The Annals of Statistics
1997, Vol. 25, No. 2, 683–714

GENERALIZED MARTINGALE-RESIDUAL PROCESSES FOR GOODNESS-OF-FIT INFERENCE IN COX'S TYPE REGRESSION MODELS

BY LESZEK MARZEC AND PAWEŁ MARZEC

University of Wrocław

In the paper a general class of stochastic processes based on the sums of weighted martingale-transform residuals for goodness-of-fit inference in general Cox's type regression models is studied. Their form makes the inference robust to covariate outliers. A weak convergence result for such processes is obtained giving the possibility of establishing the randomness of their graphs together with the construction of the formal χ^2-type goodness-of-fit tests. By using the Khmaladze innovation approach, a modified version of the initial class of processes is also defined. Weak convergence results for the processes are derived. This leads to the main application which concerns the formal construction of the Kolmogorov–Smirnov and Cramér–von Mises-type goodness-of-fit tests. This is done within the general situation considered.

1. Introduction. We consider a general class of stochastic processes for goodness-of-fit examination with the general Cox regression model of Andersen and Gill (1982) and other semiparametric regression models. In the inference we use the robust version of Cox's (1975) estimator—the maximum weighted partial likelihood estimator (MWPLE) of Bednarski (1993). It should be mentioned that Bednarski's (1993) idea of modifying the Cox (1975) partial likelihood function differs from that of Lin (1991) and Sasieni (1993a, b). In this paper, we proceed with the idea of defining the weighted martingale-transform residuals which can be treated as the robust version of Barlow and Prentice (1988) residuals [see also Fleming and Harrington (1991), Henderson and Milner (1991), Lin, Wei and Ying (1993), Segal, James, French and Mallai (1995)].

Our general class of processes for checking the adequacy of the model is based on the sums of weighted martingale-transform residuals. The class contains, as special cases, the scores [Schoenfeld (1980), Arjas (1988) and Lin, Wei and Ying (1993)] for one-parameter fixed-type processes. These previously considered processes have been mainly used to propose diagnostic graphical methods [Arjas (1988), Lin, Wei and Ying (1993)], the χ^2-type test [Schoenfeld (1980)] or the Kolmogorov–Smirnov type test for the Cox (1972) model with a single time-independent covariate [Wei (1984), Haara (1987), Thernau, Grambsch and Fleming (1990)].

Received December 1995; revised June 1996.
AMS 1991 *subject classifications*. Primary 62F12; secondary 62G10, 62G20.
Key words and phrases. Censoring, counting process, Cox-regression, goodness-of-fit, martingale, residual, innovation process, weak convergence, robust inference, survival analysis.

In this paper our main purpose is to provide the formal constructions of goodness-of-fit tests of the Kolmogorov–Smirnov (KS) and Cramér–von Mises (CM) type under a relatively weak set of restrictions concerning the general semiparametric models considered. In particular, repeated occurrences of the events of interest (e.g., failures) are possible. The asymptotic behavior of the processes based on sums of weighted martingale-transform residuals is established in Theorem 3.1 of Section 3.1. The general structure of the asymptotic limit corresponds to a zero-mean continuous Gaussian process which only in special cases appears to be a transformed Brownian motion or Brownian bridge, allowing then for the direct constructions of KS and CM-type tests. In the general situation, however, the Gaussian limit process appears to be neither a martingale nor a process with well-known properties. It should be noted that for situations of such type, Lin, Wei and Ying (1993) proposed some numerical methods based on the given data and simulated normal samples to derive p-values of tests.

In the present paper we propose an alternative approach leading to the formal constructions of the critical regions of KS and CM-type goodness-of-fit tests. To achieve this, we introduce in the second part of Section 3.1 the appropriately modified version of the initial processes based on weighted martingale residuals and establish the weak convergence of the latter processes to Gaussian martingales (Theorem 3.3). As a result of this, we obtain the convergence in distribution of the KS and CM-type test statistics to the well-known functionals of the standard Brownian motion. The behavior of the considered processes outside initial model specification together with the power function performance of the tests are investigated in Section 4. Some optimality results are also reached. Comparison of some KS and CM-type tests with respect to existing tests proposed for the Cox proportional hazards model is provided in Section 5. This is done by using Monte Carlo simulations. Section 6 is devoted to the generalizations of the asymptotic results of Section 3 with respect to the semiparametric model with general relative risk form and the multistates semiparametric regression model. All proofs are provided in Section 7.

2. Notation and preliminaries.

Notation. We use the following notation, where $x \in R$, $\beta \in R^p$:

$$Q_{k,l}(\beta, x) = \frac{1}{n} \sum_{i=1}^{n} Y_i(x) w(Z_i(x), x) \phi_i(x)^{\otimes k} Z_i(x)^{\otimes l} \exp[\beta^T Z_i(x)],$$

$$k, l = 0, 1, 2;\ k + l \leq 2,$$

where $\phi^{\otimes 0} = 1$, $\phi^{\otimes 1} = \phi$, $\phi^{\otimes 2} = \phi\phi^T$, $\phi^{\otimes 1} Z^{\otimes 1} = \phi Z^T$;

$$\overline{N}_w(x) = \sum_{i=1}^{n} \int_0^x w(Z_i(s), s)\, dN_i(s);$$

$$\overline{M}_w(x) = \overline{N}_w(x) - n \int_0^x Q_{0,0}(\beta_0, s) \lambda_0(s)\, ds;$$

$$a(\beta, x) = l(\beta, x)^T \left\{ \frac{q_{2,0}(\beta, x)}{q_{0,0}(\beta, x)} - \frac{q_{1,0}(\beta, x)^{\otimes 2}}{q_{0,0}(\beta, x)^2} \right\} l(\beta, x);$$

$$\hat{a}(\beta, x) = L(\beta, x)^T \left\{ \frac{Q_{2,0}(\beta, x)}{Q_{0,0}(\beta, x)} - \frac{Q_{1,0}(\beta, x)^{\otimes 2}}{Q_{0,0}(\beta, x)^2} \right\} L(\beta, x);$$

$$A(\beta, x) = \int_0^x a(\beta, s) q_{0,0}(\beta, s) \lambda_0(s)\, ds;$$

$$\hat{A}(\beta, x) = \int_0^x \hat{a}(\beta, s)\, d\overline{N}_w(s)/n;$$

$$\rho(\beta, x) = \frac{q_{1,1}(\beta, x)}{q_{0,0}(\beta, x)} - \frac{q_{1,0}(\beta, x) q_{0,1}(\beta, x)^T}{q_{0,0}(\beta, x)^2};$$

$$\hat{\rho}(\beta, x) = \frac{Q_{1,1}(\beta, x)}{Q_{0,0}(\beta, x)} - \frac{Q_{1,0}(\beta, x) Q_{0,1}(\beta, x)^T}{Q_{0,0}(\beta, x)^2}$$

$$b(\beta, x) = \rho(\beta, x)^T l(\beta, x); \qquad \hat{b}(\beta, x) = \hat{\rho}(\beta, x)^T L(\beta, x);$$

$$\hat{b}_1(\beta, x) = \hat{\rho}(\beta, x)^T L(\beta_0, x);$$

$$B(\beta, x) = \int_0^x b(\beta, s) q_{0,0}(\beta, s) \lambda_0(s)\, ds;$$

$$\hat{B}(\beta, x) = \int_0^x \hat{b}(\beta, s)\, d\overline{N}_w(s)/n;$$

$$\sigma(\beta, x) = \frac{q_{0,2}(\beta, x)}{q_{0,0}(\beta, x)} - \frac{q_{0,1}(\beta, x)^{\otimes 2}}{q_{0,0}(\beta, x)^2};$$

$$\hat{\sigma}(\beta, x) = \frac{Q_{0,2}(\beta, x)}{Q_{0,0}(\beta, x)} - \frac{Q_{0,1}(\beta, x)^{\otimes 2}}{Q_{0,0}(\beta, x)^2};$$

$$\Sigma(\beta, x) = \int_0^x \sigma(\beta, s) q_{0,0}(\beta, s) \lambda_0(s)\, ds;$$

$$\hat{\Sigma}(\beta, x) = \int_0^x \hat{\sigma}(\beta, s)\, d\overline{N}_w(s)/n;$$

$$\gamma_{kl}(x) = \int_x^\tau \sigma_{kl}(\beta_0, s) q_{0,0}(\beta_0, s) \lambda_0(s)\, ds;$$

$$\hat{\gamma}_{kl}(x) = \int_{[x,\tau]} \hat{\sigma}_{kl}\left(\hat{\beta}_w, s\right) d\overline{N}_w(s)/n;$$

$$\zeta(\beta, x) = \Sigma(\beta, \tau) - \Sigma(\beta, x); \qquad \hat{\zeta}(\beta, x) = \hat{\Sigma}(\beta, \tau) - \hat{\Sigma}(\beta, x-).$$

C^- stands for the generalized inverse of the matrix $C = (c_{ij})$; $\|C\| = \max_{i,j} |c_{ij}|$.

Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a complete probability space and let $\{\mathscr{F}_t\colon t \in [0, \tau]\}$ be an increasing right-continuous family of sub σ-algebras of $\mathscr{F}$. Counting processes, local martingales, predictable processes, and so on are defined with reference to these sub σ-algebras. Following Andersen and Gill (1982), let $\mathbf{N} = (N_1, \dots, N_n)$, $n \geq 1$, be the multivariate counting process defined so that N_i counts failures on the ith subject at times $t \in [0, \tau]$. Thus $\mathbf{N}$ has components N_i which are right-continuous step functions, zero at time zero, with jumps of size $+1$ only, such that no two components jump simultaneously. Assume each $N_i(\tau)$ to be almost surely finite.

The general version of Cox's regression model of Andersen and Gill (1982) postulates that N_i $(i = 1, \dots, n)$ has intensity process λ_i, that is,

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(s)\, ds, \qquad t \in [0, \tau] \tag{1}$$

is a local square integrable martingale, of the form

$$\lambda_i(t) = Y_i(t) \exp\left[\beta_0^T Z_i(t)\right] \lambda_0(t), \qquad t \in [0, \tau]. \tag{2}$$

Here β_0 is a column vector of p unknown regression coefficients, λ_0 is an arbitrary and unspecified baseline hazard function, Y_i is a predictable $\{0, 1\}$-valued process indicating that the ith individual is at risk when $Y_i = 1$ and Z_i is a p-variate column vector of processes which are assumed to be predictable and locally bounded.

The MWPLE $\hat{\beta}_w$ of β_0 is defined as a solution of

$$U(\beta, \tau) = 0, \tag{3}$$

where

$$U(\beta, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t w(Z_i(s), s) \left\{ Z_i(s) - \frac{Q_{0,1}(\beta, s)}{Q_{0,0}(\beta, s)} \right\} dN_i(s), \qquad t \in [0, \tau]. \tag{4}$$

Here a weight function $w\colon R^p \times [0, \tau] \to R_+$ is assumed to be bounded and Borel-measurable. We use the estimator $\hat{\beta}_w$ in further goodness-of-fit inference for Cox's regression model (2). Our bases are the weighted martingale residuals of the form

$$\hat{M}_i^w(t) = \int_0^t w(Z_i(s), s)\, dN_i(s) - \int_0^t w(Z_i(s), s) Y_i(s) \exp\left[\hat{\beta}_w^T Z_i(s)\right] d\hat{\Lambda}_w(s), \qquad t \in [0, \tau],\ i = 1, \dots, n,$$

where

$$\hat{\Lambda}_w(s) = \int_0^s \left[n Q_{0,0}(\hat{\beta}_w, x) \right]^{-1} d\overline{N}_w(x)$$

is a weighted version of the standard estimator for the cumulative baseline hazard function $\Lambda_0(s) = \int_0^s \lambda_0(x)\, dx$, $s \in [0, \tau]$.

Now the generalization based on the approach of Barlow and Prentice (1988) leads to the weighted martingale-transform residuals

$$\hat{\mathbb{M}}_i^w(t) = \int_0^t \phi_i(s)\, d\hat{M}_i^w(s), \qquad t \in [0, \tau], i = 1, \dots, n, \tag{5}$$

where ϕ_i are predictable and locally bounded q-variate processes. The weighted sum of $\hat{\mathbb{M}}_i^w(\cdot)$, $i = 1, \dots, n$, of (5) is our main interest. It has the form

$$\begin{aligned}\psi(\hat{\beta}_w, \cdot) &= \frac{1}{\sqrt{n}} \int_0^{(\cdot)} L(\hat{\beta}_w, s)^T d \sum_{i=1}^n \hat{\mathbb{M}}_i^w(s) \\ &= \frac{1}{\sqrt{n}} \int_0^{(\cdot)} L(\hat{\beta}_w, s)^T \sum_{i=1}^n w(Z_i(s), s) \left[\phi_i(s) - \frac{Q_{1,0}(\hat{\beta}_w, s)}{Q_{0,0}(\hat{\beta}_w, s)} \right] dN_i(s).\end{aligned} \tag{6}$$

The process $L(\beta_0, \cdot)$ is assumed to be a linear combination, with coefficients which are random variables, of predictable and locally bounded q-variate processes. If model (2) holds, $\psi(\hat{\beta}_w, \cdot)$ will fluctuate randomly around zero; we may replace dN_i by dM_i, where M_i is given by (1), in the definition of $\Psi(\beta_0, \cdot)$. Since under the correct model of (2), $\psi(\beta_0, \cdot)$ is a linear combination of local martingales, $\psi(\hat{\beta}_w, \cdot)$ expresses a balance between the suitable restricted actual count of failures and the corresponding estimated collective cumulative hazard.

