mathcal{S}_X(\Theta; t)$ is not continuous in $\theta(t)$. To obtain a well-defined root $\hat{\theta}_X(t)$ of the estimating equation (1.1), one may replace the Breslow estimate with a continuous process by connecting the midpoints of every two consecutive steps with a straight line. This simple modification has the same large sample properties as the Breslow estimator. In the Appendix, we show that $\widehat{\Theta}_X$ is uniformly consistent for $\{\theta_0(t), t \in \mathcal{I}\}$.

It follows from Andersen & Gill (1982) and Lin et al. (1994) that $\mathcal{S}_X(\Theta_0)$ is asymptotically pivotal: $S_X(\Theta_0, t)$ is asymptotically equivalent to the process

$$\tilde{W}\{\theta_0(t)\} = n^{-1/2} \sum_{i=1}^{n} \left[\int_0^{\theta_0(t)} \frac{dM_i(u)}{V^{(0)}(\beta_0, u)} + H'\{\theta_0(t)\} Q^{-1} \int_0^{\infty} \left\{ Z_i - \frac{V^{(1)}(\beta_0, u)}{V^{(0)}(\beta_0, u)} \right\} dM_i(u) \right],$$

where

$$M_i(u) = N_i(u) - \int_0^u I(T_i \ge s) e^{\beta_0' Z_i} d\Lambda(s),$$

$$V^{(r)}(\beta, u) = n^{-1} \sum_{i=1}^{n} I(T_i \ge u) e^{\beta' Z_i} Z_i^{\otimes r}, \; r = 0, 1, 2,$$

$$Z_i^{\otimes 0} = 1, \; Z_i^{\otimes 1} = Z_i, \; Z_i^{\otimes 2} = Z_i Z_i'$$

$$H(u) = -\int_0^u \frac{V^{(1)}(\beta_0, s)}{V^{(0)}(\beta_0, s)} d\Lambda(s),$$

$$Q = \int_0^{\infty} \left[\frac{V^{(2)}(\beta_0, u)}{V^{(0)}(\beta_0, u)} - \left\{ \frac{V^{(1)}(\beta_0, u)}{V^{(0)}(\beta_0, u)} \right\}^{\otimes 2} \right] V^{(0)}(\beta_0, u) d\Lambda(u).$$

It can be shown via the standard martingale central limit theorem that the above process converges weakly to a zero-mean Gaussian process $\mathcal{W}$.

We now replace $\{M_i(u), i = 1, \cdots, n\}$ in $\tilde{W}\{\theta_0(t)\}$ with $\{N_i(u)G_i\}$, where $\{G_i, i = 1, \cdots, n\}$ is a random sample from the standard normal distribution, which is independent of the data, and also replace other unknown quantities in $\tilde{W}\{\theta_0(t)\}$ with their respective estimators. This results in the process

$$W_X(t) = n^{-1/2} \sum_{i=1}^{n} \left[(\hat{v}_i^{(0)})^{-1} I\{T_i \leq \hat{\theta}_X(t)\}\Delta_i + \hat{H}'(t)\hat{Q}^{-1}\Delta_i\{Z_i - \bar{Z}_i\} \right] G_i, \qquad (3.4)$$

$$\hat{v}_i^{(r)} = n^{-1} \sum_{j=1}^{n} I(T_j \geq T_i) e^{\hat{\beta}' Z_j} Z_j^{\otimes r}, \; r = 0, 1, 2,$$

$$\bar{Z}_i = (\hat{v}_i^{(0)})^{-1} \hat{v}_i^{(1)}, \quad \hat{H}(t) = -n^{-1} \sum_{i=1}^{n} I\{T_i \leq \hat{\theta}_X(t)\}\Delta_i \bar{Z}_i (\hat{v}_i^{(0)})^{-1},$$

$$\hat{Q} = n^{-1} \sum_{i=1}^{n} \Delta_i \{\hat{v}_i^{(2)} (\hat{v}_i^{(0)})^{-1} - \bar{Z}_i \bar{Z}_i'\}.$$

It follows from the same argument as in Lin et al. (1994) that, conditional on $X = x$, the distribution of $\mathcal{W}_x$ converges weakly to $\mathcal{W}$.

In the Appendix, we show that Θ_X^* is consistent and Conditions C.1-C.3 are satisfied. It follows that the distribution of the process $(\Theta_x^* - \widehat{\Theta}_x)$ is a good approximation to that of $(\widehat{\Theta}_X - \Theta_0)$. A $(1 - \alpha)$ confidence band for Θ_0 can then be obtained via (3.1)-(3.3). Since, for any fixed t, $S_X(\Theta; t) = n^{1/2}[\hat{\Lambda}\{\theta(t)\} + \log(1 - t)]$ is an increasing function of $\theta(t)$, solving (1.1) and (1.2) numerically is a trivial task. To predict the percentiles of the failure time distribution for a specific covariate vector z, we simply replace $\{Z_i, i = 1, \cdots, n\}$ by $\{Z_i - z, i = 1, \cdots, n\}$ and use this modified dataset to obtain $\widehat{\Theta}_X$ and Θ_X^*.

We continue to use the Mayo Clinic data from § 3.1, but include covariate values from each patient to illustrate our simultaneous interval estimation procedure. Dickson et al. (1989) and Fleming & Harrington (1991) analysed this dataset extensively and established a Cox model with five baseline covariates to predict patient survival with primary biliary cirrhosis disease. The five covariates are age, log(albumin), log(bilirubin), oedema and log(protime).

