Biometrika (2000), **87,** 1, *pp.* 89–97

Printed in Great Britain

Inferences for a semiparametric model with panel data

BY S. C. CHENG

Department of Statistics, Texas A&M University, College Station, Texas 77845, U.S.A.

scheng@stat.tamu.edu

AND L. J. WEI

Department of Biostatistics, Harvard University, Boston, Massachusetts 02115, U.S.A.

wei@biostat.harvard.edu

SUMMARY

In a longitudinal study, suppose that, for each subject, repeated measurements of the response variable and covariates are collected at a set of distinct, irregularly spaced time points. We consider a semiparametric model which relates the mean of the response variable at each time point proportionally to a function of a time-dependent covariate vector to analyse such panel data. Inference procedures for regression parameters are proposed without involving any nonparametric function estimation for the nuisance mean function. A dataset from a recent AIDS clinical trial is used to illustate the new proposal.

Some key words: Counting process; Repeated measurements; Stochastic process; Time-dependent covariate.

1. INTRODUCTION

In a longitudinal study to examine the effect of a time dependent covariate vector $\{Z(t), t \geqslant 0\}$ on the response process $\{X(t), t \geqslant 0\}$ with a group of independent subjects, the realisations of these processes for each subject are often obtained at a set of distinct and irregular time points. This type of data is quite common in clinical and observational studies. Inference procedures for parametric models with such repeated measurements over time have been extensively studied, for example by Laird & Ware (1982), Liang & Zeger (1986), Diggle (1988), Diggle, Liang & Zeger (1994) and Davidian & Giltinan (1995). Methods for nonparametric and semiparametric regression models with panel data have been proposed by authors including Hart & Wehrly (1986), Altman (1990), Hart (1991), Rice & Silverman (1991), Zeger & Diggle (1994), Moyeed & Diggle (1994) and Pepe & Couper (1997). Recently, Hoover et al. (1998) and Wu, Chiang & Hoover (1998) considered nonparametric estimation procedures for a time-varying regression coefficient model.

Almost all the aforementioned nonparametric and semiparametric methods involve nonparametric function estimators. Their validity may require quite large sample sizes, especially when the dimension of the covariate $Z(.)$ is high. In this paper, a class of simple non-iterative inference procedures for the regression parameters of a semiparametric model based on observations collected at irregular time points is proposed. Our method does not need any nonparametric function estimation. Large-sample properties of the new

proposal are derived in § 2 and are illustrated with a dataset from a recent AIDS clinical study conducted by the AIDS Clinical Trials Group in § 3.

2. INFERENCE PROCEDURES FOR REGRESSION COEFFICIENTS

For the ith subject, at time t, let $X_i(t)$ be the univariate response variable, $Z_i(t)$ be the $p \times 1$ vector of bounded covariates, C_i be the follow-up time and $N_i(t)$ be the number of the observation times which are less than or equal to t, where $i = 1, \ldots, n$ and $t \in [0, \tau]$, a given finite time interval. Assume that $\{(X_i(.), Z_i(.), C_i, N_i(.)), i = 1, \ldots, n\}$ are n independent and identically distributed random elements. Let $Z_i^* = \{Z_i(t), t \in [0, \tau]\}$. We assume that the observation time points process $N_i(.)$ is independent of Z_i^*, C_i and $X_i(.)$, but C_i may depend on Z_i^*. However, conditional on Z_i^*, C_i and $X_i(.)$ are independent and $\mathrm{pr}(C_i \geqslant \tau \mid Z_i^*) > 0$. Let $E\{N_i(t)\} = \Lambda(t)$ and $\Delta_i(t) = I(C_i \geqslant t)$, where $I(.)$ is the indicator function. Without loss of generality, we assume that $X_i(.)$ is positive and

$$E\{X_i(t) \mid Z_i^*\} = \mu(t) \exp\{\beta_0^{\mathrm{T}} Z_i(t)\}, \tag{1}$$

where $\mu(t)$ is a completely unknown, positive and continuous function and β_0 is the true value of a $p \times 1$ vector β of regression parameters. Note that we only model the mean function of $X_i(.)$ at each time point. The stochastic structure of the process $X_i(.)$ is totally unspecified. If $X_i(t)$ is continuous, model (1) is equivalent to

$$\log X_i(t) = \log \xi(t) + \beta_0^{\mathrm{T}} Z_i(t) + \varepsilon_i(t), \tag{2}$$

where $\xi(t)$ is an unspecified function and $E\{\exp \varepsilon_i(t)\}$ is a function of t only, but the unspecified distribution of the error term $\varepsilon_i(t)$ may depend on Z_i^*. Model (2) is a generalisation of the partial linear model with a specific parametric error structure, studied by Zeger & Diggle (1994) and Moyeed & Diggle (1994).