A variety of examples of the processes ϕ_i and L appear in the definition of $\psi(\hat{\beta}_w, \cdot)$ of (6) which are of special interest. Most of them can be adopted from Crowley and Jones (1989, 1990) and Jones (1991) who considered $\psi(0, t)$ with $w \equiv 1$ ($p = 1$) for the problem of testing $\beta = 0$ ($p = 1$) in Cox's regression model. Their general proposition for $\phi_i(s)$ of the form $g_s(Z_i(s)|\mathscr{F}_{s-})$, where g_s is an $\mathscr{F}_{s-}$-measurable function, contains as special examples $\phi_i(s) = g(Z_i(s), s)$ with a Borel-measurable function g: $R^{p+1} \to R^q$ or $\phi_i(s) = \Psi(Z_i(s) - \bar{Z}(s))$ with $\bar{Z} = \sum Y_i Z_i / \sum Y_i$ and some downweighting function Ψ: $R^p \to R^q$. Note that the examples of g of the form $g(z, s) = z$, $g(z, s) = I\{(z, s) \in A\}$, $A \subset R^p \times [0, \tau]$, $g(z, s) = f(z) I\{z \le x\}$, $x \in R^p$, $g(z, s) = z f(s)$ lead, according to (6) with $w \equiv 1$, $L \equiv 1$, to the score process [Wei (1984), Haara (1987), Thernau, Grambsch and Fleming (1990)], Schoenfeld's (1980) process, Lin, Wei and Ying's (1993) class and Sasieni's (1993b) example, respectively. In the situation when ϕ_i has the status of the additional model covariate, $\psi(\hat{\beta}_w, \cdot)$ of (6) (with $w \equiv 1$, $L \equiv 1$) is the standard score process based on the Cox-type alternative to (2) of the form $\lambda_i = Y_i \exp(\beta^T Z_i + \gamma^T \phi_i) \lambda_0$, $\gamma \neq 0$ [see Tsatis, Rosner and Tritchler (1985), Slud (1991)]. Note also that for $\phi_i = I\{i \in J\}$, $J \subset \{1, \dots, n\}$, the process of (6) is a weighted version of the stratified process of Arjas (1988) [see also Marzec and Marzec (1993), Marzec (1993)].

By considering the weighting process L in the definition of (6), we provide on the one hand a possibility for the practitioner to exercise his choice concerning the way in which model (2) is to be fitted [cf. Sasieni (1993a)]. On the other hand, L can be used as the parameter in optimization problems (see Section 4). Following the Crowley and Jones (1989, 1990) examples of L that include (for $p = 1$, $p = q$) $L \equiv 1$ and the Gehan-type weight $L = \Sigma Y_i/n$, one may consider more general special forms $L(\beta, s) = l(s)$ or $L(\beta, s) = l(s)K(\beta, s)^\alpha$, $\alpha \geq 1$, where l is a deterministic function and for a measurable function W, $K(\beta, s) = \Sigma Y_i(s)W(\beta, Z_i(s))/n$ is the predictable and locally bounded process. Note that to reflect a decreasing willingness to rely on the model (2) assumption as time progresses, one can put $L(\beta, s) = [K(\beta, \tau) - K(\beta, s)]^\alpha$, $\alpha \in N$.

Finally, by considering the weight function w in (6), we provide a possibility of making the inference robust to the violations of measured covariates [cf. Bednarski (1993)]. In this paper we assume, for simplicity, that the weight function $w(z, s)$ takes values from $\{0, 1\}$. In particular, it corresponds to the situation when $w(z, s)$ censors large values of covariates. Then the resulting estimators and processes should become generally robust to covariate outliers. Obviously, the choice $w(z, s) \equiv 1$ leads to considerations based on the classical MPLE.

3. Asymptotic properties of residual process. In this section we investigate the asymptotic performance of the process of (6) together with its modification useful for constructing the formal KS and CM-type goodness-of-fit tests. The process is given in Section 3.1, the modification in Section 3.2.

3.1. *Weak convergence results.* The following assumptions are used in order to establish the weak convergence of the process $\psi(\hat{\beta}_w, \cdot)$ of (6).

CONDITION A. $\int_0^\tau \lambda_0(s)\, ds < \infty$.

CONDITION B. There exists a neighborhood B of β_0 and the functions $q_{k,l}$, $k, l = 0, 1, 2$, $k + l \leq 2$ such that

$$\sup_{\substack{\beta \in B \\ s \in [0, \tau]}} \| Q_{k,l}(\beta, s) - q_{k,l}(\beta, s) \| = o_p(1),$$

where $q_{k,l}(\cdot, s)$ are continuous in $\beta \in B$, uniformly in $s \in [0, \tau]$, $q_{k,l}(\cdot, \cdot)$ are bounded on $B \times [0, \tau]$, $q_{0,0}(\cdot, \cdot)$ is bounded away from zero, $(\partial/\partial\beta) q_{0,0}(\beta, s) = q_{0,1}(\beta, s)$, $(\partial/\partial\beta) q_{0,1}(\beta, s) = q_{0,2}(\beta, s)$, $s \in [0, \tau]$.

CONDITION C. There exists $\delta > 0$ such that

$$\frac{1}{\sqrt{n}} \sup_{k, s} \left[w(Z_k(s), s) Y_k(s) \{ \| Z_k(s) \| + \| \phi_k(s) \| \} I\{ \beta_0^T Z_k(s) > -\delta \| Z_k(s) \| \} \right]$$
$$= o_p(1).$$

CONDITION D. $\Sigma(\beta_0, \tau)$ is positive definite.

CONDITION E. $L(\beta_0, \cdot) = \sum_{m=1}^{r} \Theta_m(\beta_0) L_m(\beta_0, \cdot), r \geq 1$, where $\Theta_m(\beta_0), m = 1, \dots, r$ are random variables, $L_m(\beta_0, \cdot)$, $m = 1, \dots, r$, are q-variate locally bounded predictable processes for which there exist numbers $\theta_m(\beta_0)$ and (vector) functions $l_m(\beta_0, \cdot)$, $m = 1, \dots, r$, respectively such that

$$\sup_{\substack{\beta \in B \\ s \in [0, \tau]}} \| L_m(\beta_0, s) - l_m(\beta_0, s) \| = o_p(1); \qquad \Theta_m(\beta_0) - \theta_m(\beta_0) = o_p(1)$$

as $n \to \infty$,

and the function $l(\beta_0, s) = \sum_{m=1}^{r} \theta_m(\beta_0) l_m(\beta_0, s)$ is bounded on $[0, \tau]$.

CONDITION F. There exists a (matrix) function $l_1(\cdot, \cdot)$ such that $l_1(\cdot, s)$ is continuous in $\beta \in B$, uniformly in $s \in [0, \tau]$, $l_1(\beta_0, \cdot)$ is bounded and

$$\sup_{\substack{\beta \in B \\ s \in [0, \tau]}} \left\| \frac{\partial}{\partial \beta} L(\beta, s) - l_1(\beta, s) \right\| = o_p(1).$$

Conditions A–F correspond to the standard asymptotic stability and regularity assumptions commonly used in the study of Cox's regression model for counting processes [cf. Andersen and Gill (1982)]. Note that in the case of i.i.d. observations, Conditions E and F are easy to establish for the following previously mentioned examples.

EXAMPLES. (a) $L(\beta, s) = l(s)$, $s \in [0, \tau]$, where l is a q-variate bounded deterministic function.

(b) $L(\beta, s) = l(s) K(\beta, s)$, $s \in [0, \tau]$, where $K(\beta, \cdot)$ is a predictable and locally bounded process of the form $K(\beta, s) = (1/n) \sum Y_i(s) W(\beta, Z_i(s))$. Here $W(\beta, z)$ and $(\partial/\partial\beta) W(\beta, z)$ are continuous in z. If l is a scalar then K is a q-variate vector and vice versa.

(c) $L(\beta, s) = K(\beta, \tau) - K(\beta, s)$, $s \in [0, \tau]$, with K of (b).

The following theorem establishes the asymptotic behavior, under conditions A–F, of the processes based on the general class of martingale residuals of Section 2.

THEOREM 3.1. *The process* $\psi(\hat{\beta}_w, \cdot)$ *of* (6) *converges weakly in* $D[0, \tau]$ *as* $n \to \infty$ *to a zero-mean continuous Gaussian process* $\Psi(\cdot) = \Gamma_1(\cdot) - B(\beta_0, \cdot)^T \Sigma(\beta_0, \tau)^{-1} \Gamma_2(\tau)$, *say, where* $\Gamma(\cdot) = (\Gamma_1(\cdot), \Gamma_2(\cdot))$ *is a* $(p+1)$*-variate zero-mean continuous Gaussian martingale with covariance function of the form given by*

$$\begin{aligned} \operatorname{Cov}[\Gamma_1(s), \Gamma_1(t)] &= A(\beta_0, \min(s, t)), \\ \operatorname{Cov}[\Gamma_1(s), \Gamma_2(t)] &= B(\beta_0, \min(s, t)), \\ \operatorname{Cov}[\Gamma_2(s), \Gamma_2(t)] &= \Sigma(\beta_0, \min(s, t)), \qquad s, t \in [0, \tau]. \end{aligned} \tag{7}$$

Note that in view of (7) the covariance function of the limit process Ψ of Theorem 3.1 has the form

$$\operatorname{Cov}[\Psi(s),\Psi(t)] = A(\beta_0,\min(s,t)) - B(\beta_0,s)^T\Sigma(\beta_0,\tau)^{-1}B(\beta_0,t), \quad s,t\in[0,\tau].$$

Obviously, Theorem 3.1 establishes the asymptotic randomness in the graph based on the trajectory of $\psi(\hat{\beta}_w,\cdot)$ which can be informally used for model checking [cf. Arjas (1988)]. On the other hand, it gives a possibility of constructing some formal goodness-of-fit tests. Indeed, by using Theorem 3.1 one can obtain that the statistic

$$(8)\quad K^2(t) = \frac{\psi(\hat{\beta}_w,t)^2}{\hat{A}(\hat{\beta}_w,\tau) - \hat{B}(\hat{\beta}_w,t)^T\hat{\Sigma}(\hat{\beta}_w,\tau)^{-}\hat{B}(\hat{\beta}_w,t)}, \qquad t\in(0,\tau]$$

has a limiting χ^2 distribution with one degree of freedom. Note that following the considerations of Schoenfeld (1980), Hjort (1990), McKeague and Utikal (1991), Marzec (1993) and using the above theorem, other χ^2-type tests may also be constructed [see also Li and Doss (1993)]. Theorem 3.1 provides also a tool for the formal constructions of KS and CM-type goodness-of-fit tests under some model restrictions.

First, observe that if $B(\beta_0,\cdot)\equiv 0$ then the limit process Ψ of the theorem corresponds to a time transformed standard Brownian motion W. Then it can be easily shown that

$$(9)\quad \begin{gathered} S = \hat{A}(\hat{\beta}_w,\tau)^{-1/2}\sup\left\{\left|\psi(\hat{\beta}_w,t)\right| : t\in[0,\tau]\right\},\\ \int_0^\tau\left|\psi(\hat{\beta}_w,t)\right| d\hat{A}(\hat{\beta}_w,t)/\hat{A}(\hat{\beta}_w,\tau)^{3/2},\\ \int_0^\tau\psi(\hat{\beta}_w,t)^2\, d\hat{A}(\hat{\beta}_w,t)/\hat{A}(\hat{\beta}_w,\tau)^2 \end{gathered}$$

are asymptotically distributed as

$$\sup_{s\in[0,1]}|W(s)|,\ \int_0^1|W(s)|\,ds,\ \int_0^1 W(s)^2\,ds,$$

respectively. Note that $B(\beta_0,\cdot)\equiv 0$ is fulfilled when, for example, (Y_i, Z_i, ϕ_i) are independent copies of (Y, Z, ϕ), say, and ϕ and (Y, Z) are independent processes. This is the common assumption in randomized clinical trials if ϕ has a status of the added model covariate [cf. Tsatis, Rosner and Tritchler (1985), Slud (1991)]. For other possibilities see Arjas (1988) and Marzec and Marzec (1993). Moreover, both KS and CM-type test statistics can also be constructed in the one-parameter case. Given $l(\beta_0,\cdot)\equiv l(\beta_0)$, $\phi_i = Z_i$, the limit process Ψ is a time-transformation of the Brownian bridge W^0. Then

$$\sup\Big\{\big|\psi(\hat{\beta}_w, t)\big|: t \in [0, \tau]\Big\}\Big/\Big[\big|L(\hat{\beta}_w, \tau)\big|\hat{\Sigma}(\hat{\beta}_w, \tau)^{1/2}\Big],$$

$$\int_0^\tau \big|\psi(\hat{\beta}_w, t)\big|\, d\hat{\Sigma}(\hat{\beta}_w, t)\Big/\Big[\big|L(\hat{\beta}_w, \tau)\big|\hat{\Sigma}(\hat{\beta}_w, \tau)^{3/2}\Big],$$

$$\int_0^\tau \psi(\hat{\beta}_w, t)^2\, d\hat{\Sigma}(\hat{\beta}_w, t)\Big/\big[L(\hat{\beta}_w, \tau)\hat{\Sigma}(\hat{\beta}_w, \tau)\big]^2$$

are asymptotically distributed as

$$\sup_{s \in [0,1]} |W^0(s)|,\ \int_0^1 |W^0(s)|\, ds,\ \int_0^1 W^0(s)^2\, ds,$$

respectively. The above construction of the supremum-type test based on the score process ($w \equiv 1$) corresponds to the well-known results of Wei (1984), Haara (1987) and Thernau, Grambsch and Fleming (1990) obtained in the case of time independent covariates. Obviously, large values of all mentioned test statistics are significant for the rejection of model (2).

In a general situation, however, the limit process Ψ of Theorem 3.1 is neither a Gaussian martingale nor a process with well-known properties, which leads to some substantial problems in the direct construction of KS and CM-type tests. Hence in what follows, to omit these difficulties, we will use the innovation approach of Khmaladze (1981, 1988, 1993) and propose a formal method for the construction of tests of the above-mentioned type.

Now we make the following assumptions.

CONDITION A′.

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\int_{\tau-\varepsilon}^{\tau} \lambda_0(s)\, ds > 0 \quad \text{and} \quad \overline{\lim_{\varepsilon \to 0}}\, \lambda_0(\tau - \varepsilon) < \infty.$$

CONDITION B′. The functions of Condition B satisfy

$$\sup_{s \in [0,\tau]} \|Q_{k,l}(\beta_0, s) - q_{k,l}(\beta_0, s)\| = O_p\left(\frac{1}{\sqrt{n}}\right), \qquad k, l = 0, 1, 2,\ k + l \le 2.$$

CONDITION C′. There exists a constant C such that for each i,

$$w(Z_i(s), s)\big[\|\phi_i(s)\| + \|Z_i(s)\|\big] \le Cw(Z_i(s), s).$$

CONDITION D′. The matrices $\zeta(\beta_0, s)$, $s \in [0, \tau)$, $\sigma(\beta_0, \tau)$ are nonsingular.

CONDITION E′. The functions of Condition E satisfy

$$\sup_{s \in [0,\tau]} \|L(\beta_0, s) - l(\beta_0, s)\| = O_p\left(\frac{1}{\sqrt{n}}\right).$$

CONDITION E″. $\overline{\lim}_{s\to\tau-}(\|l(\beta_0, s)\|/(\tau - s)^{\varepsilon}) < \infty$ for some $\varepsilon \in (0,1]$.