Based on this model, Lin et al. (1994) constructed confidence bands for the subject-specific cumulative hazard function by generating a large number of realisations from a process similar to $\{W_x(t)\}$ defined in (3.4). For a hypothetical patient aged 51 years, with 3.4 gm/dl serum albumin, 1.8 mg/pl serum bilirubin, no oedema and 10.74 seconds of protime, Lin et al. (1994) presented a 0.95 confidence band for this patient's survival function (Lin et al., 1994, Fig. 1).

Now we construct confidence bands for the corresponding percentile function $\{\theta(t), t \in \mathcal{I}\}$ of this patient's failure time distribution. The last observed death time in the dataset is 11.47 and its estimated survival probability is 0.3. Thus, we construct a simultaneous band for the tth percentiles for $t \in \mathcal{I} = [t_1, t_2] = [0.05, 0.6]$. Figure 2 shows the point estimate $\hat{\theta}_x(t)$, given by the solid line, 0.95 pointwise confidence intervals, given by the dotted lines, and an 0.95 equal-precision confidence band, given by the outside dashed lines. Here, $g(\cdot)$ in (3.1)-(3.3) is the identity function and $v_x^{-1}(t)$ is the estimated standard error of $\hat{\theta}_X(t)$. The c_α and $v_x^{-1}(t)$ were estimated using $M = 1000$ realisations of $\{\theta_x^*(t)\}$. Note that the point and interval estimates were approximated numerically by discretising $\{\hat{\theta}_x(t)\}$ and $\{\theta_x^*(t)\}$ for t from 0.05 to 0.6 with increments of 0.001.

The plots in Fig. 2 are quite informative. For example, the estimated median failure time is 9.3 years, with a 0.95 confidence interval of $(8.2, 10.3)$ and the corresponding band of $(7.4, 11.2)$. The estimated lower quartile failure time for this patient is 5.6 years, with a 0.95 confidence interval of (4.6, 6.6) and the corresponding band of $(3.9, 7.4)$.

3.3. *Simultaneous inferences for the heteroscedastic quantile regression model*

In this section, we consider the case in which a fully observed, continuous response variable T is related to its $p \times 1$ covariate vector Z via a quantile regression model (Koenker

& Bassett, 1978). The 100tth percentile of T is $\theta_0(t)'Z$, where $\{\theta_0(t)\}$ is a function from $\mathcal{I} \subset [0,1]$ to R^p. Note that the first component of Z is 1 and that no parametric structure on the distribution of the error terms is assumed, nor is there any parametric assumption about the joint distribution of $\{T - \theta_0(t)'Z\}$ and Z. We are interested in making inferences about $\Theta_0 = \{\theta_0(t), t \in \mathcal{I}\}$ and also in predicting the tth percentiles of the distribution of T for a future subject. Recently, Koenker & Machado (1999) derived a novel inference procedure for the quantile regression model, but with a specific parametric assumption about the variance of the error distribution given Z (Koenker & Machado, 1999, Model (12)).

To estimate the regression coefficient function $\{\theta_0(t)\}$, consider the system of estimating functions

$$S_X(\Theta; t) = n^{-1/2} \sum_{i=1}^{n} Z_i \{I(T_i - \theta(t)'Z_i < 0) - t\}. \tag{3.5}$$

Although the estimating function (3.5) is not continuous in $\theta(t)$, a consistent root $\widehat{\Theta}_X$ of the corresponding equation can be obtained by minimising the convex function $\sum_{i=1}^{n} \rho_t\{T_i - \theta(t)'Z_i\}$ by the standard linear programming technique, where $\rho_t(u) = u\{t - I(u \leq 0)\}$ (Koenker & Bassett, 1978; Koenker & d'Orey, 1987).

If $\{\theta_0(t), t \in \mathcal{I}\}$ is continuously differentiable, it follows from the arguments in Lai & Ying (1988) that $\{S_X(\Theta_0; t)\}$ converges weakly to a Gaussian process. Let

$$W_X(t) = n^{-1/2} \sum_{i=1}^{n} Z_i \{I(T_i - \hat{\theta}_X(t)'Z_i < 0) - t\} G_i.$$

By Lemmas 2.6.15 and 2.6.18 of Van der Vaart & Wellner (1996), it follows that the class of functions of $(z, y, g) : \{zI(y - \theta'z < 0)g\}$, indexed by θ in a compact set, is Donsker; see Definition 2.1.1 of Van der Vaart & Wellner (1996). Therefore, one can replace $\widehat{\Theta}_X$ in the above $W_X(t)$ by Θ_0 without affecting the resulting limiting distribution. This, coupled with the conditional multiplier central limit theorem, implies that the distribution of $\mathcal{W}_x$ can be used to approximate the limiting distribution of $\mathcal{S}_X(\Theta_0)$.

