

# Efficiency Calculations for the Maximum Partial Likelihood Estimator in Nested-Case Control Sampling

Larry Goldstein

Haimeng Zhang

July 25, 2008

## Abstract

In making inference on the relation between failure and exposure histories in the Cox semi-parametric model, the maximum partial likelihood estimator (MPLE) of the finite dimensional parameter, and the Breslow estimator of the baseline survival function, are known to achieve full efficiency when data is available for all time on all cohort members, even when the covariates are time dependent. When cohort sizes become too large for the collection of complete data, sampling schemes such as nested case-control sampling must be used and, under various models, there exist estimators based on the same information as the MPLE having smaller asymptotic variance.

Though the MPLE is therefore not efficient under sampling in general, it approaches efficiency in highly stratified situations, or instances where the covariate values are increasingly less dependent upon the past, when the covariate distribution, not depending on the real parameter of interest, is unknown and there is no censoring. In particular, in such situations, when using the nested case-control sampling design, both the MPLE and the Breslow estimator of the baseline survival function achieve the information lower bound both in the distributional and the minimax senses in the limit as the number of cohort members tends to infinity.

## 1 Introduction

For many epidemiologic studies, the cohort from which failures are observed is simply too large for the collection of full exposure data, and in order to make inference on the connection between exposure history and failure it becomes a matter of practical necessity to sample. For a cohort followed over time, one of the simplest sampling schemes, termed nested case-control sampling [16], is to choose a fixed number of controls to compare to the failure at each failure time. Though it has previously been shown that the maximum partial likelihood estimator (MPLE) in the Cox semi-parametric model achieves full efficiency when data is available for all time on all cohort members, the same is no longer true in certain situations when schemes such as nested case-control sampling are used. In counterpoint to such cases, here we explore a model where the MPLE is efficient, in both the distributional and minimax senses, for the nested case-control sampling scheme. We also show that similar remarks apply as well to the Breslow estimator of the baseline hazard. Knowing

---

<sup>0</sup>AMS 2000 subject classifications. Primary 62N01, 62N02, 62B99

<sup>0</sup>Key words and phrases: Information bound; semi-parametric models; highly stratified

in which situations the MPLE is close to efficient provides some guidelines on when it may be applied with little risk of efficiency loss, and when other estimators, perhaps depending on additional modeling assumptions, should be considered as an alternative.

In the standard Cox model [6], a common but unspecified baseline hazard function  $\lambda(t)$  is assumed to apply to all cohort members. The relation between exposure and failure is the one of most interest, and is modeled by the real parameter  $\theta$  specifying the increased relative risk, having the exponential form  $e^{\theta Z}$ , say, for an individual with covariate  $Z$ . The unknown baseline is considered for the most part to be a nuisance parameter. When covariate information is available on all cohort members, the maximum partial likelihood estimator (MPLE) makes inference on the parametric component of such models by maximizing a ‘partial likelihood’, that is, the product of the conditional probabilities, over all failures  $i_j$ , that individual  $i_j$  failed given that the individuals  $\mathcal{R}_{i_j}$  were also at risk to fail when  $i_j$  failed,

$$L(\theta) = \prod_{i_j} \frac{e^{\theta Z_{i_j}}}{\sum_{k \in \mathcal{R}_{i_j}} e^{\theta Z_k}}. \quad (1)$$

We note that the unspecified baseline hazard cancels upon forming this conditional probability.

When data is only available on some sampled subset  $\tilde{\mathcal{R}}_{i_j}$  of the entire cohort  $\mathcal{R}_{i_j}$ , an estimator may be formed by replacing  $\mathcal{R}_{i_j}$  by  $\tilde{\mathcal{R}}_{i_j}$ , (see [5]), possibly then mandating the use of weights so that the MPLE remains consistent. Nested case control sampling, which does not require the use of such weights, is the instance where  $\tilde{\mathcal{R}}_{i_j}$  consists of the failure  $i_j$  and  $m - 1$  non-failed individuals to serve as controls, chosen uniformly at random for those at risk at the time of the failure.

One price to pay for the ability to estimate  $\theta$  while leaving the nonparametric baseline hazard unspecified, and the subsequent use of the MPLE, is that it is not a true likelihood being maximized, and efficiency concerns arise. In particular, it is not clear whether one can construct estimators which depend on the same data as the MPLE, but which have better performance. In the paper of Begun et al. [1], however, these concerns are put to rest in the full cohort case where the covariates are time fixed, as the authors demonstrate that in that situation the MPLE achieves the semi-parametric efficiency bound. Greenwood and Wefelmeyer [10] show the MPLE is efficient in the full cohort situation even when the covariates are allowed to depend on time. Similar remarks apply also to the Breslow estimator of the baseline hazard.

The situation is different under sampling: Robins et al. [15] has shown that for time fixed covariates the MPLE is not efficient under nested case control sampling. In this situation, there may exist modified estimators which take advantage of the time fixed nature of the covariates, in that the exposure for a control sampled in the past is still valid at a future failure time. In time varying covariate models, Chen [7], among others, have modified the MPLE to yield consistent estimators of the parametric parameter which have smaller asymptotic variance than the MPLE. The estimator proposed in [7] uses covariates sampled for other failures at time points near to that of a given failure to take advantage of already available information. Here, to realize a practical efficiency benefit, the sequence of failures times must be sufficiently dense and the covariates not varying too rapidly in time. Though in the time fixed covariate situation the modified estimator uses information from the past specifically, in both cases one relies on the dependence of the covariate values over time to realize some efficiency gain; for the time varying covariate models, such modified estimators will perform better the stronger the time dependence. Due to the various improvements on the performance of the MPLE, it becomes less clear in just which ways its performance can be improved, or, in other words, whether the MPLE fails to be efficient for reasons in addition to the ones by which these modified estimators achieve their gains.

Showing that there is some sense in which the MPLE for nested case control sampling is efficient is therefore valuable for two reasons. First, it limits the scope of the search for estimators which might improve the MPLE's performance. Secondly, it indicates the use of the simple MPLE, and not a more complex version of same, in situations which achieve or approximate those in which it can not be improved.

Based on the known instances where the MPLE fails to be efficient under sampling, to find models where it is, by contrast, efficient, we are led to consider situations where covariate information collected for one failure is not useful at any other failure time. Indeed, such situations are fairly common in epidemiologic studies, in particular, when highly stratified cohorts are followed over a short period of time. Due to the short time under study, the covariates may be considered time fixed, and there is, for that same reason, little or no censoring. Lastly, in such cases, the groups corresponding to the terms in the product of the partial likelihood are independent, or very nearly so. A continuous time covariate model where the failures are spaced far apart relative to the correlation time of the covariates will also have the property that the covariate values at one failure time will be nearly independent of those at any other. In fact, in the limit, this latter situation becomes the former, highly stratified case. Thus we are led to a time fixed covariate model  $f$  having no censoring, where we observe  $n$  independent units of information, each consisting of the observed failure from a cohort of a possibly random number  $\eta$  of individuals who are comparable to the failure, the covariate value of the failure, and the covariate values of  $m - 1$  sampled controls.

A concrete example of such a situation is the study of occupational exposure to EMF and leukemia [11], which is fairly typical of cancer registry based case-control studies. The cohort is the adult male population in mid-Sweden followed over 1983-1987 for cases of leukemia. Two controls were sampled from risk sets based on the age of the 250 leukemia cases, matching on year of birth and geographic location. In this study, with the four year follow-up and fine stratification, there is little censoring and almost all strata have at most one failure. Thus the sampling model considered here very closely approximates the circumstances of the study.

It is easy to verify that in these situations, letting  $Z$  be the distribution of the i.i.d. covariates, under the null  $\theta_0 = 0$  the information  $-E[\partial^2 \log L(\theta)/\partial\theta^2] = \sigma_{\text{MPLE}}^{-2}$ , where

$$\sigma_{\text{MPLE}}^{-2} = \left( \frac{m-1}{m} \right) \text{Var}(Z)$$

where  $L(\theta)$  is as in (1) with the set of those at risk  $\mathcal{R}_{i_j}$  replaced by the nested case-control sampled risk set  $\tilde{\mathcal{R}}_{i_j}$ . Hence, under regularity (see e.g. [8], [5], and [9]) the MPLE  $\hat{\theta}_n$  is asymptotically normal and satisfies

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \rightarrow \mathcal{N}(0, \sigma_{\text{MPLE}}^2).$$

Our main result, Theorem 2.5, shows that when considering a growing cohort size, the limiting effective information in the data,  $I_*(\theta_0)$ , equals  $\sigma_{\text{MPLE}}^{-2}$ , and that the MPLE is efficient in the limit in both the convolution lower bound, and minimax, senses. Theorem 2.6 show similar remarks apply to the Breslow estimator of the baseline survival function.

When the complete set of covariate values are observed it is unimportant whether the covariate distribution is considered known or unknown. Again, the situation when sampling is different; knowing the covariate distribution allows one to estimate large sample quantities to within some accuracy. Consequently, the hypotheses of Theorem 2.5 includes the assumption that the covariate distribution is unknown, and the subsequent analysis must therefore handle two infinite dimensional nuisance parameters, one for the unknown baseline density, the other for the unknown covariate

distribution. In particular, the results leave open the possibility of improved estimators which take advantage of a known covariate distribution. Nevertheless, such improvements must necessarily depend on having information about, and correctly modeling, the covariate distribution, and consequently invite the possibility of bias due to modeling misspecification.

