行政院國家科學委員會專題研究計畫　成果報告

廣義聯立方程模式與廣義路徑分析：存活或事件史資料(I)

計畫類別：個別型計畫
計畫編號：NSC91-2118-M-002-004-
執行期間：91年08月01日至92年07月31日
執行單位：國立臺灣大學公共衛生學院流行病學研究所

計畫主持人：胡賦強

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 93年2月18日

Draft

Is the Maximum Partial Likelihood Estimator for Cox's Proportional Hazards Model Also a General Least Squares Estimator?

Fu-Chang Hu, M.S., Sc.D.[1]

Assistant Professor
Division of Biostatistics
Graduate Institute of Epidemiology
College of Public Health
National Taiwan University
Taipei, Taiwan 100
R.O.C.
fchu@ha.mc.ntu.edu.tw

Chyong-Mei Chen, M.S.

Doctoral Candidate
Division of Biostatistics
Graduate Institute of Epidemiology
College of Public Health
National Taiwan University
Taipei, Taiwan 100
R.O.C.
d89842006@ms89.ntu.edu.tw

June 4, 2003

[1]Correspondence: Fu-Chang Hu, Division of Biostatistics, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, 1 Jen-Ai Road, Section 1, Room 1541, Taipei, Taiwan 100, R.O.C., Phone: +886 2 / 2394-2050 or +886 2 / 2312-3456 $\times$ 8349, Fax: +886 2 / 2351-1955, and Email: fchu@ha.mc.ntu.edu.tw. The first author has also served as Adjunct Assistant Research Fellow in the Department of Psychiatry, National Taiwan University Hospital, since August 1998.

SUMMARY

The equivalences in estimation between the *maximum likelihood approach* (e.g., the usual maximum likelihood estimator) and the *least squares approach* (e.g., the ordinary, weighted, generalized, and iterative reweighted least squares estimators) have been established for many well-known classes of statistical regression models such as linear regression model, logistic regression model, and generalized linear models (GLMs). However, no such connection has been discovered yet for the maximum partial likelihood estimator (MPLE) of the regression coefficients in Cox's proportional hazards model (Cox 1972, 1975). In this study, by choosing an appropriate "moment condition" of generalized method of moments (GMM) estimation, we find that with the "asymmetric orthogonal expected information approach" of adaptive estimation, the optimal martingale estimating function obtained from the minimization of the corresponding GMM quadratic form for a consistent estimator of the regression coefficients reduces to the partial score function of the Cox's proportional hazards model, which implies that the well-behaved MPLE is also a general least squares estimator. This finding is not only very interesting in its own rights, but it provides us with an opportunity to develop GLMs-type regression models locally for stochastic processes and to apply some powerful GMM-related estimating techniques such as the instrumental variables method to deal with several known statistical modeling problems including measurement error and simultaneous-equations bias in analysis of survival or time-to-event data.

Key words:

Partial score function; MPLE; Moment conditions; Generalized method of moments; GMM; Estimating functions; Martingales; Nuisance parameters; Adaptive estimation.

Contents

1 INTRODUCTION

1.1 Motivation

The equivalences in estimation between the *maximum likelihood approach*, e.g., the usual maximum likelihood estimator (MLE), and the *least squares approach*, e.g., the ordinary, weighted, generalized, and iterative reweighted least squares (OLS, WLS, GLS, and IRLS) estimators, have been established for many well-known classes of statistical regression models such as linear regression model, logistic regression model, and generalized linear models (GLMs) (Nelder and Wedderburn 1972, Wedderburn 1974, McCullagh and Nelder 1989). Godambe and Kale (1991) gave a nice discussion on this issue from the viewpoint of estimating functions. However, to the best of our knowledge, no such connection has been discovered yet for the *maximum partial likelihood estimator* (MPLE) of the regression coefficients in Cox's proportional hazards model (Cox 1972, 1975).

1.2 Objectives

In this study, by choosing an appropriate "moment condition" of *generalized method of moments* (GMM) estimation (see, e.g., Greene 2000, pp. 474-488), we find that with an aid from *adaptive estimation* to deal with the nuisance parameter, the optimal *martingale estimating function* (m.e.f.) obtained from the minimization of the corresponding GMM quadratic form for a consistent estimator of the regression coefficients reduces to the usual partial score function of the Cox's proportional hazards model, which implies that the well-behaved MPLE is also a general least squares estimator. In the derivations, we will show (1) how the optimal m.e.f. for estimating the regression coefficients can be obtained from a GMM quadratic form and (2) how the baseline hazard function as the nuisance parameter can be purposely eliminated to yield an adaptive estimation of the regression coefficients. Finally, we discuss some interesting properties and extensions associated with this class of m.e.f.'s.

2 REVIEW

2.1 Cox's Proportional Hazards Model

The semi-parametric Cox's proportional hazards model for modeling the hazard function $\lambda(t \mid \mathbf{z}_i(t))$ by a set of covariate values $\mathbf{z}_i(t)$ is

$$\lambda(t \mid \mathbf{z}_i(t)) \quad = \quad \lambda_0(t) e^{\boldsymbol{\beta}' \mathbf{z}_i(t)} \tag{1}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}$ is a vector of p unknown regression coefficients, and $\exp\{\boldsymbol{\beta}' \mathbf{z}_i(t)\} > 0$ is a risk multiplier (Cox 1972). It is by far the most popular regression model in survival analysis. Under the assumption of independent censoring (conditioning on the covariates $\mathbf{z}_i(t)$), the MPLE of $\boldsymbol{\beta}$ can be obtained by maximizing the partial likelihood function for $\boldsymbol{\beta}$ (Cox 1975), in which the infinite dimensional nuisance parameter $\lambda_0(t)$ is nicely eliminated due to the unique structure of the proportional hazards model. The asymptotic properties of the MPLE $\widehat{\boldsymbol{\beta}}$ was studied by Tsiatis (1981) (using the standard classical approach) and by Andersen and Gill (1982) and Gill (1984) (using the modern counting process martingale approach) respectively. And, as shown by Begun, et. al. (1983), one important property of the MPLE $\widehat{\boldsymbol{\beta}}$ is that without knowing $\lambda_0(t)$, it achieves the semi-parametric information bound in estimating $\boldsymbol{\beta}$. A class of hazard functions with general risk multipliers was investigated by Prentice and Self (1983). There were many theoretical and methodological research works on Cox's proportional hazards model in the past thirty years or so. However, as far as we know, no one has specifically answered the question about whether the MPLE $\widehat{\boldsymbol{\beta}}$ is equivalent to any least squares-type estimators.

2.2 Theory of Estimating Functions

On the other hand, estimating function is a flexible and robust way to estimating the parameters of interest in many general regression settings. Godambe (1960) might be the first one to formally study the properties of estimating functions. The theory of estimating functions for discrete-time stochastic processes was discussed by Godambe (1985) and extended to the continuous-time cases by Thavanewswaran and Thompson (1986). And, Godambe and Heyde (1987) and Heyde (1997) gave nice reviews of the important results in this area. Greenwood and Wefelmeyer (1991) discussed optimal m.e.f.'s for partially specified counting process models based on various optimality criteria. In particular, Chang and Hsiung (1990) studied the finite sample optimality of MPLE $\widehat{\boldsymbol{\beta}}$ for Cox's proportional hazards model within the framework of estimating functions, which was quite different from the usual partial likelihood approach and thus gave us a good insight into the problem.

