Statistics Canada | Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes auprès des entreprises

Canada

WORKING PAPER NO. BSMD-2006-003E
METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. DMEE-2006-003E
DIRECTION DE LA MÉTHODOLOGIE

ON FITTING COX'S PROPORTIONAL HAZARDS MODELS TO DATA FROM COMPLEX SURVEYS

by

Susana Rubin-Bleuer

BSMD-2006-003E

ON FITTING COX'S PROPORTIONAL HAZARDS MODELS TO DATA FROM COMPLEX SURVEYS

by

Susana Rubin-Bleuer[1]

March 2006
Business Survey Methods Division
Statistics Canada

1 Susana Rubin-Bleuer, Statistics Canada, Ottawa, Canada, K1H-0T6, rubisus@statcan.ca

ABSTRACT

We use the "survey sample" partial likelihood score function to fit the proportional hazards regression model to survey data with complex sampling designs. The survey sample maximum partial likelihood estimator is the solution of the survey sample partial likelihood score function. Many authors applied this method to fit survival survey data. Binder (1992) dealt with inference on the descriptive census population parameter, that is, design-based inference on the maximum partial likelihood estimate that could be calculated had a census been taken on the finite population. Lin (2000) gave a formal justification of Binder's method under the super-population approach and dealt with inference on the model parameter. Neither Binder nor Lin provided conditions for the respective asymptotic results to hold. Rubin-Bleuer (2003 b) used Lin's (2000) set up of the super-population approach and developed counting process methodology for a joint design-model space to obtain, under stated sufficient model and design conditions, a proof of Binder's approximation of the SPLS. In this paper, we give a rigorous proof of the weak convergence of the SPLS process and the asymptotic normality of the sample maximum partial likelihood estimator in a formally expressed joint design-model space and we propose a consistent variance estimator. Furthermore, using counting processes tools in the joint design-model space, we show that Lin's (2000) variance estimator is asymptotically unbiased and robust against misspecification of the correlation model. We also show that it is design-model consistent, and that it is less efficient than the variance estimator proposed here. Strict rates of approximation for Lin's (2000) variance estimator are given.

Key words: complex survey data, partial likelihood, proportional hazards, counting processes.

RÉSUMÉ

Nous utilisons la fonction de score de vraisemblance partielle « d'échantillon » pour ajuster un modèle de régression à risques proportionnels à des données provenant d'enquêtes à plan de sondage complexe. L'estimateur du maximum de vraisemblance partielle d'échantillon est la solution de la fonction de score de vraisemblance partielle d'échantillon. De nombreux auteurs ont appliqué cette méthode afin d'ajuster un modèle à des données d'enquête sur la survie. Binder (1992) s'est penché sur l'inférence concernant les paramètres descriptifs de population dans des conditions de recensement (population finie), c'est-à-dire l'inférence basée sur le plan de sondage concernant l'estimation du maximum de vraisemblance partielle qui pourrait être calculée si la population finie avait été recensée. Lin (2000) a cherché à justifier formellement la méthode de Binder sous l'approche de superpopulation et a traité de l'inférence concernant les paramètres du modèle. Ni Binder ni Lin n'ont précisé les conditions suffisantes pour que leurs résultats asymptotiques respectifs tiennent. Rubin-Bleuer (2003 b) a utilisé les conditions de l'approche de superpopulation de Lin (2000) et a élaboré une méthode basée sur un processus de comptage pour un espace conjoint plan de sondage-modèle afin d'obtenir, sous des conditions suffisantes clairement énoncées s'appliquant au modèle et au plan de sondage, une preuve de l'approximation du score de vraisemblance partielle d'échantillon (SPLS pour *Sample Partial Likelihood Score*) de Binder. Dans le présent article, nous donnons une preuve rigoureuse de la faible convergence du processus du SPLS et de la normalité asymptotique de l'estimateur du maximum de vraisemblance partielle d'échantillon dans un espace conjoint plan de sondage-modèle exprimé formellement, et nous proposons un estimateur convergent de la variance. En outre, au moyen d'outils pour les processus de comptage dans l'espace conjoint plan de sondage-modèle, nous montrons que l'estimateur de la variance de Lin (2000) est asymptotiquement sans biais et robuste à l'erreur de spécification du modèle de corrélation. Nous montrons aussi qu'il est convergent par rapport au plan de sondage et au modèle, quoique moins efficace que l'estimateur de la variance proposé ici. Nous donnons des taux d'approximation stricts pour l'estimateur de la variance de Lin (2000).

Mots clés : Données provenant d'enquêtes à plan complexe; espace mixte tenant compte du plan de sondage et du modèle; modèle de risques proportionnels; processus de comptage; robustesse

1. INTRODUCTION

The Cox (1972) proportional hazards regression model (PHM) provides a method for studying the effects of primary covariates on failure times, while adjusting for other variables. If we assume that no covariates vary with time and let $S(t \mid X) = 1 - P(T \le t \mid X)$ be the conditional survival function of the failure time T associated with an r-dimensional vector of covariates X, then the conditional hazard function (or instantaneous conditional failure rate) is defined by

$$\lambda(t \mid X) = \lim_{h \downarrow 0} h^{-1} P(t \le T < t + h \mid T \ge t, X).$$

The PHM specifies that the conditional hazard rate $\lambda(t \mid X)$ of the failure time T is given by

$$\lambda(t \mid X) = \lambda_0(t) \cdot \exp(\beta' \cdot X),$$

where $\lambda_0(t)$ is an unspecified baseline hazard function and β is an r-dimensional vector valued regression parameter pertaining to the log hazard ratio.

Most methods of survival analysis were developed for a random sample from a given model. In order to analyze survey sample data, survey samplers often think of it as the result of a two-phase randomization procedure (an approach introduced by Hartley and Sielken in 1975), where the infinite population (also called super-population) generates a finite population in the first phase, and the sample is selected from the finite population in the second phase. The finite population could have been completely observed, had we taken a census. The analysis of the data obtained in the first phase (that is, of the finite population) is called "the census case" from now on.

Fitting the PHM to survey data poses difficulties because complex survey data consist of dependent observations and are often subject to selection bias due to unequal selection probabilities (see for example, Pfeffermann, 1993). As a consequence, the usual asymptotic theory does not apply. The problem is to determine the properties of the "survey sample" estimator of β for inference.

The census case

The failure time T (also called lifetime) is subject to right censoring given by C. Let $I(A)$ be the indicator function of the set A. Let $\tilde{T} = \min(T, C)$ denote the censored lifetime, $\delta = I(T \le C)$ the indicator of whether the lifetime was censored or not, and $Y(t) = I(\tilde{T} \ge t)$, the indicator of whether the unit with lifetime $\tilde{T}$ was at risk or not at time t. The data consists of a realization of random triples $(\tilde{T}_i, \delta_i, X_i)$, $i = 1, \ldots, N$, independent but not necessarily identically distributed random vectors defined on a probability space $(\Omega, \Im, P)$. Under the PHM β_0 can be estimated from the "Census" Partial Likelihood Score (PLS) function

$$U(\beta) = \sum_{i=1}^{N} \delta_i \left\{ X_i - \frac{S^{(1)}(\beta, \tilde{T}_i)}{S^{(0)}(\beta, \tilde{T}_i)} \right\}, \qquad (1.1)$$

where the S-functions, given the covariates, are linear combinations of the risk functions $I(\tilde{T}_i \ge t)$:

$$S^{(0)}(\beta,t)=\frac{1}{N}\sum_{i=1}^{N} I(\tilde{T}_i \geq t)\cdot e^{\beta'\cdot X_i} \text{ and } S^{(1)}(\beta,t)=\frac{1}{N}\sum_{i=1}^{N} X_i \cdot I(\tilde{T}_i \geq t)\cdot e^{\beta'\cdot X_i}.$$

The solution to $U(\beta)=0$ yields β_N, the maximum partial likelihood estimator of the model parameter β_0. We call β_N the census parameter. Under regularity conditions, the expression $\sqrt{N}(\beta_N-\beta_0)$ is asymptotically normal with zero mean and covariance matrix that can be consistently estimated by $I^{-1}(\beta)=-\{(1/N)\partial U/\partial\beta\}^{-1}(\beta_N)$ (Andersen and Gill, 1982).

The survey data case

We assume now that the units of the finite population (census population) are classified into L strata, with N_h primary sampling units ($psus$) within stratum h, $h=1,\dots,L$, and N_{hi} secondary sampling units within psu i, $i=1,\dots,N_h$. We define $N=\sum_{h=1}^{L} N_h$, and $M=\sum_{h=1}^{L}\sum_{i=1}^{N_h} N_{hi}$, respectively, as the number of $psus$ and the number of ultimate units in the finite population. Throughout the paper we assume that the psu sizes N_{hi} remain bounded while the number of $psu's$ N increases towards infinity. Hence M and N are asymptotically equivalent. Even though here the number of ultimate units in the population is M, the solution to the Census PLS equation $U(\beta)=0$ will again be denoted by β_N.

The survey data consist of a subset s of the finite population,

$(\tilde{T}_{hik}, \delta_{hik}, X_{hik})$, $k=1,\dots,n_{hi}$, $i=1,\dots,n_h$, $h=1,\dots,L$, $n=n_1+\dots+n_L$,

which is selected according to a "without replacement" sampling design. Consider now a survey sample estimator of the Census Partial Likelihood Score function. For the sake of brevity, we will use the word "sample" rather than "survey sample". The "Sample Partial Likelihood Score" (SPLS) function is given by:

$$\hat{U}(\beta)=\sum_{hik\in s}\delta_{hik}\left\{X_{hik}-\frac{\hat{S}^{(1)}(\beta,\tilde{T}_{hik})}{\hat{S}^{(0)}(\beta,\tilde{T}_{hik})}\right\}/\pi_{hik},$$

where $hik\in s$ means the unit hik in the finite (census) population is included in the selected sample, π_{hik} is the sampling inclusion probability for the unit hik and the $\hat{S}$-functions are the Horvitz-Thompson (HT) estimators of the finite population S-functions shown in (1.1) (see Sarndal et al, 1992, p.43 for the definition of HT). The solution of the estimating equation given by the SPLS function $\hat{U}(\beta)=0$ yields $\hat{\beta}_N$, the Sample Maximum Partial Likelihood estimator (SMPLE).

Briefly, we determine the properties of $\hat{\beta}_N$ for inference, propose a variance estimator of $\hat{\beta}_N$, and compare it with Lin (2000)'s variance estimator. The motivation for this work stems from the recognition that even though the asymptotic theory developed by Andersen and Gill (1982) for the census case is intended for quite general sequences of probability spaces, their asymptotic results do not apply directly to the Sample Partial Likelihood process. The Sample Partial Likelihood process lives in a joint design-model space, and the design induces some stochastic dependencies that cannot be accommodated by the results for independent random failure times. This claim is illustrated by the following. In the census case, the finite population is assumed to be a realization from M stochastically independent censored lifetimes and the normalized Partial Likelihood Score (PLS) process can be approximated by the following sum of stochastically independent terms with mean zero:

$$\frac{U(\beta,t,\omega)}{\sqrt{M}}=\frac{1}{\sqrt{M}}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\left(X_{hik}-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right)d\mathcal{M}_{hik}(u),\ 0\le t\le\tau,\ \omega\in\Omega,$$

where the $s^{(j)}(\beta,t), j=0,1$ are non-stochastic functions, $d\mathcal{M}_{hik}(u)=\left\{\delta_{hik}dI(\tilde{T}_{hik}\le u)-I(\tilde{T}_{hik}\ge u)e^{\beta' X_{hik}}\lambda_0(u)du\right\}$ and $dI(T_{hik}\le u)$ is the measure that assigns the value $g(T_{hik})$ to a function g. Andersen and Gill (1982) have shown that from this approximation follows the weak convergence of the normalized PLS process.

In the survey case, the normalized SPLS process can also be approximated by a sum of zero mean but not necessarily independent terms:

$$\sqrt{n}\frac{\tilde{U}(\beta,t,\omega,s)}{M}=\sqrt{n}\frac{1}{M}\sum_{hik\in s}\int_0^t\left(X_{hik}-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right)d\mathcal{M}_{hik}(u)/\pi_{hik},\quad 0\le t\le\tau,\quad \omega\in\Omega. \tag{1.2}$$

There are two approaches that are used to look at this process and its asymptotic distribution: one is to view it as a random vector in a design-model "product space". The terms in the sum above are, in general, stochastically dependent in the product space, and this holds whether the design is with replacement or not. In order to calculate its asymptotic distribution in a rigorous way, using the theory of counting processes, we need to define counting processes, martingales, their associated filtrations, etc., in a particular probability space that is the joint design-model space, as well as examine the behavior of the processes associated with the proportional hazards model in that space.

The second approach used by some analysts when considering the asymptotic distribution of the SPLS process is to "ignore the design, omit the sampling weights". What we are actually doing when we analyze complex survey data without sampling weights is working with the conditional probability of the product space given the selected sample. For a simple random sample without replacement (SRSWOR) design, we can ignore the design and work in the realm of the super-population. However, the most common designs (stratified SRSWOR, stratified multi-stage, etc.) usually generate selection bias, and this may cause the original stochastic terms in (1.2) which have zero model mean, to become random variables with mean different from zero, once we condition on the selected sample. Even if the random variables remained independent under the conditional probability given the selected sample, this would not be enough to ensure the convergence of the sum (for a detailed exposition of stochastic dependence in the design-model product space and in its conditional probability spaces, see Rubin-Bleuer and Schiopu-Kratina, 2005, Section 4).

Four previous papers applied the PHM to data from complex surveys: the first paper (Binder, 1992) dealt with inference on the census population parameter β_N; a second paper by (Lin, 2000) dealt with inference on the infinite population (or model) parameter β_0; a third paper on the subject (Rubin-Bleuer, 2003) provided a theoretical justification of an approximation property used by Binder and Lin for their work; and the fourth paper (Rubin-Bleuer, 2004) dealt with the asymptotic distribution of the SPLS process and the corresponding SMPLE function.

Binder (1992) proposed a method of fitting proportional hazards models to survey data from complex designs, based on asymptotic theory in the design probability space. His method provides inference on the "descriptive" census estimator β_N that would be completely known if all the values of the finite population were known. It does not assume a super-population model and it is entirely based on a fixed finite population from which the sample is observed. He assumed that the SPLS function is asymptotically equivalent (in design) to the sample estimator of a total and from this, he derives the asymptotic normality of the SPLS and of the solution $\hat{\beta}_N$ of the SPLS estimating equation (for the asymptotic normality of a sample total see, for example, Krewski and Rao, 1981).

Lin (2000) proposed a method to perform inference on the model parameter β_0. He worked with the super-population approach of Hartley and Sielken (1975) and showed how the sample maximum partial likelihood estimator $\hat{\beta}_N$, proposed by Binder (1992), can provide inference for the model parameter β_0, with a variance that accounts for both the design and the model randomizations. Lin (2000) stated that both, the SPLS function and the sample maximum partial likelihood estimator $\hat{\beta}_N$ are asymptotically normal, provided that certain sample processes were tight. However, he did not provide design nor model conditions under which these sample processes are tight. Lin (2000) also proposed a variance estimator of $\hat{\beta}_N$.

Rubin-Bleuer (2003) used the super-population approach working on a joint design-model space in a formal way, to obtain a rigorous proof of Binder's and Lin's conjecture on the approximation of the SPLS function. To obtain this approximation result, even for Binder's apparently model free approach, there was a need to assume the correct specification of the hazard model underlying the finite population from which the sample is selected. In this paper, Rubin-Bleuer also assumes the stochastic independence of the super-population censored lifetimes and their covariates.

Rubin-Bleuer (2004) showed that the SPLS process converges to a Gaussian process and that the SMPLE is asymptotically normal and gave sufficient sampling design and super-population model conditions under which these results hold. She also gave the expression of the asymptotic variance of both the SPLS process and the SMPLE.