We consider the Cox model under the usual exponential relative risk, though the methods here may be applied for other relative risk forms, as was accomplished in [10] for the full cohort, time varying covariate model. The methods here also extend to accommodate censoring, though this generalization requires the inclusion of a third infinite dimensional parameter, the censoring density, and consequently the handling of an additional operator corresponding to the unknown censoring density.

The outline of this work is as follows. In Section 2.1 we review and slightly modify the theory in [1] for the calculation of information bounds in semi-parametric models to accommodate a pair of unknown densities. In Section 2.2 we further specialize that theory to the case at hand and formally state our model and the main results which were outlined above. Application of the theory presented in Section 2.1 for the relative risk parameter  $\theta$  requires verification of three assumptions. The first, Assumption 2.1 is that certain collections of perturbations form a subspace. The second, Assumption 2.2, is connected to the Hellinger differentiability of the observation density  $f$ , in particular, that perturbations of the nonparametric baseline and covariate density affect  $f$  by amounts given by operators  $A$  and  $B$  evaluated on the respective perturbations, and that perturbing the parametric parameter results in a score  $\rho_0$ . The third, Assumption 2.3, is that the orthogonal projection of the parametric score  $\rho_0$  is contained in a certain subspace,  $\mathbb{K}$ . In order to proceed as quickly as possible to the calculation of the information bounds in Section 4, we present in Section 3 only a subset of the properties eventually required of the operators  $A$  and  $B$  and of the score  $\rho_0$ ; the remaining properties required of  $A$  and  $B$  are shown in Sections 5 and 6 respectively. The verification of Assumption 2.2 can be found in Section 7. Assumptions 2.1 and 2.3 are verified in Section 8. Section 9 is a technical appendix which highlights the modifications made to the theory in [1] necessary for our application, and provides the proofs required for our purposes.

## 2 Information Bounds for Sampling in the Cox Model

In Section 2.1 we review and adapt the framework of [1] for the calculation of information bounds in semi-parametric models to the case where there are two unknown one-dimensional density functions. In Section 2.2 we specify the model  $f$  for nested case-control sampling and formally state our main result showing that the MPLE, and the Breslow estimator, achieve their respective efficiency lower bounds.

### 2.1 Information Bounds in Semi-parametric models

This section closely follows the treatment in [1] for deriving lower bounds for estimation in semi-parametric models; see also the text [4]. Let  $L^2(\mu)$  denote the collection of functions which are square integrable with respect to a measure  $\mu$ , and for  $u, v \in L^2(\mu)$  we let  $\langle u, v \rangle_\mu = \int uv d\mu$  and  $\|u\|_\mu^2 = \langle u, u \rangle_\mu$ . Here, as in [1], the data consists of  $n$  i.i.d. observations  $X_1, \dots, X_n$  taking values in a measurable space  $(\mathcal{X}, \mathcal{F}_\mathcal{X})$ , and the density function  $f$  of a single observation is with respect to a sigma-finite measure  $\sigma$ . We consider a model where the density  $f = f(\cdot, \theta, g, h)$  is determined by a real parameter  $\theta$ , the one of most interest, and by the infinite dimensional parameter  $p = (g, h)$ ,

a vector of two unknown densities  $g$  and  $h$ , the baseline failure time density, and the marginal covariate density, respectively.

Let  $\mathcal{D}^+$  and  $\mathcal{D}$  denote the collection of densities with respect to Lebesgue measure  $\nu^+$  and  $\nu$  on  $\mathbb{R}^+ = [0, \infty)$  and  $\mathbb{R}$ , respectively. We let the parameter space  $\mathcal{G}$  for the unknown baseline failure density be

$$\mathcal{G} = \mathcal{D}^+. \quad (2)$$

To impose growth conditions on the covariates similar to the ones typically assumed, for a covariate density  $\mathbf{h} : \mathbb{R} \rightarrow [0, \infty)$  and  $\theta \in \mathbb{R}$  let

$$M_{\mathbf{h}}(\theta) = \int \mathbf{h} d\nu_{\theta} \quad \text{where} \quad \frac{d\nu_{\theta}}{d\nu} = e^{\theta z}.$$

For some fixed  $\xi > 0$  and  $0 < \theta_{\xi} < \theta_{\kappa}$  we let the parameter space for the covariate density be

$$\mathcal{H} = \{\mathbf{h} \in \mathcal{D} : M_{\mathbf{h}}(\theta) < \infty \text{ for all } |\theta| < \theta_{\kappa} \text{ and } M_{\mathbf{h}}(\theta_{\xi}) + M_{\mathbf{h}}(-\theta_{\xi}) < \xi\}. \quad (3)$$

Hence, the parameter space  $\mathcal{P}$  for the pair  $p$  of unknown densities is given by

$$\mathcal{P} = \mathcal{G} \times \mathcal{H}.$$

Adopting slightly inconsistent notation for the sake of ease, we let  $\theta_0$  denote the null parameter in  $\mathbb{R}$ , and henceforth,  $g$  and  $h$  the null parameters in  $\mathcal{G}$  and  $\mathcal{H}$ , respectively; we label them also as  $g_0$  and  $h_0$  when convenient.

For  $\tau \in \mathbb{R}$  let  $\Theta(\tau)$  denote the collection of all real sequences  $\{\theta_n\}_{n \geq 1}$  such that

$$|\sqrt{n}(\theta_n - \theta_0) - \tau| \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad \text{and set} \quad \Theta = \bigcup \{\Theta(\tau) : \tau \in \mathbb{R}\}.$$

Let  $\Pi_{\theta} = L^2(\nu^+) \times L^2(\nu_{\theta})$ , and for  $\gamma = (\alpha, \beta) \in \Pi_{\theta}$  let  $\|\gamma\|_{\Pi_{\theta}} = \max\{\|\alpha\|_{\nu^+}, \|\beta\|_{\nu_{\theta}}\}$ , the product metric, and with  $p = (g, h)$  the null parameter let  $\mathcal{C}(p, \gamma)$  be the collection of all sequences  $\{p_n\}_{n \geq 0} = \{(g_n, h_n)\}_{n \geq 0} \subset \mathcal{P}$  such that

$$\|\sqrt{n}(p_n^{1/2} - p^{1/2}) - \gamma\|_{\Pi_{\theta}} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad \text{for all } |\theta| < \theta_{\kappa}. \quad (4)$$

Let  $\Gamma$  be the set of all  $\gamma$  such that (4) holds for some  $\{p_n\}_{n \geq 0} \subset \mathcal{P}$ , and

$$\mathcal{C}(p) = \bigcup_{\gamma \in \Gamma} \mathcal{C}(p, \gamma).$$

By considering the components of  $\{p_n\}_{n \geq 0}$  we see that  $\{g_n\}_{n \geq 0} \in \mathcal{C}_1(g, \alpha)$ , the collection of all sequences in  $\mathcal{G}$  such that

$$\|\sqrt{n}(g_n^{1/2} - g^{1/2}) - \alpha\|_{\nu^+} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and therefore  $\alpha \in L^2(\nu^+)$  satisfies  $\alpha \perp g^{1/2}$  in  $L^2(\nu^+)$ , that is,  $\langle \alpha, g^{1/2} \rangle_{\nu^+} = 0$ , or,

$$\int_0^{\infty} g^{1/2} \alpha d\nu^+ = 0. \quad (5)$$

Now let

$$\mathcal{A} = \{\alpha \in L^2(\nu^+) : \text{there exists } \{g_n\}_{n \geq 0} \subset \mathcal{G} \text{ such that } \|\sqrt{n}(g_n^{1/2} - g^{1/2}) - \alpha\|_{\nu^+} \rightarrow 0\} \quad (6)$$

and set

$$\mathcal{C}_1(g) = \bigcup_{\alpha \in \mathcal{A}} \mathcal{C}_1(g, \alpha).$$

Similarly,  $\{h_n\}_{n \geq 0} \subset \mathcal{C}_2(h, \beta)$ , the collection of all sequences in  $\mathcal{H}$  such that

$$\|\sqrt{n}(h_n^{1/2} - h^{1/2}) - \beta\|_{\nu_\theta} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \text{ for all } |\theta| < \theta_\kappa. \quad (7)$$

For  $\theta = 0$  (7) yields

$$\|\sqrt{n}(h_n^{1/2} - h^{1/2}) - \beta\|_{\nu} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (8)$$

and therefore that  $\beta$  satisfies

$$\int_{-\infty}^{\infty} h^{1/2} \beta d\nu = 0. \quad (9)$$

Now let  $\mathcal{B}$  be the collection of all  $\beta \in L^2(\nu)$  such that there exists  $\{h_n\}_{n \geq 0} \subset \mathcal{H}$  such that

$$\|\sqrt{n}(h_n^{1/2} - h^{1/2}) - \beta\|_{\nu_\theta} \rightarrow 0 \quad \text{for all } |\theta| < \theta_\kappa, \quad (10)$$

and set

$$\mathcal{C}_2(h) = \bigcup_{\beta \in \mathcal{B}} \mathcal{C}_2(h, \beta).$$

Clearly

$$\mathcal{C}(p, \gamma) = \mathcal{C}_1(g, \alpha) \times \mathcal{C}_2(h, \beta), \quad \mathcal{C}(p) = \mathcal{C}_1(g) \times \mathcal{C}_2(h) \quad \text{and} \quad \Gamma = \mathcal{A} \times \mathcal{B}.$$

The following three assumptions will be needed to demonstrate Theorems 2.1 and 2.2, and, in addition, the fourth for Theorems 2.3 and 2.4. The first is that  $\Gamma$  is a subspace of  $L^2(\nu^+) \times L^2(\nu)$ , or equivalently,

**Assumption 2.1.** *The sets  $\mathcal{A}$  and  $\mathcal{B}$  are subspaces of  $L^2(\nu^+)$  and  $L^2(\nu)$ , respectively.*

It is shown in [1] that parts of the following assumption are a consequence of the Hellinger differentiability of  $f$ ; we verify Assumption 2.2 directly.