The data from the ith subject consist of observations from $\{X_i(.), Z_i(.)\}$ at those time points where the counting process $N_i(.)$ jumps before the censoring time C_i. With these n sets of repeated measurements, we derive estimation procedures for β_0 in (1) without using any nonparametric function estimates of $\mu(.)$. To this end, consider the process $Y_i(t) = \int_0^t X_i(s)\, dN_i(s)$, which comprises cumulative sums of the observed values of $X_i(.)$ over time. This process $Y_i(.)$ can capture the entire original repeated measurements from $X_i(.)$. Note that, given Z_i^*, the mean of $Y_i(t)$ is $\int_0^t \exp\{\beta_0^{\mathrm{T}} Z_i(s)\}\, d\Gamma(s)$, where $\Gamma(t) = \int_0^t \mu(s)\, d\Lambda(s)$. It follows that the expected value of

$$\hat{\Gamma}(\beta_0;\, t) = \sum_{i=1}^{n} \int_0^t \frac{\Delta_i(s)}{\sum_j \Delta_j(s) \exp\{\beta_0^{\mathrm{T}} Z_j(s)\}}\, dY_i(s) \tag{3}$$

is $\Gamma(t)$, which motivates us to consider a class of estimating functions $U(\beta)$ to estimate β_0, where

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t) W(t) Z_i(t)\, d\left[Y_i(t) - \int_0^t \exp\{\beta^{\mathrm{T}} Z_i(s)\}\, d\hat{\Gamma}(\beta;\, s) \right],$$

and $W(t)$ is a positive weight function which converges uniformly to a deterministic function $w(t)$ in $t \in [0, \tau]$. It is straightforward to show that $U(\beta)$ is algebraically equivalent to

$$\sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t) W(t) \{Z_i(t) - \bar{Z}(\beta;\, t)\}\, dY_i(t), \tag{4}$$

where

$$\bar{Z}(\beta;\, t) = \frac{S^{(1)}(\beta;\, t)}{S^{(0)}(\beta;\, t)}, \quad S^{(l)}(\beta;\, t) = n^{-1} \sum_{i=1}^{n} \Delta_i(t) \exp\{\beta^{\mathrm{T}} Z_i(t)\} \{Z_i(t)\}^{\otimes l},$$

with $l = 0, 1, 2$, $v^{\otimes 0} = 1$, $v^{\otimes 1} = v$ and $v^{\otimes 2} = vv^{\mathrm{T}}$ for a column vector v. If $Y_i(t)$ is a counting process and $W(t) = 1$ in (4), the estimating function $U(\beta)$ is the partial likelihood score function of the Andersen–Gill proportional intensities model (Andersen & Gill, 1982). This type of estimating function is also asymptotically unbiased for a more general non-Poisson process (Pepe & Cai, 1993; Lawless & Nadeau, 1995; Lawless, Nadeau & Cook, 1997).

Now, let $s^{(l)}(\beta;\, t)$ and $\bar{z}(\beta;\, t)$ be the limits of $S^{(l)}(\beta;\, t)$ and $\bar{Z}(\beta;\, t)$, respectively. Assume that the matrix

$$A = E\left[\int_0^{\tau} \Delta_1(t) w(t) \exp\{\beta_0^{\mathrm{T}} Z_1(t)\} \{Z_1(t) - \bar{z}(\beta_0;\, t)\}^{\otimes 2}\, d\Gamma(t) \right],$$

is nonsingular. In Appendix 1, we show that under the above mild condition any solution $\hat{\beta}$ to the equation $U(\beta) = 0$ is strongly consistent. Furthermore, for large n, $\hat{\beta}$ is unique.