REMARK. Note that if (Y_i, Z_i, ϕ_i) are i.i.d. replicates of (Y, Z, ϕ), where Y corresponds to the standard indicator risk process which is generated by a random variable with continuous distribution function and possibly multiplied by a $\{0,1\}$ random variable, Z and ϕ are left-continuous piecewise constant processes according to a finite nonrandom division of the time interval $[0,\tau]$, then Condition B′ is clearly satisfied. This can be established by applying the central limit theorem of Hahn (1978). Conditions A′ and D′ are technical. Note that in the case of i.i.d. observations (N_i, Y_i, Z_i) and $p = 1$, Condition D′ is implied by $\mathrm{Var}\{Z(\tau)|Y(\tau)w(Z(\tau),\tau)\exp[\beta_0 Z(\tau)]\} > 0$. Now we give some comments on Conditions E′ and E″ according to examples (a), (b), (c) for the weighting process $L(\beta,\cdot)$. If the function l of examples (a) and (b) satisfies E″ then obviously the same holds for the corresponding function $l(\beta_0,\cdot) = l(\cdot)k(\beta_0,\cdot)$ where $k(\beta_0,\cdot)$ is the limit in probability of $K(\beta_0,\cdot)$. Condition E′, in view of the structure of $K(\beta_0,\cdot)$, is analogical to Condition B′. To discuss $L(\beta_0,\cdot)$ of the form of example (c), we consider the case of i.i.d. observations (N_i, Y_i, Z_i). Then $l(\beta_0, s) = E[Y(\tau) - Y(s)]W(\beta_0, Z(\tau)) + EY(s)[W(\beta_0, Z(\tau)) - W(\beta_0, Z(s))]$. If W is a constant then Condition E″ requires that $E|Y(\tau) - Y(s)|$ be ε-Lipschitz in a neighborhood of τ. This obviously holds with $\varepsilon = 1$ for Y discussed in this Remark, provided that the random variable generating the risk process is absolutely continuous with density bounded in a neighborhood of τ. The same holds if W is bounded with $Z(\tau) = Z(s)$ in a neighborhood of τ as in the proportional hazards model. In the most general situation, Condition E″ may be guaranteed by the Lipschitz property of $W(\beta_0,\cdot)$ together with the assumption concerning the ε-Lipschitz property of $E\|Z(\tau) - Z(s)\|$ in a neighborhood of τ.

An essential step in constructing tests of the KS and CM type is the following lemma. It provides a modified process of Ψ, being the limit in Theorem 3.1, which appears to be the transformed standard Brownian motion. The structure of the new process can be obtained by replacing Ψ by this process minus its compensator [see Khmaladze (1981), and Andersen, Borgan, Gill and Keiding (1993)]. Under Conditions A, B, D′ and (7) we have the lemma.

LEMMA 3.2. *The process*

$$G(\cdot) = \Gamma_1(\cdot) - \int_0^{(\cdot)} [\Gamma_2(\tau) - \Gamma_2(s)]^T \zeta(\beta_0, s)^{-1} \times b(\beta_0, s) q_{0,0}(\beta_0, s)\lambda_0(s)\, ds \tag{10}$$

is a zero-mean continuous Gaussian martingale with variance function

$$\mathrm{Var}[G(t)] = A(\beta_0, t), \qquad t \in [0,\tau]. \tag{11}$$

The modified version of the process of (6) is defined as follows:

$$\gamma(\hat{\beta}_w, t) = \psi(\hat{\beta}_w, t) - \int_0^t [U(\hat{\beta}_w, \tau) - U(\hat{\beta}_w, s-)]^T \hat{\zeta}(\hat{\beta}_w, s)^- \hat{b}(\hat{\beta}_w, s) \frac{d\overline{N}_w(s)}{n}, \quad t \in [0, \tau], \tag{12}$$

where the process U is given by (4).

Then under Conditions A, A′, B, B′, C′, D′, E, E′, E″ and F, we have the following theorem.

THEOREM 3.3. *The process* $\gamma(\hat{\beta}_w, \cdot)$ *of* (12) *converges weakly in* $D[0, \tau]$ *as* $n \to \infty$ *to a zero-mean continuous Gaussian martingale* $G(\cdot)$ *of* (10).

3.2. *Kolmogorov–Smirnov and Cramér–von Mises type goodness-of-fit tests.* Given $m \geq 1$, consider the weight functions $w^{(k)}(z, s)$, $(z, s) \in R^p \times [0, \tau]$, $k = 1, \dots, m$, with the corresponding quantities $U^{(k)}(\beta, \cdot)$, $\hat{\beta}^{(k)}$, $\gamma^{(k)}(\beta, \cdot)$, $Q_{0,0}^{(k)}(\beta, \cdot)$, $A^{(k)}(\beta, \cdot)$, $a^{(k)}(\beta, \cdot)$, $\Gamma^{(k)}(\cdot)$, $\overline{N}^{(k)}(\cdot)$, and so on, denoted previously for $m = 1$ as $U(\beta, \cdot)$, $\hat{\beta}_w$, $\gamma(\beta, \cdot)$, $Q_{0,0}(\beta, \cdot)$, $A(\beta, \cdot)$, $a(\beta, \cdot)$, $\Gamma(\cdot)$, $\overline{N}_w(\cdot)$ and so on, respectively, such that

$$w^{(k)} w^{(l)} = 0 \quad \text{where } k, l \in \{1, \dots, m\},\ k \neq l. \tag{13}$$

Note that when $w^{(k)}(z, s) = I\{(z, s) \in A_k\}$, $A_k \subset R^p \times [0, \tau]$, $k = 1, \dots, m$, are Borel-measurable subsets and $A_i \cap A_j = \emptyset$ for $i \neq j, i,\ j \in \{1, \dots, m\}$ then (13) is satisfied. For some practical remarks concerning the partition of the covariate space, see Schoenfeld (1980) and Marzec (1993).

Under the assumptions of Theorem 3.3, fulfilled according to each weight $w^{(k)}$, $k = 1, \dots, m$, $m \geq 1$ and (11) we have the following theorem.

THEOREM 3.4.

(a)
$$\max_{1 \leq k \leq m} \left\{ \frac{\sup\left\{ \left| \gamma^{(k)}(\hat{\beta}^{(k)}, s) \right| : s \in [0, \tau] \right\}}{\left[\hat{A}^{(k)}(\hat{\beta}^{(k)}, \tau) \right]^{1/2}} \right\} \to_D Z_1^{(m)},$$

(b)
$$\max_{1 \leq k \leq m} \left\{ \frac{\int_0^\tau \left| \gamma^{(k)}(\hat{\beta}^{(k)}, s) \right| d\hat{A}^{(k)}(\hat{\beta}^{(k)}, s)}{\left[\hat{A}^{(k)}(\hat{\beta}^{(k)}, \tau) \right]^{3/2}} \right\} \to_D Z_2^{(m)},$$

(c)
$$\max_{1 \leq k \leq m} \left\{ \frac{\int_0^\tau \gamma^{(k)}(\hat{\beta}^{(k)}, s)^2 \, d\hat{A}^{(k)}(\hat{\beta}^{(k)}, s)}{\left[\hat{A}^{(k)}(\hat{\beta}^{(k)}, \tau) \right]^2} \right\} \to_D Z_3^{(m)},$$

where

$$P\{Z_1^{(m)} \le x\} = \left[P\left\{\sup_{s\in[0,1]} |W(s)| \le x\right\}\right]^m,$$

$$P\{Z_2^{(m)} \le x\} = \left[P\left\{\int_0^1 |W(s)|\, ds \le x\right\}\right]^m,$$

$$P\{Z_3^{(m)} \le x\} = \left[P\left\{\int_0^1 W(s)^2\, ds \le x\right\}\right]^m, \qquad x \ge 0$$

and W is a standard Brownian motion.

Obviously, the above result is the basis for the formal construction of the critical regions of KS and CM-type goodness-of-fit tests.

4. Power considerations. In this section, we briefly discuss results about the distributions of $\psi(\hat{\beta}, \cdot)$ and $\gamma(\hat{\beta}_w, \cdot)$ of (6) and (12) outside model conditions, which are relevant for power function considerations. At first we derive the asymptotic distribution of $\psi(\hat{\beta}_w, \cdot)$ by considering a sequence of contiguous models (alternatives) indexed by n.

Given n, we assume that under the alternative $H_A^{(n)}$ the stochastic intensity process of $\mathbf{N}^{(n)} = (N_1^{(n)}, \dots, N_n^{(n)})$ equals $\bar{\lambda}^{(n)} = (\bar{\lambda}_1^{(n)}, \dots, \bar{\lambda}_n^{(n)})$, where

$$\bar{\lambda}_i^{(n)}(t) = Y_i(t)\exp\left[\beta_0^T Z_i(t) + n^{-1/2}\delta g_i(\beta_0, t)\right]\lambda_0(t), \tag{14}$$

$t \in [0, \tau]$, $i = 1, \dots, n$, $\delta \ne 0$. Here $g_i(\beta_0, \cdot)$ $(i = 1, \dots, n)$ are uniformly bounded predictable processes.

REMARK. Note that if $g_i(\beta_0, \cdot)$ $(i = 1, \dots, n)$ are independent of β_0, then they may have the status of additional covariates not considered in the model (2). On the other hand, if $g_i(\beta_0, \cdot) = g(\beta_0, Z_i(\cdot))$ for a Borel measurable function then (14) describes the local misspecification of (2) based on the nonlog linear hazards form. Indeed, if we consider a $(2p+1)$-dimensional real function $\bar{g}(\beta_0, \eta, z)$, where $\eta = \eta_0$ corresponds to $\beta_0^T z$, then (14) can be viewed in the Pitman-like framework as $Y_i(t)\exp[\bar{g}(\beta_0, \eta_0 + n^{-1/2}\delta, Z_i(t))]\lambda_0(t)$. Here, by a Taylor expansion $g(\beta_0, z)$, is approximately the derivative of $\bar{g}(\beta_0, \eta, z)$ with respect to η and evaluated at η_0.

Now assume that the asymptotic stability assumptions B, C, E, F are also valid under the sequence of models given by (14) [i.e., under the probability measures leading to the intensities of (14)] and that under (14),

$$T_{k,l}(\beta_0, t) = \frac{1}{n}\sum_{i=1}^n Y_i(t) w(Z_i(t), t)\phi_i(t)^{\otimes k} Z_i(t)^{\otimes 1} \exp\left[\beta_0^T Z_i(t)\right] g_i(\beta_0, t)$$

converges in probability to the bounded function $t_{k,l}(\beta_0, t)$, uniformly in $t \in [0, \tau]$ $(k + l \le 1$ or $k = 0$ and $l = 2)$. Moreover, let Z_i $(i = 1, \dots, n)$ be uniformly bounded.

Then we have the following result.

THEOREM 4.1. *Under the sequence of local alternatives* (14), $\psi(\hat{\beta}_w, \cdot)$ *of* (6) *converges weakly in* $D[0, \tau]$ *as* $n \to \infty$ *to*

$$\begin{aligned}\overline{\Psi}(\cdot) = \Psi(\cdot) + \delta\Big\{\int_0^{(\cdot)} & l(\beta_0, s)^T [t_{1,0}(\beta_0, s) - q_{1,0}(\beta_0, s) \\ & \times t_{0,0}(\beta_0, s)/q_{0,0}(\beta_0, s)]\lambda_0(s)\,ds \\ & - B(\beta_0, \cdot)^T \Sigma(\beta_0, \tau)^{-1} \int_0^{\tau} [t_{0,1}(\beta_0, s) - q_{0,1}(\beta_0, s) t_{0,0} \\ & \qquad \times (\beta_0, s)/q_{0,0}(\beta_0, s)]\lambda_0(s)\,ds\Big\},\end{aligned}$$

where Ψ *is given in Theorem* 3.1.

REMARK. If $p = 1$, $l(\beta_0, s) \equiv 1$, $g_i(\beta_0, s) \equiv \phi_i(s)$ then $\overline{\Psi}(\cdot) =_D W(A(\beta_0, \cdot)) + \delta A(\beta_0, \cdot)$ is the transformed Brownian motion with drift δ. This concerns the situation, for example, when ϕ_i is the indicator for the treatment group for the ith item and (14) describes the local alternative to the null hypothesis of (2) of the no treatment effect on survival [cf. Lagakos and Schoenfeld (1984), Tsatis, Rosner and Tritchler (1985) and Slud (1991)].

The result of Theorem 4.1 can be used to derive the optimal test within the class of χ^2-type texts indexed by L and based on the statistics $K^2(t)$ of (8). Let $q = 1$. First observe that by using Theorem 4.1 it is easy to establish that under (14), $K^2(t)$ converges in distribution to $U^2(t)$, say, where $U(t) - E\overline{\Psi}(t)/\{\mathrm{Var}[\overline{\Psi}(t)]\}^{1/2}$ has the standard normal distribution. So the problem of maximizing the local asymptotic power of the test based on $K^2(t)$ with respect to L is equivalent to that of maximizing $[E\overline{\Psi}(t)]^2/\mathrm{Var}[\overline{\Psi}(t)]$.