**Assumption 2.2.** *There exists  $\rho_\theta \in L^2(\sigma)$  and linear operators  $A : L^2(\nu^+) \rightarrow L^2(\sigma)$  and  $B : L^2(\nu) \rightarrow L^2(\sigma)$  such that for any  $(\tau, \alpha, \beta) \in \mathbb{R} \times \mathcal{A} \times \mathcal{B}$  and*

$$(\{\theta_n\}_{n \geq 0}, \{g_n\}_{n \geq 0}, \{h_n\}_{n \geq 0}) \in \Theta(\tau) \times \mathcal{C}_1(g, \alpha) \times \mathcal{C}_2(h, \beta), \quad (11)$$

*the sequence of densities given by  $f_n = f(\cdot, \theta_n, g_n, h_n)$  for  $n = 0, 1, \dots$  satisfies*

$$\|\sqrt{n}(f_n^{1/2} - f_0^{1/2}) - \zeta\|_{\sigma} \rightarrow 0 \quad \text{for } \zeta = \tau \rho_\theta + A\alpha + B\beta \text{ as } n \rightarrow \infty. \quad (12)$$

Let

$$\mathbb{H} = \{\zeta \in L^2(\sigma) : \zeta = \tau\rho_\theta + A\alpha + B\beta \text{ for some } \tau \in \mathbb{R}, \alpha \in \mathcal{A}, \beta \in \mathcal{B}\} \quad (13)$$

and

$$\mathbb{K} = \{\delta \in L^2(\sigma) : \delta = A\alpha + B\beta \text{ for some } \alpha \in \mathcal{A} \text{ and } \beta \in \mathcal{B}\}. \quad (14)$$

The classical projection theorem shows that the orthogonal projection of  $\rho_\theta$  onto the closure of  $\mathbb{K}$  is an element of the closure of  $\mathbb{K}$ . However, we consider situations satisfying the following assumption, that is, where  $\mathbb{K}$  itself contains the projection of  $\rho_\theta$ .

**Assumption 2.3.** *There exists  $\hat{\alpha} \in \mathcal{A}$  and  $\hat{\beta} \in \mathcal{B}$  such that  $\hat{\delta} = A\hat{\alpha} + B\hat{\beta}$  satisfies*

$$\rho_\theta - \hat{\delta} \perp \delta \text{ for all } \delta \in \mathbb{K}.$$

Since for any  $\delta = A\alpha + B\beta \in \mathbb{K}$ , by orthogonality,

$$\begin{aligned} \|\rho_\theta - \delta\|_\sigma^2 &= \|\rho_\theta - A\alpha - B\beta\|_\sigma^2 \\ &= \|\rho_\theta - \hat{\delta} - A(\alpha - \hat{\alpha}) - B(\beta - \hat{\beta})\|_\sigma^2 \\ &= \|\rho_\theta - \hat{\delta}\|_\sigma^2 + \|A(\alpha - \hat{\alpha}) + B(\beta - \hat{\beta})\|_\sigma^2 \\ &\geq \|\rho_\theta - \hat{\delta}\|_\sigma^2, \end{aligned}$$

hence  $\hat{\delta}$  minimizes  $\|\rho_\theta - \delta\|_\sigma^2$  over  $\delta \in \mathbb{K}$ , and thus corresponds to the worst case direction of approach to the null, that is, the one which minimizes the available information. Set the effective information  $I_*$  to be

$$I_* = 4\|\rho_\theta - \hat{\delta}\|_\sigma^2. \quad (15)$$

For  $\zeta \in \mathbb{H}$ , let  $\mathcal{F}(f, \zeta)$  be the collection of all sequences  $\{f_n\}_{n \geq 0}$  such that (12) holds, and  $\mathcal{F}(f)$  the union of  $\mathcal{F}(f, \zeta)$  over all  $\zeta \in \mathbb{H}$ . We say that an estimator  $\hat{\theta}_n$  of  $\theta_0$  is *regular* at  $f = f(\cdot, \theta_0, g, h)$  if for every sequence  $f_n(\cdot, \theta_n, g_n, h_n)$  with  $\{\theta_n\}_{n \geq 0}$ ,  $\{g_n\}_{n \geq 0}$  and  $\{h_n\}_{n \geq 0}$  as in (11), the distribution of  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  converges in distribution to  $\mathcal{L} = \mathcal{L}(f)$  which depends on  $f$  but not on the particular sequence  $f_n$ .

The setup above differs in two ways from that in [1]. First, the model considered here has two nonparametric components,  $g$  and  $h$ , while in [1] only one nonparametric component is considered. Secondly, as we specify the parameter space  $\mathcal{H}$  on the covariate density  $h$  in such a way as to accommodate more relaxed integrability conditions, the resulting space of perturbations  $\mathcal{B}$  is expressed as the intersection of subspaces (see (77)), one for each  $\theta$  in  $(-\theta_\kappa, \theta_\kappa)$ . This is so as the perturbations  $\beta$  are required to be limiting approximations to  $\sqrt{n}(h_n^{1/2} - h^{1/2})$  in  $L^2(\nu_\theta)$  for all  $|\theta| < \theta_\kappa$ , rather than in  $L^2(\nu)$ . As (7) implies (8) our condition gives rise to a smaller collection  $\mathcal{B}$  of perturbations than in [1]. Nevertheless, only minimal adaptations of the proofs of Theorems 3.1 and 3.2, and Theorems 4.1 and 4.2, of [1] are required to demonstrate Theorems 2.1 through 2.4 for our model, so these are relegated to Section 9.

**Theorem 2.1.** *Suppose that  $\hat{\theta}_n$  is a regular estimator of  $\theta_0$  in the model  $f = f(\cdot, \theta, g, h)$  with limit law  $\mathcal{L} = \mathcal{L}(f)$  and that Assumptions 2.1, 2.2 and 2.3 hold. Then  $\mathcal{L}$  is the convolution of a normal  $\mathcal{N}(0, 1/I_*)$  distribution with a distribution depending only on  $f$ , where  $I_*$  is given by (15).*

We may also adapt the asymptotic minimax result of [1]. Recall that we say a loss function  $\ell : \mathbb{R} \rightarrow \mathbb{R}^+$  is subconvex when  $\{x : \ell(x) \leq y\}$  is closed, convex, and symmetric for every  $y \geq 0$ . We will also assume our loss function satisfies

$$\int_{-\infty}^{\infty} \ell(z)\phi(az)dz < \infty \quad \text{for all } a > 0, \quad (16)$$

where  $\phi$  denotes the standard normal density function.

**Theorem 2.2.** *Suppose Assumptions 2.1, 2.2 and 2.3 hold, that  $\ell$  is subconvex and satisfies (16), and for  $c \geq 0$  let*

$$B_n(c) = \{f_n \in \mathcal{F} : \sqrt{n}\|f_n^{1/2} - f^{1/2}\|_\sigma \leq c\}. \quad (17)$$

Then

$$\lim_{c \rightarrow \infty} \liminf_{n \rightarrow \infty} \sup_{\theta_n} \sup_{f_n \in B_n(c)} E_{f_n} \ell(\sqrt{n}(\hat{\theta}_n - \theta_n)) \geq E\ell(Z_*), \quad (18)$$

where  $Z_* \sim \mathcal{N}(0, 1/I_*)$  where  $I_*$  is given by (15).

The infimum in (18) is taken over the class of “generalized procedures,” the closure of the class of randomized Markov kernel procedures (see [14], page 235). We also obtain lower bounds on the performance of regular estimators of the baseline survival function  $\bar{G}(\cdot)$  by similarly adapting Theorem 4.1 and Theorem 4.2 of [1], under the following assumption.

**Assumption 2.4.** *The linear operator  $A^*A : L^2(\nu^+) \rightarrow L^2(\nu^+)$  is invertible with bounded inverse  $(A^*A)^{-1}$ .*

We suppose also that, perhaps by a suitable map such as the probability integral transformation, the density  $g$  is supported on  $[0, 1]$ . Let

$$G_s = (I_{[0,s]} - G(s))g(s)^{1/2},$$

and define the covariance functions

$$K(s, t) = \langle G_s, (A^*A)^{-1}G_t \rangle_{\nu^+} \quad \text{and} \quad K_*(s, t) = K(s, t) + 4I_*^{-1} \int_0^s \hat{\alpha}g^{1/2} \int_0^t \hat{\alpha}g^{1/2}, \quad (19)$$

where  $I_*$  is given by (15) and  $\hat{\alpha}$  is as in Assumption 2.3. For the precise definition of a regular estimator of  $G(\cdot)$ , analogous to that for estimators of  $\theta_0$ , see [1].

**Theorem 2.3.** *Suppose that  $\widehat{G(\cdot)}_n$  is a regular estimator of  $G(\cdot) = \int_0^\cdot g d\nu^+$  in the model  $f = f(\cdot; \theta, g, h)$  with limit process  $\mathbb{S}$ , that Assumptions 2.2 to 2.4 hold, and that Assumption 2.1 holds with  $\mathcal{A}$  given by  $\{\alpha \in L^2(\nu^+) : \int \alpha g^{1/2} d\nu^+ = 0\}$ . Then*

$$\mathbb{S} =_d \mathbb{Z}_* + \mathbb{W},$$

where  $\mathbb{Z}_*$  is a mean zero Gaussian process with covariance function  $K_*(s, t)$  given by (19) and the process  $\mathbb{W}$  is independent of  $\mathbb{Z}_*$ .