To derive the large-sample approximation to the distribution of $\hat{\beta}$, we expand $U(\hat{\beta})$ around β_0. It follows that $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ is asymptotically equivalent to $-n\{\partial U(\beta_0)/\partial\beta\}^{-1} n^{-\frac{1}{2}} U(\beta_0)$. It is straightforward to show that $-n^{-1}\, \partial U(\beta_0)/\partial\beta$ converges to A, as $n \to \infty$. Now, let

$$Q_i(t) = Y_i(t) - \int_0^t \exp\{\beta_0^{\mathrm{T}} Z_i(s)\}\, d\Gamma(s).$$

Then

$$U(\beta_0) = \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t) W(t) \{Z_i(t) - \bar{Z}(\beta_0;\, t)\}\, dQ_i(t).$$

Note that $E\{Q_i(t)\} = 0$, for $t \in [0, \tau]$. In Appendix 2, we show that, if the total variation of $Z_i(.)$, for $i = 1, \ldots, n$, is bounded by a constant, then

$$n^{-\frac{1}{2}} U(\beta_0) \doteq n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t) w(t) \{Z_i(t) - \bar{z}(\beta_0;\, t)\}\, dQ_i(t).$$

It follows from the Multivariate Central Limit Theorem that $n^{-\frac{1}{2}} U(\beta_0)$ is asymptotically normal with mean zero and variance–covariance matrix

$$\Sigma = E\left(\left[\int_0^{\tau} \Delta_1(t) w(t) \{Z_1(t) - \bar{z}(\beta_0;\, t)\}\, dQ_1(t) \right]^{\otimes 2} \right).$$

A consistent estimator $\hat{\Sigma}$ for Σ can be obtained by replacing the theoretical quantities in Σ with their empirical counterparts, where

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \left[\int_0^{\tau} \Delta_i(t) W(t) \{Z_i(t) - \bar{Z}(\hat{\beta};\, t)\}\, d\hat{Q}_i(t) \right]^{\otimes 2},$$

and $\hat{Q}_i(t) = Y_i(t) - \int_0^t \exp\{\hat{\beta}^{\mathrm{T}} Z_i(s)\}\, d\hat{\Gamma}(\hat{\beta};\, s)$. The proof of the consistency is given in Appendix 3.

It follows that the distribution of $\hat{\beta}$ can be approximated by a normal with mean β_0

and variance–covariance matrix $n\{\partial U(\hat{\beta})/\partial\beta\}^{-1}\hat{\Sigma}\{\partial U(\hat{\beta})/\partial\beta\}^{-1}$. Inferences about β_0 can then be made through this large-sample approximation.

3. EXAMPLE

We use a dataset which consists of repeated measurements over time from a recent clinical trial conducted by the AIDS Clinical Trials Group to illustrate the new proposal. The trial, ACTG 144, was designed to evaluate and compare the efficacy of two doses of the drug ddI for the treatment of HIV-infected children who had AZT intolerance. One of the primary goals for the study is to use the patient's CD4 cell count values over time as an immunological response variable for comparisons. The trial started in August 1991. As of December 1994, 335 patients were randomised to the two groups. There were 169 and 166 subjects in the high and low dose group, respectively. For each patient, the CD4 counts were repeatedly taken during the treatment period. The set of observation time points and the follow-up time vary from subject to subject. Twenty-four patients had no baseline CD4 count and 20 patients did not have any post-randomisation CD4 measures. Our analysis is based on 1553 CD4 cell counts from 291 patients. Here, for the ith patient, $X_i(t)$ is the CD4 count at time t. Consider model (1) with $Z_i(t)$ being a vector of two time-independent covariates. The first covariate is the treatment indicator, with high dose being coded as 1, and the second covariate is the logarithm of the baseline CD4 value. If we use the estimating function $U(\beta)$ with $W(t)=1$ in (4), the resulting point estimate $\hat{\beta}$ is $(0{\cdot}18, 0{\cdot}85)^{\mathrm{T}}$ with standard errors 0·09 and 0·04, respectively. The $U(\beta)$ with $W(t)=1$ is analogous to the partial likelihood score function for the Cox model. If we let $W(t)=S^{(0)}(\beta; t)$ in (4), the observations at early time points are given more weight than other data points. This is similar to the case with the Gehan score for the weighted log-rank test statistic in survival analysis. If we use this weighting function, the point estimate $\hat{\beta}$ is $(0{\cdot}18, 0{\cdot}87)^{\mathrm{T}}$ with standard errors 0·08 and 0·03, respectively.