Here we look at the SPLS as a process in the product design-model space, which is the original approach of Lin (2000). The M censored lifetimes are assumed stochastically independent. The design set-up is a stratified two stage design without replacement in the first stage. It may introduce selection bias because there are unequal probabilities of selection, and sampling correlation due to selection without replacement and to the sampling correlation within the *psus*. The selection bias is taken care of with the sampling weights and the sampling correlations disappear due to the convergence of the process to the sample mean and the model independence of the hik units.

In this paper, we work with the normalized SPLS process $\left\{\sqrt{n}\,\frac{\hat{U}(\beta_0,t)}{M},\ 0 \le t \le \tau\right\}$, τ positive, in the style of the original result of Andersen and Gill (1982), rather than with the normalized SPLS function $\sqrt{n}\frac{U(\beta_0)}{M}$. Consequently, the results are more general than those obtained by the previous authors. We develop a counting process methodology for the design-model product space. We obtain a rigorous proof of the weak convergence of the SPLS process in this space, the consistency of the sample maximum partial likelihood estimator $\hat{\beta}_N(t)$ and its asymptotic normality about the model parameter β_0 and give sufficient model and design conditions. We obtain a closed form of the limiting variance of the SPLS process. We prove weak convergence of the SPLS process directly, using neither Binder's (1992) approximation nor Andersen and Gill (1982) results for the PLS. We work from "scratch" using the Central Limit Theorem for Martingales and apply it to the design-model product space. These outcomes require the correct specification of both the hazards model and correlation structure of the super-population. The above results, developed for the case where the first stage sampling rate $\frac{n}{N} \to f$, with f positive, are contained in the JSM proceedings paper by Rubin-Bleuer (2004). Here we extend them to the case $f \geq 0$, and give the proofs in detail. In addition, we propose variance estimators $\hat{\Sigma}_\pi$, $\hat{\Sigma}_m^{-1}$ and $\hat{\Sigma}_m^{-1}\hat{\Sigma}_\pi\hat{\Sigma}_m^{-1}$ of the SPLS, the PLS and $\hat{\beta}_N(t)$ respectively, and prove that they are design-model consistent (see Sarndal et al, 1992, for the definition of design-consistency and other sampling theory concepts). Furthermore, using the tools of counting process methodology, we investigate the properties of Lin's (2000) variance estimator, which is based on the approximation of the SPLS function to the sample HT estimator of a total of the finite population. For this purpose, we give a new proof for obtaining Binder's (1992) approximation to the SPLS function, which is valid under the semi-parametric model for the hazard function, even if the correlation structure of the super-population is not correctly specified. Furthermore, we show that if the underlying super-population lifetimes and covariates are stochastically independent, Lin's (2000) variance estimator of the normalized SPLS process is design-model consistent though less efficient than $\hat{\Sigma}_\pi$. On the other hand, consider SPLS with our normalizing constant rather than that used by Lin's (2000) and assume $\frac{n}{N} \to 0$, then we can argue that Lin's (2000) variance estimator is a robust variance estimator against misspecification of the correlation model.

Before going into the organization of the paper, we provide some comments about the assumptions and techniques used here. The proofs follow similar techniques used for the census case by Andersen and Gill (1982), modified for simple situations by Fleming and Harrington (1991). But there are some considerations that we must make and this is why the proofs here are carefully rendered. For example, some of the processes that are bounded almost surely in the census case, are now π-weighted processes which are only bounded in probability. Also, the limit of the information matrix does not coincide with the limiting covariance matrix of the SPLS process: new finite population "π- weighted" S processes turn up in the calculation of this variance. In addition, design conditions have to be stated for the design consistency of the HT sample estimators of the processes involved. Finally, we also use the property that convergence in design probability does imply convergence in the design-model space in which we work. The proof of this statement is not trivial, and we refer to previous work on the subject (Rubin-Bleuer, 2005, Theorem 5.1).

In the census case, there exist results for more general model assumptions than those we present here. This article was done under somewhat restricted conditions (i.e., continuous failure times, covariates constant over time and uniformly bounded, a common baseline hazard function and conditionally independent failure and censoring times) to concentrate on the added complexity of the survey-model process. Many practical situations occurring with survey data fall under these conditions and some of the restrictions arise from the practical issues in the collection of the data. Survey data is, in general, obtained from a stratified multistage design with a large number of strata and a small number of clusters per stratum, usually two or three. In each stratum, the first stage of sampling selects the clusters with probability proportional to size. Accordingly, for the asymptotic set-up, the number of strata is assumed to increase towards infinity, while the sample size stays bounded within each stratum. Also, in many situations the design clusters are specified by operational reasons and not by an assumption that the underlying super-population is clustered (for example employment in the Canadian Labor Force Survey may depend on economic regions (the strata) but analysts have in general, assumed that employment status is independent within each stratum regardless of the designed clusters). The results presented here assume that the super-population of lifetimes and censoring times are stochastically independent. As a caveat, it is noted that data obtained from a longitudinal survey could be subject to a significant amount of censoring due to attrition, and that attrition is sometimes not independent of failure times. This situation is not treated in this paper. Finally, we assume that the vector of covariates X does not depend on time. As a result, simpler model conditions are used. At this moment there is also a practical reason for working with covariates that are constant in time: there is a hitch with its application for survey data. The variance estimators proposed until now (see for example Binder, 1992 and Lin, 2000) contain terms in $X_j(\widetilde{T}_i)$, for subjects i and j respectively. These cannot not be observed if, for example, subject j died before subject i.

The paper is organized as follows. In Section 2, we give the notation used throughout the paper, and define a stratified super-population model and design. In Section 3, we formally express the joint design-model space envisaged by Lin (2000) as a "product space" containing both the model and the design probability spaces. Also, we state the counting process methodology developed by Rubin Bleuer (2001) and Rubin Bleuer (2003 b) for the design-model space, which is used to derive the results in this paper. The S and $\hat{S}$ processes are functions of the number of units at risk at time t. Tightness of many survival processes often follows from the convergence in sup norm of the S and $\hat{S}$ processes. In Section 4, we prove the uniform convergence of the $S^{(j)}(\beta,t)$, and the π-weighted $S^{(j)}(\beta,t)$, $j=0,\ 1,\ 2,$ finite population functions, their sample estimators and certain combinations of them. We also give the new proof of Binder's approximation. The weak convergence of the SPLS, the consistency of the maximum partial likelihood estimator, the calculation of the limiting variance of the maximum sample partial likelihood estimator and of the limiting sample information matrix, the consistency of their respective sample estimators and the asymptotic normality of the maximum sample partial likelihood estimator are given in Section 5. In Section 6, we investigate the properties of Lin (2000)'s variance estimator and compare it with the variance estimator proposed in this paper. Finally, in Section 7, we summarize the results and present conclusions.

2. THE MODEL AND THE DESIGN

2.1 The model

Consider right censored lifetimes defined on a probability space $(\Omega, \Im, P)$. The lifetimes are classified into L strata, and for the sake of consistency of notation with the design, we assign the censored lifetimes and respective covariates into N_h primary sampling units ($psus$) within stratum $h, h=1,...,L,$ and N_{hi} secondary sampling units within psu i, $i=1,...,N_h$. The data values will be labeled by the number of $psus$ in the finite population. The model is defined by triples:

$$\left(\tilde{T}_{hik}^N, \delta_{hik}^N, X_{hik}^N\right) \quad k=1,...,N_{hi}, \ \ i=1,...,N_h, \ \ h=1,...,L, \tag{2.1}$$

such that

a) X_{hik}^N are r-dimensional covariates constant over time, $k=1,...,N_{hi}$, $i=1,...,N_h$, $h=1,...,L$,

b) $\tilde{T}_{hik}^N = T_{hik}^N \wedge C_{hik}^N$ are censored failure times, where failure time and censoring time variables T_{hik}^N and C_{hik}^N are assumed conditionally independent given X_{hik}^N, and $k=1,...,N_{hi}$, $i=1,...,N_h$, $h=1,...,L$,

c) $\delta_{hik}^N = I\left(\tilde{T}_{hik}^N = T_{hik}^N\right)$ are indicators of whether a failure time is actually being observed or not, $k=1,...,N_{hi}$, $i=1,...,N_h$, $h=1,...,L$.

d) the triples $\left(\tilde{T}_{hik}^N, \delta_{hik}^N, X_{hik}^N\right)$ $k=1,...,N_{hi}$, $i=1,...,N_h$, $h=1,...,L$, are stochastically independent.

The Cox (1972) proportional hazards model specifies that the hazard rate $\lambda_{hik}^N(t)$ (or instantaneous failure rate) of the failure time T_{hik}^N satisfies

$$\lambda_{hik}^N(t) = \lambda_0(t) \cdot \exp(\beta_0' \cdot X_{hik}^N), k=1,...,N_{hi}, \ \ i=1,...,N_h, \ \ h=1,...,L, \tag{2.2}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, with absolutely continuous survival function $S_0(t) = 1 - F_0(t)$ and β_0 is an r-dimensional vector valued regression parameter. In the asymptotic set-up the number L of strata increases towards infinity while the number of clusters N_h within strata remains bounded. Thus, even though the failure times are not necessarily identically distributed, they share the same baseline hazard function. The differences among hazard rates are to be taken into account by the conditional distributions given the covariates. E_m and V_m denote, respectively, the expectation and variance in the space $(\Omega, \Im, P)$.

From now on, whenever there is no room for confusion, we omit the superscript N. In the following, we use standard notation for counting processes and state their properties under the proportional hazards model given above (see Fleming and Harrington, 1991).

Let $\eta(t) = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \eta_{hik}(t)$, $\eta_{hik}(t) = I(T_{hik} \le t)\delta_{hik}$ denote a counting process, which is the number of failed uncensored observations by time t. We use the notation $\eta(t)$ for a counting process, and $\mathcal{M}(t)$ for a martingale, instead of the usual $N(t)$ and $M(t)$ respectively, because in this study N denotes the number of clusters (or primary sampling units) in the finite population, and M denotes the number of ultimate units in the population.

Let the number of units at risk at time t be given by $Y(t)=\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} Y_{hik}(t),\ \ Y_{hik}(t)=I(\widetilde{T}_{hik}\geq t)$.

The symbol $\otimes 2$ denotes the outer product of the vector within brackets (i.e. $X^{\otimes 2}=X\cdot X'$). We write $X^{\otimes 1}=X$ and $X^{\otimes 0}=1$. Let $S^{(j)},\ \ j=0,\ 1,\ 2$ be respectively, a scalar, an r-dimensional vector and an $r\times r$ dimensional matrix defined by:

$$S^{(0)}(\beta,t)=\frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}}=\frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} X_{hik}^{\otimes 0}\cdot Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}}$$

$$S^{(1)}(\beta,t)=\frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} X_{hik}\cdot Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}}=\frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} X_{hik}^{\otimes 1}\cdot Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}},$$

and

$$S^{(2)}(\beta,t)=\frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} X_{hik}\cdot X'_{hik}\cdot Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}}=\frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} X_{hik}^{\otimes 2}\cdot Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}}. \tag{2.3}$$

Here we deviate from the usual model conditions and introduce "π-weighted" $S_\pi^{(j)},\ \ j=0,\ 1,\ 2$ that are respectively, a scalar, an r-dimensional vector and an $r\times r$ dimensional matrix process defined by:

$$S_\pi^{(j)}(\beta,t)=\frac{1}{M}\sum_{k=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}} \frac{1}{\pi_{hik}} X_{hik}^{\otimes j}\cdot Y_{hik}(t)\cdot e^{\beta'\cdot X_{hik}},\quad j=0,\ 1,\ 2. \tag{2.4}$$

Also let

$$e(\beta,t)=\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)},\ V(\beta,t)=\frac{S^{(2)}(\beta,t)S^{(0)}(\beta,t)-S^{(1)}\cdot S^{(1)'}(\beta,t)}{(S^{(0)}(\beta,t))^2}\quad \text{and}$$

$$V_\pi(\beta,t)=\frac{S_\pi^{(2)}(\beta,t)}{S^{(0)}(\beta,t)}-S_\pi^{(1)}(\beta,t)\frac{S^{(1)'}(\beta,t)}{\left(S^{(0)}(\beta,t)\right)^2}-\frac{S^{(1)}(\beta,t)}{\left(S^{(0)}(\beta,t)\right)^2}S_\pi^{(1)'}(\beta,t)+\frac{\left(S^{(1)}(\beta,t)\right)^{\otimes 2}}{\left(S^{(0)}(\beta,t)\right)^2}\frac{S_\pi^{(0)}(\beta,t)}{S^{(0)}(\beta,t)} \tag{2.5}$$

Let $\Im_t$ be the sigma field defined by the failure and censoring indicators, that is,

$$\Im_t=\sigma(\eta_{hik}(u),\ \ \eta_{hik}^C(u),\ k=1,\dots,N_{hi},\ \ i=1,\dots,N_h,h=1,\dots,L,\ \ 0\leq u\leq t).$$

Under the proportional hazards model (2.2), the process $\mathcal{M}_{hik}(t)=\mathcal{M}_{hik}(\beta_0,t)=\eta_{hik}(t)-\int_0^t Y_{hik}(u)\exp\{\beta_0' X_{hik}\}\cdot\lambda_0(u)du$ is a martingale with respect to the filtration $\{\Im_t : t\geq 0\}$ and has absolutely continuous compensator

$$A_{hik}=\int Y_{hik}(u)e^{\beta_0' X_{hik}}\lambda_0(u)du,\quad k=1,\ldots,N_{hi},\quad i=1,\ldots,N_h,\quad h=1,\ldots,L.$$

The continuity of the A_{hik} follows from the absolute continuity of the failure time distributions. Let $\langle\mathcal{M}_1,\mathcal{M}_2\rangle$ denote the predictable co-variation of the martingales $\mathcal{M}_1$ and $\mathcal{M}_2$. Then $\langle\mathcal{M}_{hik}(t),\mathcal{M}_{hik}(t)\rangle=\int_0^t Y_{hik}(u)\exp\{\beta' X_{hik}\}\cdot\lambda_0(u)du$. The partial likelihood score (PLS) process can be expressed by

$$U(\beta,t)=\sum_{hik}\int_0^t\{X_{hik}-e(\beta,u)\}d\eta_{hik}(u)=\sum_{hik}\int_0^t\{X_{hik}-e(\beta,u)\}d\mathcal{M}_{hik}(\beta_0,u),$$

and it is also a martingale, since it is the sum of stochastic integrals of predictable processes with respect to a martingale. If the $\mathcal{M}_{hik}(\beta_0,t)$ are independent. the predictable co-variation process is given by:

$$\langle U(\beta,t),U(\beta,t)\rangle=\sum_{hik}\int_0^t(X_{hik}-e(\beta,t))^{\otimes 2}Y_{hik}(u)\exp\{\beta_0' X_{hik}\}\cdot\lambda_0(u)du,\tag{2.6}$$

(see Theorem 2.4.3, p.40 and Theorem 2.5.2, statement 1, p.75 in Fleming and Harrington, 1991).

An apparent limitation in the theory is the requirement that the data used must be restricted to an interval $0\leq t\leq\tau$ (see Fleming and Harrington, 1991, p.307). This is to ensure that for the development of the asymptotic properties we have both that the function $s^{(0)}(\beta,t)$ is bounded away from zero and that the integral of the baseline hazard function has a finite value in such interval. Both requirements would hold if 1) the censoring variables C_i are defined in $0\leq t<\tau$ and have densities of the form $g_i(t)I(t<\tau)+(1-P(C_i<\tau\mid X_i)\delta_\tau(t)$ in $0\leq t<\infty$ where $g_i(t)$ are functions with finite integrals and $\delta_\tau(t)$ is the delta measure with mass point at $t=\tau$; and 2) the baseline lifetime distribution $F(t)=P(T\leq t)$ has support in an interval larger than $0\leq t\leq\tau$ (i.e., if $0<F(\tau)<1$). In most studies with survey data, there is a pre-fixed time C_0 when the study ends, and censoring variables with densities of the form outlined above have a positive probability of being realized at time $t=C_0$. In this case, all observed failures would fall within the interval $0\leq t\leq C_0$ and all the data could be utilized to build the Sample Partial Likelihood Score function evaluated at $\tau=C_0$. We could then say that $U(\beta)=U(\beta,\tau)=U(\beta,\infty)$.