For the local asymptotic minimax bound, we let  $\ell : C[0, 1] \rightarrow \mathbb{R}^+$  be a subconvex loss function, such as  $\ell(x) = \sup_t |x(t)|$ ,  $\ell(x) = \int |x(t)|^2 dt$ , or  $\ell(x) = \mathbf{1}(x : \|x\| \geq c)$ .



**Theorem 2.4.** *Suppose the hypotheses of Theorem 2.3 are satisfied, that  $\ell$  is subconvex, and that  $B_n(c)$  is as in (17). Then*

$$\lim_{c \rightarrow \infty} \liminf_{n \rightarrow \infty} \inf_{\widehat{G}(\cdot)_n} \sup_{f_n \in B_n(c)} E_{f_n} \ell(\sqrt{n}(\widehat{G}(\cdot)_n - G_n)) \geq E\ell(\mathbb{Z}_*), \quad (20)$$

where  $\mathbb{Z}_*$  is the mean zero Gaussian process with covariance  $K_*(s, t)$  given by (19).

The infimum over estimators  $\widehat{G}(\cdot)_n$  in (20) is taken over the class of “generalized procedures” as in [14], page 235. The proofs of Theorems 2.1 to 2.4 in the Appendix detail the modifications required for the application of the methods of [1] to the case at hand.

## 2.2 Main Results

We now specify our model  $f$  for the nested case-control sampling of  $m - 1$  controls for the failure in each group. For any integer  $k$ , let  $[k] = \{1, \dots, k\}$ , and for any set  $S$  let  $\mathcal{P}_k(S)$  be the collection of all subsets of  $S$  of size  $k$ . Groups of individuals of size  $\eta \geq m$  are observed up to the time of the first failure, at which point covariates are collected on a simple random sample of  $m - 1$  non-failed individuals, and the failure.

An observation  $X = (\eta, i, r, t, z_r)$  consists of the group size  $\eta$ , the identity  $i \in [\eta]$  of the failed individual, the group  $r \subset [\eta]$  of the  $m$  individuals whose covariates are collected, the time  $t$  of the failure, and the covariates  $z_r = \{z_j, j \in r\}$ . In particular,  $X$  takes values in the space

$$\mathcal{X} = \bigcup_{\eta \geq m} \{\eta \times [\eta] \times \mathcal{P}_m([\eta]) \times \mathbb{R}^+ \times \mathbb{R}^m\}$$

which we endow with the  $\sigma$ -finite product measure

$$\sigma = (\text{counting measure}) \times (\text{counting measure}) \times (\text{counting measure}) \times \nu^+ \times \nu^m.$$

To begin the specification of the density  $f$  of the observations, corresponding to the baseline survival density  $g$  on  $\mathbb{R}^+$  are the baseline survival and hazard functions, for  $t \geq 0$ , given by, respectively

$$\overline{G}(t) = \int_t^\infty g(u) du \quad \text{and} \quad \lambda(t) = \begin{cases} g(t)\overline{G}^{-1}(t) & \text{for } \overline{G}(t) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Under the assumed standard exponential relative risk form, the hazard function  $\lambda(t; z)$  for an individual with covariate value  $z$  is the baseline hazard scaled by the factor  $\exp(\theta z)$ , that is,  $\lambda(t; z) = \exp(\theta z)\lambda(t)$ , resulting in survival and density functions, respectively, of

$$\overline{G}_\theta(t; z) = \overline{G}^{e^{\theta z}}(t) \quad \text{and} \quad g_\theta(t; z) = \begin{cases} e^{\theta z} g(t)\overline{G}^{e^{\theta z}-1}(t) & \text{for } \overline{G}(t) > 0 \\ 0 & \text{otherwise;} \end{cases} \quad (21)$$

we note  $g_0(t; z) = g_\theta(t; 0) = g(t)$ . As the marginal covariate density is  $h$ , the survival function  $\overline{G}_\theta(t; z)$  averaged over individuals with covariate density  $h(z)$  results in the (mixture) survival function

$$\overline{G}_\theta(t) = \int \overline{G}_\theta(t; z) h(z) dz \quad (22)$$

for individuals whose covariates are not observed.

The group size  $\eta$  may vary from strata to strata, and we assume it to be random with distribution, say,  $\varrho$ . At the time  $t$  of the failure of individual  $i$ , a simple random sample of size  $m - 1$  is taken from the non-failures to serve as controls. Hence, when the group size is  $\eta$ , and the identity of the failure  $i$ , the probability that the set  $r \subset [\eta]$  is selected is given by

$$K_{\eta,m} = \binom{\eta - 1}{m - 1}^{-1}$$

for any set  $r$  of size  $m$  containing  $i$ . We assume that the individuals in  $[\eta]$  are independent, and therefore the density of the sampled covariates  $z_r$  is the product

$$h(z_r) = \prod_{j \in r} h(z_j).$$

Putting all the factors together, the density for  $X = (\eta, i, r, t, z_r)$  is given by

$$\begin{aligned} f(X; \theta, g, h) &= K_{\eta,m} e^{\theta z_i} g(t) \overline{G}(t)^{\sum_{j \in r} e^{\theta z_j} - 1} \overline{G}_\theta(t)^{\eta - m} h(z_r) \varrho(\eta) \\ &= K_{\eta,m} g(t; z_i) \left[ \prod_{j \in r \setminus \{i\}} \overline{G}(t; z_j) \right] \overline{G}_\theta(t)^{\eta - m} h(z_r) \varrho(\eta). \end{aligned} \quad (23)$$

For the sake of clarify or brevity, the density may be written with either its parameters or its variables suppressed, that is, as  $f(\eta, i, r, t, z_r)$  or  $f(\theta, g, h)$ , respectively. At the null, (23) reduces to

$$f(X; \theta_0, g, h) = K_{\eta,m} g(t) \overline{G}(t)^{\eta - 1} h(z_r) \varrho(\eta), \quad (24)$$

which, in agreement with the notation introduced in Section 2.1, may appear in the abbreviated form  $f_0$ . We may take the distribution  $\varrho$  of  $\eta$  as known when proving Theorem 2.5 since the MPLE is computed without knowledge of  $\varrho$  and already achieves the bound (26) in the limit.

We are now ready to state our main result regarding the estimation of the parametric component of the model.

**Theorem 2.5.** *Suppose that  $\eta \geq 2$  almost surely,  $E[\eta^5] < \infty$ , and at least one of the following conditions is satisfied:*

- i. Positivity: The parameter space  $\Theta = [0, \infty)$ , the covariates  $Z$  take on nonnegative values, and  $\eta \geq m$  almost surely.*
- ii. Boundedness: The covariates  $Z$  are bounded and  $\eta \geq m$  almost surely.*
- iii. Cohort size:  $1 \leq m \leq \eta - 4$  almost surely.*

*Then Theorems 2.1 and 2.2 obtain for the nested case control model given in (23) with effective information*

$$I_*^\varrho(\theta_0) = \text{Var}(Z) \left( 1 - \frac{1}{m} \right) + m \text{Var}(Z) \left( 2 \text{Var} \left( \frac{1}{\eta} \right) + \left[ E \left( \frac{1}{\eta} \right) \right]^2 \right). \quad (25)$$

In particular, under any of the above three scenarios, if  $\varrho_n$  is a sequence of distributions such that  $\eta_n \rightarrow_p \infty$  when  $\eta_n$  has distribution  $\varrho_n$ , then

$$I_*(\theta_0) = \lim_{n \rightarrow \infty} I_*^{\varrho_n}(\theta_0) = \text{Var}(Z) \left( \frac{m-1}{m} \right), \quad (26)$$

and hence the Cox MPLE is efficient for the limiting nested case-control model.

The situation where there is full cohort information is covered by the special case  $P(\eta = m) = 1$ , for which (25) reduces to the lower bound  $\text{Var}(Z)$ , recovering the result of [1] for the case of no censoring. The positivity condition is the one assumed in [4] when considering the Cox model. That one of the three conditions above is satisfied is used in the proof of Lemma 7.8, and that  $\eta \geq 2$  in Lemma 7.12.

Next, we consider lower bounds for the estimation of the non-parametric component of the model. It is shown in [9] that the Breslow estimator of the baseline survival is asymptotically normal with covariance function

$$\omega(s, t) = \bar{G}(t)\bar{G}(s) \left( \int_0^{s \wedge t} \frac{dG}{E[\eta \bar{G}(u)^{\eta+1}]} + [E(Z)]^2 (\log \bar{G}(t) \log \bar{G}(s)) [I_*(\theta_0)]^{-1} \right) \quad (27)$$

where  $I_*(\theta_0)$  is given in (26).

**Theorem 2.6.** *Let the hypotheses of Theorem 2.5 be satisfied. Then on any interval  $[0, T_0]$  for which  $\bar{G}(T_0) > 0$ , Theorems 2.3 and 2.4 obtain, with*

$$K_*(s, t) = \bar{G}(t)\bar{G}(s) \left( \int_0^{s \wedge t} \frac{dG}{E[\eta \bar{G}(u)^{\eta+1}]} + [E(Z)]^2 (\log \bar{G}(t) \log \bar{G}(s)) [I_*^e(\theta_0)]^{-1} \right). \quad (28)$$

By (26) and (27), we see that the Breslow estimator becomes asymptotically efficient as the cohort size increases, under the nested case control model considered.