One may use approaches similar to assessing the adequacy of the Cox model in survival analysis to examine if the above simple additive model (1) fits the CD4 count data well. For example, stratifying on K levels of the baseline CD4 count leads to the stratified model (1) with $\mu(t)=\mu_k(t)$ and a single covariate z, the treatment indicator, where $k=1,\ldots,K$ and $z=0, 1$. Let $\hat{\Gamma}_{kz}(t)$ be the Breslow type of estimator (3) for stratum k with treatment z. If the additive model is correctly specified, the $2K$ plots of the logarithm of $\hat{\Gamma}_{kz}(t)$ over t would be parallel to each other. Furthermore, the distance between two parallel plots within each stratum is expected to be the same across the strata. In Fig. 1, we present such plots with $K=3$. It appears that the distance between the two plots for the top stratum is much narrower than those for the bottom two strata. This suggests that there is an interaction between the baseline CD4 count and the treatment indicator. With the standard interaction term added to the additive model, the estimates with the 'log-rank' and 'Gehan' scores are reported in Table 1. The results are quite similar for these two cases.

4. REMARKS

Several simulation studies were conducted to evaluate the adequacy of the proposals for practical usage. We mimicked the set-up of the AIDS trial discussed in § 3 to examine whether or not the interval estimation procedure proposed in § 2 for the regression parameter has correct coverage probabilities. For example, in one of the studies, the chance of

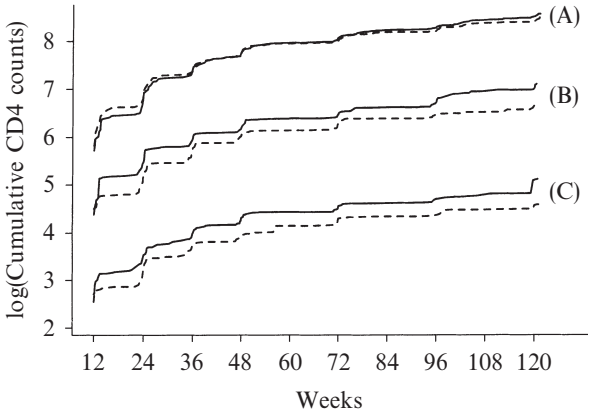

Fig. 1. AIDS Clinical Trial example. Plots of logarithms of cumulative CD4 counts over time by discretising the baseline CD4 count to three strata: (A) more than 270, (B) between 40 and 270, (C) less than 40. Solid lines correspond to high-dose drug ddI, dashed lines correspond to low-dose ddI.

Table 1: *AIDS Clinical Trial example. Regression analysis of the* CD4 *count data with two weighting functions* $W(t)$

	$W(t)=1$		$W(t)=S^{(0)}(\beta;t)$	
Covariate	Parameter estimate	Estimated SE	Parameter estimate	Estimated SE
Treatment	0·61	0·47	0·56	0·44
Baseline CD4	0·89	0·04	0·90	0·04
Interaction	−0·07	0·07	−0·06	0·07

taking an observation at each time point t, where $t = 1, \ldots, 120$, is 0·1. This gives us about 12 observations per subject. The model considered in this study, which produces observations $X(t)$ similar to the CD4 count in our example, is

$$\log\{X(t)\} = 1 - 0{\cdot}1t + 0{\cdot}2Z_1 + 0{\cdot}85Z_2 + \varepsilon(t),$$

where Z_1 is a Bernoulli random variable with success probability 0·5, Z_2 is the natural logarithm of a normal random variable with mean 4·4 and standard deviation 0·5, $\varepsilon(t) = \rho\varepsilon(t-1) + v(t)$, $\varepsilon(0) = 0$ and the $v(t)$'s are independent normal variables with mean 0 and standard deviation σ. In Table 2, we report the empirical coverage levels of the 95% confidence intervals for the regression parameters with various n, σ and ρ. Our confidence intervals generally have fairly accurate coverage probabilities. For the case with $n = 100$, the confidence levels of the intervals for the coefficient of the continuous covariate Z_2 are slightly lower than the nominal counterparts. Based on all the studies we have done, the interval estimation procedure performs well especially when $n > 200$. Each empirical level was based on 2500 iterations and was obtained with $W(t) = 1$.