2.2 The design

Consider a general stratified, without replacement, two-stage design on a finite population obtained from independent failure and censoring times and independent covariates. For an outcome $\omega\in\Omega$ of the super-population, the finite population is represented by $\left(\tilde{T}_{hik}^N(\omega),\delta_{hik}^N(\omega),X_{hik}^N(\omega)\right)$ $k=1,\ldots,N_{hi}$, $i=1,\ldots,N_h$, $h=1,\ldots,L$. Let π_{hik}^N denote the probability that the unit hik is selected in the sample. For simplicity we omit the superscript N in the notation of the inclusion probabilities. The number of primary sampling units in the sample is n and the number of ultimate sample units is $m = \sum_{h=1}^{L} \sum_{i=1}^{n_h} n_{hi}$ with $n = \sum_{h=1}^{L} n_h$. The sample selection indicators are defined by:

$$I_{hik}(s) = 1 \text{ if } hik \in s,\ I_{hik}(s) = 0 \text{ otherwise},\ k = 1,\ldots, N_{hi},$$
$$i = 1,\ldots, N_h,\ h = 1,\ldots, L\,. \quad (2.7)$$

We denote by S_N the collection of all possible samples under the sample scheme, by $C(S_N)$ the collection of subsets of S_N, and by p_{dN} a sampling probability distribution defined on $C(S_N)$. Then the design space is given by the triple $(S_N,\ C(S_N),\ p_{dN})$. In the following E_d and V_d denote, respectively, the expectation and variance with respect to the sampling design.

Remark 2.1 Traditional notation under the proportional hazards model use "S" for the random functions and their deterministic limits: $S^{(j)}(\beta,t,\omega) \to s^{(j)}(\beta,t)$ $j = 0,\ 1,\ 2$. For $\omega \in \Omega$, the respective scalar, vector and matrix functions $S^{(j)}(\beta,t)$, $j = 0,\ 1,\ 2$, are finite population parameters. Their sample estimators depend on the selected sample $s \in S_N$: $\hat{S}^{(j)}(\beta,t,s,\omega)$ $j = 0,\ 1,\ 2$. The use of "s" to denote an outcome of the sample design is also a well known convention in survey theory, and we will do so here with the caveat that the sample s should not be confused with the deterministic limit functions $s^{(j)}(\beta,t)$ $j = 0,\ 1,\ 2$. Design-unbiased estimators of the S-functions (Särndal et al, 1992), p.167) are given by their Horvitz-Thompson (HT) sample estimators:

$$\hat{S}^{(j)}(\beta,t) = \frac{1}{M} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \frac{I_{hik}(s)}{\pi_{hik}} X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' \cdot X_{hik}}, \quad j = 0,\ 1,\ 2 \quad (2.8)$$

and

$$\hat{S}_{\pi}^{(j)}(\beta,t) = \frac{1}{M} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \frac{I_{hi}(s)}{\pi_{hik}^{2}} X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' \cdot X_{hik}}, \quad j = 0,1,2. \quad (2.9)$$

Also let

$$\hat{e}(\beta,t) = \hat{S}^{(1)}(\beta,t) / \hat{S}^{(0)}(\beta,t) \text{ and let } \hat{V}(\beta,t) = \frac{\hat{S}^{(2)}(\beta,t)\hat{S}^{(0)}(\beta,t) - \hat{S}^{(1)}(\beta,t) \cdot \hat{S}^{(1)'}(\beta,t)}{(\hat{S}^{(0)}(\beta,t))^2},$$

and

(2.10)

$$\hat{V}_\pi(\beta,t)=\frac{\hat{S}_\pi^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}-\hat{S}_\pi^{(1)}(\beta,t)\frac{\hat{S}^{(1)'}(\beta,t)}{\left(\hat{S}^{(0)}(\beta,t)\right)^2}-\frac{\hat{S}^{(1)}(\beta,t)}{\left(\hat{S}^{(0)}(\beta,t)\right)^2}\hat{S}_\pi^{(1)'}(\beta,t)+\frac{\left(\hat{S}^{(1)}(\beta,t)\right)^2}{\left(\hat{S}^{(0)}(\beta,t)\right)^2}\frac{\hat{S}_\pi^{(0)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}$$

The HT estimators of the $S^{(j)}(\beta,t)$ and the $\frac{n}{N}S_\pi{}^{(j)}(\beta,t)$, $j=0,\ 1,\ 2$ functions are also design consistent under Conditions C_1, C_2 and C_3 stated at the end of this section. Under the same conditions, it is easy to show that $\hat{e}$, $\hat{V}$ and $\frac{n}{N}\hat{V}_\pi$ are design-consistent estimators of e, V and $\frac{n}{N}V_\pi$ respectively, as $n\to\infty$.

In what follows the $\hat{S}^{(j)}(\beta,t)$ and $\frac{n}{N}\hat{S}_\pi^{(j)}(\beta,t)$, $j=0,\ 1,\ 2$ can be replaced by any design-consistent estimators of the corresponding finite population processes. The sample partial likelihood score vector is defined by the sum of stochastic integrals:

$$\hat{U}(\beta,t)=\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\hat{\psi}_{hik}(\beta,u)d\eta_{hik}(u) \text{ with} \tag{2.11}$$

$$\hat{\psi}_{hik}(\beta,u)=\frac{I_{hik}(s)}{\pi_{hik}}(X_{hik}-\hat{e}(\beta,u)),\quad k=1,\ldots,N_{hi},\quad i=1,\ldots,N_h,\quad h=1,\ldots,L. \tag{2.12}$$

For a fixed sample, the process $\hat{U}(\beta,t)$ has also a martingale representation under the model given by (2.2):

$$\hat{U}(\beta,t)=\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\hat{\psi}_{hik}(\beta,u)d\mathcal{M}_{hik}(\beta_0,u), \tag{2.13}$$

which follows from subtracting from $\hat{U}(\beta,t)$ the zero expression

$$\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\hat{\psi}_{hik}(\beta,u)Y_{hik}(u)\cdot e^{\beta_0' X_{hik}}\lambda_0(u)du\,.$$

Throughout the paper we will assume subsets of the following regularity conditions on the sampling design.

In the following $\sum\limits_{hik\neq\ell jm\in s}$ denotes the 6-way sum of units selected to the sample and $\sum\limits_{hik\neq\ell jm\in Pop}$ denotes the 6-way sum of units in the finite population, that is, $k=1,\ldots,N_{hi},\quad k=1,\ldots,N_{\ell j},\quad i=1,\ldots,N_h,\quad j=1,\ldots,N_\ell,$
$h,\ell=1,\ldots,\quad L.$

$C_0: \ f = \lim_n n/N \geq 0$ as $n \to \infty$.

$C_1: \ \max_{hik} \dfrac{1}{\pi_{hik}} = O\left(\dfrac{N}{n}\right)$ as $n \to \infty$.

$C_2: \ \dfrac{M}{N} \to \mu$ as $\mathrm{N} \to \infty$.

$$C_3: \frac{1}{M} \sum_{hik \neq \ell jm \in Pop} \frac{\pi_{hik\ell jm} - \pi_{hik}\pi_{\ell jm}}{\pi_{hik}\pi_{\ell jm}} = O\left(\frac{N}{n}\right) \text{as } n \to \infty.$$

Remark 2.2 If we impose $f > 0$ in Condition C_0 we ensure that the relationship between the sample and population number of *psus* remains the same as we increase the number of *psus* in the population towards infinity. For the asymptotic properties shown in this paper we assume $f \geq 0$. Condition C_1 means that as $n \to \infty$ the selection probabilities are approximately of the same magnitude. If the first stage selection probabilities are proportional to the size of the *psus* in the stratum, and the second stage selection probabilities are SRSWOR, then C_1 means that no *psu* is of disproportionate size. Condition C_2 implies that the number of units in the *psus* in the finite population remain bounded as the number of clusters in the finite population increase ($N \to \infty$). Condition C_1 holds for SRSWOR and for a design with size measures Z_i and probabilities $\pi_i \approx nZ_i / \sum_j Z_j$ with $\overline{Z} = \sum_{j=1}^{N} Z_j / N \to z$ as $N \to \infty$ and C_3 holds for SRSWOR. C_1 and C_3 together are sufficient for obtaining design-consistency of HT sample estimators in general.

3. COUNTING PROCESS THEORY IN THE PRODUCT SPACE

In this section we show that the SPLS process is also a martingale with respect to a filtration where the sample
varies at random as well. The product space determined by the proportional hazards model given in Section 2 and one stage sampling designs is given by $(\Omega \times S_N, \ \Im \times C(S_N), \ P_{d,m})$ with probability measure defined in the elementary rectangles by:

$$P_{d,m}(s \times F) = \int_F p_{dN}(s,\omega) dP(\omega), \quad s \in C(S_N), \quad F \in \Im.$$

If the sampling design is two stage, and the model probability P_Z is the conditional probability given the design prior information $\{Z(\omega) = z\}$, then the product space determined by the model and the two stage sampling design is $(\Omega \times S_N, \ \Im \times C(S_N), \ P_{d,m})$ where probability measure is defined in the elementary rectangles by:

$$P_{d,m}(s \times F) = p_{dN}(s) \cdot P_Z(F), \quad s \in C(S_N), \quad F \in \Im.$$

See Rubin-Bleuer and Schiopu-Kratina (2005), Example 4.2 for a description of the product space where the model probability is conditional to the prior information. Next we define the tools of counting process theory in the product space.

3.1 Filtrations

Martingale and counting process theory is developed on a stochastic basis, that is, a probability space with a filtration. A filtration is an increasing family of right continuous sub-σ-algebras. With the new indexing, the filtration corresponding to the counting process $\eta\ (t)$ in the proportional hazards model is given by:

$$\Im_t = \sigma(\eta_{hik}(u),\ \eta_{hik}^C(u), 1 \le k \le N_{hi},\ 1 \le i \le N_h,\ 1 \le h \le L,\ 0 \le u \le t).$$

We define here a stochastic basis for the product space. The family of sub-σ-algebras defined by $\{\Im_t^{d,m} = C(S_N) \times \Im_t,\ t \ge 0\}$ is a filtration; it is increasing and right continuous because $\{\Im_t,\ t \ge 0\}$ is so.

3.2 Sample Counting Processes and Martingales

We define the sample counting process $\hat{\eta}(t,s,\omega) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \eta_{hik}(t,\omega) \cdot I_{hik}(s)/\pi_{hik}$. $\hat{\eta}$ is a counting process in the product space with respect to the filtration $\left\{\Im_t^{d,m} : t \ge 0\right\}$, since each term is the product of a counting process with respect to the original filtration $\left\{\Im_t : t \ge 0\right\}$ in $(\Omega, \Im, P)$ and the factor I_{hik} which is $C(S_N)$-measurable. Let $\mathcal{M}_{hik}(t) = \mathcal{M}_{hik}(\beta,t),\ k = 1,\ldots,N_{hi},\ i = 1,\ldots,N_h,\ h = 1,\ldots,L$ be the martingales associated with the η_{hik} in $(\Omega, \Im, P)$ with respect to $\{\Im_t : t \ge 0\}$. They are also product space martingales with stochastic basis

$$(\Omega_{d,m},\ C(S) \times \Im,\ \Im_t^{d,m},\ P_{d,m}).$$

Then the martingale sample process defined by:

$$\hat{\mathcal{M}}(\beta,t,s,\omega) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \mathcal{M}_{hik}(\beta,t,\omega) \cdot I_{hik}(s)/\pi_{hik} \tag{3.1}$$

is a martingale in the product space. We can easily verify the three necessary conditions (see Definition 1.2.8, Fleming and Harrington, 1991, p. 22) Each term of $\hat{\mathcal{M}}$ is a martingale since for all $hik,\ k = 1,\ldots,N_{hi},\ i = 1,\ldots,N_h,\ h = 1,\ldots,L,$ we have:

1) $\mathcal{M}_{hik}(t,\omega) I_{hik}(s)/\pi_{hik}$ is adapted to $\{\Im_t^{d,m} : t \ge 0\}$,
2) $E_{d,m}(|\ \mathcal{M}_{hik}(t,\omega) \cdot I_{hik}(s)/\pi_{hik}\ |) < \infty$ for all $t < \infty$, and
3) $E_{d,m}(\mathcal{M}_{hik}(t+u) \cdot I_{hik}(s)/\pi_{hik} \mid \Im_t^{d,m}) = (I_{hik}(s)/\pi_{hik}) E_{d,m}(\mathcal{M}_{hik}(t+u) \mid \Im_t)$
$= (I_{hik}(s)/\pi_{hik}) \mathcal{M}_{hik}(t)$ for all $t \ge 0,\ u \ge 0$.

Thus $\hat{\mathcal{M}}(t)$, the product-space martingale $\mathcal{M}_{hik}(t,\omega) I_{hik}(s)/\pi_{hik}$ and the model martingales $\mathcal{M}_{hik}(t)$ are martingales with respect to the filtration $\left\{\Im_t^{d,m} : t \ge 0\right\}$ in the product space. Now, if each martingale $\mathcal{M}_{hik}(t)$ $k=1,\ldots,N_{hi}$, $i=1,\ldots,N_h$, $h=1,\ldots,L$ has absolute continuous compensator

$\int_0^t Y_{hik}(u)\exp\{\beta_0' X_{hik}\}\lambda_0(u)du$, then the sample process $\hat{\mathcal{M}}(t)=\hat{\eta}(t)-\hat{A}(t)$ is a martingale with

$$\hat{A}(\beta,t,s,\omega)=M\cdot\int_0^t \hat{S}^{(0)}(\beta,u)\lambda_0(u)du. \tag{3.2}$$

Since $E_{d,m}(\hat{A}(t))<\infty$, $\hat{A}(0)=0$, then by the uniqueness portion of the Doob-Meyer Decomposition Theorem (see Fleming and Harrington, 1991, p.37), $\hat{A}(t)$ is the absolutely continuous compensator of the product space martingale $\hat{\mathcal{M}}(t)$.

3.3 Predictable co-variation processes

In order to complete the tool bag for counting process in the product space, we need to ensure that the counting processes yield martingales such that for any two of them their predictable co-variation processes are zero. This property is crucial for calculating co-variances of the stochastic integrals under consideration, and for applying the Central Limit Theorem for Martingales.

We say that $\{\eta_{hik}(t),\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$ are multivariate counting processes, if not two of them jump at the same time. If the sequence of processes above are multivariate counting processes, or if the processes $\{\Delta\eta_{hik}(t),\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$ are independent random variables, we have that the following predictable co-variation processes are zero:

$$\langle \mathcal{M}_{hik},\mathcal{M}_{ljm}\rangle(t)=0, \tag{3.3}$$

$h\neq l$ or $i\neq j$, or $k\neq m$, $k,m=1,\ldots,N_{hi}$, $i,j=1,\ldots,N_h$, $h,l=1,\ldots,L$, $t\geq 0$ (see Fleming and Harrington (1992) Lemma 2.6.1 p. 81). If the $\{\eta_{hik}(t),\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$ are multivariate counting processes in the super-population, then the sampling processes $\{\eta_{hik}(t)I_{hik}(s)/\pi_{hik},\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$ are also multivariate counting processes in the product space. Independent failure times in the super-population, with continuous distribution functions yield multivariate counting processes. However, if failure times are not continuous, and the $\{\eta_{hik}(t)\}$ are not multivariate counting processes, independent differences

$$\{\Delta\eta_{hik}(t),\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$$

do not yield independent differences in the product space of the form

$$\{\Delta\eta_{hik}(t)I_{hik}(s)/\pi_{hik},\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$$

Instead we have:

Lemma 3.1 If the counting processes $\{\eta_{hik}(t),\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L\}$ are stochastically independent in the super-population (model) space, and the corresponding sample martingales are $\{\mathcal{M}_{hik}(t) I_{hik}(s) / \pi_{hik},\ k = 1,...,N_{hi},\ i = 1,...,N_h,\ h = 1,...,L\}$ then the predictable co-variation sample-processes satisfy (3.3).