Theorem 2.5 follows from Theorems 2.1 and 2.2. The application of these theorems is a consequence of Theorem 4.1, which provides the effective information  $I_*^e(\theta_0)$ , and the verification of Assumptions 2.1, 2.2 and 2.3. Lemma 8.1 shows that Assumption 2.1 is satisfied with  $\mathcal{A}$  and  $\mathcal{B}$  given by (77). The verification of Assumption 2.2 is somewhat involved. The relevant quantities,  $A, B, \hat{\alpha}, \hat{\beta}$  and  $\rho_0$ , are given in (30), (31), Lemma 3.2, Lemma 3.4, and (29) respectively. The remainder of the verification of Assumption 2.2, that is, the convergence to zero in (12), is shown in Lemma 3.1. Assumption 2.3, following in a fairly straightforward manner from (77), is proved in Lemma 8.2.

Theorem 2.6 follows similarly from Theorems 2.3 and 2.4. In addition to Assumptions 2.1, 2.2 and 2.3, the application of these theorems follow from Theorem 4.2, which verifies the covariance lower bound (28), Lemma 5.2, from which Assumption 2.4 on  $[0, T_0]$  follows easily, and (77), which shows that  $\mathcal{A}$  is of the form required by Theorem 2.3. Regarding the restriction of the result to  $[0, T_0]$ , see Example 4 in [1], page 450 in particular, and the proof of Lemma 2 in [17].

### 3 Operators $A$ and $B$ : Properties

The following lemma provides the parametric score  $\rho_0$  and the operators  $A$  and  $B$  required by Assumption 2.2 and needed for the computation of the effective information  $I_*$  in (15). Sums over  $r$  denote a sum over all  $r \subset [\eta]$  of size  $m$ , and sums over  $\eta, i, r$  is short for the sum over all  $\eta \in \mathbb{Z}^+$ ,  $i \in [\eta]$  and  $r \subset [\eta]$  of size  $m$  with  $r \ni i$ .

**Lemma 3.1.** *Assumption 2.2 is satisfied for the nested case control model (23) with*

$$\rho_0 = \frac{1}{2} \left[ z_i + \log \bar{G}(t) \sum_{j \in r} (z_j - EZ) + \eta EZ \log \bar{G}(t) \right] f_0^{1/2}, \quad (29)$$

$$A\alpha = \left( g^{-1/2}(t)\alpha(t) + \frac{(\eta - 1) \int_t^\infty g^{1/2} \alpha d\nu}{\bar{G}(t)} \right) f_0^{1/2}, \quad (30)$$

and

$$B\beta = \left( \sum_{j \in r} h^{-1/2}(z_j) \beta(z_j) \right) f_0^{1/2}. \quad (31)$$

Lemma 3.1 is proved in Section 7.

### 3.1 A Operator: Properties

Regarding the definition and calculation of adjoint operators such as  $A^*$  in the following lemma, the reader is referred to [13]. The proof of the following lemma appears in Section 5.

**Lemma 3.2.** *Let  $\rho_0$  and  $A$  be given by (29) and (30), respectively. Then the function*

$$\hat{\alpha} = \frac{EZ}{2} [1 + \log \bar{G}(t)] g^{1/2}(t)$$

*is the solution to the normal equation  $A^*A\alpha = A^*\rho_0$ , and the projection of  $\rho_0$  onto the range of  $A$  is given by*

$$A\hat{\alpha} = \frac{EZ}{2} [1 + \eta \log \bar{G}(t)] f_0^{1/2}. \quad (32)$$

### 3.2 B Operator: Properties

Let  $r \subset [\eta]$  of size  $m$  be fixed. For  $s \subset r$  let  $z_s = \{z_j : j \in s\}$  and  $z_{-s} = \{z_j : j \in r \setminus s\}$ , and denote integration over  $z_s$  and  $z_{-s}$  with respect to the measures  $\nu^{|s|}$  and  $\nu^{m-|s|}$  by  $dz_s$  and  $dz_{-s}$ , respectively. When  $s = \{j\}$ , we identify that  $j^{\text{th}}$  variable  $z_j$  with  $z$ . Integration with respect to  $\nu^+$  is often indicated by  $dt$ , but may also be indicated by other notations such as  $du$ , or suppressed, when clear from context.

**Lemma 3.3.** *The adjoint  $B^* : L^2(\sigma) \rightarrow L^2(\nu)$  of the operator  $B$  in (31) is given by*

$$B^*\mu = h^{-1/2}(z) \sum_{\eta, i, r, j \in r} \int_{z_{-j}} \int_0^\infty f_0^{1/2} \mu dt dz_{-j}. \quad (33)$$

**Proof:** As  $B = \sum_{j \in r} B_j$  with

$$B_j\beta = h^{-1/2}(z_j) f_0^{1/2} \beta(z_j) \quad \text{for } \beta \in L^2(\nu),$$

by linearity one need only sum the adjoints  $B_j^*$  of  $B_j$  over  $j \in r$  to obtain  $B^*$ . For  $\mu \in L^2(\sigma)$ , the calculation

$$\begin{aligned}
\langle B_j \beta, \mu \rangle_\sigma &= \int_{\mathcal{X}} B_j \beta \mu d\sigma \\
&= \sum_{\eta, i, r} \int_{z_r} \int_0^\infty h^{-1/2}(z_j) f_0^{1/2} \beta(z_j) \mu dt dz_r \\
&= \int_z \beta(z) \left( h^{-1/2}(z) \sum_{\eta, i, r} \int_{z_{-j}} \int_0^\infty f_0^{1/2} \mu dt dz_{-j} \right) dz \\
&= \langle \beta, B_j^* \mu \rangle_\nu
\end{aligned}$$

provides the desired conclusion. □

The proof of the following lemma appears in Section 6.

**Lemma 3.4.** *The function*

$$\hat{\beta} = \frac{1}{2} h^{1/2}(z) E \left[ \frac{\eta - m}{m\eta} \right] (z - EZ) \quad (34)$$

is the solution to the normal equation  $B^* B \hat{\beta} = B^* \rho_0$ , and the projection of  $\rho_0$  onto the range of  $B$  is given by

$$B \hat{\beta} = \frac{1}{2} \left( E \left[ \frac{\eta - m}{m\eta} \right] \sum_{j \in r} (z_j - EZ) \right) f_0^{1/2}. \quad (35)$$

## 4 Lower Bound Calculations

We begin the computation of the information bound by showing that the two operators  $A$  and  $B$  have orthogonal ranges.

**Lemma 4.1.** *Let  $A$  and  $B$  be the operators given by (30) and (31) respectively. Then*

$$B^* A = 0 \quad \text{and} \quad A^* B = 0.$$

**Proof:** Since  $(A^* B)^* = B^* A$  it suffices to prove only the first claim. By (30) and (33),

$$\begin{aligned}
B^* A \alpha &= B^* \left( g^{-1/2}(t) \alpha(t) + \frac{(\eta - 1) \int_t^\infty g^{1/2} \alpha}{\bar{G}(t)} \right) f_0^{1/2} \\
&= h^{-1/2}(z) \sum_{\eta, i, r, j \in r} \int_{z_{-j}} \int_0^\infty \left( g^{-1/2}(t) \alpha(t) + \frac{(\eta - 1) \int_t^\infty g^{1/2} \alpha}{\bar{G}(t)} \right) f_0 dt dz_{-j} \\
&= h^{-1/2}(z) \sum_{\eta} K_{\eta, m} \varrho(\eta) \\
&\quad \times \left[ \sum_{i, r, j \in r} \int_{z_{-j}} h(z_r) \int_0^\infty \left( g^{-1/2}(t) \alpha(t) + \frac{(\eta - 1) \int_t^\infty g^{1/2} \alpha}{\bar{G}(t)} \right) g(t) \bar{G}^{\eta-1}(t) dt dz_{-j} \right].
\end{aligned}$$

Integrating the inner integral by parts,

$$\begin{aligned}
& \int_0^\infty \left( g^{-1/2}(t)\alpha(t) + (\eta - 1) \frac{\int_t^\infty g^{1/2}\alpha}{\bar{G}(t)} \right) g(t)\bar{G}^{\eta-1}(t)dt \\
&= \int_0^\infty \left( g^{1/2}(t)\bar{G}^{\eta-1}(t)\alpha(t) + (\eta - 1)g(t)\bar{G}^{\eta-2}(t) \int_t^\infty g^{1/2}\alpha \right) dt \\
&= \int_0^\infty g^{1/2}(t)\bar{G}^{\eta-1}(t)\alpha(t)dt - \int_0^\infty (\bar{G}^{\eta-1}(t))' \left( \int_t^\infty g^{1/2}\alpha \right) dt \\
&= \int_0^\infty g^{1/2}(t)\bar{G}^{\eta-1}(t)\alpha(t)dt - \left[ \bar{G}^{\eta-1}(t) \int_t^\infty g^{1/2}\alpha \right]_0^\infty - \int_0^\infty g^{1/2}(t)\bar{G}^{\eta-1}(t)\alpha(t)dt \\
&= \int_0^\infty g^{1/2}\alpha
\end{aligned}$$

which equals zero by (5).  $\square$

The perpendicularity relation which holds between  $A$  and  $B$  allows for the application of the following lemma which simplifies the calculation of the information bound.