Our estimation involves a weighting function. It is important to know how to choose this function for the efficiency of $\hat{\beta}$. In this paper, we used two weighting functions, which are similar to the most popular score functions in survival analysis, namely log-rank and Gehan, to analyse the data. For the AIDS example, the results of the analysis are quite

Table 2: *Simulation study. Empirical coverage levels of* 0·95 *confidence intervals for regression parameters*

		$\sigma = 0{\cdot}10$			$\sigma = 0{\cdot}25$		
n	Covariate	$\rho = 0{\cdot}00$	$\rho = 0{\cdot}25$	$\rho = 0{\cdot}50$	$\rho = 0{\cdot}00$	$\rho = 0{\cdot}25$	$\rho = 0{\cdot}50$
100	Treatment	0·94	0·94	0·95	0·94	0·94	0·94
	Baseline CD4	0·91	0·91	0·91	0·91	0·91	0·91
200	Treatment	0·95	0·95	0·95	0·94	0·94	0·94
	Baseline CD4	0·93	0·93	0·93	0·93	0·93	0·93
300	Treatment	0·95	0·95	0·95	0·95	0·95	0·94
	Baseline CD4	0·94	0·94	0·94	0·94	0·94	0·93
400	Treatment	0·94	0·94	0·94	0·95	0·95	0·95
	Baseline CD4	0·94	0·94	0·94	0·94	0·94	0·95

similar between the two estimation procedures. We plan to investigate this important problem in future research.

By similar techniques, inference procedures for the regression parameters can be easily obtained for the case when the exponential function in (1) is replaced by any positive link function.

If the response variable $X_i(t)$ is dichotomous, (1) is the so-called proportional rate model, which is a useful alternative to the logistic regression model for analysing binary data. Recently, in an unpublished technical report from the Department of Statistics, University of Missouri, T. Sun and J. Hu used the estimating function $U(\beta)$ for the case when $X_i(.)$ is a counting process.

If one is interested in estimating the nuisance function $\mu(t)$, one-sample nonparametric function estimates, such as those proposed by Hart & Wehrly (1986), Altman (1990) and Rice & Silverman (1991), may be used based on repeated measurements from the process $\{\exp\{-\hat{\beta}^{\mathrm{T}}Z_i(t)\}X_i(t), i = 1, \ldots, n\}$ at the jump points of $\{N_i(.), i = 1, \ldots, n\}$.

Appendix 1

Consistency of $\hat{\beta}$

To show that $\hat{\beta}$ is consistent and for large n there is a unique solution $\hat{\beta}$ to the equation $U(\beta) = 0$, consider

$$D(\beta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Delta_i(t)W(t)\left[(\beta-\beta_0)^{\mathrm{T}}Z_i(t) - \log\left\{\frac{S^{(0)}(\beta;\, t)}{S^{(0)}(\beta_0;\, t)}\right\}\right] dY_i(t).$$

Note that $\partial D(\beta)/\partial\beta = n^{-1}U(\beta)$ and

$$-\frac{\partial^2 D(\beta)}{\partial\beta^2} = \tilde{A}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Delta_i(t)W(t)\exp\{\beta^{\mathrm{T}}Z_i(t)\}\{Z_i(t) - \bar{Z}(\beta;\, t)\}^{\otimes 2}\, d\hat{\Gamma}(\beta;\, t).$$

By the strong law of large numbers, $D(\beta)$ converges almost surely to

$$d(\beta) = E\left[\int_0^\tau \Delta_1(t)w(t)(\beta-\beta_0)^{\mathrm{T}}Z_1(t)\, dY_1(t) - \int_0^\tau \Delta_1(t)w(t)\log\left\{\frac{s^{(0)}(\beta;\, t)}{s^{(0)}(\beta_0;\, t)}\right\} dY_1(t)\right],$$

and $\tilde{A}(\beta_0)$ converges to A almost surely, as $n \to \infty$. It is straightforward to show that $d(\beta)$ is concave with $\partial d(\beta_0)/\partial\beta = 0$ and $\partial^2 d(\beta_0)/\partial\beta^2 = -A$, a negative definite matrix. It follows that β_0 is the unique maximiser for $d(\beta)$. Since $\tilde{A}(\beta)$ is positive semidefinite, $D(\beta)$ is also concave. Therefore, $D(\beta)$ con-

verges uniformly in a compact set of β (Rockafellar, 1970, Theorem 10.8) and $\hat{\beta}$ converges to β_0, almost surely, as $n \to \infty$ (Newey & McFadden, 1994). Furthermore, since the covariate process is assumed to be bounded, all the components of $\partial \tilde{A}(\beta)/\partial \beta$ are bounded. This implies that, for large n, $\tilde{A}(\beta)$ is positive definite for β in a compact set around β_0. It follows that, for large n, $\hat{\beta}$ is the unique solution to the equation $U(\beta) = 0$.