Proof: We follow the argument of Lemma 2.6.1 in Fleming and Harrington (1991), taking into consideration that calculations are with respect to the product space filtration $\{\Im_t^{d,m} = C(S_N) \times \Im_t,\ t \ge 0\}$.

4. A REPRESENTATION OF THE SPLS

The sup-norm convergence in probability of a sequence of bounded monotonic random functions has been previously used by many authors and we state it below in Lemma 4.1, for a sequence of probability spaces so we can use it for problems posed in the joint design-model space. One consequence to Lemma 4.1 for example, is the sup-norm convergence of the sample empirical distribution functions, (Rubin Bleuer, 2003), which in turn yields tightness of the sample weighted log-rank statistics (Rubin Bleuer, 2001). Of interest to this article is another consequence of the lemma, Theorem 4.1, the sup-norm convergence of the risk functions involved in the sample partial likelihood score under the proportional hazards model.

Lemma 4.1 Let $\{G_N(t) : -\infty < t < \infty\}$ be a sequence of random functions defined on probability spaces $(\Omega_N, \Im_N, P_N)$ with sample paths that are monotonic bounded functions. In addition, they are right-continuous if non-decreasing, and left-continuous if non-increasing. Let $\{g(t) : -\infty < t < \infty\}$ be the non-stochastic bounded monotonic limit in probability of $\{G_N(t) : -\infty < t < \infty\}$, such that $G_N(t) - g(t) \to 0\ in\ P_N$ for all t and either $G_N(t-) - g(t-) \to 0\ in\ P_N$ for all t if non-decreasing, or $G_N(t+) - g(t+) \to 0\ in\ P_N$ for all t if non-increasing. Then $\sup_t | G_N(t) - g(t) | \to 0\ in\ P_N$.

Proof:
We use the same argument in the proof of the Glivenko-Cantelli theorem for the census case but applied to convergence in the product space (see the uniform convergence of the sample empirical distribution function in Rubin-Bleuer (2003 a), also the argument used for the S functions in Fleming and Harrington, 1991, p.305).

Theorem 4.1 We assume the proportional hazards model given in (2.1) a), b) and c) and (2.2), the design conditions C_0, C_1, C_2, C_3 and the following model conditions:

M_1: The covariate vectors X_{hik} are constant (in time) and bounded: $\sup_{h,i,k} \left| X_{hik} \right| \le B\ a.s.\ as\ N \to \infty$.

M_2: There exists a neighborhood Λ of β_0 and, respectively, scalar, vector and matrix functions $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ defined on $\Lambda \times [0, \tau]$ such that for $j = 0,1,2,$ and for $0 \le t \le \tau$, $\beta \in \Lambda$ we have:

i) $s^{(j)}(\beta,t) = \lim\limits_{N\to\infty} \frac{1}{M} \sum\limits_{h=1}^{L} \sum\limits_{i=1}^{N_h} \sum\limits_{k=1}^{N_{hi}} E_m \left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' X_{hik}} \right\}.$

ii) $s^{(j)}(\beta,t+) = \lim\limits_{N\to\infty} \frac{1}{M} \sum\limits_{h=1}^{L} \sum\limits_{i=1}^{N_h} \sum\limits_{k=1}^{N_{hi}} E_m \left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t+) \cdot e^{\beta' X_{hik}} \right\}.$

M_3: There exists a neighborhood Λ of β_0 and, respectively, scalar, vector and matrix functions $s_\pi^{(0)}$, $s_\pi^{(1)}$ and $s_\pi^{(2)}$ defined on $\Lambda \times [0,\tau]$ such that for $j = 0,1,2,$ and for $0 \le t \le \tau$, $\beta \in \Lambda$ we have:

i) $s_\pi^{(j)}(\beta,t) = \lim\limits_{N\to\infty} \frac{1}{M} \sum\limits_{h=1}^{L} \sum\limits_{i=1}^{N_h} \sum\limits_{k=1}^{N_{hi}} \frac{n}{N} E_m \left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' X_{hik}} \right\} / \pi_{hik}.$

ii) $s_\pi^{(j)}(\beta,t+) = \lim\limits_{N\to\infty} \frac{1}{M} \sum\limits_{h=1}^{L} \sum\limits_{i=1}^{N_h} \sum\limits_{k=1}^{N_{hi}} \frac{n}{N} E_m \left\{ X_{hik}^{\otimes j} \cdot Y_{hik}(t+) \cdot e^{\beta' X_{hik}} \right\} / \pi_{hik}.$

M_4: The time τ is such that $\int_0^\tau \lambda_0(t)dt < \infty$.

M_5: $\liminf\limits_{N\to\infty} \frac{1}{N} \sum\limits_h \sum\limits_i \sum\limits_k P(C_{hik} \ge \tau) \rangle 0.$

We also define the matrix functions v and v_π on $\Lambda \times [0,\tau]$ by

$$v(\beta,t) = \frac{s^{(2)}(\beta,t) s^{(0)}(\beta,t) - s^{(1)}(\beta,t) \cdot s^{(1)'}(\beta,t)}{(s^{(0)}(\beta,t))^2} \quad \text{and}$$

$$v_\pi(\beta,t) = \frac{s_\pi^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} - s_\pi^{(1)}(\beta,t) \frac{s^{(1)'}(\beta,t)}{\left(s^{(0)}(\beta,t)\right)^2} - \frac{s^{(1)}(\beta,t)}{\left(s^{(0)}(\beta,t)\right)^2} s_\pi^{(1)'}(\beta,t) + \frac{\left(s^{(1)}(\beta,t)\right)^2}{\left(s^{(0)}(\beta,t)\right)^2} \frac{s_\pi^{(0)}(\beta,t)}{s^{(0)}(\beta,t)}. \tag{4.1}$$

Then we have:

$$S^{(j)}(\beta,t) \xrightarrow{P} s^{(j)}(\beta,t),\; S^{(j)}(\beta,t+) \xrightarrow{P} s^{(j)}(\beta,t+),\; \frac{n}{N} S_\pi^{(j)}(\beta,t) \xrightarrow{P} s_\pi^{(j)}(\beta,t),$$
$$\frac{n}{N} S_\pi^{(j)}(\beta,t+) \xrightarrow{P} s_\pi^{(j)}(\beta,t+) \tag{4.2}$$

as $N \to \infty$, $j = 0,\ 1,\ 2,$ for each $t \ge 0$.

$$\sup_{0\le t\le\tau,\ \beta\in\Lambda} | S^j(\beta,t) - s^j(\beta,t) | \xrightarrow{P} 0 \quad j=0,\ 1,\ 2 \quad as\ N\to\infty, \tag{4.3}$$

$$\sup_{0\le t\le\tau,\ \beta\in\Lambda} | \hat{S}^j(\beta,t) - S^j(\beta,t) | \xrightarrow{P_{d,m}} 0 \quad j=0,\ 1,\ 2 \quad as\ n\to\infty, \tag{4.4}$$

$$\sup_{0\le t\le\tau,\beta\in\Lambda} \left| \frac{n}{N} S_\pi^{(j)}(\beta,t) - s_\pi^{(j)}(\beta,t) \right| \xrightarrow{P} 0, \quad j=0,\ 1,\ 2 \quad as\ N\to\infty, \tag{4.5}$$

$$\sup_{0\le t\le\tau,\beta\in\Lambda} \left| \frac{n}{N} \hat{S}_\pi^{(j)}(\beta,t) - s_\pi^{(j)}(\beta,t) \right| \xrightarrow{P_{d,m}} 0, \quad j=0,\ 1,\ 2, \quad as\ n\to\infty, \tag{4.6}$$

$$\frac{\partial}{\partial\beta} s^{(0)}(\beta,t) = s^{(1)}(\beta,t) \text{ and } \frac{\partial^2}{\partial\beta^2} s^{(0)}(\beta,t) = s^{(2)}(\beta,t) \text{ for } 0\le t\le\tau,\quad \beta\in\Lambda, \tag{4.7}$$

$s^{(0)}(\beta,t) \ge a(\tau) \rangle\ 0$ for all $0\le t\le\tau,\ \beta\in\Lambda$ and the families of functions $s^{(j)}(\cdot,t)$ and $s_\pi{}^{(j)}(\cdot,t)$ $0\le t\le\tau,\ j=0,1,2,$ are equicontinuous at β_0, (4.8)

$$\sup_{0\le t\le\tau,\beta\in\Lambda} \left| \hat{V}(\beta,t) - v(\beta,t) \right| \xrightarrow{P_{d,m}} 0, \quad as\ n\to\infty, \tag{4.9}$$

$$\sup_{0\le t\le\tau,\beta\in\Lambda} \left| \frac{n}{N} \hat{V}_\pi(\beta,t) - v_\pi(\beta,t) \right| \xrightarrow{P_{d,m}} 0, \quad as\ n\to\infty. \tag{4.10}$$

Remark 4.1. Under SRSWOR, $s_\pi^{(j)}(\beta,t) = s^{(j)}(\beta,t),\ j=0,\ 1,\ 2$. Condition M_1 implies that $E_m\left\{\left| X_{hik}^{\otimes j} \cdot Y_{hik}(t) \cdot e^{\beta' X_{hik}} \right|\right\} \le B_1 < \infty$ for all h,i,k,N. This enable us to establish equations (4.3), (4.5) and (4.6) and the equicontinuity of the $s^{(j)}(\beta,t)$ and $s_\pi^{(j)}(\beta,t),\ j=0,\ 1,\ 2$ functions. We note also that Condition M_4 follows if $0 < P(T\le\tau \mid X=0) < 1)$. Condition M_5 is necessary to ensure that the $s^{(0)}(\beta,t)$ function is bounded away from zero and it would follow, for example, from a study of fixed length of time τ, where the censoring variables are continuous until time $t=\tau$ and have densities of the form outlined in Section 2.1. The proof of Theorem 4.1 is in the Appendix.

Next we state below Lenglart's inequality, first proved by Andersen and Gill (1982), and shown in Lemma 8.2.1 of Fleming and Harrington (1991), since it is an essential part of the proof of Corollary 4.1, Theorem 5.1 and lemma 6.1 below. The proof can be found in Fleming and Harrington (1991).

Lenglart's inequality. Let η be a univariate counting process with continuous compensator A. Let $\mathcal{M} = \eta - A$ and let H be a locally bounded predictable process. Then for all δ, ρ positive and any $t \geq 0$,

$$P_{d,m}\left(\sup_{0 \le u \le t} \left| \int_0^u H d\mathcal{M} \right| \geq \rho \right) \leq \frac{\delta}{\rho^2} + P_{d,m}\left(\int_0^t H^2 dA \geq \delta \right).$$

Corollary 4.1 Approximation for the normalized sample partial likelihood score process under the proportional hazards model. We consider the SPLS as a process in the product space, where both the sample s and the outcome $\omega \in \Omega$ of the model variables are random. We assume the proportional hazards model given in (2.1) a), b) and c) (we do not assume d), i.e., the stochastic independence of the censored lifetimes) and (2.2), as well as model conditions M_1, M_2, M_4 and M_5. We assume the design described in section 2 along with design conditions C_0, C_1, C_2 and C_3. We also assume that τ is the upper bound of the censoring variables, so $\hat{U}(\beta,\infty) = \hat{U}(\beta,\tau)$. We express the SPLS function as in equation (2.13), that is, as a sum of terms which are martingales under the model given in (2.2):

$$\hat{U}(\beta,\tau) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \int_0^\tau \hat{\psi}_{hik}(\beta,u) d\mathcal{M}_{hik}(u),$$

where for brevity, $\mathcal{M}_{hik}(u) = \mathcal{M}_{hik}(\beta_0,u)$ and $\hat{\psi}(\beta,u)$ is as in (2.12). Now let the process $\tilde{U}(\beta,\tau)$ be defined by

$$\tilde{U}(\beta,\tau) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} \int_0^\tau \psi_{hik}(\beta,u) d\mathcal{M}_{hik}(u) \tag{4.11}$$

where

$$\psi_{hik}(\beta,u) = \frac{I_{hik}(s)}{\pi_{hik}} (X_{hik} - e(\beta,u)), \quad k = 1,\ldots,N_{hi}, \quad i = 1,\ldots,N_h, \quad h = 1,\ldots,L. \tag{4.12}$$

Then we have:

$$\frac{\sqrt{n}}{M}\left(\hat{U}(\beta,\tau) - \tilde{U}(\beta,\tau)\right) = o_{P_{d,m}}(1), \quad n \to \infty. \tag{4.13}$$

The proof of Corollary 4.1 is in the Appendix.

Remark 4.2. The approximation (4.13) holds for every $0 \le t \le \tau$, and this means equivalence of the finite dimensional distributions. Here we have not proven tightness of the process in $0 \le t \le \tau$, and thus we cannot conclude from Corollary 4.1 that the processes $\left\{ \hat{U}(\beta,t),\ 0 \le t \le \tau \right\}$ and $\left\{ \tilde{U}(\beta,t),\ 0 \le t \le \tau \right\}$ are equivalent in distribution. However, under the correct assumption of a correlation structure in the model, tightness would follow from the same arguments we use in Section 5 to show tightness of the process $\{\hat{U}(\beta,t),\ 0 \le t \le \tau\}$ directly. Also we remark that this is a different proof of that in Rubin-Bleuer (2003) done anew with the aim to obtain the approximation without assuming a correlation structure in the super-population.

5. WEAK CONVERGENCE OF THE SPLS

For each t, $0 \le t \le \infty$, we denote by $\hat{\beta}_N(t)$ the solution of the estimating equation derived from the SPLS process $\hat{U}(\beta,t) = 0$. If $t = \infty$, we write $\hat{\beta}_N = \hat{\beta}_N(\infty)$ as established in the introduction. In the developments below, for simplicity of notation we write $\hat{\beta} = \hat{\beta}_N(t)$. Let $\hat{\Sigma}_\pi(\beta,t) = \int_0^t \frac{n}{N}\hat{V}_\pi(\beta,u)\frac{d\hat{\eta}(u)}{M}$, and let

$\mathrm{I}(\beta,\mathrm{t}) = \hat{\Sigma}_m(\beta,t) = \int_0^t \hat{V}(\beta,u)\frac{d\hat{\eta}(u)}{M}$.