**Lemma 4.2.** *Let  $\mathbb{K}$  be given by (13). Then under the perpendicularity relations provided by Lemma 4.1, the function*

$$\hat{\delta} = A\hat{\alpha} + B\hat{\beta} \quad \text{minimizes } \|\rho_0 - \delta\|_\sigma \text{ over } \delta \in \mathbb{K},$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are the solutions to the normal equations  $A^*A\hat{\alpha} = A^*\rho_0$  and  $B^*B\hat{\beta} = B^*\rho_0$ , respectively. Consequently, the effective information (15) is given by

$$I_*(\theta_0) = 4\|\rho_0 - A\hat{\alpha} - B\hat{\beta}\|_\sigma^2. \quad (36)$$

**Proof:** Since  $A^*B = 0$  we have  $A^*\rho_0 = A^*A\hat{\alpha} = A^*\hat{\delta}$  and similarly  $B^*\rho_0 = B^*B\hat{\beta} = B^*\hat{\delta}$ . Therefore  $(A + B)^*\rho_0 = (A + B)^*\hat{\delta}$ , or  $(A + B)^*(\rho_0 - \hat{\delta}) = 0$ . Hence we have

$$\rho_0 - \hat{\delta} \perp \mathbb{K} \quad \text{and} \quad \hat{\delta} \in \mathbb{K},$$

showing  $\hat{\delta}$  is the claimed minimizer.  $\square$

We pause to record a simple calculation which will be used frequently in what follows.

**Lemma 4.3.** *Let  $s(t)$  be any density on  $\mathbb{R}^+$  and  $\bar{S}(t)$  the corresponding survival function. Then for all integers  $\eta$  and  $k$  satisfying  $\eta \geq k$ , and  $j = 1, 2, \dots$ ,*

$$\int_0^\infty s(t)\bar{S}(t)^{\eta-k}[\log \bar{S}(t)]^j dt = (-1)^j(\eta - k + 1)^{-(j+1)}j!.$$

In particular, as  $\log S(t) \leq 0$  for all  $t \in \mathbb{R}^+$ , if  $k$  and  $j$  are fixed then for any constant  $C > 1$  there exists  $\eta_C$  such that

$$\int_0^\infty s(t)\bar{S}(t)^{\eta-k}|\log \bar{S}(t)|^j dt \leq \frac{Cj!}{\eta^{j+1}} \quad \text{for all } \eta \geq \eta_C.$$

**Proof:** Rewriting the integral and then applying the change of variables  $u = \bar{S}(t)^{\eta-k+1}$  followed by  $u = e^{-x}$  we have

$$\begin{aligned}
\int_0^\infty s(t)\bar{S}(t)^{\eta-k}[\log \bar{S}(t)]^j dt &= (\eta - k + 1)^{-j} \int_0^\infty s(t)\bar{S}(t)^{\eta-k}[\log \bar{S}(t)^{\eta-k+1}]^j dt \\
&= (\eta - k + 1)^{-(j+1)} \int_0^1 [\log u]^j du \\
&= (-1)^j (\eta - k + 1)^{-(j+1)} \Gamma(j + 1) \\
&= (-1)^j (\eta - k + 1)^{-(j+1)} j!.
\end{aligned}$$

Taking absolute value and noting that  $(\eta - k + 1)/\eta \rightarrow 1$  suffices to prove the final claim.  $\square$

**Theorem 4.1.** *The effective information for the nested case control model (23) is given by (25).*

**Proof:** Substituting (29), (32), and (35) into (36) we obtain

$$\begin{aligned}
I_*^g(\theta_0) &= \left\| \left[ (z_i - EZ) + \log \bar{G}(t) \sum_{j \in r} (z_j - EZ) - E \left[ \frac{\eta - m}{\eta m} \right] \sum_{j \in r} (z_j - EZ) \right] f_0^{1/2} \right\|_\sigma^2 \\
&= \left\| \left[ (z_i - EZ) + \sum_{j \in r} (z_j - EZ) \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right) \right] f_0^{1/2} \right\|_\sigma^2 \\
&= \left\| \left[ \left( 1 + \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right) (z_i - EZ) \right. \right. \\
&\quad \left. \left. + \sum_{j \in r \setminus \{i\}} (z_j - EZ) \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right) \right] f_0^{1/2} \right\|_\sigma^2,
\end{aligned}$$

which, by the independence of  $Z_i$  and  $\{Z_j, j \in r \setminus \{i\}\}$ , equals

$$\begin{aligned}
&\left\| \left[ \left( 1 + \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right) (z_i - EZ) \right] f_0^{1/2} \right\|_\sigma^2 \\
&+ \left\| \left[ \sum_{j \in r \setminus \{i\}} (z_j - EZ) \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right) \right] f_0^{1/2} \right\|_\sigma^2.
\end{aligned}$$

Squaring and integrating against the null density (24) we obtain

$$\begin{aligned}
& \int_{\mathcal{X}} \left[ \left( 1 + \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 (z_i - EZ)^2 + \sum_{j \in r \setminus \{i\}} \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 (z_j - EZ)^2 \right] f_0 d\sigma \\
&= \text{Var}(Z) \sum_{\eta} K_{\eta, m} \varrho(\eta) \\
&\times \left[ \sum_{i, r} \int_0^{\infty} \left[ \left( 1 + \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 + (m-1) \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 \right] g(t) \bar{G}^{\eta-1}(t) dt \right] \\
&= \text{Var}(Z) \sum_{\eta} \varrho(\eta) \\
&\times \left[ \int_0^{\infty} \left[ \left( 1 + \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 + (m-1) \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 \right] \eta g(t) \bar{G}^{\eta-1}(t) dt \right] \\
&= \text{Var}(Z) \sum_{\eta} \varrho(\eta) \\
&\times \left[ \int_0^{\infty} \left[ 1 + 2 \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right) + m \left( \log \bar{G}(t) - E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 \right] \eta g(t) \bar{G}^{\eta-1}(t) dt \right] \\
&= \text{Var}(Z) E \left[ \left( 1 - 2 \left( \frac{1}{\eta} + E \left[ \frac{\eta - m}{\eta m} \right] \right) + m \left( \frac{2}{\eta^2} + 2 \frac{1}{\eta} E \left[ \frac{\eta - m}{\eta m} \right] + \left( E \left[ \frac{\eta - m}{\eta m} \right] \right)^2 \right) \right) \right],
\end{aligned}$$

by applying Lemma 4.3. Simplifying we obtain

$$I_*^{\varrho}(\theta_0) = \text{Var}(Z) \left( 1 - \frac{1}{m} \right) + m \text{Var}(Z) \left( 2 \text{Var} \left( \frac{1}{\eta} \right) + \left[ E \left( \frac{1}{\eta} \right) \right]^2 \right).$$

which is (25). □

We now calculate the lower bound for the estimation of the baseline survival.

**Theorem 4.2.** *The covariance function  $K_*(s, t)$  in (19) specializes to (28) for the nested case-control model (23).*

**Proof:** Lemma 5.2 shows that  $A^*A$  is given by (42) with  $M_0(t) = E[\eta \bar{G}^{\eta}(t)]$ . Now (6.8) of [1] yields

$$K(s, t) = \bar{G}(t) \bar{G}(s) \int_0^{s \wedge t} \frac{dG}{M_0(u) \bar{G}(u)} = \bar{G}(t) \bar{G}(s) \int_0^{s \wedge t} \frac{dG}{E[\eta \bar{G}(u)^{\eta+1}]}.$$

Regarding the integral in (19), using the form  $\hat{\alpha}$  given in Lemma 3.2, we have

$$\begin{aligned}
\int_0^t \hat{\alpha} g^{1/2} d\nu &= \int_0^t \frac{EZ}{2} [1 + \log \bar{G}(u)] g(u) du \\
&= \frac{EZ}{2} \int_{\bar{G}(t)}^1 (1 + \log x) dx = -\frac{EZ}{2} \bar{G}(t) \log \bar{G}(t).
\end{aligned}$$

Substitution into (19) now yields (28).



## 5 A Operator

In this section we provide the proof to Lemma 3.2. We begin by calculating the adjoint  $A^* : L^2(\sigma) \rightarrow L^2(\nu^+)$  of the operator  $A$  given in Lemma 3.1.

**Lemma 5.1.** *For the operator  $A : L^2(\nu^+) \rightarrow L^2(\sigma)$  given in (30), write*

$$A = A_1 + A_2$$

where

$$A_1\alpha = g^{-1/2}f_0^{1/2}\alpha \quad \text{and} \quad A_2\alpha = \frac{(\eta-1)\int_t^\infty g^{1/2}\alpha}{\overline{G}(t)}f_0^{1/2}. \quad (37)$$

Then the adjoint of  $A$  is given by  $A^* = A_1^* + A_2^*$  where

$$A_1^*\mu = g^{-1/2}(t)\sum_{\eta,i,r}\int_{z_r}f_0^{1/2}\mu dz_r \quad \text{and} \quad A_2^*\mu = g^{1/2}(t)\sum_{\eta,i,r}(\eta-1)\int_0^t\int_{z_r}\frac{f_0^{1/2}}{\overline{G}(u)}\mu dz_r du. \quad (38)$$

**Proof:** Let  $\alpha \in L^2(\nu^+)$  and  $\mu \in L^2(\sigma)$ . Then

$$\begin{aligned} \langle A_1\alpha, \mu \rangle_\sigma &= \langle g^{-1/2}f_0^{1/2}\alpha, \mu \rangle_\sigma \\ &= \sum_{\eta,i,r}\int_0^\infty\int_{z_r}g^{-1/2}(t)\alpha(t)f_0^{1/2}\mu dz_r dt \\ &= \int_0^\infty\alpha(t)\left(g^{-1/2}(t)\sum_{\eta,i,r}\int_{z_r}f_0^{1/2}\mu dz_r\right) dt \\ &= \langle \alpha, A_1^*\mu \rangle_{\nu^+} \end{aligned}$$

when  $A_1^*$  is as given in (38).