Appendix 2

Weak convergence of $U(\beta_0)$

Define

$$M(t) = \sum_{i=1}^{n} \int_0^t \Delta_i(s) W(s)\, dQ_i(s), \quad M_z(t) = \sum_{i=1}^{n} \int_0^t \Delta_i(s) W(t) Z_i(t)\, dQ_i(s).$$

Note that $U(\beta_0) = M_z(\tau) - \int_0^\tau \bar{Z}(\beta_0; t)\, dM(t)$. Furthermore, for each t, $M(t)$ and $M_z(t)$ are sums of independent and identically distributed zero-mean terms. By the Multivariate Central Limit Theorem, $n^{-\frac{1}{2}}(M(t), M_z(t))$ converges in finite-dimensional distributions to a zero-mean Gaussian process, denoted by $(\mathscr{W}(t), \mathscr{W}_z(t))$. Without loss of generality, we assume that $Z_i(t) \geqslant 0$. Then the processes $\{\int_0^t \Delta_i(s) W(s)\, dQ_i(s)\}$ and $\{\int_0^t \Delta_i(s) W(s) Z_i(s)\, dQ_i(s)\}$ can be written as sums of monotone functions in t and are therefore 'manageable' (Pollard, 1990, p. 38; Bilias, Gu & Ying, 1997, Theorem 2.1). It follows from a functional central limit theorem (Pollard, 1990, p. 53) that $(n^{-\frac{1}{2}} M, n^{-\frac{1}{2}} M_z)$ converges weakly to $(\mathscr{W}, \mathscr{W}_z)$ with continuous sample paths, as $n \to \infty$. By the strong embedding theorem (Shorack & Wellner, 1986, p. 47), one can show that there exists a new probability space such that $(n^{-\frac{1}{2}} M(t), n^{-\frac{1}{2}} M_z(t), S^{(1)}(\beta_0; t), S^{(0)}(\beta_0; t))$ converges almost surely to $(\mathscr{W}(t), \mathscr{W}_z(t), s^{(1)}(\beta_0; t), s^{(0)}(\beta_0; t))$.

For the case that covariates are time-independent, $S^{(0)}(\beta_0; t)$ and $S^{(1)}(\beta_0; t)$ are monotone functions in t. Thus, one can use Helly's Theorem (Serfling, 1980, p. 352) and Lemma 8.2.3 in Chow & Teicher (1988, p. 265) to show that

$$n^{-\frac{1}{2}} \int_0^t \frac{1}{S^{(0)}(\beta_0; s)}\, dM(s) \to \int_0^t \frac{1}{s^{(0)}(\beta_0; s)}\, d\mathscr{W}(s), \tag{A1}$$

almost surely and uniformly in t. Furthermore,

$$n^{-\frac{1}{2}} \int_0^t \frac{S^{(1)}(\beta_0; s)}{S^{(0)}(\beta_0; s)}\, dM(s) \to \int_0^t \bar{z}(\beta_0, s)\, d\mathscr{W}(s), \tag{A2}$$

almost surely and uniformly in t. It follows that $n^{-\frac{1}{2}} U(\beta_0)$ converges almost surely to $\mathscr{W}_z(\tau) - \int_0^\tau \bar{z}(\beta_0; s)\, d\mathscr{W}(s)$ and thus weakly in the original probability space.

In general, when the covariates are time-dependent, $S^{(l)}(\beta_0; t)$, for $l = 0, 1$, may not be monotone. However, if the total variations of $\{Z_i(t), i = 1, \ldots, n, t \in [0, \tau]\}$, are bounded, then $S^{(l)}(\beta_0; t)$ has a bounded variation and (A1) and (A2) still hold. The derivation of the limiting variance–covariance matrix of $n^{-\frac{1}{2}} U(\beta_0)$ is straightforward.