3 MAIN RESULTS

Suppose that there are n independent subjects indexed by i in a biomedical study. Let T_i and C_i be the event time and the right censoring time respectively for subject i. Due to the censoring, we can only observe the realizations of $X_i = T_i \wedge C_i = \min(T_i, C_i)$ and

$\Delta_i = I\{T_i \leq C_i\}$. And, $\mathbf{z}_i$ is a $(p \times 1)$ vector containing the observed values of p covariates $\mathbf{Z}_i$, which can be time-dependent in general cases, for each subject i. As usual, we assume independent censoring conditioning on the covariates, i.e., $(T_i \perp C_i) \mid \mathbf{z}_i$, for all subjects.

Next, we define a *counting process* $N_i(t) = I\{X_i \leq t, \Delta_i = 1\}$, an status indicator $\kappa_i(t) = I\{X_i \geq t\}$, and a right continuous *filtration* (or "history") $\{\mathcal{F}(t) : \ t \geq 0\}$, where

$$\mathcal{F}(t) \ = \ \sigma\{N_i(u), \kappa_i(u+), \mathbf{Z}_i(u) : \ 0 \leq u \leq t \text{ and } i = 1, 2, \cdots, n\}$$

is the smallest σ-algebra (or contains the information) generated by $N_i(u)$, $\kappa_i(u+)$, and $\mathbf{Z}_i(u)$ inside the bracket. Then, given a correctly specified Cox's proportional hazards model $\lambda(t \mid \mathbf{z}_i) = \lambda_0(t)\exp\{\boldsymbol{\beta}^{'}\mathbf{z}_i\}$, the *intensity process* for an increment of $N_i(t)$ is

$$E\{dN_i(t) \mid \mathcal{F}(t-)\} \ = \ \kappa_i(t)\lambda_0(t)e^{\boldsymbol{\beta}^{'}\mathbf{z}_i}dt.$$

Thus,

$$dM_i(t) \ = \ dN_i(t) \ - \ \kappa_i(t)\lambda_0(t)e^{\boldsymbol{\beta}^{'}\mathbf{z}_i}dt$$

is an increment of the *martingale* $M_i(t)$ associated with the above filtration $\{\mathcal{F}(t), t \geq 0\}$. Finally, we shall denote a bounded *predictable process* with respective to the same filtration $\{\mathcal{F}(t), t \geq 0\}$ by $\mathbf{H}_i(t; \boldsymbol{\beta})$.

3.1 Generalized Method of Moments Estimation

A natural choice of the *moment condition* for obtaining a GMM estimator $\widetilde{\boldsymbol{\beta}}$ of the regression coefficients $\boldsymbol{\beta}$ in Cox's proportional hazards model (Eq. (1)) is to use $dM_i(t)$ (for $i = 1, 2, \cdots, n$), which are usually called *martingale residuals* after being properly estimated from the data (see, e.g., Fleming and Harrington 1991, pp. 163-165). This is *the* source of empirical information from which we can possibly get the right estimates of the unknown parameters over n subjects. Any other moment conditions (e.g., the conditions from higher moments), if available, can also be incorporated into the framework of the GMM estimation by stacking them together.

Since $M_i(t)$ is an $\mathcal{F}(t)$-martingale, we have

$$Var\{dN_i(t) \mid \mathcal{F}(t-)\} = Var\{dM_i(t) \mid \mathcal{F}(t-)\} = \kappa_i(t)\lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{z}_i}dt$$

which is denoted as $d < M >_i (t)$, so that locally at time t, $dN_i(t) \mid \mathcal{F}(t-)$ behaves just like a *Poisson* random variable (McCullagh and Nelder 1989, Sec. 6.4, pp. 209-214, Fleming and Harrington 1991, p. 87, Andersen, et. al. 1993, pp. 54-55, 223, and 482). Then, following the standard procedure for obtaining a GMM estimator $\widetilde{\boldsymbol{\beta}}$ (based on the L_2-norm), we construct the following quadratic form from the chosen moment condition

$$Q_T(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^T \frac{[dM_i(u)]^2}{\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du} \tag{2}$$

where T is a stopping time.

3.2 The Estimating Functions

However, direct minimization of the above quadratic form $Q_T(\boldsymbol{\beta})$ would usually yield an inconsistent estimator of the unknown parameter $\boldsymbol{\beta}$ because the conditional variance of $dN_i(t)$ given $\mathcal{F}(t-)$, i.e., $Var(dN_i(t) \mid \mathcal{F}(t-))$, contains $\boldsymbol{\beta}$ (see, e.g., Greene 2000, p. 477). Thus, to derive an unbiased estimating function (e.f.) from the GMM quadratic form for a consistent estimation of $\boldsymbol{\beta}$, we suggest making the following remedy. As discussed by Godambe and Kale (1991, Sec. 1.5, pp. 9-10), Desmond (1991, p. 140), and McCullagh

(1991, Sec. 11.3, esp., Eq. (3.2), pp. 269-271) in similar settings, we should avoid taking the derivative of $\kappa_i(u)\lambda_0(u)\exp\{\boldsymbol{\beta}'\mathbf{z}_i\}du$ in the denominator with respective to $\boldsymbol{\beta}$ to preserve $\partial Q_T(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ to have a zero mean (at the true value of $\boldsymbol{\beta}$). Hence, the remedy is to treat the $\boldsymbol{\beta}$ inside the weight function $\left[\kappa_i(u)\lambda_0(u)\exp\{\boldsymbol{\beta}'\mathbf{z}_i\}du\right]^{-1}$ as if it were known in taking the derivative of the quadratic form $Q_T(\boldsymbol{\beta})$ with respective to $\boldsymbol{\beta}$.

Then, although there exists another unknown parameter $\lambda_0(t)$, we obtain the following $(p\times 1)$ e.f. $\mathbf{U}_T(\boldsymbol{\beta})$ from the GMM quadratic form $Q_T(\boldsymbol{\beta})$ (Eq. (2)) for estimating the unknown parameter of interest $\boldsymbol{\beta}$

$$\begin{aligned}
\mathbf{U}_T(\boldsymbol{\beta}) &\equiv \sum_{i=1}^{n}\int_0^T \left(\frac{\partial\,[dM_i(u)]}{\partial\boldsymbol{\beta}}\right)\left[\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du\right]^{-1} dM_i(u) \\
&= \sum_{i=1}^{n}\int_0^T d\dot{M}_i(u)\left[\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du\right]^{-1} dM_i(u) \\
&= -\sum_{i=1}^{n}\int_0^T \left(\frac{\partial}{\partial\boldsymbol{\beta}}\left[\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du\right]\right)\left[\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du\right]^{-1} dM_i(u) \\
&= -\sum_{i=1}^{n}\int_0^T \mathbf{z}_i dM_i(u) \qquad (3)
\end{aligned}$$

which, conditioning on the filtration $\mathcal{F}(t-)$ locally (at any given time $t\le T$), has exactly the same form as the optimal e.f. (i.e., the quasi-score function) of the GLM for the conditional mean of a *Poisson*-like response, $E\{dN_i(t)\mid\mathcal{F}(t-)\}$, with the canonical link function ($\log$) (McCullagh and Nelder 1989, Eq. (9.5), p. 327) and as an optimal m.e.f. too (Desmond 1991, p. 143, Greenwood and Wefelmeyer 1991, p. 153). In this e.f.,

$$\mathbf{H}_i(t;\boldsymbol{\beta}) \equiv \mathbf{z}_i$$

is a $(p\times 1)$ bounded predictable process with respective to the same filtration $\{\mathcal{F}(t), t\ge 0\}$. For simplicity, we shall drop the "$-$" sign of $\mathbf{U}_T(\boldsymbol{\beta})$ in the sequel.