Next, writing  $A_2$  as

$$A_2\alpha = L\int_t^\infty g^{1/2}\alpha \quad \text{for} \quad L = (\eta-1)\overline{G}^{-1}(t)f_0^{1/2}, \quad (39)$$

we have

$$\begin{aligned} \langle A_2\alpha, \mu \rangle_\sigma &= \langle L\int_t^\infty g^{1/2}\alpha, \mu \rangle_\sigma \\ &= \sum_{\eta,i,r}\int_0^\infty\int_{z_r}\int_t^\infty Lg^{1/2}\alpha\mu dudz_r dt \\ &= \int_0^\infty\alpha(t)\left(g^{1/2}(t)\sum_{\eta,i,r}\int_0^t\int_{z_r}L\mu dz_r du\right) dt \\ &= \langle \alpha, A_2^*\mu \rangle_{\nu^+} \end{aligned}$$

when

$$A_2^*\mu = g^{1/2}(t)\sum_{\eta,i,r}\int_0^t\int_{z_r}L\mu dz_r du.$$

Substituting  $L$  from (39) now yields the stated conclusion.  $\square$

To help express the solution to the normal equations in  $A$ , for  $\alpha \in L^2(\nu^+)$  define the operator  $R$  as in [1] by the first equality in

$$R\alpha = g^{-1/2}(t)\alpha(t) - \frac{\int_t^\infty g^{1/2}\alpha}{\overline{G}(t)} = g^{-1/2}(t)\alpha(t) + \frac{\int_0^t g^{1/2}\alpha}{\overline{G}(t)}; \quad (40)$$

the second equality follows from (5). Also, set

$$M_0(t) = E[\eta\overline{G}^\eta(t)] \quad \text{and} \quad M_1(t) = E[Z]E[\eta\overline{G}(t)^\eta]. \quad (41)$$

**Lemma 5.2.** *Let the operator  $A$  be given by (30). Then, for  $\alpha \in L^2(\nu^+)$ ,*

$$A^*A\alpha = \left[ R\alpha(t)\frac{M_0(t)}{\overline{G}(t)} - \int_0^t R\alpha\frac{M_0}{\overline{G}(u)}\frac{dG}{\overline{G}(u)} \right] g^{1/2}, \quad (42)$$

with inverse given by

$$(A^*A)^{-1}\alpha = \left[ R\alpha(t)\frac{\overline{G}(t)}{M_0(t)} - \int_0^t R\alpha\frac{\overline{G}(u)}{M_0(u)}\frac{dG}{\overline{G}(u)} \right] g^{1/2}. \quad (43)$$

**Proof:** Using the decompositions  $A = A_1 + A_2$  and  $A^* = A_1^* + A_2^*$  given in Lemma 5.1, write

$$A\alpha = \mu_1 + \mu_2 \quad \text{where} \quad \mu_i = A_i\alpha \quad i = 1, 2,$$

so that

$$A^*A\alpha = (A_1^* + A_2^*)(A_1 + A_2)\alpha = A_1^*\mu_1 + A_1^*\mu_2 + A_2^*\mu_1 + A_2^*\mu_2.$$

Consider  $A_1^*\mu_1$ . From (37) and (38),

$$\begin{aligned} A_1^*\mu_1 &= g^{-1/2}(t) \sum_{\eta,i,r} \int_{z_r} f_0 g^{-1/2} \alpha dz_r \\ &= g^{-1}(t) \alpha(t) \sum_{\eta,i,r} \int_{z_r} f_0 dz_r \\ &= g^{-1}(t) \alpha(t) \sum_{\eta,i,r} \varrho(\eta) \left[ \int_{z_r} K_{\eta,m} g(t) \overline{G}^{\eta-1}(t) h(z_r) dz_r \right] \\ &= \alpha(t) \sum_{\eta} \varrho(\eta) \left[ \overline{G}^{\eta-1}(t) K_{\eta,m} \sum_{i,r} \int_{z_r} h(z_r) dz_r \right] \\ &= \alpha(t) \sum_{\eta} \varrho(\eta) \left[ \overline{G}^{\eta-1}(t) K_{\eta,m} \sum_{i,r} 1 \right] \\ &= \alpha(t) E[\eta \overline{G}^{\eta-1}(t)]. \end{aligned}$$

In a similar fashion,

$$\begin{aligned}
A_1^* \mu_2 &= \sum_{\eta} (\eta - 1) g^{-1/2}(t) \sum_{i,r} \int_{z_r} f_0 \frac{\int_t^{\infty} g^{1/2} \alpha}{\overline{G}(t)} dz_r \\
&= \sum_{\eta} \varrho(\eta) \left[ (\eta - 1) g^{-1/2}(t) K_{\eta,m} \sum_{i,r} \int_{z_r} g(t) \overline{G}^{\eta-2}(t) \int_t^{\infty} g^{1/2} \alpha h(z_r) dz_r \right] \\
&= E \left[ (\eta - 1) g^{1/2}(t) \overline{G}^{\eta-2}(t) \int_t^{\infty} g^{1/2} \alpha K_{\eta,m} \sum_{i,r} \int_{z_r} h(z_r) dz_r \right] \\
&= E \left[ \eta(\eta - 1) g^{1/2}(t) \overline{G}^{\eta-2}(t) \int_t^{\infty} g^{1/2} \alpha \right] \\
&= E \left[ -\eta(\eta - 1) g^{1/2}(t) \overline{G}^{\eta-2}(t) \int_0^t g^{1/2} \alpha \right],
\end{aligned}$$

where the final equality follows from (5).

Now moving on to the terms involving  $A_2^*$ ,

$$\begin{aligned}
A_2^* \mu_1 &= \sum_{\eta} (\eta - 1) g^{1/2}(t) \sum_{i,r} \int_0^t \int_{z_r} \frac{f_0}{\overline{G}} g^{-1/2} \alpha dz_r du \\
&= \sum_{\eta} \varrho(\eta) \left[ (\eta - 1) g^{1/2}(t) \int_0^t g^{1/2} \overline{G}^{\eta-2} \alpha du K_{\eta,m} \sum_{i,r} \int_{z_r} h(z_r) dz_r \right] \\
&= E \left[ \eta(\eta - 1) g^{1/2}(t) \int_0^t g^{1/2} \overline{G}^{\eta-2} \alpha du \right],
\end{aligned}$$

and lastly,

$$\begin{aligned}
A_2^* \mu_2 &= \sum_{\eta} (\eta - 1)^2 g^{1/2}(t) \sum_{i,r} \int_0^t \int_{z_r} \frac{f_0}{\overline{G}(u)^2} \int_u^{\infty} g^{1/2} \alpha dz_r du \\
&= \sum_{\eta} \varrho(\eta) (\eta - 1)^2 g^{1/2}(t) \int_0^t g(u) \overline{G}^{\eta-3}(u) \int_u^{\infty} g^{1/2} \alpha K_{\eta,m} \sum_{i,r} \int_{z_r} h(z_r) dz_r \\
&= E \left[ -\eta(\eta - 1)^2 g^{1/2}(t) \int_0^t g(u) \overline{G}^{\eta-3}(u) \int_0^u g^{1/2}(v) \alpha(v) dv du \right] \\
&= E \left[ -\eta(\eta - 1)^2 g^{1/2}(t) \int_0^t g^{1/2}(v) \alpha(v) \int_v^t g(u) \overline{G}^{\eta-3}(u) du dv \right] \\
&= E \left[ \frac{-\eta(\eta - 1)^2}{\eta - 2} g^{1/2}(t) \int_0^t g^{1/2}(v) \alpha(v) [\overline{G}^{\eta-2}(v) - \overline{G}^{\eta-2}(t)] \right] \\
&= E \left[ \frac{\eta(\eta - 1)^2}{\eta - 2} g^{1/2}(t) \left( \overline{G}^{\eta-2}(t) \int_0^t g^{1/2} \alpha - \int_0^t g^{1/2} \overline{G}^{\eta-2} \alpha \right) \right].
\end{aligned}$$

Combining terms, we arrive at

$$\begin{aligned}
A^*\mu &= E \left[ \eta g^{1/2} \left( g^{-1/2} \bar{G}^{\eta-1} \alpha - (\eta-1) \bar{G}^{\eta-2} \int_0^t g^{1/2} \alpha + (\eta-1) \int_0^t g^{1/2} \bar{G}^{\eta-2} \alpha \right. \right. \\
&\quad \left. \left. + \frac{(\eta-1)^2}{\eta-2} \left( \bar{G}^{\eta-2} \int_0^t g^{1/2} \alpha - \int_0^t g^{1/2} \bar{G}^{\eta-2} \alpha \right) \right) \right] \\
&= E \left[ \eta g^{1/2} \left( g^{-1/2} \bar{G}^{\eta-1} \alpha - (\eta-1) \left( 1 - \frac{\eta-1}{\eta-2} \right) \left( \bar{G}^{\eta-2} \int_0^t g^{1/2} \alpha - \int_0^t g^{1/2} \bar{G}^{\eta-2} \alpha \right) \right) \right] \\
&= g^{1/2} E \left[ g^{-1/2} \alpha \eta \bar{G}^{\eta-1} + \left( \frac{\eta-1}{\eta-2} \right) \left( \eta \bar{G}^{\eta-2} \int_0^t g^{1/2} \alpha - \int_0^t g^{1/2} \alpha \eta \bar{G}^{\eta-2} \right) \right].
\end{aligned}$$