In the Appendix, we show the consistency of $\widehat{\Theta}_X$ and Θ^*_X and the validity of Conditions C.1-C.3 under rather mild assumptions that the covariate vector Z is bounded, $E(ZZ')$ is a strictly positive definite matrix, and the density function $f_t(\cdot; Z)$ of $\{T - \theta_0(t)'Z\}$ given Z is continuous and positive on its entire support. It follows that the distribution of $(\widehat{\Theta}_X - \Theta_0)$ can be approximated by the conditional distribution of $(\Theta^*_x - \widehat{\Theta}_x)$. Note that, in practice, the solution Θ^*_x of equation (1.2) can be obtained by minimising the convex function

$$\sum_{i=1}^{n} \rho_t\{T_i - \theta(t)'Z_i\} + \rho_t\{T_{n+1} - \theta(t)'Z_{n+1}(t)\}$$

with the linear programming technique, where T_{n+1} is an extremely large, artificially generated number and $Z_{n+1}(t) = n^{1/2}W_x(t)/t$. This minimisation procedure has been successfully used by Parzen et al. (1994) for a finite-dimensional parameter under a median regression model. Suppose that we are interested in constructing confidence bands for a specific regression coefficient of $\{\theta_0(t), t \in \mathcal{I}\}$. Such bands can be obtained via formulae (3.1)-(3.3). Moreover, for predicting the percentile process $\{\theta_0(t)'z, t \in \mathcal{I}\}$ of T for a future subject with the covariate vector z, this process can be consistently estimated by $\{\hat{\theta}_X(t)'z\}$. The distribution of $\{\hat{\theta}_X(t)'z - \theta_0(t)'z, t \in \mathcal{I}\}$ can be approximated by that of $\{\theta^*_x(t)'z - \hat{\theta}_x(t)'z, t \in \mathcal{I}\}$, and a $(1-\alpha)$ confidence band for $\theta_0(t)'z$ can then be obtained by replacing each $\theta(t)$ quantity in (3.1)-(3.3) with its corresponding $\theta(t)'z$ counterpart.

We illustrate the above method using a dataset from a recent HIV clinical trial conducted by the AIDS Clinical Trials Group (Hammer et al., 1997). This randomised, double-blind trial compared a three-drug combination therapy of indinavir, zidovudine and lamivudine, $n = 423$, to a two-drug combination of zidovudine and lamivudine, $n = 429$. For illustration, we consider an additive, heteroscedastic quantile regression model with week 24 CD4 cell count as the response T, and the covariate vector Z consisting of the baseline CD4 count, a treatment indicator equal to 1 for the three-drug group, months of prior zidovudine treatment, age, a gender indicator equal to 1 for male, and two race indicators, equal to (1,0) for 'black' and equal to (0,1) for 'Hispanic'. In our analysis, each covariate was standardised by subtracting its empirical mean and dividing by its sample standard deviation.

Figure 3 gives the point estimates, denoted by the solid lines, and 0.95 equal-precision confidence bands, represented by the dashed lines, for each regression coefficient of the vector $\theta_0(t)$ for $t \in [0.1, 0.9]$. These bands were constructed from M=1000 realisations of Θ_x^*. Note that $\hat{\theta}_x(t)$ and $\theta_x^*(t)$ were approximated numerically by discretising t between 0.1 and 0.9 with increments of 0.01. From Fig. 3 it appears that the standard location-shift, linear regression model does not fit the data well. The coefficients of the baseline CD4 count and the treatment indicator increase significantly over $0.1 \leq t \leq 0.9$.

We also used this fitted quantile regression model to predict the tth percentiles of the distribution of week 24 CD4 count for a 38-year-old, white male patient with a baseline CD4 count of 71 and twenty-two months of prior zidovudine usage. Figure 4(a) shows the point estimates, given by the solid line, and 0.95 equal-precision confidence band, given by the dashed lines, for $\{\theta_0(t)'z\}$ when the patient is treated with the three-drug combination. In Fig. 4(b), we present the corresponding estimates when that patient is treated with the two-drug combination. Together, Figs. 4(a) and (b) indicate substantial heterogeneity of treatment effects between subjects with the above covariate values. For example, the estimated 20th percentile of week 24 CD4 count for the above subject in the two-drug combination group is 59, with 0.95 confidence band (48, 69), and in the three-drug combination group the estimate is 89, with 0.95 confidence band (72, 106), indicating a significant but relatively small treatment effect. In comparison, the estimated median week 24 CD4 count in the two-drug combination group is 79, with 0.95 confidence band (70, 88), whereas the estimated median in the three-drug combination group is 144, with 0.95 confidence band (123, 165).

4. Discussion

When one is interested in drawing inferences about Θ_0 based on a non-smooth estimating function $\mathcal{S}_X(\Theta)$, a common approach is to 'perturb' $\mathcal{S}_x(\Theta)$ directly via, for example, the standard bootstrap method, and to solve the resulting estimating equation $\mathcal{S}_x^*(\Theta) = 0$ repeatedly, thereby obtaining a large number of Θ which can be used to approximate the unconditional distribution of $(\widehat{\Theta} - \Theta_0)$ (Efron & Tibshirani, 1993; Hu & Kalbfleisch, 2000). In their recent technical report, Wellner and Zhan elegantly justified the validity of such an approximation when $\mathcal{S}_X(\Theta)$ is a sum of independent and identically distributed quantities. Our resampling procedure, which does not alter $\mathcal{S}_x(\Theta)$, rather perturbs $\mathcal{W}_x$, whose distribution is an approximation to that of $\mathcal{S}_X(\Theta_0)$, and solves the equation $\mathcal{S}_x(\Theta) = W_x$ repeatedly. For the cases in which the bootstrap is applicable, we find empirically that the results from both approaches are practically identical.

It is important to note that the new resampling method works well in practice when there are efficient numerical algorithms available for solving (1.1) and (1.2).

Acknowledgement

The authors are grateful to the referee, the associate editor and the editor for the extremely useful and detailed comments on the paper. This research was partially supported by grants from the U.S. National Institutes of Health.