It can be seen now that the ”orthogonality” between the derivatives of the intensity process $\kappa_i(u)\lambda_0(u)\exp\{\boldsymbol{\beta}'\mathbf{z}_i\}du$ with respective to $\boldsymbol{\beta}$ and the martingale residuals $dM_i(u)$ is the key to obtaining a good estimate of $\boldsymbol{\beta}$. This fact may imply that *orthogonality* is also a more fundamental property than *minimization* (or *maximization*) required in statistical estimation in the dependent case.

3.3 The Nuisance Parameter Problem

However, the unspecified infinite dimensional nuisance parameter $\lambda_0(t)$ sitting inside the intensity process $\kappa_i(u)\lambda_0(u)\exp\{\boldsymbol{\beta}'\mathbf{z}_i\}du$ of $dM_i(u)$ in the e.f. $\mathbf{U}_T(\boldsymbol{\beta})$ (Eq. (3)) causes a major difficulty in the estimation of $\boldsymbol{\beta}$. Even though a consistent estimator for $\lambda_0(t)$, $\widetilde{\lambda}_0(t)$ say, may be available, the asymptotic properties of the resulting GMM estimator $\widetilde{\boldsymbol{\beta}}$ would generally depend on $\widetilde{\lambda}_0(t)$. Nonetheless, following the idea of Williams (1982) and Moore (1986) developed in a different setup, we find that one plausible way to get around this nuisance parameter problem is to search for a suitable transformation on $\mathbf{z}_i$ in the crude e.f. $\mathbf{U}_T(\boldsymbol{\beta})$ to derive an e.f. $\mathbf{U}^{\star}_T(\boldsymbol{\beta})$ which satisfies the following *key condition* for an adaptive estimation of $\boldsymbol{\beta}$

$$E\left[\frac{\partial \mathbf{U}^{\star}_T(\boldsymbol{\beta})}{\partial \lambda_0(t)}\right] = \mathbf{0}. \tag{4}$$

The rationale behind this approach is that if the expectation of the cross derivative of the e.f. $\mathbf{U}^{\star}_T(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$ with respective to $\lambda_0(t)$ is zero, then the inverse of the joint expected quasi-information matrix for $(\boldsymbol{\beta}, \lambda_0(t))$ has a zero in its right-top corner so that the asymptotic distribution of $\widetilde{\boldsymbol{\beta}}^{\star}$, which is the root of the derived e.f. $\mathbf{U}^{\star}_T(\boldsymbol{\beta})$, would *not* depend on $\widetilde{\lambda}_0(t)$ (see **Appendix**).

To this end, we replace the $\mathbf{z}_i$ in the crude e.f. $\mathbf{U}_T(\boldsymbol{\beta})$ (Eq. (3)) with the following $(p \times 1)$ bounded $\mathcal{F}(t)$-predictable process

$$\mathbf{H}_i^\star(t;\boldsymbol{\beta}) \equiv \mathbf{z}_i - \bar{\mathbf{z}}(t) = \mathbf{z}_i - \frac{\sum_{j=1}^n \kappa_j(t) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j}{\sum_{j=1}^n \kappa_j(t) e^{\boldsymbol{\beta}'\mathbf{z}_j}} \tag{5}$$

where $\bar{\mathbf{z}}(t)$ is an empirical weighted mean of $\mathbf{z}_j$'s at time $t \le T$. Certainly, $\mathbf{H}_i^\star(t;\boldsymbol{\beta})$ itself should not contain the nuisance parameter $\lambda_0(t)$. Then, the derived e.f. for $\boldsymbol{\beta}$ is

$$\mathbf{U}_T^\star(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^T \mathbf{H}_i^\star(u;\boldsymbol{\beta}) dM_i(u) = \sum_{i=1}^n \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, dM_i(u). \tag{6}$$

In taking derivatives of the above e.f.'s for $\boldsymbol{\beta}$ with respective to the infinite dimensional parameter $\lambda_0(t)$, we develop a three-step approach to avoid the potential technical complexity:

1. At first, we assume a *constant* baseline hazard function λ_0 (for all $t \in [0, T]$).

2. Next, we consider a *piecewise constant* baseline hazard function $\lambda_0(k)$ (for $1 \le k \le K$), where K is the total number of small time intervals within $[0, T]$.

3. Finally, we extend the result to a *general* baseline hazard function $\lambda_0(t)$ (for $t \in [0, T]$) by letting $K \to \infty$.

Now, assuming a constant baseline hazard function $\lambda_0(t) = \lambda_0$ (for all $t \in [0, T]$), we first check

$$\begin{aligned}
E\left[\frac{\partial \mathbf{U}_T(\boldsymbol{\beta})}{\partial \lambda_0}\right] &= E\left[\frac{\partial \sum_{i=1}^n \int_0^T \mathbf{z}_i dM_i(u)}{\partial \lambda_0}\right] \\
&= E\left[\frac{\partial \sum_{i=1}^n \int_0^T \mathbf{z}_i dN_i(u)}{\partial \lambda_0}\right] - E\left[\frac{\partial \sum_{i=1}^n \int_0^T \mathbf{z}_i \kappa_i(u)\lambda_0 e^{\boldsymbol{\beta}'\mathbf{z}_i} du}{\partial \lambda_0}\right] \\
&= -E\left[\sum_{i=1}^n \int_0^T \mathbf{z}_i \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i} du\right] \\
&\ne \mathbf{0}
\end{aligned}$$

unless it is properly *centered*. Thus, even in the simplest case, the above-mentioned key condition does *not* hold for the crude e.f. $\mathbf{U}_T(\boldsymbol{\beta})$.

In contrast, assuming a constant baseline hazard function $\lambda_0(t) = \lambda_0$ (for all $t \in [0, T]$) again, we have

$$
\begin{aligned}
E\left[\frac{\partial \mathbf{U}_T^\star(\boldsymbol{\beta})}{\partial \lambda_0}\right] &= E\left[\frac{\partial \sum_{i=1}^n \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, dM_i(u)}{\partial \lambda_0}\right] \\
&= -E\left[\sum_{i=1}^n \int_0^T \left(\mathbf{z}_i - \frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i} du\right] \\
&= E\int_0^T \left[\sum_{i=1}^n \left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i} - \sum_{i=1}^n \mathbf{z}_i \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i}\right] du \\
&= E\int_0^T \left[\left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) \sum_{i=1}^n \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i} - \sum_{i=1}^n \mathbf{z}_i \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i}\right] du \\
&= E\int_0^T \left[\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j - \sum_{i=1}^n \mathbf{z}_i \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i}\right] du \\
&= \mathbf{0}.
\end{aligned}
$$

By the same token, considering a piecewise constant baseline hazard function $\lambda_0(t) = \lambda_0(k)$ (for $1 \leq k \leq K$), we obtain the same result for the derived e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$. Finally, to extend this result to any general baseline hazard functions $\lambda_0(t)$ (for $t \in [0, T]$), we let $K \to \infty$.

Notice that as just shown in the case of Cox's proportional hazards model, the nice algebraic structure which makes the key condition to be true for the preferred e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ is

$$
\sum_{i=1}^n \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i} du = \int_0^T \left[\sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, \kappa_i(u) e^{\boldsymbol{\beta}'\mathbf{z}_i}\right] du = \mathbf{0} \quad (7)
$$

which is exactly the same crucial equality required for the martingale representation of the partial score function of $\boldsymbol{\beta}$ in the partial likelihood approach (see, e.g., Fleming and Harrington

1991, p. 150). And, it can be seen clearly that *centering* the covariate values by their empirical weighted means at time t (Eq. (5)) provides us with a solution to the failure of the key condition for adaptive estimation in the crude e.f. $\mathbf{U}_T(\boldsymbol{\beta})$.