Theorem 5.1. Assume the PHM given in (2.1) a), b), c) and d) and (2.2) and the design stated in Section 2. Assume the conditions of Theorem 4.1 and assume that the matrices

$$\Sigma_\pi(\beta,t) = \int_0^t v_\pi(\beta,u)s^{(0)}(\beta,u)\lambda_0(u)du \text{ and } \Sigma_m(\beta,t) = \int_0^t v(\beta,u)s^{(0)}(\beta,u)\lambda_0(u)du \qquad (5.1)$$

are positive definite, for $0 \le t \le \tau$. Then the normalized vector sample partial likelihood score process (SPLS) $\{\sqrt{n}\frac{\hat{U}(\beta_0,t)}{M} : 0 \le t \le \tau\}$ whose value at time t is $\sqrt{n}\frac{\hat{U}(\beta_0,t)}{M} = \frac{1}{\sqrt{n}}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\{\hat{\varphi}_{hik}(\beta_0,u)\}d\eta_{hik}(u)$,

- converges weakly in $D[0,\tau]^r$ to mean zero r-dimensional Gaussian process such that each component process has independent increments and the covariance function at t for components ℓ and ℓ' is $\frac{1}{\mu}\Sigma_\pi(\beta_0,t)_{\ell,\ell'}$ with

$$\Sigma_\pi(\beta_0,t)_{\ell,\ell'} = \int_0^t v_\pi(\beta_0,u)_{\ell,\ell'}s^{(0)}(\beta_0,u)\lambda_0(u)du, \qquad (5.2)$$

- $\hat{\beta} = \hat{\beta}(t)$, the solution of the estimating equation given by the SPLS function $\sqrt{n}\frac{\hat{U}(\beta_0,t)}{M}$, is a consistent estimator of β_0, for $0 \le t \le \tau$,

$$(5.3)$$

- $\Sigma_\pi(\beta,t)$ is the asymptotic variance of the SPLS and

$$\sup_{0\le t\le\tau}\left|\hat{\Sigma}_\pi(\hat{\beta},t) - \Sigma_\pi(\beta_0,t)\right| \xrightarrow{P_{d,m}} 0 \text{ as } n \to \infty, \qquad (5.4)$$

- $\Sigma_m(\beta,t)$ is the asymptotic variance of the Partial Likelihood Score process in the census case and

$$\sup_{0\le t\le\tau}\left|\mathrm{I}(\hat{\beta},t) - \Sigma_m(\beta_0,t)\right| \xrightarrow{P_{d,m}} 0 \text{ as } n \to \infty. \qquad (5.5)$$

Statement (5.5) means that the information matrix converges to the same limit as that of the census case, but it does not coincide with the variance of the SPLS. The proof of Theorem 5.1 is in the Appendix.

Remark 5.1 Note that the model conditions imposed by Theorem 4.1 and Theorem 5.1 are not too different from the conditions imposed on the model by Fleming and Harrington, 1991, Theorem 8.4.1, for the partial likelihood score function to converge. The extra model conditions M_2 and M_3 stem from the generalization of that theorem to the case of non-identically distributed failure times and to the joint design-model product space.

Remark 5.2 We have two comments about the proof. First, we do not use the approximation to the SPLS of Corollary 4.1 to show weak convergence of the SPLS process, as Lin (2000) does. Second, in the course of the proof, depending on what is most convenient, we will use either that the model martingales $\mathcal{M}_{hik}$ associated with the counting processes η_{hik} are also martingales in the product space with respect to the filtration $\left\{\Im_t^{d,m} : t \geq 0\right\}$ or that both product space processes $\hat{\eta}$ and $\hat{\mathcal{M}}$ are respectively a counting process and its associated martingale with respect to the filtration $\left\{\Im_t^{d,m} : t \geq 0\right\}$.

Corollary 5.1
The solution $\hat{\beta} = \hat{\beta}_N(t)$ be the solution of the SPLS estimating equation evaluated at t, which we call maximum sample partial likelihood estimator, is asymptotically normal and

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \Rightarrow \mathcal{N}(0, \frac{1}{\mu}\Sigma_m^{-1}(\beta_0, t)\Sigma_\pi(\beta_0, t)\Sigma_m^{-1}(\beta_0, t)) \tag{5.6}$$

where $\hat{\Sigma}_m$ and $\hat{\Sigma}_\pi$ are defined in (5.1). The proof is standard and given in the appendix.

6. A ROBUST VARIANCE ESTIMATOR

When working within the design-model space, we can define at least two different variance estimators, , denoted here by v_{eff} (defined in (A.6.13)) and v_{robust} (defined in (A.6.12)), for sample estimating equations when the terms have zero model expectation and are stochastically independent in the model space. Both variance estimators are consistent if the correlation structure is correct. As their names indicate, v_{eff} has the smaller variance while v_{robust} is robust against misspecification of the correlation structure of the model space. The corresponding definitions and a detailed illustration are given in the Appendix (see "Robust versus efficient variance estimators").

Lin (2000) considered the SPLS function $\hat{U}(\beta_0) = \hat{U}(\beta_0, \tau)$ normalized by the constant $\frac{1}{\sqrt{M}}$, where M is the number of units in the finite population. He sets

$$\frac{\hat{U}(\beta_0)}{\sqrt{M}} = \frac{\hat{U}(\beta_0) - U(\beta_0)}{\sqrt{M}} + \frac{U(\beta_0)}{\sqrt{M}}, \tag{6.1}$$

and uses the approximation

$$\frac{\hat{U}(\beta_0)-U(\beta_0)}{\sqrt{M}} \approx \frac{1}{\sqrt{M}} \sum_{i=1}^{M} (\frac{I_i(s)}{\pi_i} - 1) U_i(\beta_0), \tag{6.2}$$

where

$$U_i(\beta_0) = \int_0^\tau \ (X_i(\omega) - \frac{S^{(1)}(\beta_0,t)}{S^{(0)}(\beta_0,t)}) \{d\eta_i - Y_i(t)\exp(\beta_0' X_i)\lambda_0(t)dt\}. \tag{6.3}$$

In the following, we use Lin's (2000) notation $i = 1,2,\ldots,M$, for the unit labels, without the design classification into stratum, *psu*, and secondary sampling unit identification. Equation (2.13) and Corollary 4.1 shows that the approximation (6.2) is valid if the semi-parametric model for the hazard function is correct, and it does not require specification of the correlation structure of the model. Lin (2000) proposed as variance estimator of (6.2) the statistic

$$v_{Lin} = \frac{1}{M} \sum_{i \in s} \sum_{j \in s} \frac{\Delta_{ij}}{\pi_{ij}} \hat{U}_i(\hat{\beta}) \hat{U}_j(\hat{\beta}), \tag{6.4}$$

where

$$\hat{U}_i(\hat{\beta}) = \int_0^\tau \left(X_i - \hat{e}(\hat{\beta},t)\right) \left(d\eta_i - Y_i(t)\exp(\hat{\beta}' X_i) \frac{d\hat{\eta}}{M\hat{S}^{(0)}(\hat{\beta},t)} \right). \tag{6.5}$$

In Lemma 6.1, we show that v_{Lin} is asymptotically unbiased.

Lemma 6.1. We assume the conditions of Corollary 4.1, as well $\lim \frac{n}{N} > 0$ ($f > 0$ in Condition C_0) and $\hat{\beta} - \beta_0 = O_{P_{d,m}}(\frac{1}{\sqrt{M}})$ as $M \to \infty$, then v_{Lin} is asymptotically unbiased.

The proof of Lemma 6.1, which uses Lenglart's inequality, can be found in the Appendix.

Remark 6.1. Note that with $\frac{1}{\sqrt{M}}$ as the normalizing constant, the result is invalid if the first stage sampling rate is negligible (i.e., if $\lim \frac{n}{M} = 0$).

On the other hand, Theorem 5.1 shows that if in addition, the censored lifetimes and covariates are stochastically independent, the limiting variance of each of the terms in (6.1) is $\frac{1}{f}\Sigma_\pi$, $\frac{1}{f}\Sigma_\pi - \Sigma_m$ and Σ_m respectively. Moreover, $\frac{N}{n}\hat{\Sigma}_\pi$, $\frac{N}{n}\hat{\Sigma}_\pi - \hat{\Sigma}_m$ and $\hat{\Sigma}_m$, as defined in (5.4) and (5.5) are the respective design-model consistent estimators. Two questions arise:

1) how does v_{Lin} compares with $\frac{N}{n}\hat{\Sigma}_\pi - \hat{\Sigma}_m$ when the correlation structure of the model is correct, and
2) is v_{Lin} robust as a variance estimator of the normalized SPLS function?

Lemma 6.2 below answers the first question and we can make the robustness argument if we work with the normalizing constant $\frac{\sqrt{n}}{M}$ (instead of $\frac{1}{\sqrt{M}}$) and assume $\lim \frac{n}{M} = 0$ as $n \to \infty$. Indeed, let us express the normalized SPLS function as the sum of two terms:

$$\sqrt{n}\left[\frac{\hat{U}(\beta_0)}{M}\right]=\sqrt{n}\left[\frac{\hat{U}(\beta_0)-U(\beta_0)}{M}\right]+\sqrt{\frac{n}{M}}\left[\frac{U(\beta_0)}{\sqrt{M}}\right]. \tag{6.6}$$

The limiting variance of the normalized SPLS in (6.6) is now $\frac{1}{\mu}\Sigma_\pi(\beta_0,t)$ whereas the limiting variance of the first term in (6.6) is $\frac{1}{\mu}\{\Sigma_\pi(\beta_0,t)-f\cdot\Sigma_m(\beta_0,t)\}$. Now we write Lin's (2000) variance estimator as

$$v_{Lin}=v_{eff}+v_{resid} \tag{6.7}$$

where

$$v_{eff}=\frac{n}{M^2}\sum_{i\in s}(\frac{1}{\pi_i}-1)\frac{\hat{U}_i^2(\hat{\beta})}{\pi_i}, \tag{6.8}$$

and

$$v_{resid}=\frac{n}{M^2}\sum_{i\neq j\in s}\frac{\Delta_{ij}}{\pi_{ij}}\hat{U}_i(\hat{\beta})\hat{U}_j(\hat{\beta}). \tag{6.9}$$

Lemma 6.2. We assume the conditions of Theorem 5.1. Then we have:

$$v_{eff}=\frac{N}{M}\hat{\Sigma}_\pi(\beta_0,t)-\frac{n}{M}\hat{\Sigma}_m(\beta_0,t)+O_{P_{d,m}}(\frac{1}{\sqrt{n}})\text{ as } n\to\infty. \tag{6.10}$$

If we also assume

$$C_4:\quad \max_{i\in Pop,j\in Pop}\frac{1}{\pi_{ij}}=O(\frac{N}{n})^2\ \text{ as } n\to\infty,$$

then

$$v_{resid}=O_{P_{d,m}}(\frac{1}{\sqrt{n}})\ \text{ as } n\to\infty. \tag{6.11}$$

Note that with the new normalizing constant we have, by Corollary 5.1, $\hat{\beta}-\beta_0=O_{P_{d,m}}(\frac{1}{\sqrt{n}})$. The proof of Lemma 6.2 is in the Appendix.

Lin's (2000) variance estimator is of the type v_{robust} whereas the variance estimator proposed here is of the type v_{eff}. Lemma 6.2 implies that the variance of the first term in (6.6) proposed by Lin (2000) is consistent:

$$v_{Lin}\to\frac{1}{\mu}\Sigma_\pi(\beta_0,t)-\frac{f}{\mu}\Sigma_m(\beta_0,t)\ \text{ as } n\to\infty.$$

From the discussion of robust versus efficient estimators given in the Appendix, the variance of v_{Lin} is always larger than the variance of $\frac{N}{M}\hat{\Sigma}_\pi(\beta_0,t)-\frac{n}{M}\hat{\Sigma}_m(\beta_0,t)$, under the correct model for the correlations of the units in the super-population, and hence v_{Lin} is less efficient. On the other hand, if the second term in (6.6) is negligible we could say that v_{Lin} is "robust" in the sense that it is an asymptotically unbiased estimator of the variance of the normalized

SPLS function even if the correlation structure of the super-population model is not correctly specified.

7. CONCLUSIONS

Data from most complex surveys are subject to selection bias and clustering due to the sampling design. Hence, results developed for a random sample from a super-population model do not apply. Ignoring the survey sampling weights may cause the divergence of the involved processes. The data is subject to a two phase randomization and accounting for this means working in a "product space" that includes the model and the design.

The design conditions we assume refer to the order of magnitude of the sequence of selection probabilities $\left\{\pi_{hik}^{N}\right\}$ and of the joint selection probabilities $\left\{\pi_{hik\ell jm}^{N}\right\}$ as $n \to \infty$ (see conditions C_0 to C_3 in Section 2.2 and C_4 in Lemma 6.2). They are mild conditions in the sense that are verified by a SRSWOR design and that are minimum sufficient conditions for the Horvitz-Thompson estimators to be design-consistent. We distinguish between three sets of model conditions:

General regularity conditions: A series of baseline model conditions, e.g., (2.1) conditions a), b) and c), and M_1, M_2, M_3, M_4 and M_5.
Proportional hazards model: The model specification for the hazard function:

$$\lambda_{hik}(t \mid X_{hik}) = \lambda_0(t) \cdot \exp(\beta_0' \cdot X_{hik}),$$

$k = 1,\ldots,N_{hi}$, $i = 1,\ldots,N_h$, $h = 1,\ldots,L$ (given in equation (2.2)).

Correlation model: The model specification for the correlation structure, that is, the censored lifetimes and covariates are stochastically independent (given in equation (2.1) d)).

Given the censored lifetimes $\tilde{T}_{hik}$ with corresponding covariates X_{hik}, recall that the $S^{(j)}$ and the π-weighted $S_{\pi}^{(j)}$ are processes which are functions of the units at risk at time t, and that $\hat{\eta}$ is the sample counting process associated with the proportional hazards model and the design.

Under design conditions C_0 to C_3, and *general regularity conditions* we obtained the following:

Result 1. Assuming in addition *the proportional hazards model*, counting processes tools were developed in the joint design-model "product space".

Result 2. Assuming in addition *the proportional hazards model,* the normalized "survey sample" partial likelihood score process (SPLS) function

$$\frac{\sqrt{n}}{M}\hat{U}(\beta) = \frac{\sqrt{n}}{M}\sum_{hik \in s} \delta_{hik}\left\{X_{hik} - \frac{\hat{S}^{(1)}(\beta,\tilde{T}_{hik})}{\hat{S}^{(0)}(\beta,\tilde{T}_{hik})}\right\} / \pi_{hik}$$

is approximated by the Horvitz-Thompson sample estimator of a total:

$$\frac{\sqrt{n}}{M}\hat{U}(\beta)=\frac{\sqrt{n}}{M}\sum_{hik\in s}\delta_{hik}\left\{X_{hik}-\frac{S^{(1)}(\beta,\widetilde{T}_{hik})}{S^{(0)}(\beta,\widetilde{T}_{hik})}\right\}/\pi_{hik}+o_{P_{d,m}}(1).$$

Result 3. Assuming in addition *the proportional hazards model* and *the correlation model,* the normalized SPLS process converges weakly to a Gaussian process in converges weakly in $D[0,\tau]^r$ to mean zero r-dimensional Gaussian process with covariance function at t given by:

$$\frac{1}{\mu}\Sigma_\pi(\beta_0,t)=\frac{1}{\mu}\int_0^t v_\pi(\beta_0,u)s^{(0)}(\beta_0,u)\lambda_0(u)du,$$

where the v_π function is the point-wise limit of a combination of $S^{(j)}$ and π-weighted $S_\pi^{(j)}$ functions and $\mu=\lim\frac{M}{N}$ as $n\to\infty$.

Result 4. Assuming in addition *the proportional hazards model* and *the correlation model,* the matrix

$$\frac{N}{M}\hat{\Sigma}_\pi(\beta_0,t)=\frac{n}{M}\int_0^t\hat{V}_\pi(\beta_0,u)d\hat{\eta}(u)/M$$

is a design-model consistent estimator of the variance of the normalized SPLS process.