Appendix

Justification of conditions

Justification of conditions for § 3.2. Since we obtained the estimator using a continuous version of the estimating function, C.1 and C.3 are trivially satisfied. Next, if the estimator is not uniformly consistent, we show that $\|\mathcal{S}_X(\widehat{\Theta}_X)\| \to \infty$. To this end, suppose that there is Θ such that $\|\Theta - \Theta_0\| > \delta > 0$. Then we can find a $t \in \mathcal{I}$ such that $|\theta(t) - \theta_0(t)| \geq \delta > 0$. It follows that

$$\begin{aligned} |S_X(\Theta, t)| &= \left| n^{1/2}[\hat{\Lambda}\{\theta(t)\} - \Lambda\{\theta(t)\}] + n^{1/2}[\Lambda\{\theta(t)\} - \Lambda\{\theta_0(t)\}] \right| \\ &\geq -\left| n^{1/2}[\hat{\Lambda}\{\theta(t)\} - \Lambda\{\theta(t)\}] \right| + n^{1/2} c\delta = O_{P_X}(n^{1/2}), \end{aligned}$$

where $c > 0$ and the hazard function $\lambda(t)$ is assumed to be bounded away from 0. Thus, $\|\mathcal{S}_X(\Theta)\| \geq O_{P_X}(n^{1/2}) \to \infty$. Since $\|\mathcal{S}_X(\widehat{\Theta}_X)\| = 0$, this implies that $\widehat{\Theta}_X$ is uniformly consistent with respect to probability measure P_X. Using similar arguments, one can show that Θ_X^* is also uniformly consistent for Θ_0 with respect to probability measure P.

Lastly, we show that Condition C.2 is satisfied. It follows from the tightness of the process $n^{1/2}\{\hat{\Lambda}(s) - \Lambda(s)\}$ that, for any sequence of $\{\epsilon_n\} \to 0$,

$$\sup_{|s_2 - s_1| \leq \epsilon_n} |\hat{\Lambda}(s_2) - \hat{\Lambda}(s_1) - \Lambda(s_2) + \Lambda(s_1)| = o_{P_X}(n^{-1/2}).$$

Therefore, if $\|\Theta_l - \Theta_0\| \leq \epsilon_n$, $l = 1, 2$,

$$\sup_{t \in \mathcal{I}} \left| \hat{\Lambda}\{\theta_2(t)\} - \hat{\Lambda}\{\theta_1(t)\} - \Lambda\{\theta_2(t)\} + \Lambda\{\theta_1(t)\} \right| = o_{P_X}(n^{-1/2}).$$

Furthermore, since

$$\sup_{t \in \mathcal{I}} \left| \Lambda\{\theta_2(t)\} - \Lambda\{\theta_1(t)\} - \lambda\{\theta_0(t)\}\{\theta_2(t) - \theta_1(t)\} \right| = o(\|(\Theta_2 - \Theta_1)\|),$$

it follows that for $l = 1, 2$,

$$\sup_{\|\Theta_l - \Theta_0\| \leq \epsilon_n} \frac{\|n^{1/2}\{\hat{\Lambda}(\Theta_2) - \hat{\Lambda}(\Theta_1)\} - \lambda(\Theta_0) n^{1/2}(\Theta_2 - \Theta_1)|}{1 + n^{1/2}\|\Theta_2 - \Theta_1\|} = o_{P_X}(1).$$

Justification of conditions for § 3.3. To show that C.1 is satisfied, recall that we define $\hat{\theta}_X(t) = \arg\min \sum_{i=1}^n \rho_t\{T_i - \theta(t)'Z_i\}$. With some elementary algebraic arguments, one can show that, for any fixed $t \in \mathcal{I}$,

$$||S_X(\widehat{\Theta};t)||_{R^p} \leq cn^{-1/2} \sum_{i=1}^n I\{T_i - \hat{\theta}(t)'Z_i = 0\},$$

where c is a positive constant. Furthermore, since it is assumed that $E(ZZ')$ is strictly positive definite and $f_t(\cdot;Z)$ is continuous, the right-hand side of the above inequality is $o_{P_X}(1)$. Condition C.3 can be verified with the same argument.

Next, we show the consistency of $\widehat{\Theta}_X$. Suppose that there is a Θ such that $\|\Theta - \Theta_0\| > \delta > 0$. Then we can find a $t \in \mathcal{I}$ such that $|\theta(t) - \theta_0(t)| \geq \delta > 0$. For this specific t,

$$\begin{aligned} |S_X(\Theta;t)| &= |S_X(\Theta;t) - S_X(\Theta_0;t) + S_X(\Theta_0;t)| \\ &\geq |n^{-1/2} \sum_{i=1}^n Z_i \left[I\{T_i - \theta(t)'Z_i \leq 0\} - I\{T_i - \theta_0(t)'Z_i \leq 0\}\right]| \\ &\quad -|n^{-1/2} \sum_{i=1}^n Z_i \left[I\{T_i - \theta_0(t)'Z_i \leq 0\} - t\right]|. \end{aligned}$$

By the central limit theorem, the second term on the right-hand side of the above inequality is $O_p(1)$. Furthermore, using the central limit theorem again, one can show that the corresponding first term is $\left|n^{-1/2} \sum_{i=1}^n f_t(0;Z_i)Z_iZ_i' \{\theta(t) - \theta_0(t)\}\right| + o_{P_X}(1)$, which is greater than $n^{1/2}c\delta + o_{P_X}(1)$, where $c > 0$. It follows that $\|\mathcal{S}_X(\Theta)\| \geq O_{P_X}(n^{1/2}) \to \infty$. Therefore, with C.1, $\widehat{\Theta}_X$ is uniformly consistent with respect to probability measure P_X. Using similar arguments, one can show the uniform consistency of Θ_X^* with respect to probability measure P.