In fact, according to this equality (Eq. (7)), the proposed e.f. $\mathbf{U}_T^{\star}(\boldsymbol{\beta})$ (Eq. (6)) reduces to a rather simple form

$$
\begin{aligned}
\mathbf{U}_T^{\star}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, dM_i(u) \\
&= \sum_{i=1}^{n} \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, dN_i(u) \; - \sum_{i=1}^{n} \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, \kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i} du \\
&= \sum_{i=1}^{n} \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, dN_i(u) - \int_0^T \lambda_0(u) \left[\sum_{i=1}^{n} (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, \kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i} \right] du \\
&= \sum_{i=1}^{n} \int_0^T (\mathbf{z}_i - \bar{\mathbf{z}}(u))\, dN_i(u) \qquad (8)
\end{aligned}
$$

which does *not* contain the nuisance parameter $\lambda_0(t)$ (see, e.g., Fleming and Harrington 1991, Eq. (3.24), p. 149). It becomes too simple to see that it is actually a least squares-type normal equation. In this case, the equality (Eq. (7)) helps us remove the infinite dimensional nuisance parameter $\lambda_0(t)$ from the derived e.f. $\mathbf{U}_T^{\star}(\boldsymbol{\beta})$ completely so that the estimation of $\boldsymbol{\beta}$ from it does *not* depend on $\lambda_0(t)$ at all.

Finally, we must point out that the above e.f. $\mathbf{U}_T^{\star}(\boldsymbol{\beta})$ (esp., Eq. (8)) is identical to the partial score function derived from the partial likelihood function (based on *multinomial* probabilities at each event time t) for obtaining the MPLE of $\boldsymbol{\beta}$ in Cox's proportional hazards model (see, e.g., Fleming and Harrington 1991, Eqs. (2.4), (2.5), (3.24), and (3.25), pp. 12-13 and 149-150, Andersen, et. al. 1993, pp. 105 and 483-487). Hence, the root $\widetilde{\boldsymbol{\beta}}^{\star}$ of the proposed e.f. $\mathbf{U}_T^{\star}(\boldsymbol{\beta})$ is not only the MPLE but also an adaptive GMM estimator of $\boldsymbol{\beta}$.

3.4 Properties

To examine the unbiasedness properties of the crude e.f. $\mathbf{U}_T(\boldsymbol{\beta})$ (Eq. (3)) and the derived e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ (Eqs. (6) and (8)), we first compute

$$
\begin{aligned}
E\{\mathbf{U}_T(\boldsymbol{\beta})\} &= E\left[\sum_{i=1}^{n}\int_0^T \mathbf{z}_i dM_i(u)\right] \\
&= \sum_{i=1}^{n} E\left[\int_0^T \mathbf{z}_i E\left(dM_i(u) \mid \mathcal{F}(u-)\right)\right] \\
&= \sum_{i=1}^{n} E\left\{\int_0^T \mathbf{z}_i \Big[E\left(dN_i(u) \mid \mathcal{F}(u-)\right) \;-\; \kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du\Big]\right\} \\
&= \mathbf{0}
\end{aligned}
$$

since $\mathbf{z}_i$ is bounded and predictable given the filtration $\mathcal{F}(u-)$. By the same token, we have

$$
E\{\mathbf{U}_T^\star(\boldsymbol{\beta})\} \;=\; E\left[\sum_{i=1}^{n}\int_0^T \left(\mathbf{z}_i - \bar{\mathbf{z}}(u)\right) dM_i(u)\right] \;=\; \mathbf{0}.
$$

Equivalently, according to the **Theorem 1.5.1** of Fleming and Harrington (1991, pp. 46-47 and 49), both e.f.'s $\mathbf{U}_T(\boldsymbol{\beta})$ and $\mathbf{U}_T^\star(\boldsymbol{\beta})$ are m.e.f.'s.

For any vector $\mathbf{v}$, let the outer product $\mathbf{v}\mathbf{v}'$ be denoted as $\mathbf{v}^{\otimes 2}$. Then, assuming that the cumulative baseline hazard function $\Lambda_0(t)$, where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$, is continuous (if not, then an alternative assumption is needed), we obtain the $(p \times p)$ *predictable covariation process* of the m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$, which is the compensator of $[\mathbf{U}_T^\star(\boldsymbol{\beta})]^{\otimes 2}$, that

$$
\begin{aligned}
< \mathbf{U}^\star(\boldsymbol{\beta}) >_T &= \sum_{i=1}^{n} < \mathbf{U}_i^\star(\boldsymbol{\beta}) >_T \;+\; \sum_{i \neq j} < \mathbf{U}_i^\star(\boldsymbol{\beta}), \mathbf{U}_j^\star(\boldsymbol{\beta}) >_T \\
&= \sum_{i=1}^{n}\int_0^T [\mathbf{z}_i - \bar{\mathbf{z}}(u)]^{\otimes 2}\, \kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du \qquad (9)
\end{aligned}
$$

due to the *orthogonality* property that $< M_i, M_j >_T = 0$, i.e., $Cov(dM_i(u), dM_j(u) \mid \mathcal{F}(u-))$

= 0, for all $i \neq j$ (Fleming and Harrington 1991, pp. 42, 81-82, 86-87, and 150, Andersen, et. al. 1993, pp. 69 and 74).

On the other hand, since the bounded $\mathcal{F}(t)$-predictable process $\mathbf{H}_i^\star(t;\boldsymbol{\beta})$ is a function of $\boldsymbol{\beta}$, the $(p \times p)$ *derivative process* of the m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ with respective to $\boldsymbol{\beta}$ (based on Eq. (6)), denoted as $\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta})$, is a sum of two terms

$$\begin{aligned}\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta}) &\equiv \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sum_{i=1}^n \int_0^T \mathbf{H}_i^\star(u;\boldsymbol{\beta})dM_i(u)\right] \\ &= \sum_{i=1}^n \int_0^T \dot{\mathbf{H}}_i^\star(u;\boldsymbol{\beta})dM_i(u) \;+\; \sum_{i=1}^n \int_0^T \mathbf{H}_i^\star(u;\boldsymbol{\beta})d\dot{M}_i(u) \qquad (10)\end{aligned}$$

where the "dot" put on the tops of $\mathbf{H}_i^\star(u;\boldsymbol{\beta})$ and $dM_i(u)$ respectively indicates the derivatives with respective to $\boldsymbol{\beta}$. Since $\dot{\mathbf{H}}_i^\star(u;\boldsymbol{\beta})$ is also a bounded $\mathcal{F}(t)$-predictable process, the first term on the right-hand side of Eq. (10) is a martingale (Fleming and Harrington 1991, Theorem 1.5.1, pp. 46-47 and 49). Thus, the second term on the right-hand side of Eq. (10)

$$\sum_{i=1}^n \int_0^T \mathbf{H}_i^\star(u;\boldsymbol{\beta})d\dot{M}_i(u) \;=\; -\sum_{i=1}^n \int_0^T \left(\mathbf{z}_i - \bar{\mathbf{z}}(u)\right)\mathbf{z}_i^{'}\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}^{'}\mathbf{z}_i}du \qquad (11)$$