Note that by Theorem 4.1 we can write

$$\begin{aligned}\left[E\overline{\Psi}(t)\right]^2 = \delta^2\Big\{\int_0^t & l_0(s)\nu_0(s)\lambda_0(s)\,ds \\ & - \left[\int_0^t l_0(s)\gamma_0(s)^T \lambda_0(s)\,ds\right] \\ & \times \Sigma(\beta_0, \tau)^{-1} \int_0^{\tau} \alpha_0(s)\lambda_0(s)\,ds\Big\}^2,\end{aligned}$$

$$\mathrm{Var}\left[\overline{\Psi}(t)\right] = \langle l_0, l_0 \rangle,$$

where $\langle \cdot, \cdot \rangle$ stands for a bona fide inner product of the form

$$\begin{aligned}\langle f, g\rangle = \int_0^t & f(s) g(s)\eta_0(s)\lambda_0(s)\,ds \\ & - \left[\int_0^t f(s)\gamma_0(s)^T\lambda_0(s)\,ds\right] \\ & \times \Sigma(\beta_0, \tau)^{-1} \int_0^t g(s)\gamma_0(s)\lambda_0(s)\,ds\end{aligned}$$

and

$$
\begin{aligned}
l_0(s) &= l(\beta_0, s),\\
\eta_0(s) &= q_{2,0}(\beta_0, s) - q_{1,0}(\beta_0, s)^2/q_{0,0}(\beta_0, s),\\
\nu_0(s) &= t_{1,0}(\beta_0, s) - q_{1,0}(\beta_0, s)t_{0,0}(\beta_0, s)/q_{0,0}(\beta_0, s),\\
\gamma_0(s)^T &= q_{1,1}(\beta_0, s) - q_{1,0}(\beta_0, s)q_{0,1}(\beta_0, s)^T/q_{0,0}(\beta_0, s),\\
\alpha_0(s) &= t_{0,1}(\beta_0, s) - q_{0,1}(\beta_0, s)t_{0,0}(\beta_0, s)/q_{0,0}(\beta_0, s).
\end{aligned} \tag{15}
$$

An inspection shows that $E\overline{\Psi}(t) = \delta\langle l_0, \nu_0/\eta_0 - \kappa_0(t)^T\gamma_0/\eta_0\rangle$, where

$$
\begin{aligned}
\kappa_0(t) = &\left[\Sigma(\beta_0, \tau) - \int_0^t \gamma_0(s)^{\otimes 2}\lambda_0(s)/\eta_0(s)\,ds\right]^{-1}\\
&\times\left[\int_0^\tau \alpha_0(s)\lambda_0(s)\,ds - \int_0^t \gamma_0(s)\nu_0(s)\lambda_0(s)/\eta_0(s)\,ds\right].
\end{aligned} \tag{16}
$$

Thus by the Cauchy–Schwarz inequality $[E\overline{\Psi}(t)]^2/\mathrm{Var}[\overline{\Psi}(t)]$ is maximal when $l_0 \propto \nu_0/\eta_0 - \kappa_0(t)^T\gamma_0/\eta_0$. This leads to the optimal weight process

$$
L(\hat{\beta}_w, s) = \hat{\nu}(s)/\hat{\eta}(s) - \hat{\kappa}(t)^T\hat{\gamma}(s)/\hat{\eta}(s),
$$

where the forms of $\hat{\nu}, \hat{\eta}, \hat{\gamma}, \hat{\alpha}$ correspond to that of $\nu_0, \eta_0, \gamma_0, \alpha_0$ of (15) with $q_{k,l}(\beta_0, s), t_{k,l}(\beta_0, s)$ replaced by $Q_{k,l}(\hat{\beta}_w, s), T_{k,l}(\hat{\beta}_w, s)$. By replacing, in the form of (16), the quantities $\Sigma(\beta_0, \tau)$, ν, γ_0, η_0, α_0, $\lambda_0(s)\,ds$ by $\hat{\Sigma}(\hat{\beta}_w, \tau)$, $\hat{\nu}$, $\hat{\gamma}$, $\hat{\alpha}$, $[nQ_{0,0}(\hat{\beta}_w, s)]^{-1}\,d\overline{N}_w(s)$, respectively, $\hat{\kappa}(t)$ is defined.

EXAMPLES. (a) If $\phi_i = Z_i$ then $\alpha_0 = \nu_0$, $\eta_0 = \gamma_0$ and the optimal $l_0 \propto \nu_0/\eta_0 - [\int_t^\tau \eta_0(s)\lambda_0(s)\,ds]^{-1}\int_t^\tau \nu_0(s)\lambda_0(s)\,ds$.

(b) When $\phi_i \in \{0,1\}$, $g_i(\beta_0, \cdot) = \phi_i(\cdot)$ as in the previous Remark, then $\alpha_0 = \gamma_0$, $\eta_0 = \nu_0$ so that the optimal $l_0 \propto 1 - [\Sigma(\beta_0, \tau) - \int_0^t \gamma_0(s)^{\otimes 2}\lambda_0(s)/\eta_0(s)\,ds][\int_t^\tau \gamma_0(s)\lambda_0(s)\,ds]\gamma_0/\eta_0$. If, moreover, (Y_i, Z_i, ϕ_i) are independent replicates of (Y, Z, ϕ) and ϕ_i and (Y_i, Z_i) are independent, then $\gamma_0 = 0$ and consequently our $l_0 \propto 1$.

REMARK. One can use the methods of the proofs of Theorems 3.3 and 4.1 to establish that under the sequence of local alternatives given by (14), the process $\gamma(\hat{\beta}_w, \cdot)$ of (12) tends in distribution to $G(\cdot) + \delta\omega(\cdot)$, where

$$
\begin{aligned}
\omega(t) = &\int_0^t l(\beta_0, s)^T[t_{1,0}(\beta_0, s) - q_{1,0}(\beta_0, s)t_{0,0}(\beta_0, s)/q_{0,0}(\beta_0, s)]\lambda_0(s)\,ds\\
&- \int_0^t \left\{\int_s^\tau [t_{0,1}(\beta_0, u) - q_{0,1}(\beta_0, u)t_{0,0}(\beta_0, u)/q_{0,0}(\beta_0, u)]\lambda_0(u)\,du\right\}^T\\
&\times \zeta(\beta_0, s)^{-1} b(\beta_0, s) q_{0,0}(\beta_0, s)\lambda_0(s)\,ds, \qquad t \in [0, \tau],
\end{aligned}
$$

and G is the limiting process of Theorem 3.3. In view of the highly complex form of $\omega(\cdot)$, the optimization of the power functions of the KS and CM-type tests of Theorem 3.4 becomes here an extremely difficult problem.

Now consider the case of a fixed alternative and assume that in the true state of affairs $\exp[\beta_0^T Z_i(t)]\lambda_0(t)$ is replaced by a random function $h_i(t)$. This means that the true intensity process of N_i equals $Y_i h_i$, $i = 1, \dots, n$. Our additional notation specifies $Q_{k,l}(t)$ to be of the form corresponding to $Q_{k,l}(\beta, t)$ but with $\exp[\beta^T Z_i(t)]$ replaced by $h_i(t)$. Moreover, $q_{k,l}(t)$ denotes the limit in probability of $Q_{k,l}(t)$. It should be noted [Hjort (1992), Sasieni (1993b)] that under a possibly misspecified Cox-type model of (2), the MWPLE $\hat{\beta}_w$ converges in probability to a p-vector of constants β^* (depending on w), which is the unique solution to the system of p equations

$$\int_0^\tau [q_{0,1}(s) - q_{0,1}(\beta^*, s)q_{0,0}(s)/q_{0,0}(\beta^*, s)]\, ds = 0. \tag{17}$$

This can be proved by using the condition of the form corresponding to that of Condition B but with β_0 replaced by β^*, where we also postulate the assumptions for $Q_{k,l}(\cdot)$ and $q_{k,l}(\cdot)$, quite analogous to these for $Q_{k,l}(\beta, \cdot)$ and $q_{k,l}(\beta, \cdot)$ and the condition corresponding to D which concerns the positive-definiteness of the matrix $\Sigma^*(\beta^*, \tau)$. Here

$$\Sigma^*(\beta, t) = \int_0^t \left[q_{0,2}(\beta, s)/q_{0,0}(\beta, s) - q_{0,1}(\beta, s)^{\otimes 2}/q_{0,0}(\beta, s)^2\right] q_{0,0}(s)\, ds, \qquad t \in [0, \tau].$$

Under the assumptions defined analogously to $A' - E'$, E, F, but with $\lambda_0(s)$ and β_0 replaced by $q_{0,0}(s)$ and β^* [with $\zeta^*(\beta^*, s) = \Sigma^*(\beta^*, \tau) - \Sigma^*(\beta^*, s)$ in the condition corresponding to D'], under the requirement of the convergence of $Q_{k,l}(\cdot)$ to $q_{k,l}(\cdot)$ of the form specified in B' for $Q_{k,l}(\beta_0, \cdot)$ and $q_{k,l}(\beta_0, \cdot)$, and in view of (2), (12) and (17), we have the following result.

THEOREM 4.2. *Uniformly in* $t \in [0, \tau]$, $n^{-1/2}\gamma(\hat{\beta}_w, t)$ *converges in probability, as* $n \to \infty$ *to* $\gamma(t)$, *where*

$$\begin{aligned} \gamma(t) &= \int_0^t \Big[l(\beta^*, s)^T f(\beta^*, s) \\ &\qquad + u(\beta^*, s)^T \zeta^*(\beta^*, s)^{-1} b(\beta^*, s)\Big] q_{0,0}(s)\, ds, \\ f(\beta^*, s) &= q_{1,0}(s)/q_{0,0}(s) - q_{1,0}(\beta^*, s)/q_{0,0}(\beta^*, s), \\ u(\beta^*, s) &= \int_0^s [q_{0,1}(t) - q_{0,1}(\beta^*, t)q_{0,0}(t)/q_{0,0}(\beta^*, t)]\, dt. \end{aligned} \tag{18}$$

REMARK. For $n^{-1/2}\psi(\hat{\beta}_w, t)$ of (6), the similar result holds if we only use the assumptions corresponding to B, E and F. Then the limit is also of the form of (18), where we put $b(\beta^*, s) = 0$.

Now by using Theorem 4.2, we can state that the tests of KS and CM type of Theorem 3.4 (let $m = 1$ without loss of generality) are consistent against any model misspecification under which $\gamma(t)$ of (18) is nonzero for some $t \in (0, \tau]$. By combining the two complex structures of (17) and (18) it seems that the above condition is generally satisfied. We shall briefly examine this for the special (but most common in real applications) situations of the monotone departures from the proportional hazards assumption and for the misspecification based on the added covariate [cf. Lin (1991), Lagakos and Schoenfeld (1984), Struthers and Kalblfleisch (1986), Slud (1991)].

EXAMPLES. (a) Let (Y_i, Z_i) be independent replicates of (Y, Z). Given the null hypothesis of the form (2) with $p = 1$, let, under the alternative, the hazard rate structure be of the form $h_i(t) = \exp[\beta(t)Z_i(t)]\lambda_0^*(t)$, where β is an unspecified strictly monotone function of t [see Lin and Wei (1991), Lin (1991), Lin, Wei and Ying (1993)]. Since $q_{0,1}(\beta, t)/q_{0,0}(\beta, t)$ increases in β we have that $q_{0,1}(t)/q_{0,0}(t) = q_{0,1}(\beta^*, t)/q_{0,0}(\beta^*, t)$ at the point where $\beta(t) = \beta^*$. By monotonicity of $\beta(\cdot)$ and in view of (17), this point is unique and belongs to the interval $(0, \tau)$. Now observe that for the modified weighted score-type process $\gamma(\hat{\beta}_w, \cdot)$ of (12) (i.e., when $\phi_i = Z_i$), we have that $f(\beta^*, 0) \neq 0$ (here $q_{0,1}/q_{0,0} = q_{1,0}/q_{0,0}$) and since $u(\beta^*, 0) = 0$, the argument of continuity automatically guarantees that $\gamma(t) \neq 0$ in a neighborhood of zero (assuming that $l(\beta^*, t)q_{0,0}(t) \neq 0$ for t near zero). For the general case of $\phi_i(t) = g(Z_i(t), t)$, where $g\colon R \times [0, \tau] \to R$, consider the binary random variable $Z \in \{0, 1\}$ such that $P\{Z = 1\} = \rho$, $\rho \in (0, 1)$ [Lin and Wei (1991)]. We have again that $f(\beta^*, 0) \neq 0$ since otherwise $\beta(0) = \beta^*$, which is impossible.

(b) Let (Y_i, Z_i, ϕ_i) be independent replicates of (Y, Z, Φ). Suppose that the time hazard rate structure is of the form $h_i(t) = \exp[\beta^T Z_i(t) + \delta\phi_i(t)]\lambda_0^*(t)$, $\delta \neq 0$. Also assume that $\Phi(0)$ and $Z(0)$ are independent and $\Phi(0)$ is the binary variable: $P\{\Phi(0) = 1\} = q = 1 - P\{\Phi(0) = 0\}$, $q \in (0, 1)$ [Lagakos and Schoenfeld (1984), Tsatis, Rosner and Tritchler (1985), Slud (1991)]. Then $f(\beta^*, 0) = [\exp(\delta) - 1]/[q + (1 - q)\exp(\delta)]$. Obviously it is different from zero and consequently we conclude that γ of (18) is nonzero in a neighborhood of zero.

Now we shall discuss the problem of the optimality in the class of the KS-type tests ($m = 1$, $q = 1$) of Theorem 3.4(a). Similarly to the previous considerations, we take the weight process L as the parameter. Unfortunately, the asymptotic distribution of the supremum-type test statistic is not normal and consequently the asymptotic Pitman efficacy approach does not work here. We therefore consider the approximate Bahadur efficacy approach [see Bahadur (1960)]. By Theorem 3.4(a), the KS type-test statistic has the asymptotic distribution function G of $\sup\{|W(t)|\colon t \in [0, 1]\}$. Hence it satisfies the required condition [cf. Aki (1986)] of the form

$$\log[1 - G(x)] = -2^{-1}x^2[1 + o(1)]$$

as $x \to \infty$. On the other hand, it can be easily obtained, by using Theorem 4.2, that under the Cox model misspecification the supremum-type test statistic of Theorem 3.4(a) when multiplied by $n^{-1/2}$ converges in probability as $n \to \infty$ to b_l, where

$$b_l = \frac{\sup\left\{\left|\int_0^t l(\beta^*, s)\bar{f}(\beta^*, s) q_{0,0}(s)\, ds\right| : t \in [0, \tau]\right\}}{\left[\int_0^t l(\beta^*, s)^2 \eta(\beta^*, s) q_{0,0}(s)\, ds\right]^{1/2}}. \tag{19}$$

Here $\bar{f} = f + u(\zeta^*)^{-1}\rho$, $\eta = q_{2,0}/q_{0,0} - (q_{1,0}/q_{0,0})^2$. Thus the approximate Bahadur slope of the test statistic of Theorem 3.4(a) is equal to b_l^2. To maximize b_l^2 with respect to the function l, note that by applying the Cauchy–Schwarz inequality we have in view of (19) that

$$b_l \le \left[\int_0^t \left\{\bar{f}(\beta^*, s)^2 / \eta(\beta^*, s)\right\} q_{0,0}(s)\, ds\right]^{1/2}.$$

This upper bound is attained for the weight function $l^*(\beta^*, s)$, where $l^* \propto \bar{f}/\eta$. Obviously, the natural candidate for the estimator of $l^*(\beta^*, s)$ is

$$L^*(\hat{\beta}_w, s) = \left[\hat{f}(\hat{\beta}_w, s) + n^{-1/2} U(\hat{\beta}_w, s) \hat{\zeta}(\hat{\beta}_w, s)^{-1} \hat{\rho}(\hat{\beta}_w, s)\right] \Big/ \hat{\eta}(\hat{\beta}_w, s),$$

where

$$\hat{f}(\hat{\beta}_w, s) = \hat{q}_{1,0}(s)/\hat{q}_{0,0}(s) - Q_{1,0}(\hat{\beta}_w, s)/Q_{0,0}(\hat{\beta}_w, s),$$

$$\hat{\eta}(\hat{\beta}_w, s) = Q_{2,0}(\hat{\beta}_w, s)/Q_{0,0}(\hat{\beta}_w, s) - \left[Q_{1,0}(\hat{\beta}_w, s)/Q_{0,0}(\hat{\beta}_w, s)\right]^2.$$

Here $\hat{q}_{1,0}(s)/\hat{q}_{0,0}(s)$ denotes the estimator of $q_{1,0}(s)/q_{0,0}(s)$ of the general form $\sum Y_i(s) w(Z_i(s), s)\phi_i(s)\hat{h}_i(s) / \sum Y_i(s) w(Z_i(s), s)\hat{h}_i(s)$, where $\hat{h}_i$ is defined for the special misspecifications of the model (2) as follows:

EXAMPLES. Let λ_0^* be an unspecified baseline hazard function.