Lastly, C.2 can be justified using the results presented by Lai & Ying (1988). For any $\epsilon_n \to 0$, $l = 1, 2$,

$$\sup_{\|\Theta_l - \Theta_0\| \leq \epsilon_n} \|\mathcal{S}_X(\Theta_2) - \mathcal{S}_X(\Theta_1) - E\{\mathcal{S}_X(\Theta_2)\} + E\{\mathcal{S}_X(\Theta_1)\}\| = o_{P_X}(1).$$

Moreover,

$$E\{S_X(\Theta_2;t)\}-E\{S_X(\Theta_1;t)\} = \left\{n^{-1}\sum_{i=1}^{n} f_t(0;Z_i)Z_iZ_i'\right\} n^{1/2}\{\theta_2(t)-\theta_1(t)\}+o(n^{1/2}\|\Theta_2-\Theta_1\|).$$

Therefore, for any $\epsilon_n \to 0$, $l = 1, 2$,

$$\sup_{\|\Theta_l-\Theta_0\|\leq\epsilon_n} \frac{\|\mathcal{S}_X(\Theta_2)-\mathcal{S}_X(\Theta_1)-\{n^{-1}\sum_{i=1}^{n} f_t(0;Z_i)Z_iZ_i'\} n^{1/2}(\Theta_2-\Theta_1)\|}{1+n^{1/2}\|\Theta_2-\Theta_1\|} = o_{P_X}(1).$$

References

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-20.

Arcones, M. & Gine, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the Limits of Bootstrap*, Ed. R. LePage and L. Billard, pp. 13-47. Wiley, New York.

Bickel, P., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins Univ. Press.

Breslow, N. (1972). Discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **34**, 216-7.

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

Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis.* New York: Wiley Press.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals (with Discussion). *Ann. Statist.* **16**, 927-85.

HAMMER, S. M., SQUIRES, K. E., HUGHES, M. D., GRIMES, J. M., DEMETER L. M., CURRIER, J. S., ERON J. J., FEINBERG, J. E., BALFOUR H. H., DEYTON L. R., CHODAKEWITZ, J. A. & FISCHL, M. A. (1997). A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and cd4 cell counts of 200 per cubic millimeter or less. *New Engl. J. Med.* **337**, 725-33.

HU, F. & KALBFLEISCH, J. D. (2000). The estimating function bootstrap (with Discussion). *Can. J. Statist.* **28**, 449-99.

JIN, Z., YING, Z. & WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-90.

KOENKER, R. & BASSETT, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

KOENKER, R. & D'OREY, V. (1987). Algorithm AS 229: Computing regression quantiles. *Appl. Statist.* **36**, 383-93.

KOENKER, R. & MACHADO, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Am. Statist. Assoc.* **94**, 1296-310.

LAI, T. L. & YING, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Mult. Anal.* **27**, 334-58.

LIN, D. Y., FLEMING, T. R. & WEI, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73-81.

LIN, D. Y., WEI, L. J. & YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-72.

PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-50.

VAN DER VAART, A. W. (1995). Efficiency of infinite dimensional M-estimators. *Statist. Neer.* **49**, 9-30.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics.* New York: Springer-Verlag Press.

YING, Z., JUNG, S. H. & WEI, L. J. (1995). Survival analysis with median regression models. *J. Am. Statist. Assoc.* **90**, 178-84.

Figure 1: Estimation of the survival function with the Mayo Clinic data. Point estimate shown by the middle solid line; 0.95 pointwise intervals, the dotted lines; 0.95 simultaneous band, the dashed lines.

Figure 2: Prediction of the percentile process of the failure time distribution for a patient aged 51 years, with 3.4 gm/dl albumin, 1.8 mg/pl bilirubin, 10.7 seconds of protime, and no oedema. Point estimate, the middle solid line; 0.95 pointwise confidence intervals, the dotted lines; 0.95 confidence band, the outside dashed lines.

Figure 3: Simultaneous estimation for quantile regression coefficients with the HIV dataset. Point estimates, the solid lines; 0.95 confidence bands, the dashed lines. (a) baseline CD4, (b) treatment, (c) months of prior zidovudine use, (d) age, (e) gender, (f) black vs. white, (g) Hispanic vs. white

Figure 4: Simultaneous prediction of percentiles of the week 24 CD4 distribution for a white male subject, aged 38.45 years, with 70.75 baseline CD4 cells/mm^3, and 22 months of prior zidovudine use. Point estimates, the middle solid lines; 0.95 confidence bands, the dashed lines. (a) for three-drug combination therapy, (b) for two-drug combination therapy

Figure 1:

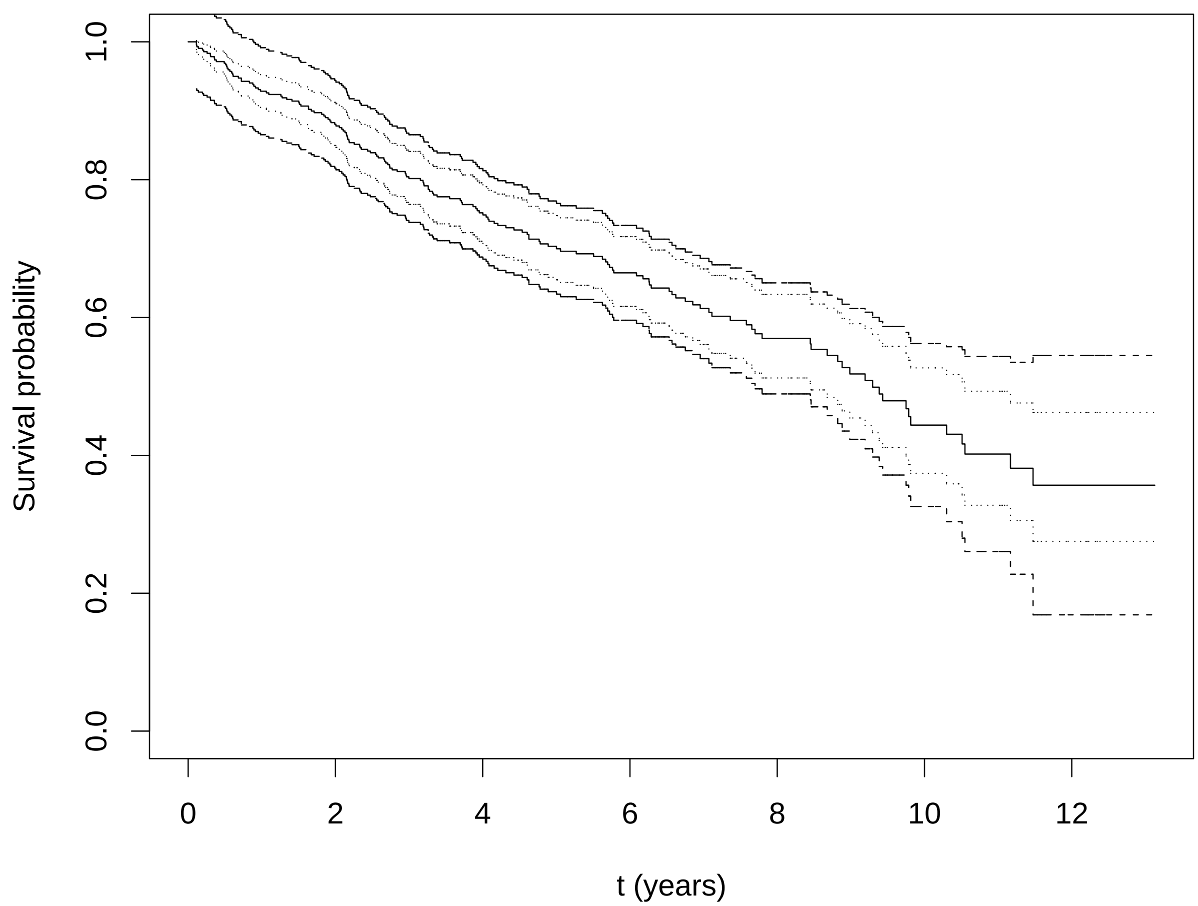


Figure 2:

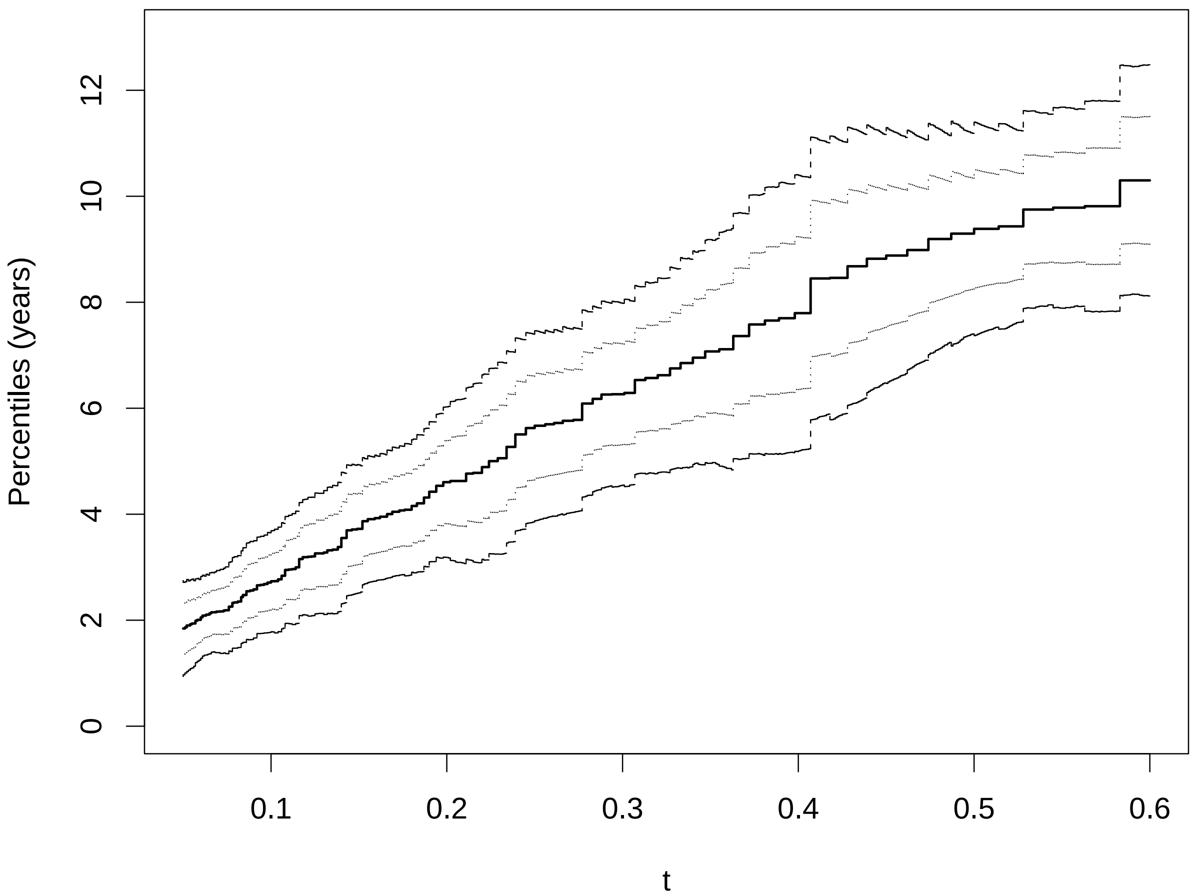


Figure 3:

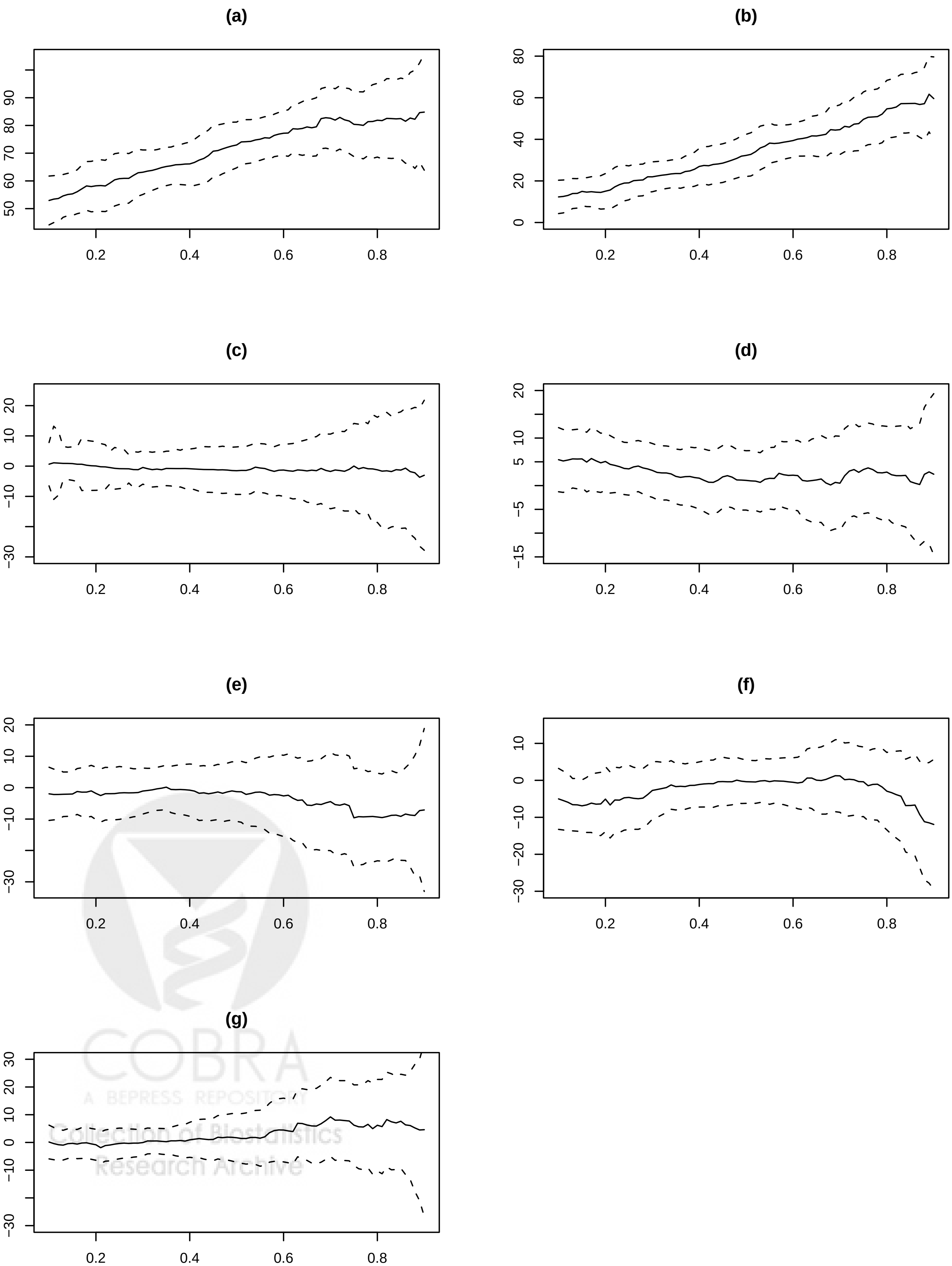
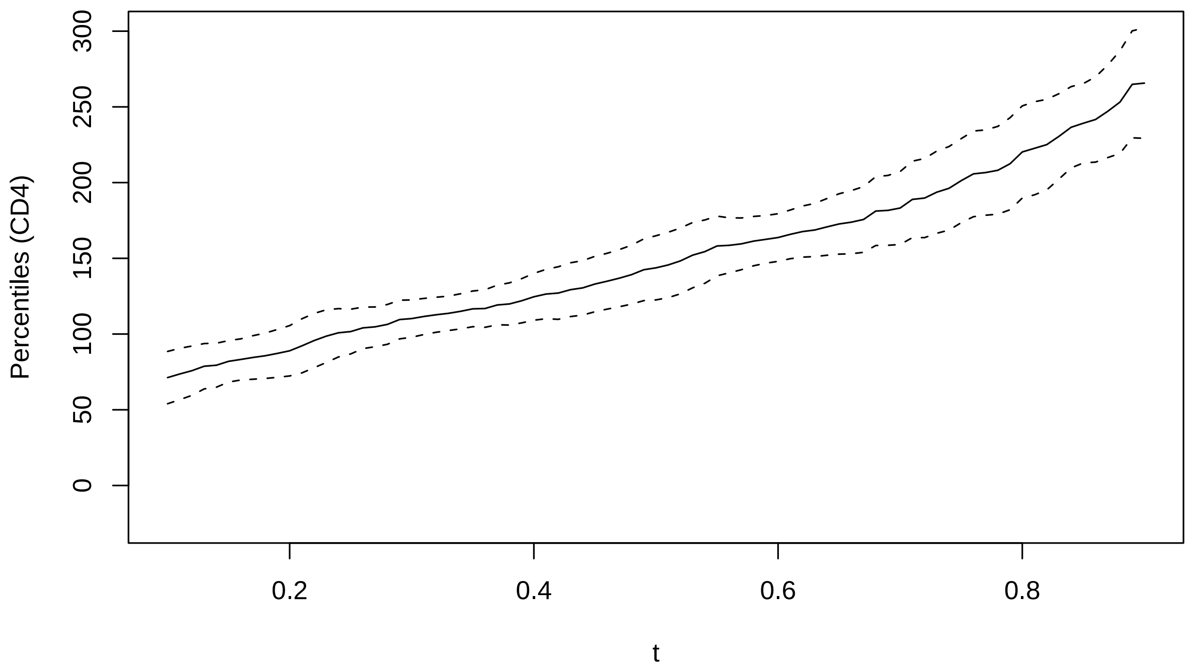