is the compensator of the derivative process $\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta})$. However, as shown below, Eqs. (11) and (9) are closely related:

$$\begin{aligned}-\left[\sum_{i=1}^n \int_0^T \mathbf{H}_i^\star(u;\boldsymbol{\beta})d\dot{M}_i(u)\right]^{'} &- <\mathbf{U}^\star(\boldsymbol{\beta})>_T \\ &= \sum_{i=1}^n \int_0^T \mathbf{z}_i\left(\mathbf{z}_i - \bar{\mathbf{z}}(u)\right)^{'}\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}^{'}\mathbf{z}_i}du \;- \\ &\quad \sum_{i=1}^n \int_0^T \left(\mathbf{z}_i - \bar{\mathbf{z}}(u)\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}(u)\right)^{'}\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}^{'}\mathbf{z}_i}du \\ &= \int_0^T \bar{\mathbf{z}}(u)\left[\sum_{i=1}^n \left(\mathbf{z}_i - \bar{\mathbf{z}}(u)\right)^{'}\kappa_i(u)e^{\boldsymbol{\beta}^{'}\mathbf{z}_i}\right]\lambda_0(u)du \\ &= \mathbf{0}\end{aligned}$$

due to the equality of Eq. (7). Hence, a version of the so-called *quasi-information equality* holds for the derived m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ in the estimation of $\boldsymbol{\beta}$ alone (see, e.g., Andersen, et. al. 1993, pp. 102-103).

Alternatively, by computing the $(p \times p)$ *derivative process* $\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta})$ of Eq. (10) directly from Eq. (8), we have

$$\begin{aligned}
\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta}) &\equiv \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sum_{i=1}^n \int_0^T \mathbf{H}_i^\star(t;\boldsymbol{\beta}) dN_i(u)\right] \\
&= -\sum_{i=1}^n \int_0^T \frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) dN_i(u) \\
&= -\sum_{i=1}^n \int_0^T \left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j\right]^{\otimes 2}}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) - \left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right)^{\otimes 2} dN_i(u) \\
&= -\sum_{i=1}^n \int_0^T \left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j\right]^{\otimes 2}}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) - \left[\bar{\mathbf{z}}(u)\right]^{\otimes 2} dN_i(u). \qquad (12)
\end{aligned}$$

Since

$$\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j - \bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}} = \frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\left\{\left[\mathbf{z}_j\right]^{\otimes 2} - \mathbf{z}_j\bar{\mathbf{z}}'(u) - \bar{\mathbf{z}}(u)\mathbf{z}_j' + \left[\bar{\mathbf{z}}(u)\right]^{\otimes 2}\right\}}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}$$

and

$$\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j\bar{\mathbf{z}}'(u)\right]}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}} = \frac{\left\{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\mathbf{z}_j\right\}\bar{\mathbf{z}}'(u)}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}} = \left[\bar{\mathbf{z}}(u)\right]^{\otimes 2}$$

we obtain from Eq. (12) that

$$\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta}) = -\sum_{i=1}^n \int_0^T \left(\frac{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j - \bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^n \kappa_j(u) e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right) dN_i(u) \qquad (13)$$

where the thing inside the big parentheses is an empirical weighted mean of the covariances of $\mathbf{z}_j$'s at time u. Again, according to the fact that the first term on the right-hand side of Eq. (10) is a martingale, the compensator of the derivative process $\dot{\mathbf{U}}_T^\star(\boldsymbol{\beta})$ based on Eq. (13)

is simply

$$-\sum_{i=1}^{n}\int_{0}^{T}\left(\frac{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j-\bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right)\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du. \tag{14}$$

Then, as shown below, Eqs. (14) and (9) are also closely related:

$$\begin{aligned}
&\sum_{i=1}^{n}\int_{0}^{T}\left(\frac{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j-\bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right)\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du \\
&\qquad\qquad -\sum_{i=1}^{n}\int_{0}^{T}\left[\mathbf{z}_i-\bar{\mathbf{z}}(u)\right]^{\otimes 2}\kappa_i(u)\lambda_0(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du \\
&= \int_{0}^{T}\left[\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j-\bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right)\kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}\right. \\
&\qquad\qquad \left. -\sum_{i=1}^{n}\left[\mathbf{z}_i-\bar{\mathbf{z}}(u)\right]^{\otimes 2}\kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}\right]\lambda_0(u)du \\
&= \int_{0}^{T}\left[\left(\frac{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j-\bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right)\sum_{i=1}^{n}\kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}\right. \\
&\qquad\qquad \left. -\sum_{i=1}^{n}\left[\mathbf{z}_i-\bar{\mathbf{z}}(u)\right]^{\otimes 2}\kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}\right]\lambda_0(u)du \\
&= \int_{0}^{T}\left[\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j-\bar{\mathbf{z}}(u)\right]^{\otimes 2}-\sum_{i=1}^{n}\left[\mathbf{z}_i-\bar{\mathbf{z}}(u)\right]^{\otimes 2}\kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}\right]\lambda_0(u)du \\
&= \mathbf{0}
\end{aligned}$$

which implies that $-E[\dot{\mathbf{U}}^{\star}_T(\boldsymbol{\beta})] = E\left\{[\mathbf{U}^{\star}_T(\boldsymbol{\beta})]^{\otimes 2}\right\}$. And, analogous to Eq. (7), we get

$$\sum_{i=1}^{n}\int_{0}^{T}\left(\left[\mathbf{z}_i-\bar{\mathbf{z}}(u)\right]^{\otimes 2}-\frac{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}\left[\mathbf{z}_j-\bar{\mathbf{z}}(u)\right]^{\otimes 2}}{\sum_{j=1}^{n}\kappa_j(u)e^{\boldsymbol{\beta}'\mathbf{z}_j}}\right)\kappa_i(u)e^{\boldsymbol{\beta}'\mathbf{z}_i}du = \mathbf{0}. \tag{15}$$

Hence, an equivalent version of the *quasi-information equality* also holds for the derived m.e.f. $\mathbf{U}^{\star}_T(\boldsymbol{\beta})$ in the estimation of $\boldsymbol{\beta}$ alone (see, e.g., Fleming and Harrington 1991, Eq. (3.27), p. 151).

Finally, Eq. (15) may suggest that

$$\sum_{i=1}^{n} \int_0^T \left([\mathbf{z}_i - \bar{\mathbf{z}}(u)]^{\otimes 2} - \frac{\sum_{j=1}^{n} \kappa_j(u) e^{\boldsymbol{\beta}' \mathbf{z}_j} [\mathbf{z}_j - \bar{\mathbf{z}}(u)]^{\otimes 2}}{\sum_{j=1}^{n} \kappa_j(u) e^{\boldsymbol{\beta}' \mathbf{z}_j}} \right) dN_i(u) \quad = \quad \mathbf{0} \tag{16}$$

be used as the basis for constructing an information equality-type model specification test, as long as the thing inside the big parentheses is bounded and $\mathcal{F}(t)$-predictable, since Eq. (16), like Eq. (8), does *not* involve the nuisance parameter $\lambda_0(t)$ at all.

3.5 Some Extensions

In this study, we use only one moment condition $dM_i(t)$ for estimating the p regression coefficients $\boldsymbol{\beta}$ in Cox's proportional hazards model (Eq. (1)). However, when there is more information available to be used as moment conditions in some cases, the GMM approach actually provides us with a natural way to combine the information about $\boldsymbol{\beta}$ from all available moment conditions by constructing a properly stacked quadratic form (see, e.g., Greene 2000, Secs. 11.5-11.6, pp. 474-496). Thus, the most efficient estimate of $\boldsymbol{\beta}$ can be obtained in such a systematic way. And, by doing so, the resulting m.e.f. for $\boldsymbol{\beta}$ would still consist of exact p equations although there are more than one moment condition used in the estimation of $\boldsymbol{\beta}$.