Result 5. Assuming in addition *the proportional hazards model* and *the correlation model,* and denoting by $\hat{\beta}$, the maximum sample partial likelihood estimator of β_0, then $\sqrt{n}(\hat{\beta}-\beta_0)$ is asymptotically normal with mean zero and variance $\frac{1}{\mu}\Sigma_m^{-1}(\beta_0)\Sigma_\pi(\beta_0)\Sigma_m^{-1}(\beta_0)$, where

$$\Sigma_m(\beta_0)=\int_0^t v(\beta_0,u)s^{(0)}(\beta_0,u)\lambda_0(u)du.$$

Result 6. Assuming in addition *the proportional hazards model* and $\sqrt{n}(\hat{\beta}-\beta_0)=O_{P_{d,m}}(1)$ then v_{Lin}, the variance estimator proposed by Lin (2000) is asymptotically unbiased. The variance estimator v_{Lin}, when the SPLS function is normalized by $\frac{\sqrt{n}}{M}$, is robust against misspecification of the correlation structure when the first stage sampling rate is negligible $\left(\frac{n}{N}\to 0\right)$ and the PLS is bounded in probability.

Result 7. Assuming in addition *the proportional hazards model, the correlation model,* and design condition C_4 then the variance estimator of the SPLS proposed by Lin (2000), when normalized by $\frac{\sqrt{n}}{M}$, attain the approximation

$$v_{Lin} = \frac{N}{M}\hat{\Sigma}_{\pi}(\beta_0,t) + O_{P_{d,m}}\left(\frac{1}{\sqrt{n}}\right) \text{ as } n \to \infty .$$

Result 8. Assuming in addition *the proportional hazards model* and *the correlation model,* then $\frac{N}{M}\hat{\Sigma}_{\pi}(\beta_0,t)$ is more efficient than v_{Lin}.

Result 9. All of the above results are shown for a particular estimator of PLS, an "approximate" HT estimator of the PLS process (with HT estimators $\hat{S}$ of the S processes), which is design consistent under design conditions C_1 to C_3. But the proofs hold for any design-consistent estimators of both, the S and the PLS processes, under a "without replacement" design.

Result 10. All of the above results are valid for both the unconditional and the conditional (given the covariates) proportional hazards model, since we assume that the covariates are uniformly bounded.

8. REFERENCES

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

Binder, D. A.(1992). "Fitting Cox's proportional hazards models from survey data", *Biometrika* **79**, 139-147.

Chung, K. L. (1974). *A course in probability theory*. Academic Press, New York and London, second ed.

Cox, D.A. (1992). "Regression models and life-tables" (with Discussion), *J.R. Statist. Soc.* B **34**, 187-220.

Fleming,T. and Harrington, D.(1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Krewski, D. and Rao, JNK. (1981). "Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication methods", *Ann. Statist.* **9**, 1010–1019.

Hartley, H.O.& Sielken, R.L.(1975). "A 'super-population viewpoint' for finite population sampling", *Biometrics* **31**, 411-422

Lin, D.Y.(2000). "On fitting Cox's proportional hazards models to survey data", *Biometrika* **87**, 37-47.

Pfeffermann, D.(1993). "The role of sampling weights when modeling survey data", *International Statistical Review* **61**, 317-337.

Rubin Bleuer, S. (2001). "A test for survival distributions using data from a complex sample", *Proceedings of the Survey Methods Section, SSC Annual meeting, 2001,* 103-110.

Rubin Bleuer, S. (2003 a). "On the weak convergence of the sample empirical distribution process", *Statistics Canada Series BSMD – 2003 -003E.*

Rubin Bleuer, S. (2003 b). "An approximation of the partial likelihood score in a joint design-model space", *Proceedings of the Survey Methods Section, SSC Annual Meeting, 2003.*

Rubin Bleuer, S. (2004)."On Fitting the Proportional Hazards Model to Data from Complex Surveys," *2004 Proceedings of the American Statistical Association*, Section on Survey Research Methods [CD-ROM], Alexandria, VA: American Statistical Association: 4283-4290.

Rubin Bleuer, S. and Schiopu Kratina, I. (2005). "On the two-phase framework for joint model and design-based inference", *The Annals of Statistics, 2005,* Vol 33, No 5, 2789-2810.

Särndal, C-E, Swensson, B. and Wretman, J. (1992). "*Model Assisted Survey Sampling*", Springer-Verlag, New York.

Shao, J. (1999). "*Mathematical Statistics*", Springer-Verlag, New York.

Sutradhar and Kovacevic (2000). "Analysing ordinal longitudinal survey data: Generalised estimating equations approach", *Biometrika*, **87**, 4, 837-848.

APPENDIX

Proof of Theorem 4.1:
The point-wise convergence of the S processes (Statement (4.2) of the theorem) follows from assumptions M_2 and M_3, the independence of the units of the super-population, the WLLN's (see Chung, 1974, Theorem 5.1.1) and the fact that the terms in the $S^{(j)}(\beta,t)$, in the π- weighted $S_\pi^{(j)}(\beta,t)$ and respective evaluations at $t+$ are uniformly bounded in $0 \le t < \infty$ and a compact neighborhood of β_0. Under design conditions C_1, C_2 and C_3, the HT sample estimators $\hat{S}^{(j)}(\beta,t)$, $\hat{S}^{(j)}(\beta,t+)$, $\frac{n}{N}\hat{S}_\pi^{(j)}(\beta,t)$ and $\frac{n}{N}\hat{S}_\pi^{(j)}(\beta,t+)$ are design consistent. Indeed, for ease of notation, set $z_{hik} = X_{hik}^{\otimes j} Y_{hik}(t)\exp\{\beta' X_{hik}\}$ and $\sum_{hik \neq \ell jm}$ to denote the 6-way sum over the indices. Then $E_d\left(\frac{n}{N}\hat{S}_\pi^{(j)}(\beta,t) - \frac{n}{N}S_\pi^{(j)}(\beta,t)\right) = 0$ and its design variance is equal to

$$\frac{1}{M^2}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\left(\frac{1}{\pi_{hik}}-1\right)\frac{1}{\pi_{hik}{}^2}\left(\frac{n}{N}\right)^2 z_{hik}^2 + \frac{1}{M^2}\sum_{hik\neq \ell jm}\frac{\pi_{hik\ell jm}-\pi_{hik}\pi_{\ell jm}}{\pi_{hik}\pi_{\ell jm}}\frac{1}{\pi_{hik}\pi_{\ell jm}}\left(\frac{n}{N}\right)^2 z_{hik}z_{\ell jm}$$

$$= O\left(\frac{1}{n}\right)$$

as $n \to \infty$, bounded this way by the design conditions given above and by the fact that the z_{hik} are uniformly bounded. The design consistency of the other HT sample estimators follow from the same considerations. The design consistency and (4.2) imply the point-wise convergence of the HT estimators to the non-stochastic s functions in the probability of the product space ($P_{d,m}$): convergence in design probability implies convergence in the probability of the product space (Rubin-Bleuer and Schiopu-Kratina, 2005, Theorem 5.1).

The uniform convergence in probability of the S- processes and their sample estimators, $j = 0,\ 1,\ 2,$ is established along the same lines. We first note that as functions of $t,\ 0 \le t < \infty$, all of these functions are left-continuous and non-increasing, hence Lemma 4.1 and the point-wise convergence in probability P for the S processes and in probability $P_{d,m}$ for the $\hat{S}$ processes, implies uniform convergence in $0 \le t \le \tau$ in their respective probability. The uniform convergence in a compact neighborhood of β_0 for (4.5) follows because the functions

$$\sup_{0 \le t \le \tau} \left| \frac{n}{N} S_{\pi}^{(j)}(\beta,t) - s_{\pi}^{(j)}(\beta,t) \right|, \quad j = 0,\ 1,\ 2$$

converge to zero as $N \to \infty$, and are continuously differentiable with respect to β with uniformly bounded derivates (if both β and the covariates X_{hik} are uniformly bounded). The same argument holds to obtain the uniform convergence in a compact neighborhood of β_0 in (4.3) to (4.6).

Statement (4.7) follows from the Dominated Convergence Theorem and the uniform bounded character of the covariates X_{hik} which are sufficient to justify the interchange of differentiation and expectation.

That the function $s^{(0)}(\beta,t)$ is uniformly bounded away from zero (statement (4.8)) follows from adapting the argument in Fleming and Harrington (1991), Theorem 8.4.1, p.306, to the case of non-identically distributed model variables. The equicontinuity of the functions at $\beta = \beta_0$ follow, also using similar arguments to those in Fleming and Harrington (1991), Theorem 8.4.1, p.306:

$$\sup_{0 \le t\, \tau} \left| s^{(j)}(\beta_m,t) - s^{(j)}(\beta_0,t) \right| \le \lim_{N \to \infty} \frac{1}{M} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{k=1}^{N_{hi}} E_m \left\{ \left| X_{hik}^{\otimes j} \right| \cdot \left(e^{\left|\beta_m - \beta_0\right|' \left|X_{hik}\right|} - 1 \right) \right\}$$

$$\le O(1)\left(e^{\left|\beta_m - \beta_0\right|' |B|} - 1\right) \to 0 \text{ as } \beta_m - \beta_0 \to 0,$$

since the covariates are uniformly bounded. The same arguments hold for the functions $s_{\pi}^{(j)}(\beta,t), \quad j = 0,1,2,$ under design condition C_1.

Statements (4.9) and (4.10) follow from the uniform convergence of the S-processes and from the fact that $s^{(0)}(\beta,t)$ is uniformly bounded away from zero.

Proof of Corollary 4.1:

The functions defined in (4.11) are locally bounded $\Im_t^{d,m}$-predictable processes (note that in the model space, the $(X_{hik} - e(\beta,u))$, $k = 1,\dots,N_{hi}$, $i = 1,\dots,N_h$, $h = 1,\dots,L$, are locally bounded $\Im_t$ -predictable processes).

$$\sqrt{n}\frac{\hat{U}(\beta,\tau)-\tilde{U}(\beta,\tau)}{M} = \frac{\sqrt{n}}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\frac{I_{hik}(s)}{\pi_{hik}}\int_0^{\tau}\{\frac{S^1(\beta,u)}{S^0(\beta,u)} - \frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)}\}d\mathcal{M}_{hik}(u)\cdot$$

$$= \frac{\sqrt{n}}{M}\int_0^{\tau}\{\frac{S^1(\beta,u)}{S^0(\beta,u)} - \frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)}\}d\hat{\mathcal{M}}(u)\cdot \qquad \text{(A.4.1)}$$

The process $\hat{\mathcal{M}}(u)$ in (A.4.1) is the martingale defined in (3.1) with compensator $\hat{A}(u)$ as in (3.2).

Design conditions C_1, C_2 and C_3 imply that point-wise $\hat{S}^{(j)}(\beta,t) - S^{(j)}(\beta,t) \to 0$ and $\hat{S}^{(j)}(\beta,t+) - S^{(j)}(\beta,t+) \to 0$ in $P_{d,m}$. Lemma 4.1 imply that this convergence is uniform in $0 \le t \le \tau$ in $P_{d,m}$. If $s^{(0)}(\beta,t)$ is bounded away from zero uniformly, then $S^{(0)}(\beta,t)$ and $\hat{S}^{(0)}(\beta,t)$ are bounded away from zero uniformly on $0 \le t \le \tau$. Hence

$$\sup_{0\le u\le\tau} |\{\frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)} - \frac{S^1(\beta,u)}{S^0(\beta,u)}\}| \xrightarrow{P_{d,m}} 0 \text{ as } n \to \infty. \qquad \text{(A.4.2)}$$

Now we can show that the martingale in (A.4.1) converges to zero in the probability of the product space. Using Lenglart's inequality, and since $\hat{\mathcal{M}}(u)$ is a martingale and the integrand is a bounded predictable process, we have:

$$P_{d,m}\left(\left|\frac{\sqrt{n}}{M}\int_0^{\tau}\{\frac{S^1(\beta,u)}{S^0(\beta,u)} - \frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)}\}d\hat{\mathcal{M}}(u)\right| \ge \rho\right) \le \frac{\delta}{\rho^2} + P_{d,m}\left(\frac{n}{M^2}\int_0^{\tau}\{\frac{S^1(\beta,u)}{S^0(\beta,u)} - \frac{\hat{S}^1(\beta,u)}{\hat{S}^0(\beta,u)}\}^2 d\hat{A}(u) \ge \delta\right) \qquad \text{(A.4.3)}$$

By design condition C_2 we have

$$\frac{n}{M^2}d\hat{A}(t) \approx \frac{n}{N\mu}\cdot\hat{S}^{(0)}(\beta_0,t)\lambda_0(t)dt = O_{P_{d,m}}(\frac{n}{N}) \text{ as } n \to \infty. \qquad \text{(A.4.4)}$$

Equations (A.4.2), (A.4.3), (A.4.4) and condition M_4 imply that the second term of the right hand side of (A.4.3) converges to zero as $n \to \infty$. Hence the first side of (A.4.3) also converges to zero and the corollary is proven.

Proof of Theorem 5.1: we follow arguments similar, but not identical, to those used in the version of the census case theorem shown in Fleming and Harrington (Theorems 8.2.1 in p.290, 8.3.1, p. 297 and 8.4.1, p. 305) for the weak convergence of the partial likelihood score process and the consistency of the maximum partial likelihood estimator.

Proof of (5.2): Weak convergence of the SPLS process.

In the same manner of Theorem 8.2.1 mentioned above, we must show that the process $\left\{\sqrt{n}\frac{\hat{U}(\beta_0,t)}{M},\ 0\le t\le\tau\right\}$ verifies the conditions of the Central Limit Theorem for Martingales given in Theorem 5.3.5, Fleming and Harrington (1991). Consider again the martingale representation:

$$\sqrt{n}\frac{\hat{U}(\beta_0,t)}{M}=\frac{\sqrt{n}}{M}\sum_{hik}\int_0^t\hat{\psi}_{hik}(u)d\mathcal{M}_{hik}(u),$$

with the functions $\psi_{hik}(u),\ k=1,\ldots,N_{hi},\ i=1,\ldots,N_h,\ h=1,\ldots,L$ defined in (2.12) with $\beta=\beta_0$ and $\mathcal{M}_{hik}(u)=\mathcal{M}_{hik}(\beta_0,u)$. Let $\hat{\psi}_{hik\ell}$ the ℓ-the element of the vector of r components $\hat{\psi}_{hik}$ (the number of components of $\hat{\psi}_{hik}$ is the number of components of the vector of covariates X_{hik}).

The ℓ-the element of the SPLS vector process is given by

$$\sqrt{n}\frac{\hat{U}_\ell(\beta_0,t)}{M}=\frac{\sqrt{n}}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\hat{\psi}_{hik\ell}(u)d\mathcal{M}_{hik}(u)\cdot$$

For positive ε we define $\frac{\sqrt{n}\hat{U}_{\ell,\varepsilon}(\beta_0,t)}{M}=\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\frac{\sqrt{n}}{M}\hat{\psi}_{hik\ell}(u)I\left\{(s,\omega):\sqrt{n}\frac{\hat{\psi}_{hik\ell}(u)}{M}\ge\varepsilon\right\}d\mathcal{M}_{hik}(u).$

To prove the statements in (5.2) we use the Central Limit Theorem for Martingales and we show the following statements:

a) $\frac{n}{M^2}\langle\hat{U}_{\ell,\varepsilon}(\beta_0,\cdot),\hat{U}_{\ell,\varepsilon}(\beta_0,\cdot)\rangle(t)\to 0$ in $P_{d,m}$ $n\to\infty,\ \ \ell,\ell'=1,\ldots,r,\ \ \varepsilon\rangle 0,\ \ 0\le t\le\tau$,

b) $\frac{n}{M^2}\langle\hat{U}_\ell(\beta_0,\cdot),\hat{U}_{\ell'}(\beta_0,\cdot)\rangle(t)\to\frac{1}{\mu}\Sigma_\pi(\beta,t)_{\ell,\ell'}$ $as\ n\to\infty,\ \ \ell,\ell'=1,\ldots,r,\ \ 0\le t\le\tau$.