(a) $h_i(t) = \lambda(Z_i(t), t)\lambda_0^*(t)$, where λ is a completely specified function. Then $\hat{h}_i(t) = \lambda(Z_i(t), t)$.

(b) $h_i(t) = \exp[h(\beta, Z_i(t)]\lambda_0^*(t)$, where h is a completely specified function. Then $\hat{h}_i(t) = \exp[h(\hat{\beta}, Z_i(t)]$. Here $\hat{\beta}$ denotes the MPLE of β.

(c) $h_i(t) = \exp[\beta^T Z_i(t) + \delta^T W_i(t)]\lambda_0^*(t)$, where W_i is the additional covariate. Then $\hat{h}_i(t) = \exp[\tilde{\beta}_w^T Z_i(t) + \tilde{\delta}_w^T W_i(t)]$, where $(\tilde{\beta}_w, \tilde{\delta}_w)$ is the MPLE of (β, δ) based on the weighted partial likelihood function with covariates (Z_i, W_i).

The main drawback of the above optimality result is that it is not obvious why $l^*(\beta^*, s)$ should generally satisfy a condition corresponding to E″ required for Theorem 4.2. Note, however, that when we confine ourselves in our considerations to the time interval $[0, \tau - \delta]$ for some $\delta > 0$, that is, if we consider the weight $L(\hat{\beta}_w, s)I\{s \in [0, \tau - \delta]\}$, the problem disappears. Then the optimal weight is of the form $L^*(\hat{\beta}_w, s)I\{s \in [0, \tau - \delta]\}$.

REMARK. If under the null hypothesis of (2), $\rho = 0$ ($b = 0$), then one may consider the supremum-type test based on S of (9). Then in view of the Remark following Theorem 4.2, we have that the optimal weight $L^*(\hat{\beta}_w, s) = \hat{f}(\hat{\beta}_w, s)/\hat{\eta}(\hat{\beta}_w, s)$.

5. Simulation and example. A Monte Carlo study was undertaken for the special situation in which the counting processes jump at most once. To reveal the general behavior of the KS and CM-type goodness-of-fit tests we confined ourselves to the same hypothesis and alternatives as those considered by Lin and Wei (1991). Following Lin and Wei (1991), the hazard function under the null hypothesis H concerning the Cox proportional hazards model was of the form $\lambda(t, Z) = \exp(0.2Z)$. Z was taken to have the standard normal law and the censorship was imposed by the generation of independent uniform random variables on the interval $[0, 4]$ which was also chosen as our $[0, \tau]$ finite time interval. Similarly, the two alternatives A_1 and A_2 were specified to be of the form $\lambda(t, Z, Z^*) = \exp(2Z + 0.5Z^*)$ and $\lambda(t, Z) = 1 + 0.5Z$, respectively. In the case of A_1, Z^* was generated independently of Z and so that Z^* took values -1 and $+1$ with probability 0.5. In the case of A_2, Z was truncated at ± 1.98. Moreover, the censoring times came from the uniform distributions on $[0, 7]$ (A_1) and on $[0, 5]$ (A_2). For the robustness study, the covariates appearing in the definitions of the Cox-type estimator and tests were generated from $Z + 5J$ with $J \in \{0, 1\}$ being independent of Z and so that $P\{J = 1\} = 0.3$.

The KS and CM-type goodness-of-fit tests of Theorem 3.4(a) and (b) were chosen to have simple forms: $\phi = Z$, $w = 1$, $L(s) = [\bar{Y}(\tau) - \bar{Y}(s)]^2$ (KS^1, CM^1 tests) and $\phi = Z^*$, $w = 1$, $L = 1$ (KS^2, CM^2 tests). In our study we also considered the KS_w^i and CM_w^i tests being the counterparts of the KS^i and CM^i tests, respectively, defined with the use of nontrivial weight $w = I\{Z \leq 2.5\}$. For comparison, the rejection rates of the tests proposed by Cox (1972), Schoenfeld (1980), Wei (1984) and Lin and Wei (1991) (designated as Cox, Sch, Wei and LW in the tables) were also presented. All of these previously defined tests were proposed to assess the adequacy of the classical Cox's model defined in survival analysis context. For Schoenfeld's test, the time axis and the covariate range were partitioned at the sample median of survival times and at the mean of covariate, respectively. The levels of significance of 0.01 and 0.05 were considered. The experiments were carried out using sample sizes of $n = 100$ and 200. One thousand repetitions were used in each instance. The results from these Monte Carlo studies are summarized in Tables 1 and 2.

Table 1 shows that the KS^i, KS_w^i, CM^i, CM_w^i, $i = 1, 2$, tests maintain their sizes near nominal levels, which reflects the appropriateness of approximations given in Theorem 3.4 for practical use. On the other hand, the omnibus property of these tests is reflected by generally quite adequate powers for detecting the two violations of Cox's model given by A_1 and A_2. It should also be noted that, as expected, the KS^2, KS_w^2, CM^2 and CM_w^2 tests are highly powerful with respect to the alternative A_1. Moreover, the KS_w^1 and CM_w^1

TABLE 1
Empirical sizes and powers of goodness-of-fit tests

Test	***H***				***A*₁**				***A*₂**			
	α = 0.01		**α = 0.05**		**α = 0.01**		**α = 0.05**		**α = 0.01**		**α = 0.05**	
	n		***n***		***n***		***n***		***n***		***n***	
	100	**200**	**100**	**200**	**100**	**200**	**100**	**200**	**100**	**200**	**100**	**200**
LW	0.012	0.009	0.058	0.044	0.095	0.099	0.221	0.254	0.149	0.258	0.290	0.447
Cox	0.015	0.014	0.061	0.059	0.018	0.020	0.074	0.093	0.020	0.022	0.085	0.125
Sch	0.017	0.009	0.061	0.045	0.013	0.024	0.061	0.089	0.013	0.013	0.062	0.056
Wei	0.009	0.008	0.036	0.043	0.011	0.018	0.052	0.076	0.006	0.009	0.034	0.049
KS^1	0.014	0.012	0.061	0.057	0.014	0.021	0.058	0.076	0.019	0.024	0.076	0.086
KS^1_w	0.012	0.011	0.051	0.056	0.023	0.031	0.076	0.095	0.227	0.353	0.412	0.577
KS^2	0.006	0.009	0.041	0.046	0.890	0.998	0.971	0.999	0.010	0.012	0.051	0.056
KS^2_w	0.007	0.011	0.042	0.061	0.841	0.999	0.952	0.999	0.011	0.018	0.071	0.082
CM^1	0.004	0.009	0.062	0.071	0.013	0.025	0.060	0.071	0.018	0.021	0.080	0.085
CM^1_w	0.008	0.009	0.061	0.058	0.022	0.029	0.075	0.089	0.118	0.279	0.305	0.481
CM^2	0.004	0.009	0.040	0.053	0.774	0.981	0.930	0.996	0.013	0.014	0.053	0.052
CM^2_w	0.008	0.011	0.055	0.063	0.841	0.984	0.902	0.999	0.012	0.025	0.049	0.079

tests have significantly larger powers with respect to A_2 then their counterparts KS^1 and CM^1. Table 1 shows that the good competitors, as compared to the LW, Cox, Sch and Wei tests, are for A_1, the KS^2, KS^2_w, CM^2, CM^2_w tests and for A_2, the KS^1_w, CM^1_w tests. Table 2 shows that the weight w, which censors "large" observations among the covariates, makes the tests robust to covariate outliers. The empirical sizes are then more stable and closer to the nominal significance levels. The same remark concerns the powers of the KS^i_w and CM^i_w tests ($i = 1, 2$).

TABLE 2
Empirical sizes and powers of goodness-of-fit tests with covariates subject to measurement error

Test	***H***				***A*₁**				***A*₂**			
	α = 0.01		**α = 0.05**		**α = 0.01**		**α = 0.05**		**α = 0.01**		**α = 0.05**	
	n		***n***		***n***		***n***		***n***		***n***	
	100	**200**	**100**	**200**	**100**	**200**	**100**	**200**	**100**	**200**	**100**	**200**
KS^1	0.037	0.031	0.098	0.089	0.129	0.165	0.207	0.311	0.009	0.012	0.052	0.049
KS^1_w	0.018	0.012	0.069	0.055	0.033	0.051	0.101	0.163	0.157	0.284	0.372	0.521
KS^2	0.021	0.009	0.093	0.080	0.249	0.562	0.470	0.769	0.014	0.014	0.049	0.063
KS^2_w	0.013	0.008	0.054	0.046	0.701	0.912	0.795	0.939	0.012	0.016	0.068	0.089
CM^1	0.029	0.021	0.099	0.079	0.024	0.070	0.092	0.208	0.012	0.012	0.045	0.053
CM^1_w	0.006	0.009	0.095	0.052	0.031	0.042	0.123	0.228	0.125	0.226	0.351	0.404
CM^2	0.024	0.013	0.074	0.081	0.256	0.564	0.478	0.791	0.012	0.023	0.051	0.048
CM^2_w	0.012	0.009	0.054	0.048	0.635	0.890	0.710	0.893	0.015	0.027	0.057	0.081

In the simulation study we considered the simplest examples of the weights L discussed in Section 2. The resulting tests performed quite well in general. Simulation indicated that the KS and CM-type tests with $L(s) = [\bar{Y}(\tau) - \bar{Y}(s)]^\alpha$, $\alpha \geq 0$, $\alpha \in N$, or more generally with the weights discussed in Section 2, should contain representatives which are powerful for concrete alternatives. Moreover, the simulation study confirmed that the true sizes of the KS and CM-type tests for moderate sample sizes and under different levels of censorship are indeed accurately approximated by the nominal significance level based upon the asymptotic distribution results of Section 3.

We have also applied the KS^1 and CM^1 tests to the Stanford heart transplant data given in Miller and Halpern (1982). Lin, Wei and Ying (1993) note that the fit of Cox's model to these data, using the single covariate age—"age at transplant," is not satisfactory. Our tests confirm this. We obtained the p-values of the KS^1 and CM^1 tests of 0.025 and 0.004, respectively. On the other hand, Lin, Wei and Ying (1993) note that considering the two covariates age and $(\text{age})^2$ provides a satisfactory description of the data. Our tests also confirm this phenomenon. The p-values of the KS^1 and CM^1 tests are now equal to 0.412 and 0.526.

6. Generalizations. In this section we shall outline the extensions of the results of Section 3 for the two generalizations of model (2). Since the concepts are similar to that of Sections 2 and 3, the proofs will be omitted and only the main results will be presented. The first generalization discussed concerns the semiparametric model which allows the stochastic intensity to have the general form. The model postulates that the counting process N_i has the stochastic intensity

$$\lambda_i(t) = Y_i(t)\exp[h(\beta_0, Z_i(t))]\lambda_0(t), \qquad t \in [0, \tau], \tag{20}$$

where h is a completely specified Borel-measurable real-valued function, $h(\beta, Z_i(\cdot))$ and $(\partial/\partial\beta)h(\beta, Z_i(\cdot))$ are predictable and locally bounded processes. The above form corresponds to Cox's (1972) intuitive approach and for $h(\beta, z) = h(\beta^T z)$ leads to the Prentice and Self (1983) model. The MWPLE $\hat{\beta}_{w,h}$ is defined by the equality $U_h(\beta, \tau) = 0$, where

$$\begin{aligned} U_h(\beta, t) = \frac{1}{\sqrt{n}} \int_0^t \sum_{i=1}^n w(Z_i(s), s) \\ \times \left\{ \frac{\partial}{\partial\beta} h(\beta, Z_i(s)) - \frac{S_{0,1}(\beta, s)}{S_{0,0}(\beta, s)} \right\} dN_i(s). \end{aligned} \tag{21}$$

Here

$$\begin{aligned} S_{k,l}(\beta, u) = \frac{1}{n} \sum_{i=1}^n w(Z_i(u), u) Y_i(u) \phi_i(u)^{\otimes k} \left[\frac{\partial}{\partial\beta} h(\beta, Z_i(u)) \right]^{\otimes l} \\ \times \exp[h(\beta, Z_i(u))], \end{aligned}$$