Appendix 3

Consistency of $\hat{\Sigma}$

By the uniform strong law of large numbers (Pollard, 1990, p. 41),

$$S^{(0)}(\beta; t) \to s^{(0)}(\beta; t), \quad n^{-1} \sum_{i=1}^{n} \int_0^t \Delta_i(u)\, dY_i(u) \to \int_0^t s^{(0)}(\beta_0; u)\, d\Gamma(u)$$

uniformly in t and β. It follows that $\hat{\Gamma}(\beta; t)$ converges uniformly to $\int_0^t \{s^{(0)}(\beta_0; u)/s^{(0)}(\beta; u)\}\, d\Gamma(u)$. Since $\partial \hat{\Gamma}(\beta; t)/\partial \beta$ is uniformly bounded for large n and for β in a compact set, the strong consistency

of $\hat{\beta}$ implies that $\hat{\Gamma}(\hat{\beta}; t)$ converges almost surely to $\Gamma(t)$. This, coupled with the uniform convergence of $W(t)\bar{Z}(\beta_0; t)$, entails that

$$\frac{1}{n}\sum_{i=1}^{n}\left\| \int_0^\tau \Delta_i(t)W(t)\{Z_i(t)-\bar{Z}(\hat{\beta}; t)\}\, d\hat{Q}_i(t) - \int_0^\tau \Delta_i(t)w(t)\{Z_i(t)-\bar{z}(\beta_0; t)\}\, dQ_i(t)\right\|^2 \to 0,$$

almost surely. In addition, by the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \Delta_i(t)w(t)\{Z_i(t)-\bar{z}(\beta_0; t)\}\, dQ_i(t)\right]^{\otimes 2} \to \Sigma,$$

almost surely. It follows that

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \Delta_i(t)W(t)\{Z_i(t)-\bar{Z}(\hat{\beta}; t)\}\, d\hat{Q}_i(t)\right]^{\otimes 2} \to \Sigma,$$

almost surely.

REFERENCES

ALTMAN, N. S. (1990). Kernel smoothing of data with correlated errors. *J. Am. Statist. Assoc.* **85**, 749–59.

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–20.

BILIAS, Y., GU, M. & YING, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25**, 662–82.

CHOW, Y. S. & TEICHER, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*, 2nd ed. New York: Springer-Verlag.

DAVIDIAN, M. & GILTINAN, D. M. (1995). *Nonlinear Models for Repeated Measurement Data.* London: Chapman and Hall.

DIGGLE, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959–71.

DIGGLE, P. J., LIANG, K. Y. & ZEGER, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

HART, J. D. (1991). Kernel regression estimation with time series errors. *J. R. Statist. Soc.* B **53**, 173–87.

HART, J. D. & WEHRLY, T. E. (1986). Kernel regression estimation using repeated measurements data. *J. Am. Statist. Assoc.* **81**, 1080–8.

HOOVER, D. R., RICE, J. A., WU, C. O. & YANG, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–22.

LAIRD, N. M. & WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–74.

LAWLESS, J. F. & NADEAU, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158–68.

LAWLESS, J. F., NADEAU, C. & COOK, R. J. (1997). Analysis of mean and rate functions for recurrent events. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, Ed. D. Y. Lin and T. R. Fleming, pp. 37–49. New York: Springer-Verlag.

LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

MOYEED, R. A. & DIGGLE, P. J. (1994). Rates of convergence in semi-parametric modelling of longitudinal data. *Aust. J. Statist.* **36**, 75–93.

NEWEY, W. & MCFADDEN, D. L. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, **4**, Ed. R. F. Engle and D. L. McFadden, pp. 2111–245. Amsterdam: North Holland.

PEPE, M. S. & CAI, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Am. Statist. Assoc.* **88**, 811–20.

PEPE, M. S. & COUPER, D. (1997). Modeling partly conditional means with longitudinal data. *J. Am. Statist. Assoc.* **92**, 991–8.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* Hayward, CA: Institute of Mathematical Statistics.

RICE, J. A. & SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B **53**, 233–43.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton: Princeton University Press.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

SHORACK, G. R. & WELLNER, J. A. (1986). *Empirical Process with Applications to Statistics*. New York: Wiley.
WU, C. O., CHIANG, C. T. & HOOVER, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a time-varying coefficient model with longitudinal data. *J. Am. Statist. Assoc.* **88**, 1388–402.
ZEGER, S. L. & DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689–99.

[*Received July* 1998. *Revised May* 1999]