Moreover, in view of the algebraic structure in the equality of Eq. (7), we may take more general forms of the $(p \times 1)$ bounded $\mathcal{F}(t)$-predictable process than $\mathbf{H}_i^{\star}(t;\boldsymbol{\beta}) \equiv \mathbf{z}_i - \bar{\mathbf{z}}(t)$ for estimating $\boldsymbol{\beta}$ in Cox's proportional hazards model or other similar regression models for hazard functions. For example, letting $\mathbf{h}_1(t), \mathbf{h}_2(t), \cdots, \mathbf{h}_n(t)$ be any $(p \times 1)$ bounded $\mathcal{F}(t)$-predictable processes and $g_1(\cdot), g_2(\cdot), \cdots, g_n(\cdot)$ be known deterministic functions, we can choose

$$\mathbf{H}_i^{\star\star}(t;\boldsymbol{\beta}) \quad \equiv \quad \frac{g_i(e^{\boldsymbol{\beta}' \mathbf{z}_i(t)}) \mathbf{h}_i(t)}{e^{\boldsymbol{\beta}' \mathbf{z}_i(t)}} - \frac{\sum_{j=1}^{n} \kappa_j(t) g_j(e^{\boldsymbol{\beta}' \mathbf{z}_j(t)}) \mathbf{h}_j(t)}{\sum_{j=1}^{n} \kappa_j(t) e^{\boldsymbol{\beta}' \mathbf{z}_j(t)}}$$

which does not contain the nuisance parameter $\lambda_0(t)$ and holds the same kind of equality as Eq. (7) by direct calculations. When $g_i(\exp\{\boldsymbol{\beta}'\mathbf{z}_i(t)\}) = \exp\{\boldsymbol{\beta}'\mathbf{z}_i\}$ and $\mathbf{h}_i(t) = \mathbf{z}_i$ for all subjects, it reduces to $\mathbf{H}_i^\star(t;\boldsymbol{\beta})$.

Finally, as proposed by Prentice and Self (1983), we can also extend the above-derived results for Cox's proportional hazards model to the following more general class of regression models for hazard functions

$$\lambda(t \mid \mathbf{z}_i(t)) \;=\; \lambda_0(t) r(\boldsymbol{\beta}'\mathbf{z}_i(t))$$

in which the risk multiplier $r(\cdot) > 0$ is a twice differentiable known function of the linear combination $\boldsymbol{\beta}'\mathbf{z}_i(t)$. In this setting, the log partial likelihood function to be maximized for obtaining the MPLE of $\boldsymbol{\beta}$ is

$$\mathcal{L}_T(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} \int_0^T \left\{ \log\left[r(\boldsymbol{\beta}'\mathbf{z}_i(u)) \right] \;-\; \log\left[\sum_{j=1}^{n} \kappa_j(u) r(\boldsymbol{\beta}'\mathbf{z}_j(u)) \right] \right\} dN_i(u)$$

and thus the partial score function for estimating $\boldsymbol{\beta}$ becomes

$$\frac{\partial \mathcal{L}_T(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \;=\; \sum_{i=1}^{n} \int_0^T \left(\frac{\dot{r}(\boldsymbol{\beta}'\mathbf{z}_i(u))\mathbf{z}_i(u)}{r(\boldsymbol{\beta}'\mathbf{z}_i(u))} \;-\; \frac{\sum_{j=1}^{n} \kappa_j(u) \dot{r}(\boldsymbol{\beta}'\mathbf{z}_j(u))\mathbf{z}_j(u)}{\sum_{j=1}^{n} \kappa_j(u) r(\boldsymbol{\beta}'\mathbf{z}_j(u))} \right) dN_i(u)$$

which can also be derived as an m.e.f. for obtaining an adaptive GMM estimator of $\boldsymbol{\beta}$ as long as the thing inside the big parentheses is bounded and $\mathcal{F}(t)$-predictable.

4 DISCUSSION

4.1 Summary

In brief, to estimate the unknown regression coefficients $\boldsymbol{\beta}$ in Cox's proportional hazards model (Eq. (1)) without specifying the baseline hazard function $\lambda_0(t)$, we use the moment condition $dM_i(t)$ and the corresponding quadratic form $Q_T(\boldsymbol{\beta})$ (Eq. (2)) from the GMM estimation together as a useful device to constructing a class of m.e.f.'s for estimating $\boldsymbol{\beta}$ consistently. In particular, by ignoring purposely the $\boldsymbol{\beta}$ inside the weight function of $Q_T(\boldsymbol{\beta})$ as if it were known in the minimization of $Q_T(\boldsymbol{\beta})$, we maintain the unbiasedness of the crude m.e.f. $\mathbf{U}_T(\boldsymbol{\beta})$ (Eq. (3)).

Next, we derive the proposed m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ (Eq. (6)), which satisfies the key condition for adaptive estimation that $E\left[\partial\mathbf{U}_T^\star(\boldsymbol{\beta})/\partial\lambda_0(t)\right] = \mathbf{0}$ (Eq. (4)), to get rid of the impact of the unknown nuisance parameter $\lambda_0(t)$ on the asymptotic distribution of its root $\widetilde{\boldsymbol{\beta}}^\star$, although we need to plug a consistent estimator of $\lambda_0(t)$ into it directly or iteratively in the estimation of $\boldsymbol{\beta}$. In fact, it happens in the case of Cox's proportional hazards model that by doing so, we get rid of the unknown nuisance parameter $\lambda_0(t)$ completely from the derived m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ (Eq. (6)) due to the equality of Eq. (7), and thus the estimation of $\boldsymbol{\beta}$ using the derived m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ in its reduced form (Eq. (8)) does *not* depend on $\lambda_0(t)$ at all. Notice that under regularity conditions, this *asymmetric orthogonal expected information approach* to dealing with nuisance parameters is valid for the unbiased joint e.f.'s for the parameters of interest and the nuisance parameters with either a *symmetric* or *non-symmetric* expected joint derivative matrix (see **Appendix**) so that it is more general than the usual *orthogonal parameter approach*, which requests orthogonality (i.e., zero covariances) between the two sets of unbiased e.f.'s for the parameters of interest and the nuisance parameters respectively (Chang and Hsiung 1990, 1991, Small and McLeish 1994, Sec. 5.2, pp. 109-110, Pagan and Ullah 1999, Sec. 5.4, pp. 217-225), for adaptive estimation.