$k, l \in \{0, 1, 2\}$, $k + l \leq 2$, and we also define $s_{k,l}(\beta, u)$ as the function being the limit in probability of $S_{k,l}(\beta, u)$, uniformly for β in a neighborhood of β_0 and $u \in [0, \tau]$. Then under some additional standard asymptotic stability and regularity assumptions, including, for example, the asymptotic stability of the observed information matrix $n^{-1/2}(\partial/\partial\beta)U_h(\beta, \tau)$ for β in a neighborhood of β_0 [cf. Arjas and Haara (1988), Marzec (1996)], we have the decomposition

$$\sqrt{n}\left(\hat{\beta}_{w,h} - \beta_0\right) = \Sigma_h(\beta_0, \tau)^{-1} U_h(\beta_0, \tau) + o_p(1), \tag{22}$$

corresponding to the asymptotic normality of $\hat{\beta}_{w,h}$. In the above equality the matrix $\Sigma_h(\beta_0, \tau)$ is of the form given by $\Sigma(\beta_0, \tau)$ but with $q_{0,0}, q_{0,1}, q_{0,2}$ replaced by $s_{0,0}, s_{0,1}, s_{0,2}$, respectively. In the same manner, $\sigma_h(\beta, \tau)$, $\zeta_h(\beta, s)$, $A_h(\beta, s)$, $B_h(\beta, s)$ are defined as the counterparts of the previously used quantities $\sigma(\beta, \tau)$, $\zeta(\beta, s)$, $A(\beta, s)$, $B(\beta, s)$. Similarly, $\hat{\sigma}_h(\beta, \tau)$, $\hat{\zeta}_h(\beta, s)$, $\hat{b}_h(\beta, s)$ and so on are now defined analogously to $\hat{\sigma}(\beta, \tau)$, $\hat{\zeta}(\beta, s)$, $\hat{b}(\beta, s)$ and so on, but with $Q_{k,l}$ replaced here by $S_{k,l}$. For the goodness-of-fit inference with the model (20) we consider the process $\psi_h(\hat{\beta}_{w,h}, \cdot)$ and its modification $\gamma_h(\hat{\beta}_{w,h}, \cdot)$ corresponding to $\psi(\hat{\beta}_w, \cdot)$ of (6) and $\gamma(\hat{\beta}_w, \cdot)$ of (12), respectively, which are now based on the newly defined quantities $S_{0,0}, S_{1,0}, U_h, \hat{\zeta}_h, \hat{b}_h, \hat{\beta}_{w,h}, \psi_h$ put in the place of $Q_{0,0}, Q_{1,0}, U, \hat{\zeta}, \hat{b}, \hat{\beta}_w, \psi$. Starting from (22) we present a list of assumptions that will guarantee the analogous results previously obtained for $\psi(\hat{\beta}_w, \cdot)$ and $\gamma(\hat{\beta}_w, \cdot)$ under the Cox regression model. The counterpart of Theorem 3.1 states that under the model (20), the process $\psi(\hat{\beta}_{w,h}, \cdot)$ converges weakly to a zero-mean continuous Gaussian process, with covariance function of the form

$$A_h(\beta_0, \min(s, t)) - B_h(\beta_0, s)^T \Sigma_h(\beta_0, \tau)^{-1} B_h(\beta_0, t), \quad s, t \in [0, \tau].$$

The following conditions are sufficient for this result. Conditions A, E and F remain unchanged. Conditions B and D are adapted to the present situation as follows. The former concerns the quantities $S_{k,l}$ and their corresponding limits $s_{k,l}$, while the latter relates to $\Sigma_h(\beta_0, \tau)$.

Now Condition C should have the slightly stronger form

$$\frac{1}{\sqrt{n}} \sup_{k,s} \left\{ w(Z_i(s), s) Y_i(s) \left[\left\| \frac{\partial}{\partial\beta} h(\beta, Z_k(s)) \right\| + \| \phi_k(s) \| \right] \right\} = o_p(1).$$

The weak convergence of the process $\gamma_h(\hat{\beta}_{w,h}, \cdot)$ to a zero-mean continuous Gaussian martingale with variance function $A_h(\beta_0, t)$, $t \in [0, \tau]$, that is, the counterpart of Theorem 3.3, can be established by using the following assumptions: Conditions A′, E′, E″, the counterparts of Conditions B′ and D′, where we only replace $Q_{k,l}$, $q_{k,l}$ $\zeta(\beta_0, \cdot)$, $\sigma(\beta_0, \tau)$ by $S_{k,l}$, $s_{k,l}$, $\zeta_h(\beta_0, \cdot)$, $\sigma_h(\beta_0, \tau)$, respectively, the condition corresponding to Condition C′ of the form: there exist a constant C and a neighborhood B of β_0 such that

$$w(Z_i(s),s)\left\{\sup_{\beta\in B}\left\|\frac{\partial^2}{\partial\beta^2}h(\beta,Z_i(s))\right\|+\left\|\frac{\partial}{\partial\beta}h(\beta_0,Z_i(s))\right\|\right.$$
$$\left.+\|h(\beta_0,Z_i(s))\|+\|Z_i(s)\|+\|\phi_i(s)\|\right\}\le Cw(Z_i(s),s).$$

The following assumption is also required: there exists a matrix-valued function defined in a neighborhood of β_0 and on $[0,\tau]$ which is the limit in probability of the observed information matrix-valued function $n^{-1/2}(\partial/\partial\beta)U_h(\beta,x)$, uniformly in β from a neighborhood of β_0 and $x\in[0,\tau]$. Obviously, in the statement of the counterpart of Theorem 3.4, one should replace $\hat{A}^{(k)}$, $\hat{\beta}^{(k)}$ by $\hat{A}_h^{(k)}$, $\hat{\beta}_h^{(k)}$.

Another useful generalization of the model (2) is the following multistate Cox's-type regression model of Andersen, Hansen and Keiding (1991) [see also Andersen and Borgan (1985), Marzec and Marzec (1996)]. Given n processes $X_i(t)$, $t\in[0,\tau]$, $i=1,\dots,n$, with state space $\{1,\dots,M\}$, $M\ge 2$, let $N_{hji}(t)$ be the observed number of direct transitions from h to j of X_i during the time interval $[0,\tau]$ and let $Y_{hi}(t)$ indicate if X_i was observed to be in state h at time $t-$. The model of interest postulates that the multivariate counting process $(N_{hji}\colon h,j=1,\dots,M,\ h\ne j,\ i=1,\dots,n)$ has the stochastic intensity process $(\lambda_{hji}\colon h,j=1,\dots,M,\ h\ne j,\ i=1,\dots,n)$, where

$$\lambda_{hji}(t)=Y_{hi}(t)\exp\left[\beta_{hj}^TZ_i(t)\right]\lambda_{hj0}(t),\qquad t\in[0,\tau]. \tag{23}$$

Here the parameters λ_{hj0}, β_{hj} correspond to λ_0, β_0 of model (2) and Z_i is the covariate process associated with the ith individual. In our present notation $Q_{k,l}^{(h)}$ and $q_{k,l}^{(h)}$ are defined similarly to $Q_{k,l}$ and $q_{k,l}$ but with Y_i replaced here by Y_{hi}. Moreover, let U_{hj}, ζ_{hj}, ρ_{hj}, $\hat{\zeta}_{hj}$, and so on be the counterparts of U, ζ, ρ, $\hat{\zeta}$ and so on, obtained by substituting $Q_{k,l}^{(h)}$, λ_{hj0}, $q_{k,l}^{(h)}$, N_{hji} in place of $Q_{k,l}$, λ_0, $q_{k,l}$, N_i respectively. Thus according to (4) the possibly weighted Cox type estimator $\hat{\beta}_{hj}$ of β_{hj} is defined by the equality $U_{hj}(\beta,\tau)=0$, $h,j=1,\dots,M$, $h\ne j$. The goodness-of-fit procedures for the model (23) are now based on the processes $(\psi_{hj}(\hat{\beta}_{hj},\cdot)\colon h,j=1,\dots,M,\ h\ne j)$ and $(\gamma_{hj}(\hat{\beta}_{hj},\cdot)\colon h,j=1,\dots,M,\ h\ne j)$, where $\psi_{hj}(\hat{\beta}_{hj},\cdot)$ and $\gamma_{hj}(\hat{\beta}_{hj},\cdot)$ have structures like $\psi(\hat{\beta}_w,\cdot)$ of (6) and $\gamma(\hat{\beta}_w,\cdot)$ of (12), respectively. They are constructed by using $L_{hj}(\hat{\beta}_{hj},\cdot)$, $Q_{0,0}^{(h)}$, $Q_{1,0}^{(h)}$, U_{hj}, $\hat{\zeta}_{hj}$, $\hat{b}_{hj}$, N_{hji}, ψ_{hj} instead of $L(\hat{\beta}_w,\cdot)$, $Q_{0,0}$, $Q_{1,0}$, U, $\hat{\zeta}$, $\hat{b}$, N_i, ψ. Then under the conditions analogous to that of Section 3 specified now separately with respect to each pair of states (h,j), $h,j=1,\dots,M$, $h\ne j$, one can obtain the multivariate generalizations of Theorems 3.1 and 3.3. In their statements we have the asymptotic independence of the components of the limiting Gaussian processes $(\Psi_{hj}(\cdot)\colon h,j=1,\dots,M,\ h\ne j)$ and $(G_{hj}(\cdot)\colon h,j=1,\dots,M,\ h\ne j)$, where Ψ_{hj} and G_{hj} are the counterparts of Ψ and G of Theorems 3.1 and 3.3, respectively, defined by using $A_{hj}(\beta_{hj},\cdot)$, $B_{hj}(\beta_{hj},\cdot)$ and $\Sigma_{hj}(\beta_{hj},\cdot)$. This leads to the obvious constructions of the KS and CM-type goodness-of-fit tests. In the statement of Theorem 3.4, one should only replace $\gamma^{(k)}(\hat{\beta}^{(k)}, s)$, $\hat{A}^{(k)}(\hat{\beta}^{(k)}, s)$ by $\gamma_{hj}(\hat{\beta}_{hj}, s)$, $\hat{A}_{hj}(\hat{\beta}_{hj}, s)$ and consider the maxima with respect to all pairs of states (h, j), $h,\ j = 1, \dots, M,\ h \neq j$. Consequently we have $M(M-1)$ instead of m.

7. Proofs.

PROOF OF THEOREM 3.1. If we denote by

$$V(\beta, t) = \int_0^t \sum_{i=1}^n w(Z_i(s), s)\left[\phi_i(s) - \frac{Q_{1,0}(\beta, s)}{Q_{0,0}(\beta, s)}\right] dN_i(s), \qquad t \in [0, \tau],$$

then by Taylor's expansion we have that

$$\begin{aligned}
\psi\big(\hat{\beta}_w, t\big) &= \frac{1}{\sqrt{n}} \int_0^t L\big(\hat{\beta}_w, s\big)^T V\big(\hat{\beta}_w, ds\big) \\
&= \psi(\beta_0, t) - \frac{1}{\sqrt{n}} \int_0^t \hat{b}_1(\bar{\beta}, s)^T\, d\bar{N}_w(s)\big(\hat{\beta}_w - \beta_0\big) \\
&\quad + \big(\hat{\beta}_w - \beta_0\big)^T \frac{1}{\sqrt{n}} \int_0^t \frac{\partial}{\partial \beta} L(\beta^*, s) V(\beta_0, ds) \\
&\quad - \big(\hat{\beta}_w - \beta_0\big)^T \frac{1}{\sqrt{n}} \int_0^t \frac{\partial}{\partial \beta} L(\beta^*, s)\hat{\rho}(\bar{\beta}, s)\, d\bar{N}_w(s)\big(\hat{\beta}_w - \beta_0\big),
\end{aligned} \tag{24}$$

where $\max\{\|\beta^* - \beta_0\|, \|\bar{\beta} - \beta_0\|\} \leq \|\hat{\beta}_w - \beta_0\|$. By using Conditions A, B, E, F, it can be easily shown that the third and fourth terms in the above expansion of $\psi(\hat{\beta}_w, \cdot)$ are asymptotically negligible. For the second term we obtain, by using the decomposition

$$\sqrt{n}\,\big(\hat{\beta}_w - \beta_0\big) = \Sigma(\beta_0, \tau)^{-1} U(\beta_0, \tau) + o_p(1)$$

and Conditions A, B, D, E that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \int_0^t \hat{b}_1(\bar{\beta}, s)^T\, d\bar{N}_w(s)\big(\hat{\beta}_w - \beta_0\big) \\
&\qquad = \frac{1}{n} \int_0^t b(\beta_0, s)^T\, d\bar{N}_w(s) \Sigma(\beta_0, \tau)^{-1} U(\beta_0, \tau) + o_p(1) \\
&\qquad = B(\beta_0, t)^T \Sigma(\beta_0, \tau)^{-1} U(\beta_0, \tau) + o_p(1).
\end{aligned} \tag{25}$$

Here the last equality is a direct consequence of Lenglart's (1977) inequality since obviously $U(\beta_0, \tau) = O_p(1)$. Equations (24) and (25), Condition E and again Lenglart's inequality lead to the final asymptotic representation of the form

$$\psi\big(\hat{\beta}_w, t\big) = \psi_1(\beta_0, t) - B(\beta_0, t)^T \Sigma(\beta_0, \tau)^{-1} U(\beta_0, \tau) + o_p(1), \tag{26}$$

where

$$\psi_1(\beta_0, t) = \frac{1}{\sqrt{n}} \int_0^t l(\beta_0, s)^T \sum_{i=1}^n w(Z_i(s), s) \times \left[\phi_i(s) - \frac{Q_{1,0}(\beta_0, s)}{Q_{0,0}(\beta_0, s)} \right] dM_i(s), \quad t \in [0, \tau] \tag{27}$$

and $U(\beta_0, \cdot)$ is given by (4). By using the multivariate martingale central limit theorem [cf. Fleming and Harrington (1991)] one can conclude that the $(p+1)$-variate local square integrable martingale $(\psi_1(\beta_0, \cdot), U(\beta_0, \cdot))$ converges weakly in $(D[0,\tau])^{p+1}$ to the process $\Gamma(\cdot) = (\Gamma_1(\cdot), \Gamma_2(\cdot))$ specified in Theorem 3.1. This can be established in view of Conditions A, B, C, and E. Hence we have the weak convergence of $(\psi_1(\beta_0, \cdot), U(\beta_0, \tau))$. Now observe that the process of (26) may be written as $H(\psi_1(\beta_0, \cdot), U(\beta_0, \tau)) + o_p(1)$, where H is a function from $D[0,\tau] \times R^p$ to $D[0,\tau]$ of the form $H(x, u) = x - B(\beta_0, \cdot)^T \Sigma(\beta_0, \tau)^{-1} u$. Note that H is continuous on $C[0,\tau] \times R^p$. Hence the continuous mapping theorem [cf. Billingsley (1968), Theorem 5.1] completes the proof. □