To show a), consider a positive number ε and $\ell=1,\ldots,r$, we follow the argument of Theorem 8.2.1, part 1, though the bounded covariates make some steps trivial. Using the inequality $|a-b|^2 I\{|a-b|\rangle\varepsilon\}\le 4|a|^2 I\{|a|\rangle\varepsilon/2\}+4|b|^2 I\{|b|\rangle\varepsilon/2\}$,
valid for any real numbers a,b, we have

$$\frac{n}{M^2}\langle\hat{U}_{\ell,\varepsilon}(\beta_0,t),\hat{U}_{\ell,\varepsilon}(\beta_0,t)\rangle=\frac{n}{M^2}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\hat{\psi}^2_{hik\ell}(u)I\left(\sqrt{n}\frac{\hat{\psi}_{hik\ell}(u)}{M}\ge\varepsilon\right)Y_{hik}(u)e^{\beta_0 X_{hik\ell}}\lambda_0(u)du$$

$$\le\frac{4n}{M^2}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\frac{I_{hik}(s)}{\pi^2_{hik}}X^2_{hik\ell}I\left(\frac{\sqrt{n}}{M}\frac{X_{hik\ell}I_{hik}(s)}{\pi_{hik}}\ge\varepsilon/2\right)Y_{hik}(u)e^{\beta_0 X_{hik\ell}}\lambda_0(u)du$$

$$+\frac{4n}{M^2}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\int_0^t\frac{I_{hik}(s)}{\pi^2_{hik}}\hat{e}^2_\ell(u)I\left(\left|\frac{\sqrt{n}}{M}\frac{\hat{e}_\ell(u)I_{hik}(s)}{\pi_{hik}}\right|\ge\varepsilon/2\right)Y_{hik}(u)e^{\beta_0 X_{hik\ell}}\lambda_0(u)du. \quad (A.5.1)$$

Since $X_{hik\ell}$ is uniformly bounded and design condition C_2 implies that $\frac{\sqrt{n}}{M} \approx \frac{\sqrt{n}}{N} \le O\left(\frac{1}{\sqrt{N}}\right)$ as $n \to \infty$, we have $E_{d,m}\left\{\left|\frac{\sqrt{n}}{M} X_{hik\ell} \frac{I_{hik}(s)}{\pi_{hik}}\right|\right\} \to 0$ uniformly in hik as $n \to \infty$. A similar argument used to obtain (A.4.2) yields $\hat{e}_\ell = \frac{\hat{S}_\ell^{(1)}}{\hat{S}^{(0)}} \to \frac{s_\ell^{(1)}}{s^{(0)}}$ uniformly in t, $0 \le t \le \tau$ in the probability of the product space. Thus we have $E_{d,m}\left\{\left|\frac{\sqrt{n}}{M} \hat{e}_\ell(u) \frac{I_{hik}(s)}{\pi_{hik}}\right|\right\} \to 0$ uniformly in hik as $n \to \infty$.

Hence both

$$I\left\{\left|\frac{\sqrt{n}}{M} X_{hik\ell} \frac{I_{hik}(s)}{\pi_{hik}}\right| \ge \varepsilon/2\right\} = o_{P_{d,m}}(1) \text{ and } I\left\{\left|\frac{\sqrt{n}}{M} \hat{e}_\ell(\beta_0, u) \frac{I_{hik}(s)}{\pi_{hik}}\right| \ge \varepsilon/2\right\} = o_{P_{d,m}}(1) \text{ as } n \to \infty,$$

uniformly for all $hik\ell$ and t, $0 \le t \le \tau$. Hence the two terms in (A.5.1) are bounded by

$$o_{P_{d,m}}(1) \frac{4n}{M} \int_0^t \hat{S}_{\pi\ell}^{(2)}(u)\lambda_0(u)du + o_{P_{d,m}}(1) \frac{4n}{M} \int_0^t \frac{\hat{S}_\ell^{(1)}(u)}{\hat{S}^{(0)}(u)} \hat{S}_\pi^{(2)}(u)\lambda_0(u)du = o_{P_{d,m}}(1) \text{ as } n \to \infty,$$

since $\frac{n}{M} \hat{S}_{\pi\ell}^{(2)}(t) = O_{P_{d,m}}(1)$, $\frac{n}{M} \frac{\hat{S}_\ell^{(1)}(t)}{\hat{S}^{(0)}(t)} \hat{S}_\pi^{(0)}(t) = O_{P_{d,m}}(1)$ uniformly in t, $0 \le t \le \tau$ and the baseline function is integrable in $0 \le t \le \tau$. Thus we obtained a).

To show b), let $\ell, \ell' = 1, \ldots, r$, $0 \le t < \infty$, then the predictable co-variation in the product space yields:

$$\frac{n}{M^2} \langle \hat{U}_\ell(\beta_0, \cdot), \hat{U}_{\ell'}(\beta_0, \cdot) \rangle(t) = \frac{n}{M^2} \sum_{hik} \int_0^t \hat{\psi}_{hik\ell}(u) \cdot \hat{\psi}_{hik\ell'}(u) \cdot Y_{hik}(u) \cdot e^{\beta_0' X_{hik}} \lambda_0(u)du$$

$$= \int_0^t \frac{n}{M^2} \sum_{hik} \frac{I_{hik}(s)}{\pi_{hik}^2} (X_{hik\ell} - \hat{e}_\ell(\beta_0, u)) \cdot (X_{hik\ell'} - \hat{e}_{\ell'}(\beta_0, u)) \cdot Y_{hik}(u) \cdot e^{\beta_0' X_{hik}} \lambda_0(u)du$$

$$= \frac{N}{M} \int_0^t \frac{n}{N} \hat{V}_\pi(\beta_0, u)_{\ell,\ell'} \hat{S}^{(0)}(\beta_0, u)\lambda_0(u)du. \tag{A.5.2}$$

Statement b) above follows from the uniform convergence of the $\frac{n}{N}\hat{V}_\pi(\beta_0,u)$ matrix in the product space probability (statement (4.10) of Theorem 4.1) the bounded character of the function $v_\pi(\beta_0,\cdot)_{\ell,\ell'} s^{(0)}(\beta_0,\cdot)$ and Condition M_4 of Theorem 4.1.

Proof of (5.3): Consistency of $\hat{\beta}$

Here we also follow the proof of Fleming and Harrington (1991) for the consistency of the maximum partial likelihood estimator, but we deviate from the proof slightly at the end. We base the proof, as Fleming and Harrington (1991) do, in the property that states that if a sequence of concave functions $g_N(\beta,t,s)$ of β, converge in probability $(P_{d,m})$ to a concave function $g(\beta,t)$ with unique maximum β_0 and for each N, $\hat{\beta}_N(t)$ is the unique maximum of the concave function $g_N(\beta,t)$ then $\beta_N(t) \xrightarrow{P_{d,m}} \beta_0$ (see Fleming and Harrington,1991, p297).

Let $g_N(\beta,t) = \frac{\hat{\ell}(\beta,t) - \hat{\ell}(\beta_0,t)}{N}$ be difference between the sample log partial likelihoods, evaluated at t and β, and at t and β_0 respectively for t, $0 \le t \le \tau$:

$$g_N(\beta,t) = \int_0^t (\beta-\beta_0)'\left[\frac{1}{M}\sum_{hik\in s} X_{hik}\, d\eta_{hik}(u)/\pi_{hik}\right] - \log\left(\frac{\hat{S}^{(0)}(\beta,u)}{\hat{S}^{(0)}(\beta_0,u)}\right)\left[\frac{1}{M}\sum_{hik\in s} d\eta_{hik}(u)/\pi_{hik}\right].$$

Let

$$h_N(\beta,t) = \int_0^t (\beta-\beta_0)'\hat{S}^{(1)}(\beta_0,u)\lambda_0(u)du - \log\left(\frac{\hat{S}^{(0)}(\beta,u)}{\hat{S}^{(0)}(\beta_0,u)}\right)\hat{S}^{(0)}(\beta_0,u)\lambda_0(u)du\,.$$

It is easy to see that $h_N(\beta,t,s) \xrightarrow{P_{d,m}} h(\beta,t)$ as $n\to\infty$, $0\le t\le\tau$, with

$$h(\beta,t) = \int_0^t \left\{(\beta-\beta_0)'s^{(1)}(\beta_0,u) - \log\left[\frac{s^{(0)}(\beta,u)}{s^{(0)}(\beta_0,u)}\right]s^{(0)}(\beta_0,u)\right\}\lambda_0(u)du\,.$$

Now, the martingale $g_N(\beta,t) - h_N(\beta,t) \to 0$, in the probability of the product space, for $0\le t\le\tau$ as $n\to\infty$. Indeed,

$$g_N(\beta,t,s) - h_N(\beta,t,s) = \frac{1}{M}\sum_{hik\in s}\int_0^t \left\{(\beta-\beta_0)'X_{hik} - \log\left(\frac{\hat{S}^{(0)}(\beta,u)}{\hat{S}^{(0)}(\beta_0,u)}\right)\right\}d\mathcal{M}_{hik}(u)/\pi_{hik}$$

is a locally square integrable martingale. Its variance according to the formula in Theorem 2.6.2 in p.82, Fleming and Harrington, 1991 is:

$$M \cdot V_{d,m}(g_N(\beta,t,s)-h_N(\beta,t,s)) = E_{d,m}\int_0^t \frac{1}{M}\sum_{hik \in s}\left\{(\beta-\beta_0)' X_{hik} - \log\left[\frac{\hat{S}^{(0)}(\beta,u)}{\hat{S}^{(0)}(\beta_0,u)}\right]\right\}^2 Y_{hik}(u)e^{\beta_0' X_{hik}} / \pi_{hik}^2 \cdot \lambda_0($$

The integrand above is bounded and thus

$V_{d,m}(g_N(\beta,t,s)-h_N(\beta,t,s)) = O(\frac{1}{M})$ as $n \to \infty$, hence $g_N(\beta,t,s) \overset{P_{d,m}}{\rightarrow} h(\beta,t)$ as $n \to \infty,\ 0 \le t \le \tau$.

As it is with the difference between the logs of the census partial likelihoods, $g_N(\beta,t,s)$ is a sequence of random concave functions of β with a unique maximum $\hat{\beta}_N(t)$, which is the solution of the SPLS process.

We also note that since are bounded, the Dominated Convergence Theorem justifies the interchange of differentiation and expectation for all the s-functions, and furthermore, they constitute equicontinuous families at $\beta = \beta_0$ for $0 \le t \le \tau$ (see Fleming and Harrington, 1991, p.298). Thus $h(\beta,t)$ has a unique maximum at $\beta = \beta_0$ and hence $\hat{\beta}_N(t) \overset{P_{d,m}}{\rightarrow} \beta_0$.

Proof of (5.4): Uniform convergence of the variance estimator of the SPLS process (Consistency)

To prove this part, we again follow a similar reasoning of Theorem 8.2.1, p.295 in Fleming and Harrington (1991); we bound the difference $\hat{\Sigma}_\pi(\hat{\beta}) - \Sigma_\pi(\beta_0,t)$ respectively by four terms and show that each term converge to zero:

$$\left|\hat{\Sigma}_\pi(\hat{\beta},t) - \Sigma_\pi(\beta_0,t)\right| \le \left|\int_0^t \left\{\frac{n}{N}\hat{V}_\pi(\hat{\beta},u) - v_\pi(\hat{\beta},u)\right\} d\frac{\hat{\eta}(u)}{M}\right| + \left|\int_0^t \left\{v_\pi(\hat{\beta},u) - v_\pi(\beta_0,u)\right\} d\frac{\hat{\eta}(u)}{M}\right|$$

$$+$$

$$\left|\int_0^t v_\pi(\beta_0,u)\left\{d\frac{\hat{\eta}(u)}{M} - \hat{S}^{(0)}(\beta_0,u)\lambda_0(u)du\right\}\right| + \left|\int_0^t v_\pi(\beta_0,u)\left\{\hat{S}^{(0)}(\beta_0,u) - s^{(0)}(\beta_0,u)\right\}\lambda_0(u)du\right|.$$

By (4.10) $\sup_{\beta \in \Lambda, 0 \le t \le \tau}\left|\frac{n}{N}\hat{V}_\pi(\hat{\beta},u) - v_\pi(\hat{\beta},u)\right| = o_{P_{d,m}}(1)$ as $n \to \infty$ and the process $\hat{\eta}(t)/M$ is bounded in the probability $(P_{d,m})$ since $E_d\left\{\left|\frac{\hat{\eta}(t)}{M}\right|\right\} \le 1$ for all $0 \le t < \infty,\ M \ge 1$.

Hence

$$\left|\int_0^t \left\{\frac{n}{N}\hat{V}_\pi(\hat{\beta},u) - v_\pi(\hat{\beta},u)\right\} d\frac{\hat{\eta}(u)}{M}\right| \le \sup_{\beta \in \Lambda, 0 \le t \le \tau}\left|\frac{n}{N}\hat{V}_\pi(\hat{\beta},u) - v_\pi(\hat{\beta},u)\right| \cdot \frac{\hat{\eta}(t)}{M} = o_{P_{d,m}}(1) \text{ as } n \to \infty.$$

That is, the first term converges to zero uniformly in $0 \le t \le \tau$ in the probability of the product space. The equicontinuity of the family of functions $s_\pi^{(j)}$, $s^{(j)}$, $j = 0, 1, 2$ implies the equicontinuity of v_π and thus the integrand of the second term converges to zero uniformly in $0 \le u \le \tau$ if $\hat{\beta}$ is consistent for β_0. This convergence and the fact that $\hat{\eta}(t)/M = O_{P_{d,m}}(1)$ also implies that the second term converges to zero in the probability of the product space. Similarly, the fourth term above converges to zero in probability $(P_{d,m})$ uniformly in $0 \le t \le \tau$. Finally, for the third term, we use Lenglart's inequality. We look at each element of the v-matrix $v_{\ell\ell'}(\beta_0,u)$, $\ell,\ell' = 1,\ldots,r$, and set:

$$\int_0^t v_{\pi\ell,\ell'}(\beta_0,u)(d\frac{\hat{\eta}(u)}{M} - \hat{S}^{(0)}(\beta_0,u)\lambda_0(u)du) = \int_0^t v_{\pi\ell,\ell'}(\beta_0,u)d\frac{\hat{\mathcal{M}}(u)}{M}.$$

Using Lenglart's inequality we have:

$$P_{d,m}(\sup_{0\le t\le\tau}\left|\int_0^t v_{\pi\ell\ell'}(\beta_0,u)d\frac{\hat{\mathcal{M}}(u)}{M}\right| \ge \rho) \le \frac{\delta}{\rho^2} + P_{d,m}(\frac{1}{M}\int_0^\tau v_{\pi\ell\ell'}(\beta_0,u)^2\hat{S}^{(0)}(\beta_0,u)\lambda_0(u)du) \ge \delta)$$

thus as $n \to \infty$, the second term above goes to zero, and since δ, ρ are arbitrary, we have the left hand side above converging to zero in probability $(P_{d,m})$.

Proof of (5.5): Consistency of the Information matrix

Now calculate the information matrix given by $\mathrm{I}(\hat{\beta},t)$:

$$\mathrm{I}(\hat{\beta},t) = \frac{-1}{M}(\partial\hat{U}/\partial\beta)(\hat{\beta},t) = \frac{1}{M}\int_0^t(\frac{\partial}{\partial\beta}\hat{e})(\hat{\beta},u)d\hat{\eta} = \int_0^t \hat{V}(\hat{\beta},u)d\frac{\hat{\eta}(u)}{M}.$$

The sup norm convergence of $\mathrm{I}(\hat{\beta},t)$ to $\Sigma_m(\beta,t)$ follows from the same arguments used in proving (5.3).

This finalizes the proof of Theorem 5.1.