(a)
50
60
70
80
90
0.2
0.4
0.6
0.8
(b)
0
20
40
60
80
0.2
0.4
0.6
0.8
(c)
−30
−10
0
10
20
0.2
0.4
0.6
0.8
(d)
−15
−5
5
10
20
0.2
0.4
0.6
0.8
(e)
−30
−10
0
10
20
0.2
0.4
0.6
0.8
(f)
−30
−20
−10
0
10
0.2
0.4
0.6
0.8
(g)
−30
−10
0
10
20
30
0.2
0.4
0.6
0.8

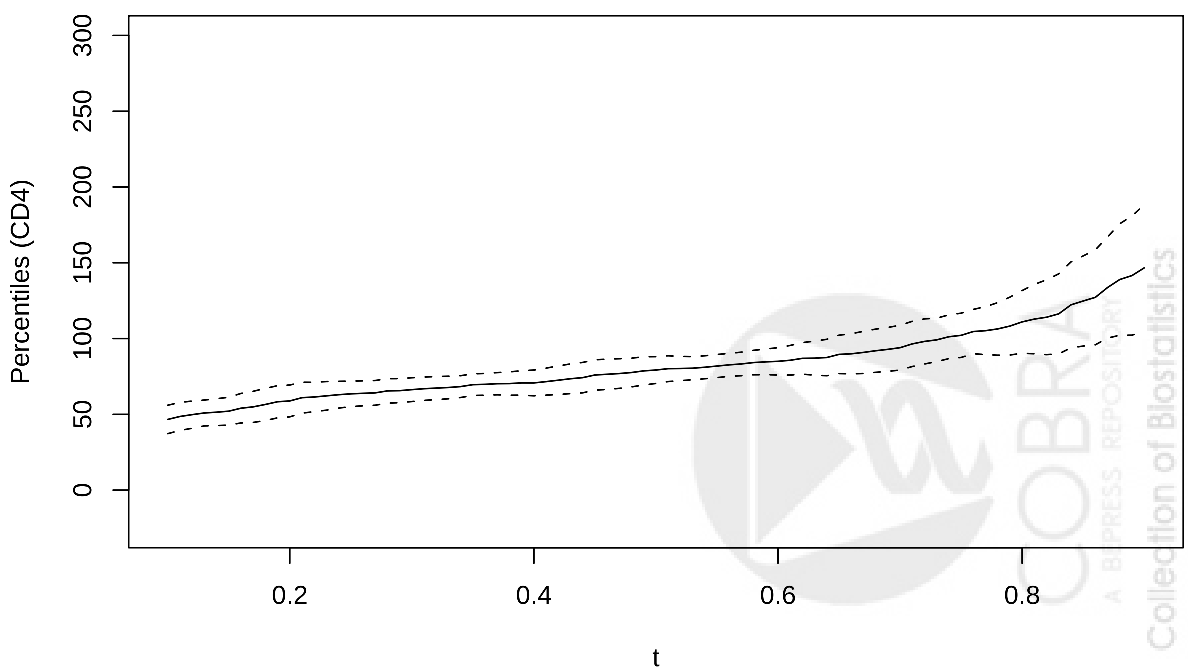
Figure 4:

(a)



Percentiles (CD4)
0
50
100
150
200
250
300
0.2
0.4
0.6
0.8
t

(b)



Percentiles (CD4)
0
50
100
150
200
250
300
0.2
0.4
0.6
0.8
t