PROOF OF LEMMA 3.2. Obviously, by (7) and Conditions A, B, D′, $G(\cdot)$ of (10) has mean zero and is a Gaussian process with continuous sample paths. We shall find the expression of $\mathrm{Cov}[G(t), G(u)]$, $t,\ u \in [0,\tau]$. Let $u \le t$ and denote by $\mu(s) = b(\beta_0, s) q_{0,0}(\beta_0, s) \lambda_0(s)$, $s \in [0,\tau]$. Then

$$\begin{aligned} EG(t)G(u) &= E\Gamma_1(t)\Gamma_1(u) - E\Gamma_1(t)\int_0^u [\Gamma_2(\tau) - \Gamma_2(s)]^T \zeta(\beta_0, s)^{-1}\mu(s)\,ds \\ &\quad - E\Gamma_1(u)\int_0^t [\Gamma_2(\tau) - \Gamma_2(s)]^T \zeta(\beta_0, s)^{-1}\mu(s)\,ds \\ &\quad + E\int_0^t [\Gamma_2(\tau) - \Gamma_2(s)]^T \zeta(\beta_0, s)^{-1}\mu(s)\,ds \\ &\qquad \times \int_0^u [\Gamma_2(\tau) - \Gamma_2(y)]^T \zeta(\beta_0, y)^{-1}\mu(y)\,dy \\ &= L_1 - L_2 - L_3 + L_4 \quad \text{(say)}. \end{aligned} \tag{28}$$

By applying Fubini's theorem and (7) we obtain

$$\begin{aligned} L_1 &= A(\beta_0, u), \\ L_2 &= B(\beta_0, t)^T \int_0^u \zeta(\beta_0, s)^{-1}\mu(s)\,ds \\ &\quad - \int_0^u B(\beta_0, s)^T \zeta(\beta_0, s)^{-1}\mu(s)\,ds, \\ L_3 &= B(\beta_0, u)^T \int_0^u \zeta(\beta_0, s)^{-1}\mu(s)\,ds \\ &\quad - \int_0^u B(\beta_0, s)^T \zeta(\beta_0, s)^{-1}\mu(s)\,ds. \end{aligned} \tag{29}$$

To deal with L_4 note that

$$E[\Gamma_2(\tau) - \Gamma_2(s)][\Gamma_2(\tau) - \Gamma_2(y)]^T = \zeta(\beta_0, \max(s, y)),$$

$$\int_0^u \int_s^u \mu(s)^T \zeta(\beta_0, y)^{-1} \mu(y)\, dy\, ds = \int_0^u B(\beta_0, y)^T \zeta(\beta_0, y)^{-1} \mu(y)\, dy.$$

Hence one can deduce that

$$\begin{aligned} L_4 = {} & [B(\beta_0, t) + B(\beta_0, u)]^T \int_0^u \zeta(\beta_0, s)^{-1} \mu(s)\, ds \\ & - 2\int_0^u B(\beta_0, s)^T \zeta(\beta_0, s)^{-1} \mu(s)\, ds. \end{aligned}$$

The above equality together with (28) and (29) imply that $EG(t)G(u) = A(\beta_0, u)$, $u \le t$ and from the symmetry of considerations we finally have that $G(\cdot)$ is a zero-mean Gaussian process with independent increments and hence a Gaussian martingale with respect to its natural filtration. This completes the proof. □

PROOF OF THEOREM 3.3. Since by a Taylor series expansion,

$$\begin{aligned} \left[U(\hat\beta_w, \tau) - U(\hat\beta_w, s-)\right]^T = {} & [U(\beta_0, \tau) - U(\beta_0, s-)]^T \\ & - \sqrt{n}\left(\hat\beta_w - \beta_0\right)^T \hat\zeta(\beta^*, s), \end{aligned}$$

where $\|\beta^* - \beta_0\| \le \|\hat\beta_w - \beta_0\|$, (12) and (26) lead to the decomposition

$$\begin{aligned} \gamma(\hat\beta_w, t) = {} & \psi_1(\beta_0, t) - \int_0^t [U(\beta_0, \tau) - U(\beta_0, s-)]^T \\ & \times \hat\zeta(\hat\beta_w, s)^- \hat b(\hat\beta_w, s) \frac{d\overline{N}_w(s)}{n} + o_p(1), \end{aligned} \tag{30}$$

provided

$$\begin{aligned} & U(\beta_0, \tau)^T \Sigma(\beta_0, \tau)^{-1} B(\beta_0, t) \\ & - \sqrt{n}\left(\beta_w - \beta_0\right)^T \int_0^t \hat\zeta(\beta^*, s) \hat\zeta(\hat\beta_w, s)^- \hat b(\hat\beta_w, s) \frac{d\overline{N}_w(s)}{n} = o_p(1). \end{aligned} \tag{31}$$

By using Conditions F, B′, C′ and E′, one can obtain that

$$\sup_{s \in [0, \tau]} \left\| \hat b(\hat\beta_w, s) - b(\beta_0, s) \right\| = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{32}$$

Moreover, by Lenglart's inequality and Conditions A and B we have

$$\sup_{t \in [0, \tau]} \left\| \int_0^t \hat b(\hat\beta_w, s) \frac{d\overline{N}_w(s)}{n} - B(\beta_0, t) \right\| = o_p(1), \tag{33}$$

which yields the equality

$$
\begin{aligned}
&\sqrt{n}\left(\hat{\beta}_w-\beta_0\right)^T\int_0^t\hat{b}\left(\hat{\beta}_w,s\right)\frac{d\overline{N}_w(s)}{n}\\
(34)\qquad &=\sqrt{n}\left(\hat{\beta}_w-\beta_0\right)^T B(\beta_0,t)+o_p(1)\\
&=U(\beta_0,\tau)^T\Sigma(\beta_0,\tau)^{-1}B(\beta_0,t)+o_p(1).
\end{aligned}
$$

Now observe that it is possible to replace the integrand $\hat{\zeta}(\beta^*,s)$ in (31) by $\hat{\zeta}(\hat{\beta}_w,s)$. This can be established by observing that under conditions B, B′, C′, $\sup_{s\in[0,\tau]}\|\hat{\zeta}(\beta^*,s)-\hat{\zeta}(\hat{\beta}_w,s)\|=O_p(1/\sqrt{n})$ and by considering the approximation of the integral in (31) with respect to the intervals $[0,\tau-\varepsilon_n]$ and $(\tau-\varepsilon_n,\tau]$ separately, where $\varepsilon_n=n^{q-1/2}$ with $0<q<\varepsilon/(2+2\varepsilon)$ and ε given by Condition E″. According to the interval $[0,\tau-\varepsilon_n]$ we may confine ourselves to the set where $\hat{\zeta}(\hat{\beta}_w,s)^-=\hat{\zeta}(\hat{\beta}_w,s)^{-1}$ and use the fact that $\|\hat{\zeta}(\hat{\beta}_w,s)^{-1}\|\le O(1)(1/\varepsilon_n)$. This follows by an application of the following facts that are technical but quite straightforward in establishing, in view of Conditions B, A′ and D′,

$$
\begin{aligned}
&P\left\{\det\left[\hat{\zeta}\left(\hat{\beta}_w,s\right)\right]\ge\tfrac{1}{2}\det[\zeta(\beta_0,s)],\ s\in[0,\tau-\varepsilon_n]\right\}\to 1,\\
(35)\quad &P\left\{n^c(\tau-s)\left\|\hat{\zeta}\left(\hat{\beta}_w,s\right)^{-1}-\zeta(\beta_0,s)^{-1}\right\|\le 1,\ s\in[0,\tau-\varepsilon_n]\right\}\to 1\\
&\hspace{22em}\text{as } n\to\infty,
\end{aligned}
$$

where $0<c<q$. On the other hand, by using (32), Condition E″ and the fact that under Conditions A′ and B′,

$$
(36)\quad P\left\{C_1 n^{q+1/2}\le\overline{N}_w(\tau)-\overline{N}_w(\tau-\varepsilon_n-)\le C_2 n^{q+1/2}\right\}\to 1\quad\text{as } n\to\infty
$$

for some constants $0<C_1<C_2$, the desired approximation can be established on the interval $(\tau-\varepsilon_n,\tau]$. Thus (30) holds and the next step is to show that

$$
\begin{aligned}
&\int_0^t[U(\beta_0,\tau)-U(\beta_0,s-)]^T\hat{\zeta}\left(\hat{\beta}_w,s\right)^-\hat{b}\left(\hat{\beta}_w,s\right)\frac{d\overline{N}_w(s)}{n}\\
(37)\qquad &=\int_0^t[U(\beta_0,\tau)-U(\beta_0,s-)]^T\zeta(\beta_0,s)^{-1}b(\beta_0,s)\frac{d\overline{N}_w(s)}{n}\\
&\quad+o_p(1).
\end{aligned}
$$

By applying (32) we may replace $\hat{b}(\hat{\beta}_w,s)$ in the left-hand side of (37) by $b(\beta_0,s)$ provided we show that

$$
O_p\left(\frac{1}{\sqrt{n}}\right)\int_0^\tau\|U(\beta_0,\tau)-U(\beta_0,s-)\|\left\|\hat{\zeta}\left(\hat{\beta}_w,s\right)^-\right\|\frac{d\overline{N}_w(s)}{n}=o_p(1).
$$

This, however, can be established again by considering the above integral separately on the intervals $[0,\tau-\varepsilon_n]$ and $(\tau-\varepsilon_n,\tau]$, respectively, and using

Condition C$'$ together with (35) and (36). By Conditions A$'$, E$''$, $\|l(\beta_0, s)\| \le c_1(\tau - s)^{\varepsilon}$, $\lambda_0(s) \le c_2$ for some $\gamma, c_1, c_2 > 0$ and $s \in (\tau - \gamma, \tau]$. Hence in view of (35), by the fact that $\sup\{\|U(\beta_0, s)\|: s \in [0, \tau]\} = O_p(1)$ and finally by applying Lenglart's inequality, the quantity

$$\int_0^{\tau - \varepsilon_n} \|U(\beta_0, \tau) - U(\beta_0, s-)\| \left\| \hat{\zeta}(\hat{\beta}_w, s)^{-} - \zeta(\beta_0, s)^{-1} \right\| \times \|b(\beta_0, s)\| \frac{d\overline{N}_w(s)}{n}$$

is asymptotically negligible. On the other hand, by Conditions A$'$, B, D$'$,

$$\begin{aligned} &\int_{\tau - \varepsilon_n}^{\tau} \|U(\beta_0, \tau) - U(\beta_0, s-)\| \left\| \zeta(\beta_0, s)^{-1} \right\| \|b(\beta_0, s)\| \frac{d\overline{N}_w(s)}{n} \\ &\quad \le O_p(1) \int_{\tau - \varepsilon_n}^{\tau} \frac{\|l(\beta_0, s)\|}{\tau - s} \frac{d\overline{N}_w(s)}{n}. \end{aligned} \tag{38}$$

Since for $c_n = 1/(n \log n)$, $P\{\overline{N}_w(\tau) = \overline{N}_w(\tau - c_n)\} \to 1$ as $n \to \infty$, and by the Markov inequality

$$P\left\{ \varepsilon_n^{-\varepsilon/2} \int_{\tau - \varepsilon_n}^{\tau - c_n} \frac{\|l(\beta_0, s)\|}{\tau - s} \frac{d\overline{N}_w(s)}{n} \ge 1 \right\} \le O(1) \varepsilon_n^{\varepsilon/2},$$

the quantity in the left-hand side of (38) is asymptotically negligible. It is also asymptotically negligible if we replace $\zeta(\beta_0, s)^{-1}$ by $\hat{\zeta}(\hat{\beta}_w, s)^{-}$ in this quantity. This follows by an application of the inequality

$$\|U(\beta_0, \tau) - U(\beta_0, s-)\| \le 2C \int_{[s, \tau]} \frac{d\overline{N}_w(x)}{\sqrt{n}},$$

which is valid under Condition C$'$, and the facts that

$$P\left\{ \det\left[\hat{\zeta}(\hat{\beta}_w, s)\right] \ge \frac{1}{4} \det[\sigma(\beta_0, s)] \left\{ \int_{[s, \tau]} \frac{d\overline{N}_w(x)}{n} \right\}^p, \ s \in [\tau - \varepsilon_n, \tau] \right\} \to 1,$$

$$P\left\{ \left\| \hat{\zeta}(\hat{\beta}_w, s)^{-1} \right\| \le O_p(1) \left\{ \int_{[s, \tau]} \frac{d\overline{N}_w(x)}{n} \right\}^{-1}, \ s \in [\tau - \varepsilon_n, \tau] \right\} \to 1 \quad \text{as } n \to \infty.$$

The above convergence can be established by using Condition B, D$'$ and in view of Theorem F.2.a of Marshall and Olkin (1979) applied for $s \in [\tau - \varepsilon_n, \tau]$ to the quantity $\det[\int_{[s, \tau]} \sigma(\beta_0, x)\, d\overline{N}_w(x)/n]$ which approximates $\det[\hat{\zeta}(\hat{\beta}_w, s)]$. Consequently,

$$\int_0^{\tau} \|U(\beta_0, \tau) - U(\beta_0, s-)\| \left\| \hat{\zeta}(\hat{\beta}_w, s)^{-} - \zeta(\beta_0, s)^{-1} \right\| \|b(\beta_0, s)\| \frac{d\overline{N}_w(s)}{n} = o_p(1).$$

Thus (37) follows and as a result of the above considerations we have that

$$\gamma(\hat{\beta}_w, t) = \psi_1(\beta_0, t) - \int_0^t [U(\beta_0, \tau) - U(\beta_0, s-)]^T \times \zeta(\beta_0, s)^{-1} b(\beta_0, s) \frac{d\overline{N}_w(s)}{n} + o_p(1).$$

Now by using Lenglart's inequality and Conditions A′, B, E″, it can be easily shown that

$$\sup_{t \in [0,\tau]} \left| \int_0^t [U(\beta_0, \tau) - U(\beta_0, s-)]^T \zeta(\beta_0, s)^{-1} b(\beta_0, s) \frac{d\overline{M}_w(s)}{n} \right| = o_p(1)$$

and finally that

$$\gamma(\hat{\beta}_w, t) = \psi_1(\beta_0, t) - \int_0^t [U(\beta_0, \tau) - U(\beta_0, s)]^T \times \zeta(\beta_0, s)^{-1} b(\beta_0, s) q_{0,0}(\beta_0, s) \lambda_0(s)\, ds + o_p(1). \tag{39}$$