Since the partial score function for estimating $\boldsymbol{\beta}$ in Cox's proportional hazards model can be represented as an adaptive GMM m.e.f. $\mathbf{U}_T^\star(\boldsymbol{\beta})$ (Eqs. (6) and (8)), the MPLE of $\boldsymbol{\beta}$

has a GMM interpretation, and thus it can be considered as a general least squares estimator of $\boldsymbol{\beta}$. Then, with this result in mind, we may rewrite the Cox's proportional hazards model in the form of a nonlinear regression model:

For $i = 1, 2, \cdots, n$,

$$E[dN_i \mid \mathcal{F}(t-)] \; = \; \lambda_0(t) e^{\boldsymbol{\beta}' \mathbf{z}_i} dt$$

$$\log \left(E[dN_i \mid \mathcal{F}(t-)] \right) \; = \; \log(\lambda_0(t)) dt \; + \; \boldsymbol{\beta}' \mathbf{z}_i$$

or equivalently

$$dN_i \mid \mathcal{F}(t-) \; = \; \lambda_0(t) e^{\boldsymbol{\beta}' \mathbf{z}_i} dt \; + \; \varsigma_i(t) dt$$

$$N_i \; = \; \int_0^{X_i} \lambda_0(t) e^{\boldsymbol{\beta}' \mathbf{z}_i} dt \; + \; \int_0^{X_i} \varsigma_i(t) dt \; = \; \int_0^{X_i} \lambda_0(t) e^{\boldsymbol{\beta}' \mathbf{z}_i} dt \; + \; \epsilon_i(X_i)$$

where $X_i = T_i \wedge C_i = \min(T_i, C_i)$ and $0 \leq t \leq X_i$.

4.2 Future Work

Given the moment condition $dM_i(t)$ for estimating the regression coefficients $\boldsymbol{\beta}$ in Cox's proportional hazards model (Eq. (1)), we have learned from this study the important roles of $\mathbf{z}_i$ and $\bar{\mathbf{z}}(t)$ in the optimal m.e.f. $\mathbf{U}_T^{\star}(\boldsymbol{\beta})$ (Eqs. (6) and (8)) respectively: (1) $\mathbf{z}_i$ comes from the derivatives $\partial \left[dM_i(t) \right] / \partial \boldsymbol{\beta}$, and thus it is expected to be orthogonal to $dM_i(t)$ given the unbiasedness of the m.e.f.; and (2) centering $\mathbf{z}_i$ by $\bar{\mathbf{z}}(t)$ helps us get rid of the nuisance parameter $\lambda_0(t)$. Since the statistical methodologies adopted in this study including the GMM estimation method, theory of e.f.'s, adaptive estimation, and martingales have not yet exerted their full powers in our case, the finding of this study is not only very interesting in its own rights, but it provides us with an opportunity to develop GLMs-type regression models locally (at each time t) for more general stochastic processes and to apply some powerful GMM-related estimating techniques such as the *instrumental variables* (IV) method for nonlinear equations to deal with several known statistical modeling problems in analysis of survival or time-to-event data.

In particular, Bowden and Turkington (1984, esp., Sec. 1.2, pp. 10-16), Greene (2000, Sec. 9.5 and Subsecs. 10.3.2, 11.5.5, 13.7.3, and 16.5.2, pp. 370-387, 430-438, 483-488, 550-552, and 680-690), and Newey (1990) among others gave nice reviews and discussions on the IV estimation method and its applications in econometric models. In the linear equation settings, the IV method has proved its power in solving the estimation problems occurring when some of the covariates in the equation are *correlated* with equation's error. Thus, it can be used to deal with four kinds of statistical modeling problems: (1) the error-in-variable (or measurement error) problem, (2) the self-selection problem, (3) the simultaneous-equations bias, and (4) the time series problem (Bowden and Turkington 1984, pp. 3-10). Then, based on the result of this study, we are specifically interested in developing IV estimators for the regression coefficients $\boldsymbol{\beta}$ in Cox's proportional hazards model to deal with the measurement error problem and simultaneous-equations bias respectively. It is our understanding that when these problems arise in Cox's proportional hazards model, some of the covariates $\mathbf{z}_i$ would *not* be orthogonal to $dM_i(t)$. And, for this purpose, our IV estimation method for GLMs (Hu, et. al. 2002) may be applied to Cox's proportional hazards model locally (at each time t).

5 ACKNOWLEDGEMENTS

This paper is based on the first part of the second author's dissertation under the first author's advice. And, this research was financially supported by the National Science Council of the Republic of China (NSC 91-2118-M-002-004).

6 APPENDIX

We shall give a sketch of the rationale behind the proposed *asymmetric orthogonal expected information approach* to dealing with the nuisance parameter $\lambda_0(t)$ for an adaptive estimation of the interested parameters $\boldsymbol{\beta}$. Yet, this general approach can possibly be applied to many different settings. Without loss of generality, we assume that $\lambda_0(t)$ can be parameterized by a finite number of parameters for simplicity. Let the estimating functions for $\boldsymbol{\beta}$ and $\lambda_0(t)$ be denoted as $\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \lambda_0(t))$ and $\mathrm{U}_{\lambda_0}(\boldsymbol{\beta}, \lambda_0(t))$ respectively, of which both are functions of $(\boldsymbol{\beta}, \lambda_0(t))$. And, given a random sample of size n, we assume that both $\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \lambda_0(t))$ and $\mathbf{U}_{\lambda_0}(\boldsymbol{\beta}, \lambda_0(t))$ are unbiased, the solutions $(\widetilde{\boldsymbol{\beta}}^{\star}, \widetilde{\lambda}_0(t))$ to these two estimating equations exist, and they are consistent estimates of $(\boldsymbol{\beta}, \lambda_0(t))$.

The *key condition* required for adaptively estimating $\boldsymbol{\beta}$ (see Eq. (4) in the text) is that at the true values of $\boldsymbol{\beta}$,

$$E\Big[\mathbf{I}^{\star}_{\boldsymbol{\beta}\lambda_0}\Big] \;=\; -\,E\left[\frac{\partial \mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \lambda_0(t))}{\partial \lambda_0(t)}\right] \;=\; \mathbf{0}.$$

Then, symbolically, the expected values of the minus mixed derivatives of the joint estimating functions $\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \lambda_0(t))$ and $\mathbf{U}_{\lambda_0}(\boldsymbol{\beta}, \lambda_0(t))$ with respect to the parameters $(\boldsymbol{\beta}, \lambda_0(t))$ are

$$E\begin{bmatrix} \mathbf{I}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}^{\star}_{\boldsymbol{\beta}\lambda_0} \\ \mathbf{I}_{\lambda_0\boldsymbol{\beta}} & \mathbf{I}_{\lambda_0\lambda_0} \end{bmatrix} = \begin{bmatrix} \mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{i}_{\lambda_0\boldsymbol{\beta}} & \mathbf{i}_{\lambda_0\lambda_0} \end{bmatrix}$$

which may be *non-symmetric*. And, with a zero on the upper right-hand corner of the above partitioned matrix, it is straightforward to show that

$$\begin{bmatrix} \mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{i}_{\lambda_0\boldsymbol{\beta}} & \mathbf{i}_{\lambda_0\lambda_0} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1} & \mathbf{0} \\ \mathbf{i}^{\lambda_0\boldsymbol{\beta}} & (\mathbf{i}_{\lambda_0\lambda_0})^{-1} \end{bmatrix}$$

where $\mathbf{i}^{\lambda_0\boldsymbol{\beta}} \neq (\mathbf{i}_{\lambda_0\boldsymbol{\beta}})^{-1}$ and it is in a complicated form (see, e.g., Khuri 1993, p. 34).