Proof of Corollary 5.1:

This proof is standard. We express the SPLS statistic by

$$\hat{U}(\beta_0,t) = \hat{U}(\beta_0,t) - \hat{U}(\hat{\beta},t) = \frac{\partial\hat{U}}{\partial\beta}(\beta^*,t)(\beta_0 - \hat{\beta}),$$

where $\left|\beta^* - \beta_0\right| \le \left|\hat{\beta} - \beta_0\right|$. Hence,

$$\sqrt{n}(\hat{\beta} - \beta_0) = (-\frac{1}{M}\frac{\partial\hat{U}}{\partial\beta}(\beta^*,t))^{-1}\frac{\sqrt{n}}{M}\hat{U}(\beta_0,t).$$

The corollary follows from the asymptotic normality of the normalized SPLS process, the consistency of $\hat{\beta}(t)$, $\hat{\Sigma}_\pi(\beta_0,t)$ and $\frac{1}{M}\frac{\partial\hat{U}}{\partial\beta}(\beta,t)$, and Slutsky's theorem.

Proof of Lemma 6.1:
We first show that

$$\max_i \left|\hat{U}_i(\hat{\beta}) - \tilde{U}_i(\beta_0)\right| = O_{P_{d,m}}(\frac{1}{\sqrt{M}}), \qquad \text{as } M \to \infty, \tag{A.6.1}$$

where

$$\tilde{U}_i(\beta_0) = \int_0^\tau (X_i - \hat{e}(\beta_0, t)) d\mathcal{M}_i(\beta_0, t), \ \ i = 1, ..., M.$$

We recall that the covariates are uniformly bounded, $\hat{S}^{(0)}$ is bounded away from zero and $\hat{\beta} - \beta_0 = O_{P_{d,m}}(\frac{1}{\sqrt{M}})$. Hence subtracting and adding

$$\int_0^\tau (X_i - \hat{e}(\hat{\beta}, t)) Y_i(t) \frac{\exp(\beta_0' X_i)}{\hat{S}^{(0)}(\beta_0, t)} (d\hat{\eta}(t)/M)$$

to $\hat{U}_i(\hat{\beta})$ we obtain:

$$\hat{U}_i(\hat{\beta}) = \int_0^\tau (X_i - \hat{e}(\hat{\beta}, t)) \left\{ d\eta_i - Y_i(t)\exp(\beta_0' X_i) \frac{d\hat{\eta}}{M\hat{S}^{(0)}(\beta_0, t)}) \right\} + O_{P_{d,m}}(\frac{1}{\sqrt{M}}),$$

uniformly in i.

Similarly, subtracting and adding $\int_0^\tau (X_i - \hat{e}(\hat{\beta}, t)) Y_i(t)\exp(\beta_0' X_i)\lambda_0(t)dt$ we obtain:

$$\hat{U}_i(\hat{\beta}) = \int_0^\tau (X_i - \hat{e}(\hat{\beta}, t)) d\mathcal{M}_i(\beta_0, t) + \int_0^\tau (X_i - \hat{e}(\hat{\beta}, t)) Y_i(t)\exp(\beta_0' X_i)(\lambda_0(t)dt - \frac{d\hat{\eta}}{M\hat{S}^{(0)}(\beta_0, t)}) \tag{A.6.2}$$

The second term in (A.6.2) can be written as:

$$\int_0^\tau \frac{(X_i - \hat{e}(\hat{\beta}, t)) Y_i(t)\exp(\hat{\beta}' X_i)}{\hat{S}^{(0)}(\hat{\beta}, t)} d\hat{\mathcal{M}}(\beta_0, t)/M \tag{A.6.3}$$

Since the integrand in (A.6.3) is a bounded predictable process and $-\hat{\mathcal{M}}(\beta_0, t)$ is the martingale given in (A.4.1) with compensator $\hat{A}(t) = M \cdot \int_0^t \hat{S}^{(0)}(u)\lambda_0(u)du$, we use Lenglart's inequality to obtain that the stochastic integral in (A.6.3) converges to zero as $O_{P_{d,m}}(\frac{1}{\sqrt{M}})$ as $M \to \infty$, uniformly in i.

Finally the $\sqrt{M}$ - consistency of $\hat{\beta}_N$ also imply that $\sup_{0 \le t \le \tau} \left|\hat{e}(\hat{\beta}) - \hat{e}(\beta_0)\right| = O_{P_{d,m}}(\frac{1}{\sqrt{M}})$ as $M \to \infty$, hence the first term in (A.6.2) is equal to $\tilde{U}_{hik}(\beta_0) + O_{P_{d,m}}(\frac{1}{\sqrt{M}})$ and (A.6.1) is obtained.

Equation (A.6.1), design conditions C_1 and C_3 and $\lim \frac{n}{N} > 0$ yield

$$v_{Lin} = \frac{1}{M}\left\{ \sum_{i \in s}\sum_{j \in s} \frac{\Delta_{ij}}{\pi_{ij}} \tilde{U}_i(\beta_0)\tilde{U}_j(\beta_0) \right\} + O_{P_{d,m}}(\frac{1}{\sqrt{M}}). \tag{A.6.4}$$

Also, (A.4.2) imply that

$$\widetilde{U}_i(\beta_0) = \int_0^\tau (X_i - e(\beta_0,t)) d\mathcal{M}_i(\beta_0,t) + o_{P_{d,m}}(1), \quad i = 1,\ldots,M \text{ as } M \to \infty,$$

and hence we have:

$$E_{d,m}(v_{Lin}) = \frac{1}{M}\left\{\sum_{i=1}^{M}\sum_{j=1}^{M} \Delta_{ij} E_m \{U_i(\beta_0) U_j(\beta_0)\}\right\} + o(1),$$

which proves the Lemma.

Proof of Lemma 6.2:

Equation (A.6.1) is now valid with the rate $O_{P_{d,m}}(\frac{1}{\sqrt{n}})$ as $n \to \infty$. This and condition C_1 yield:

$$v_{eff} = \frac{n}{M^2}\left\{\sum_{i \in s}(\frac{1}{\pi_i} - 1)\frac{\widetilde{U}_i^2(\beta_0)}{\pi_i}\right\} + O_{P_{d,m}}(\frac{1}{\sqrt{n}}). \tag{A.6.5}$$

In order to obtain (6.10) we note that $\widetilde{U}_i(\beta_0)$ are right-continuous local square integrable martingales in $0 \le t < \infty$, and hence $\widetilde{U}_i^2(\beta_0) - \langle \widetilde{U}_i(\beta_0), \widetilde{U}_i(\beta_0)\rangle$ are right-continuous local martingales with predictable co-variation function $\langle \widetilde{U}_i(\beta_0), \widetilde{U}_i(\beta_0)\rangle = \int_0^\tau (X_i - \hat{e}(\beta_0,t))^{\otimes 2} Y_i(t)\exp\{\beta_0' X_i\}\lambda_0(t)dt$. Then the first term in (A.6.5) can be written as:

$$\frac{n}{M^2}\sum_{i \in s}(\frac{1}{\pi_i} - 1)\frac{\{\widetilde{U}_i^2(\beta_0) - \langle \widetilde{U}_i(\beta_0), \widetilde{U}_i(\beta_0)\rangle\}}{\pi_i} + \frac{n}{M^2}\sum_{i \in s}(\frac{1}{\pi_i} - 1)\frac{\langle \hat{U}_i(\beta_0), \hat{U}_i(\beta_0)\rangle}{\pi_i}. \tag{A.6.6}$$

The first term in (A.6.6) is a mean of M martingales in the product space and Lenglart's inequality together with design condition C_1 imply that it is of order $O_{P_{d,m}}(\frac{1}{\sqrt{n}})$. Equation (2.6) implies that the second term in (A.6.6) is equal to:

$$\frac{n}{M^2}\sum_{i \in s}(\frac{1}{\pi_i} - 1)\int_0^\tau (X_i - \hat{e}(\beta_0,t))^{\otimes 2} Y_i(t)\exp\{\beta_0' X_i\}\lambda_0(t)dt / \pi_i$$

$$= \frac{N}{M}\int_0^\tau \left\{\frac{n}{N}\hat{V}_\pi(\beta_0,t) - \frac{n}{N}\hat{V}_m(\beta_0,t)\right\}\hat{S}^{(0)}(\beta_0,t)\lambda_0(t)dt, \tag{A.6.7}$$

Now we revert the argument used for the consistency of $\hat{\Sigma}_\pi(\hat{\beta},t)$: equation (A.6.7) is equal to:

$$= \frac{N}{M}\int_0^\tau \left\{\frac{n}{N}\hat{V}_\pi(\beta_0,t) - \frac{n}{N}\hat{V}_m(\beta_0,t)\right\}\left\{\hat{S}^{(0)}(\beta_0,t)\lambda_0(t)dt - d\hat{\eta}(t)/M\right\} + \frac{N}{M}\left\{\hat{\Sigma}_\pi(\beta_0,t) - \frac{n}{N}\hat{\Sigma}_m(\beta_0,t)\right\}$$

and use Lenglart's inequality to obtain:

$$= O_{P_{d,m}}(\frac{1}{\sqrt{M}}) + \frac{N}{M}\left\{\hat{\Sigma}_{\pi}(\beta_0,t) - \frac{n}{N}\hat{\Sigma}_m(\beta_0,t)\right\},$$

which yields (6.10) of Lemma 6.2.

To obtain (6.11) we look at v_{resid}, recall that the $\tilde{U}_i(\beta_0)$ are uniformly bounded and apply (A.6.1) to obtain:

$$v_{resid} = \frac{n}{M^2}\sum_{i \neq j \in s}\frac{\Delta_{ij}}{\pi_{ij}}\tilde{U}_i(\beta_0)\tilde{U}_j(\beta_0) + O_{P_{d,m}}(\frac{1}{\sqrt{n}})\frac{n}{M^2}\sum_{i \neq j \in s}\frac{\Delta_{ij}}{\pi_{ij}} \tag{A.6.8}$$

Now $-1 \leq \Delta_{ij} \leq 0$ for all i and j, by the usual restriction imposed to any πps design, which implies that Condition C_3 is equivalent to

$$\frac{1}{M}\sum_{i \neq j}\left|\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}\right| = O\left(\frac{N}{n}\right), \tag{A.6.9}$$

which implies that $E_d\left(\left|\frac{n}{M^2}\sum_{i \neq j \in s}\frac{\Delta_{ij}}{\pi_{ij}}\right|\right) \leq \frac{n}{M^2}\sum_{i \neq j}\left|\Delta_{ij}\right| = O(1)$. This in turn implies that the last term in (A.6.8) is an $O_{P_{d,m}}\left(\frac{1}{\sqrt{n}}\right)$.

The design-model expectation in the first term of (A.6.8) is equal to zero, since the $\tilde{U}_i(\beta_0)$ and $\tilde{U}_j(\beta_0)$, $i \neq j$, though not stochastically independent, are martingales with predictable co-variation equal to zero. The variance of the first term in (A.6.8) is the expected value of the M^2 cross-products of terms of the form $\frac{\Delta_{ij}}{\pi_{ij}}\tilde{U}_i(\beta_0)\tilde{U}_j(\beta_0)$.

All the cross-product terms with at least one single label i different from the others have model expectation equal to zero, and the terms in $\tilde{U}_i^2(\beta_0)\tilde{U}_j^2(\beta_0)$ are uniformly bounded, so design conditions C_2, C_3 and C_4 imply that the variance of the first term in (A.6.8) is bounded by:

$$O(1)\frac{n^2}{M^4}\sum_{i \neq j \in Pop}\frac{(\Delta_{ij})^2}{\pi_{ij}} = O\left(\frac{1}{n}\right)$$

and thus $v_{resid} = O\left(\frac{1}{\sqrt{n}}\right)$ as $n \to \infty$, which is equation (6.11).

Robust versus efficient variance estimators

Keeping the classic *hik* sample label notation, consider a sample estimating equation of the form:

$$\frac{\sqrt{n}}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\left(\frac{I_{hik}(s)}{\pi_{hik}}-1\right)U_{hik}=0 \tag{A.6.10}$$

with $E_d(I_{hik}(s))=\pi_{hik}$, $E_m(U_{hik})=0$ and model independent variables U_{hik}, $k=1,\dots,N_{hi}$, $i=1,\dots,N_h$, $h=1,\dots,L$. The design-model variance of the estimating equation function in (A.6.10) is

$$V_{d,m}=E_mV_d\left\{\frac{\sqrt{n}}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{k=1}^{N_{hi}}\left(\frac{I_{hik}(s)}{\pi_{hik}}-1\right)U_{hik}\right\}=E_m\left\{\frac{n}{M^2}\sum_{hik}\sum_{\ell jm\in Pop}\Delta_{hik\ell jm}U_{hik}U_{\ell jm}\right\} \tag{A.6.11}$$

with $\Delta_{hik\ell jm}=\dfrac{\pi_{hik\ell jm}-\pi_{hik}\pi_{\ell jm}}{\pi_{hik}\pi_{\ell jm}}$, where $\pi_{hik\ell jm}$ is the joint inclusion probability of units hik and ℓjm. A design-model asymptotically unbiased estimator of $V_{d,m}$ is given by

$$v_{robust}=\frac{n}{M^2}\sum_{hik\in s}\sum_{\ell jm\in s}\frac{\Delta_{hik\ell jm}}{\pi_{hik\ell jm}}U_{hik}U_{\ell jm}$$

$$=\frac{n}{M^2}\left\{\sum_{hik\in s}\left(\frac{1}{\pi_{hik}}-1\right)\frac{U_{hik}^2}{\pi_{hik}}+\sum_{hik\neq\ell jm\in s}\frac{\Delta_{hik\ell jm}}{\pi_{hik\ell jm}}U_{hik}U_{\ell jm}\right\}. \tag{A.6.12}$$

Under the model for independence of the terms U_{hik}, for $k=1,\dots,N_{hi}$, $i=1,\dots,N_h$, $h=1,\dots,L$, the terms with joint probabilities of selection disappear in the variance $V_{d,m}$:

$$V_{d,m}=E_m\left\{\frac{n}{M^2}\sum_{hik\in Pop}\left(\frac{1}{\pi_{hik}}-1\right)U_{hik}^2+\frac{n}{M^2}\sum_{hik\neq}\sum_{\ell jm\in Pop}\Delta_{hik\ell jm}U_{hik}U_{\ell jm}\right\}$$

$$=E_m\left\{\frac{n}{M^2}\sum_{hik\in Pop}\left(\frac{1}{\pi_{hik}}-1\right)U_{hik}^2\right\}$$

since $E_m\{U_{hik}U_{\ell jm}\}=E_m\{U_{hik}\}E_m\{U_{\ell jm}\}=0$ for $hik\neq\ell jm$.

In this case, a design-model consistent estimator of the variance $V_{d,m}$ is given by

$$v_{eff}=\frac{n}{M^2}\left\{\sum_{hik\in s}\left(\frac{1}{\pi_{hik}}-1\right)\frac{U_{hik}^2}{\pi_{hik}}\right\}. \tag{A.6.13}$$

Moreover, under the independence assumption, v_{eff} is more efficient than v_{robust}. Indeed, setting

$$v_{resid}=\sum_{hik\neq\ell jm\in s}\frac{\Delta_{hik\ell jm}}{\pi_{hik\ell jm}}U_{hik}U_{\ell jm}, \tag{A.6.14}$$

we have

$$V_{d,m}(v_{robust}) = V_{d,m}(v_{eff}) + V_{d,m}(v_{resid}) \geq V_{d,m}(v_{eff}) \text{ for all } n, \tag{A.6.15}$$

since the covariance term in (A.6.15) is zero: it is the design-model expectation of a weighted sum of products of four U_{hik} functions with at least one index hik different from the others, which makes the model expectation of each term equal to zero.

Note that the main reason for the efficiency is that when the model consists of stochastically independent units, the terms with joint probabilities of selection in the variance formula disappear. This last property, though not formally expressed, has been used previously in the estimation of the variance (see for example Sudhratar and Kovacevic, 2000, for the GEE approach to the analysis of longitudinal ordinal survey data).