Now observe that the function h: $D^{p+1}[0, \tau] \to D[0, \tau]$ defined by

$$h(x_1, x_2) = x_1 - \int_0^{(\cdot)} [x_2(\tau) - x_2(s)]^T \times \zeta(\beta_0, s)^{-1} b(\beta_0, s) q_{0,0}(\beta_0, s) \lambda_0(s)\, ds \tag{40}$$

is continuous on $C^{p+1}[0, \tau]$ relative to the product Skorohod topology. This is guaranteed by the fact that

$$\int_0^\tau \| \zeta(\beta_0, s)^{-1} \| \, \| b(\beta_0, s) \| q_{0,0}(\beta_0, s) \lambda_0(s)\, ds < \infty,$$

in view of Conditions A, B, E″. Obviously, by (39) and (40) we have that $\gamma(\hat{\beta}_w, \cdot) = h(\psi_1(\beta_0, \cdot), U(\beta_0, \cdot)) + o_p(1)$. Since from the proof of Theorem 3.1 it follows that $(\psi_1(\beta_0, \cdot), U(\beta_0, \cdot))$ converges in $D^{p+1}[0, \tau]$ to the continuous Gaussian process $(\Gamma_1(\cdot), \Gamma_2(\cdot))$ specified by (7), the continuous mapping theorem and Lemma 3.2 complete the proof. □

PROOF OF THEOREM 3.4. It could be easily obtained that under the asymptotic stability and regularity condition, $\hat{A}(\hat{\beta}^{(k)}, \tau)$ converges in probability to $A^{(k)}(\beta_0, \tau)$. Moreover, if $k > 1$, then by using, for each $\gamma^{(k)}(\hat{\beta}^{(k)}, \cdot)$, the asymptotic representation analogous to that of (39) and (13), one can deduce that $(\gamma^{(k)}(\hat{\beta}^{(k)}, \cdot)$: $k = 1, \dots, m)$ converges weakly in $D^m[0, \tau]$ to the process $(W^{(k)}(A^{(k)}(\beta_0, \cdot))$: $k = 1, \dots, m)$, where $W^{(k)}(\cdot)$, $k = 1, \dots, m$, are independent standard Brownian motions. The direct application of the continuous mapping theorem together with the scale-change property of Brownian motion completes the proof. □

PROOF OF THEOREM 4.1. The proof is similar to that of Theorem 3.1. First observe that in view of the asymptotic stability and regularity assumptions the process $U(\beta_0, \cdot)$ given by (4) is asymptotically equivalent to the process

$$\tilde{U}(\beta_0, \cdot) + \delta_0 \int_0^{(\cdot)} \left[t_{0,1}(\beta_0, s) - \frac{q_{0,1}(\beta_0, s)}{q_{0,0}(\beta_0, s)} t_{0,0}(\beta_0, s) \right] \lambda_0(s)\, ds,$$

where

$$\tilde{U}(\beta_0, \cdot) = n^{-1/2} \int_0^{(\cdot)} \sum_{i=1}^{n} w(Z_i(s), s) \left[Z_i(s) - \frac{Q_{0,1}(\beta_0, s)}{Q_{0,0}(\beta_0, s)} \right] d\tilde{M}_1(s)$$

and

$$\tilde{M}_i(\cdot) = N_i(\cdot) - \int_0^{(\cdot)} \bar{\lambda}_i^{(n)}(s)\, ds, \qquad i = 1, \dots, n,$$

are mutually orthogonal local square integrable martingales. By repeating the Andersen and Gill (1982) argumentation, one can easily deduce that under (14) the MWPLE $\hat{\beta}_w$ also converges in probability to β_0. Thus by using Taylor's expansion we conclude that

$$\begin{aligned} &n^{1/2}\left(\hat{\beta}_w - \beta_0\right) - \Sigma(\beta_0, \tau)^{-1} \\ &\quad \times \left\{ \tilde{U}(\beta_0, \tau) + \delta \int_0^{\tau} \left[t_{0,1}(\beta_0, s) - \frac{q_{0,1}(\beta_0, s)}{q_{0,0}(\beta_0, s)} t_{0,0}(\beta_0, s) \right] \lambda_0(s)\, ds \right\} \end{aligned}$$

is asymptotically negligible. In view of the considerations similar to those of Theorem 3.1, we obtain that under (14) the process $n^{-1/2}\psi(\hat{\beta}_w, \cdot)$ is asymptotically equivalent to the process

$$\begin{aligned} &\tilde{\psi}(\beta_0, \cdot) + \delta \int_0^{(\cdot)} l(\beta_0, s)^T \left[t_{1,0}(\beta_0, s) - \frac{q_{1,0}(\beta_0, s)}{q_{0,0}(\beta_0, s)} t_{0,0}(\beta_0, s) \right] \lambda_0(s)\, ds \\ &\quad - B(\beta_0, \cdot)^T \Sigma(\beta_0, \tau)^{-1} \\ &\quad \times \left\{ \tilde{U}(\beta_0, \tau) + \delta \int_0^{\tau} \left[t_{0,1}(\beta_0, s) - \frac{q_{0,1}(\beta_0, s)}{q_{0,0}(\beta_0, s)} t_{0,0}(\beta_0, s) \right] \lambda_0(s)\, ds \right\}. \end{aligned}$$

Here the process $\tilde{\psi}(\beta_0, \cdot)$ is defined in a similar manner as $\psi_i(\beta_0, \cdot)$ of (27) with M_i replaced by $\tilde{M}_i$. An application of Rebolledo's (1980) theorem shows that $[\tilde{\psi}(\beta_0, \cdot), \tilde{U}(\beta_0, \cdot)]$ converges weakly to the process $[\Gamma_1(\cdot), \Gamma_2(\cdot)]$ given in Theorem 3.1. Thus the final result is a direct consequence of the application of the continuous mapping theorem. □

Proof of Theorem 4.2. The result of the theorem could be easily established by mimicking the method of the proof of Theorem 3.1. Therefore we shall present only the main steps of the proof. Taylor's series expansion of

$L(\hat{\beta}_w, s)$ and $Q_{1,0}(\hat{\beta}_w, s)/Q_{0,0}(\hat{\beta}_w, s)$ at the point β^* together with the asymptotic stability and regularity assumptions lead to the equality

$$\frac{1}{\sqrt{n}}\psi(\hat{\beta}_w, t) = \int_0^t L(\beta^*, s)^T \left[Q_{1,0}(s) - \frac{Q_{1,0}(\beta^*, s)}{Q_{0,0}(\beta^*, s)} Q_{0,0}(s) \right] ds + o_p(1)$$

and finally to

$$\frac{1}{\sqrt{n}}\psi(\hat{\beta}_w, t) = \int_0^t l(\beta^*, s)^T \left[q_{1,0}(s) - \frac{q_{1,0}(\beta^*, s)}{q_{0,0}(\beta^*, s)} q_{0,0}(s) \right] ds + o_p(1).$$

Thus $n^{-1/2}\psi(\hat{\beta}_w, t)$ converges in probability to $\int_0^t l(\beta^*, s)^T f(\beta^*, s) q_{0,0}(s)\, ds$, uniformly in $t \in [0, \tau]$. Similarly

$$\begin{aligned}\frac{1}{\sqrt{n}}U(\hat{\beta}_w, t) &= \int_0^t \left[q_{0,1}(s) - \frac{q_{0,1}(\beta^*, s)}{q_{0,0}(\beta^*, s)} q_{0,0}(s) \right] ds + o_p(1)\\ &= u(\beta^*, t) + o_p(1).\end{aligned}$$

By applying an assumption corresponding to B' together with Lenglart's inequality, we conclude that $U(\beta^*, \tau) = O_p(1)$ and consequently that $n^{1/2}(\hat{\beta}_w - \beta^*) = O_p(1)$. Then by using the counterparts of the steps of the proof of Theorem 3.3 conducted with the use of the assumptions preceding Theorem 4.2, we obtain the following equalities:

$$\begin{aligned}&\int_0^t \left[\frac{1}{\sqrt{n}}U(\hat{\beta}_w, \tau) - \frac{1}{\sqrt{n}}U(\hat{\beta}_w, s-) \right]^T \hat{\zeta}(\hat{\beta}_w, s)^{-} \hat{b}(\hat{\beta}_w, s) \frac{d\overline{N}_w(s)}{n}\\
&\quad = \int_0^t \left[\frac{1}{\sqrt{n}}U(\beta^*, \tau) - \frac{1}{\sqrt{n}}U(\beta^*, s-) \right]^T\\
&\qquad \times \hat{\zeta}(\hat{\beta}_w, s)^{-} \hat{b}(\hat{\beta}_w, s) \frac{d\overline{N}_w(s)}{n} + o_p(1)\\
&\quad = \int_0^t \left[\frac{1}{\sqrt{n}}U(\beta^*, \tau) - \frac{1}{\sqrt{n}}U(\beta^*, s-) \right]^T\\
&\qquad \times \zeta^*(\beta^*, s)^{-1} b(\beta^*, s) \frac{d\overline{N}_w(s)}{n} + o_p(1)\\
&\quad = \int_0^t \left[\frac{1}{\sqrt{n}}U(\beta^*, \tau) - \frac{1}{\sqrt{n}}U(\beta^*, s-) \right]^T\\
&\qquad \times \zeta^*(\beta^*, s)^{-1} b(\beta^*, s) q_{0,0}(s)\, ds + o_p(1)\\
&\quad = \int_0^t [u(\beta^*, \tau) - u(\beta^*, s)]^T\\
&\qquad \times \zeta^*(\beta^*, s)^{-1} b(\beta^*, s) q_{0,0}(s)\, ds + o_p(1).\end{aligned}$$

By (17), $u(\beta^*, \tau) = 0$. Thus in view of (12) the proof is complete. □

Acknowledgments. The authors express their thanks to an Associate Editor and the referees for their valuable comments and suggestions which led to a substantial improvement of the original manuscript.

REFERENCES

AKI, S. (1986). Some test statistics based on the martingale term of the empirical distribution function. *Ann. Inst. Statist. Math.* **38** 1–21.

ANDERSEN, P. K. and BORGAN, Ø. (1985). Counting process models for life history data: a review. *Scand. J. Statist.* **12** 97–158.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

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

ANDERSEN, P. K., HANSEN, L. S. and KEIDING, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scand. J. Statist.* **18** 153–167.

ARJAS, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *J. Amer. Statist. Assoc.* **83** 204–212.

ARJAS, E. and HAARA (1988). A note on the asymptotic normality in the Cox regression model. *Ann. Statist.* **16** 1133–1140.

BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295.

BARLOW, W. E. and PRENTICE, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75** 65–74.

BEDNARSKI, T. (1993). Robust estimation in Cox's regression model. *Scand. J. Statist.* **20** 213–225.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

CROWLEY, J. J. and JONES, M. P. (1989). A general class of nonparametric tests for survival analysis. *Biometrics* **45** 157–170.

CROWLEY, J. J. and JONES, M. P. (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Ann. Statist.* **18** 1203–1220.

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

HAARA, (1987). A note on the asymptotic behaviour of the empirical score in Cox's regression model for counting processes. In *Proceedings of the 1st World Congress of the Bernoulli Society* 139–142. VNU Science Press, Tashkent, Soviet Union.

HAHN, M. G. (1978). Central limit theorems in $D[0,1]$. *Z. Wahrsch. Verw. Gebiete* **44** 89–101.

HENDERSON, R. and MILNER, M. (1991). On residual plots for relative risk regression. *Biometrika* **78** 631–636.

HJORT, N. L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18** 1221–1258.

HJORT, N. L. (1992). On inference in parametric survival data models. *Internat. Statist. Rev.* **60** 355–387.

JONES, M. P. (1991). Robust tests for survival data involving a single continuous covariate. *Scand. J. Statist.* **18** 323–332.

KHMALADZE, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.

KHMALADZE, E. V. (1988). An innovation approach to goodness-of-fit tests in R^m. *Ann. Statist.* **16** 1503–1516.

KHMALADZE, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.* **21** 798–829.

LAGAKOS, S. W. and SCHOENFELD, D. A. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* **40** 1037–1048.

LENGLART, E. (1977). Relation de domination entre deux processes. *Ann. Inst. H. Poincaré* **13** 171–179.

LI, G. and DOSS, H. (1993). Generalized Pearson–Fischer chi-square goodness-of-fit tests with application to models with life history data. *Ann. Statist.* **21** 772–797.

LIN, D. Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *J. Amer. Statist. Assoc.* **86** 725–728.

LIN, D. Y. and WEI, L. J. (1991). Goodness-of-fit tests for the general Cox regression model. *Statist. Sinica* **1** 1–17.

LIN, D. Y., WEI, L. J. and YING, Z. (1993). Checking the Cox model with cumulative sums of martingale residuals. *Biometrika* **80** 557–572.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MARZEC, L. (1996). Asymptotic inference for semiparametric model with general relative risk form. Unpublished manuscript.

MARZEC, L. and MARZEC, P. (1993). Goodness of fit inference based on stratification in Cox's regression model. *Scand. J. Statist.* **20** 227–238.

MARZEC, L. and MARZEC, P. (1996). Tests for semiparametric model based on non-homogeneous Markov process. *Statist. Probab. Lett.* **27** 137–143.

MARZEC, P. (1993). A process based on stratification for Cox's regression model. *Probab. Math. Statist.* **14** 115–124.

MCKEAGUE, I. W. and UTIKAL, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand. J. Statist.* **18** 177–195.

MILLER, R. and HALPERN, J. (1982). Regression with censored data. *Biometrika* **69** 521–531.

PRENTICE, and SELF (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286.

SASIENI, P. (1993a). Maximum weighted partial likelihood estimates for the Cox model. *J. Amer. Statist. Assoc.* **88** 144–152.

SASIENI, P. (1993b). Some new estimators for Cox regression. *Ann. Statist.* **21** 1721–1759.

SCHOENFELD, D. (1980). Chi-squared goodness-of-fit tests for proportional hazards regression model. *Biometrika* **67** 145–153.

SEGAL, M. R., JAMES, I. R., FRENCH, M. A. and MALLAL, S. A. (1995). Statistical issues in the evaluation of markers of HIV progression. *Internat. Statist. Rev.* **63** 179–197.

SLUD, E. (1991). Relative efficiency of the log rank test within a multiplicative intensity model. *Biometrika* **78** 621–630.

STRUTHERS, C. A. and KALBLFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73** 363–369.

THERNEAU, T. M., GRAMBSCH, P. M. and FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–160.

TSATIS, A. A., ROSNER, G. L. and TRITCHLER, D. L. (1985). Group sequential tests with censored survival data adjusting for covariates. *Biometrika* **72** 365–376.

WEI, L. J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *J. Amer. Statist. Assoc.* **79** 649–652.

MATHEMATICAL INSTITUTE
UNIVERSITY OF WROCŁAW
PL. GRUNWALDZKI 2 / 4
50-384 WROCŁAW
POLAND
E-MAIL: lmarzec@hera.math.uni.wroc.pl