Now, expanding the joint estimating functions $(\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \lambda_0(t)), \mathbf{U}_{\lambda_0}(\boldsymbol{\beta}, \lambda_0(t)))$ about the true values $(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t))$ of the parameters $(\boldsymbol{\beta}, \lambda_0(t))$ by the **mean value theorem** yields

$$\begin{pmatrix} \mathbf{U}^{\star}_{\boldsymbol{\beta}}(\widetilde{\boldsymbol{\beta}}^{\star}, \widetilde{\lambda}_0(t)) \\ \mathbf{U}_{\lambda_0}(\widetilde{\boldsymbol{\beta}}^{\star}, \widetilde{\lambda}_0(t)) \end{pmatrix} = \begin{pmatrix} \mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \\ \mathbf{U}_{\lambda_0}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \end{pmatrix} - \left. \begin{bmatrix} \mathbf{I}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{I}^{\star}_{\boldsymbol{\beta}\lambda_0} \\ \mathbf{I}_{\lambda_0\boldsymbol{\beta}} & \mathbf{I}_{\lambda_0\lambda_0} \end{bmatrix} \right|_{\boldsymbol{\beta}^{*}, \lambda_0^{*}(t)} \begin{pmatrix} \widetilde{\boldsymbol{\beta}}^{\star} - \boldsymbol{\beta}^{\bullet} \\ \widetilde{\lambda}_0(t) - \lambda_0^{\bullet}(t) \end{pmatrix}.$$

Then, under the regularity conditions,

$$\begin{aligned} n^{1/2} \begin{pmatrix} \widetilde{\boldsymbol{\beta}}^{\star} - \boldsymbol{\beta}^{\bullet} \\ \widetilde{\lambda}_0(t) - \lambda_0^{\bullet}(t) \end{pmatrix} &= \left. \begin{bmatrix} \frac{\mathbf{I}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}}{n} & \frac{\mathbf{I}^{\star}_{\boldsymbol{\beta}\lambda_0}}{n} \\ \frac{\mathbf{I}_{\lambda_0\boldsymbol{\beta}}}{n} & \frac{\mathbf{I}_{\lambda_0\lambda_0}}{n} \end{bmatrix}^{-1} \right|_{\boldsymbol{\beta}^{*}, \lambda_0^{*}(t)} n^{-1/2} \begin{pmatrix} \mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \\ \mathbf{U}_{\lambda_0}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \end{pmatrix} \\ &\cong \begin{bmatrix} \frac{\mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}}{n} & \mathbf{0} \\ \frac{\mathbf{i}_{\lambda_0\boldsymbol{\beta}}}{n} & \frac{\mathbf{i}_{\lambda_0\lambda_0}}{n} \end{bmatrix}^{-1} n^{-1/2} \begin{pmatrix} \mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \\ \mathbf{U}_{\lambda_0}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \end{pmatrix} \\ &= \begin{bmatrix} \left(\frac{\mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}}{n}\right)^{-1} & \mathbf{0} \\ n\,\mathbf{i}^{\lambda_0\boldsymbol{\beta}} & \left(\frac{\mathbf{i}_{\lambda_0\lambda_0}}{n}\right)^{-1} \end{bmatrix} n^{-1/2} \begin{pmatrix} \mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \\ \mathbf{U}_{\lambda_0}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \end{pmatrix}. \end{aligned}$$

Thus, by picking up the first p elements in the vector on the left-hand side,

$$\begin{aligned} n^{1/2}(\widetilde{\boldsymbol{\beta}}^{\star} - \boldsymbol{\beta}^{\bullet}) &\cong \left(\frac{\mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}}{n}\right)^{-1} n^{-1/2}\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) + \mathbf{0} \cdot n^{-1/2}\mathbf{U}_{\lambda_0}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \\ &= \left(\frac{\mathbf{i}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}}{n}\right)^{-1} n^{-1/2}\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)). \end{aligned}$$

According to the **multivariate central limit theorem**,

$$n^{-1/2}\mathbf{U}^{\star}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t)) \xrightarrow{\mathcal{D}} \mathbf{N}\left(\mathbf{0}, \lim_{n\to\infty} \left[\frac{\mathbf{I}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda_0^{\bullet}(t))}{n}\right]\right).$$

Therefore,

$$n^{1/2}(\widetilde{\boldsymbol{\beta}}^{\star} - \boldsymbol{\beta}^{\bullet}) \quad \xrightarrow{\mathcal{D}} \quad \mathbf{N}\left(\mathbf{0}, \ \lim_{n\to\infty}\left[\frac{\mathbf{I}^{\star}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}^{\bullet}, \lambda^{\bullet}_{0}(t))}{n}\right]^{-1}\right)$$

which is exactly the same asymptotic distribution of $\widetilde{\boldsymbol{\beta}}^{\star}$ as if $\lambda_0(t)$ were known.

7 REFERENCES

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

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

Begun, J. M., Hall, W. J., Huang, W. & Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432-53.

Bowden, R. J. & Turkington, D. A. (1984). *Instrumental Variables*. Cambridge: Cambridge University Press.

Chang, I. -S. & Hsiung, C. A. (1990). Finite sample optimality of maximum partial likelihood estimation in Cox's model for counting process. *J. Statist. Plan. Inf.* **25**, 35-42.

Chang, I. -S. & Hsiung, C. A. (1991). Applications of estimating function theory to replicates of generalized proportional hazards models. In: V. P. Godambe (Ed), *Estimating Functions*, Oxford: Clarendon Press, pp. 23-33.

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**, 269-76.

Desmond, A. F. (1991). Quasi-likelihood, stochastic processes, and optimal estimating functions. In: V. P. Godambe (Ed), *Estimating Functions*, Oxford: Clarendon Press, pp. 133-46.

Fleming, T. R. & Harrington, D. D. (1991). *Counting Processes and Survival Analysis*. New York, NY: John Wiley & Sons.

Gill, R. D. (1984). Understanding Cox's regression model: A martingale approach. *J. Amer. Statist. Assoc.* **79**, 441-7.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208-12.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **72**, 419-28.

Godambe, V. P. & Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *Intern. Statist. Rev.* **55**, 231-44.

Godambe, V. P. & Kale, B. K. (1991). Estimating functions: An overview. In: V. P. Godambe (Ed), *Estimating Functions*, Oxford: Clarendon Press, pp. 3-20.

Greene, W. H. (2000). *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice-Hall.

Greenwood, P. E. & Wefelmeyer, W. (1991). On optimal estimating functions for partially specified counting process models. In: V. P. Godambe (Ed), *Estimating Functions*, Oxford: Clarendon Press, pp. 147-60.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. New York, NY: Springer-Verlag.

Hu, F. -C., Lai, S. -H., Tsai, T. -L. & Shau, W. -Y. (2002). The method of instrumental variables for generalized linear models. Technical report. Division of Biostatistics, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan, R.O.C..

Khuri, A. I. (1993). *Advanced Calculus with Applications in Statistics*. New York, NY: John Wiley & Sons.

McCullagh, P. (1991). Quasi-likelihood and estimating functions. In: D. V. Hinkley, N. Reid & E. J. Snell (Eds), *Statistical Theory and Modelling: In Honour of Sir David Cox, FRS*, London: Chapman and Hall, pp. 265-86.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

Moore, D. F. (1986). Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika* **73**, 583-8.

Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc., Ser. A* **135**, 370-84.

Newey, W. K. (1990). Efficient instrumental variable estimation of nonlinear models. *Econometrica* **58**, 809-38.

Pagan, A. & Ullah, A. (1999). *Nonparametric Econometrics*. New York, NY: Cambridge University Press.

Prentice, R. L. & Self, S. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11**, 804-13.

Small, C. G. & McLeish, D. L. (1994). *Hilbert Space Methods in Probability and Statistical Inference*. New York, NY: John Wiley & Sons.

Thavanewswaran, A. & Thompson, M. E. (1986). Optimal estimation for semimartingales. *J. Appl. Probab.* **23**, 409-17.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93-108.

Wedderburn, R. W. M. (1974). Quasilikelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439-47.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31**, 144-8.