Recalling  $R\alpha$  from (40), we may write

$$\begin{aligned}
A^*\mu &= g^{1/2} E \left[ g^{-1/2} \alpha \frac{\eta \bar{G}^\eta}{G} + \left( \frac{\eta-1}{\eta-2} \right) \left( \frac{\eta \bar{G}^\eta}{G^2} \int_0^t g^{1/2} \alpha - \int_0^t g^{-1/2} \alpha \frac{\eta \bar{G}^\eta}{G} \frac{d\bar{G}}{G} \right) \right] \\
&= g^{1/2} E \left[ R\alpha \frac{\eta \bar{G}^\eta}{G} + \frac{1}{\eta-2} \frac{\eta \bar{G}^\eta}{G^2} \int_0^t g^{1/2} \alpha - \int_0^t g^{-1/2} \alpha \frac{\eta \bar{G}^\eta}{G} \frac{d\bar{G}}{G} - \frac{1}{\eta-2} \int_0^t g^{-1/2} \alpha \frac{\eta \bar{G}^\eta}{G} \frac{d\bar{G}}{G} \right].
\end{aligned}$$

Rewriting the third term using

$$\int_0^t g^{-1/2}(u) \alpha(u) \frac{\eta \bar{G}^\eta(u)}{G(u)} \frac{d\bar{G}}{G} = \int_0^t \left( R\alpha(u) + \frac{\int_u^\infty g^{1/2} \alpha}{G} \right) \frac{\eta \bar{G}^\eta(u)}{G(u)} \frac{d\bar{G}}{G},$$

we find

$$\begin{aligned}
A^*\mu &= g^{1/2} E \left[ R\alpha(t) \frac{\eta \bar{G}^\eta(t)}{G(t)} - \int_0^t R\alpha(u) \frac{\eta \bar{G}^\eta(u)}{G(u)} \frac{d\bar{G}}{G} \right. \\
&\quad \left. + \frac{1}{\eta-2} \left( \frac{\eta \bar{G}^\eta}{G^2} \int_0^t g^{1/2} \alpha - \int_0^t g^{-1/2} \alpha \frac{\eta \bar{G}^\eta}{G} \frac{d\bar{G}}{G} - (\eta-2) \int_0^t \left( \int_u^\infty g^{1/2} \alpha \right) \frac{\eta \bar{G}^\eta}{G^2} \frac{d\bar{G}}{G} \right) \right]. \tag{44}
\end{aligned}$$

But now we see that the term on second line of (44) vanishes, since

$$\begin{aligned}
(\eta-2) \int_0^t \left( \int_u^\infty g^{1/2} \alpha \right) \frac{\eta \bar{G}^\eta}{G^2} \frac{d\bar{G}}{G} &= -\eta \int_0^t \left( \int_u^\infty g^{1/2} \alpha \right) d\bar{G}^{\eta-2} \\
&= -\eta \left( \int_u^\infty g^{1/2} \alpha \right) \bar{G}^{\eta-2} \Big|_0^t + \eta \int_0^t \bar{G}^{\eta-2} g^{1/2} \alpha \\
&= -\eta \bar{G}^{\eta-2} \int_t^\infty g^{1/2} \alpha + \int_0^t g^{1/2} \alpha \eta \bar{G}^{\eta-2} \\
&= \eta \bar{G}^{\eta-2} \int_0^t g^{1/2} \alpha - \int_0^t g^{-1/2} \alpha \eta \bar{G}^{\eta-2} d\bar{G} \\
&= \frac{\eta \bar{G}^\eta}{G^2} \int_0^t g^{1/2} \alpha - \int_0^t g^{-1/2} \alpha \frac{\eta \bar{G}^\eta}{G} \frac{d\bar{G}}{G}.
\end{aligned}$$

Hence,  $A^*A\alpha$  is given by the first line of (44), and taking the expectation inside the integral completes the proof of (42).

Lastly as  $A^*A$  is of the form (42), the form (43) of the inverse follows as in [1], page 449.  $\square$

We are now in position to prove Lemma 3.2, giving the solution  $\hat{\alpha}$  to the normal equations  $A^*A\alpha = A^*\rho_0$ , and the projection of  $\rho_0$  onto the range of  $A$ .

**Proof of Lemma 3.2** With  $\rho_0$  as in (29), we first claim

$$A^*\rho_0 = \frac{EZ}{2}g^{1/2}(t)E \left[ \frac{\eta}{\eta-1} (\eta\bar{G}(t)^{\eta-1} - 1) \right]. \quad (45)$$

From (38) we obtain directly that

$$A_1^*\rho_0 = \frac{EZ}{2}g^{1/2}(t)E \left[ \eta(1 + \eta \log \bar{G}(t)) \bar{G}^{\eta-1}(t) \right],$$

and

$$\begin{aligned} A_2^*\rho_0 &= \frac{EZ}{2}g^{1/2}(t)E \left[ \eta \int_0^t ((1 + \eta \log \bar{G}(u)) (\eta-1)g(u)\bar{G}^{\eta-2}(u))du \right] \\ &= \frac{EZ}{2}g^{1/2}(t)E \left[ \eta \left( \frac{1}{\eta-1} \bar{G}^{\eta-1}(u) - \eta \log \bar{G}(u) \bar{G}^{\eta-1}(u) \right)_0^t \right] \\ &= \frac{EZ}{2}g^{1/2}(t)E \left[ \eta \left( -\frac{1}{\eta-1} + \frac{1}{\eta-1} \bar{G}^{\eta-1}(t) - \eta \log \bar{G}(t) \bar{G}^{\eta-1}(t) \right) \right], \end{aligned}$$

and adding these two contributions yields the result (45).

From (40) and (45)

$$\begin{aligned} R(A^*\rho_0) &= \frac{EZ}{2}E \left[ \frac{\eta}{\eta-1} \left( \eta\bar{G}(t)^{\eta-1} - 1 - \frac{1}{\bar{G}(t)} \int_t^\infty [\eta\bar{G}^{\eta-1} - 1] dG \right) \right] \\ &= \frac{EZ}{2}E \left[ \frac{\eta}{\eta-1} \left( \eta\bar{G}(t)^{\eta-1} - 1 + \frac{1}{\bar{G}(t)} (-\bar{G}^\eta(t) + \bar{G}(t)) \right) \right] \\ &= \frac{EZ}{2}E [\eta\bar{G}(t)^{\eta-1}] = \frac{1}{2} \frac{M_1(t)}{\bar{G}(t)}, \end{aligned}$$

where  $M_1(t) = E[Z]E [\eta\bar{G}(t)^\eta]$ , in accordance with (41).

Hence, by (43), the solution  $\hat{\alpha}$  to the normal equations  $A^*A\alpha = A^*\rho_0$  is given by

$$\begin{aligned} \hat{\alpha} = (A^*A)^{-1}A^*\rho_0(t) &= \frac{1}{2} \left[ \frac{M_1(t)}{M_0(t)} - \int_0^t \frac{M_1(s)}{M_0(s)} \frac{dG}{\bar{G}} \right] g^{1/2}(t) \\ &= \frac{1}{2} \left[ E(Z) - \int_0^t E(Z) \frac{dG}{\bar{G}(s)} \right] g^{1/2}(t) = \frac{E(Z)}{2} [1 + \log \bar{G}(t)] g^{1/2}(t), \end{aligned}$$

where we have used  $M_1(t)/M_0(t) = EZ$ . To calculate the projection  $A\hat{\alpha}$  of  $\rho_0$  onto the range of  $A$ , note

$$\int_t^\infty \hat{\alpha}(s)g^{1/2}(s)ds = \frac{E(Z)}{2} \int_t^\infty (1 + \log \bar{G}(s))dG(s) = \frac{E(Z)}{2}\bar{G}(t) \log \bar{G}(t),$$

and hence

$$A\hat{\alpha} = \left[ g^{-1/2}(t)\hat{\alpha}(t) + (\eta-1) \frac{\int_t^\infty g^{1/2}\hat{\alpha}}{\bar{G}(t)} \right] f_0^{1/2} = \frac{E(Z)}{2} [1 + \eta \log \bar{G}(t)] f_0^{1/2}.$$

$\square$

## 6 $B$ Operator

In this section we prove Lemma 3.4, providing the solution to the normal equations for the operator  $B$ . Parallel to Section 5, we begin by deriving an expression for  $B^*B$ .

**Lemma 6.1.** *Let the operator  $B$  be given by (31). Then*

$$B^*B\beta = m\beta(z).$$

**Proof:** Applying formulas (33), (31) and (24),

$$\begin{aligned} B^*B\beta &= h^{-1/2}(z) \sum_{\eta, i, r, j \in r} \int_{z_{-j}} \int_0^\infty f_0 \left( \sum_{k \in r} h^{-1/2}(z_k) \beta(z_k) \right) dt dz_{-j} \\ &= \sum_{\eta} \varrho(\eta) K_{\eta, m} h^{-1/2}(z) \sum_{i, r, j \in r} \int_0^\infty g(t) \bar{G}(t)^{\eta-1} dt \int_{z_{-j}} \prod_{l \in r} h(z_l) \left( \sum_{k \in r} h^{-1/2}(z_k) \beta(z_k) \right) dz_{-j} \\ &= h^{-1/2}(z) E \left[ \int_0^\infty \eta g(t) \bar{G}(t)^{\eta-1} dt \right] \int_{z_{-j}} \sum_{j \in [m]} \prod_{l \in [m]} h(z_l) \left( \sum_{k \in [m]} h^{-1/2}(z_k) \beta(z_k) \right) dz_{-j} \\ &= h^{-1/2}(z) \sum_{j \in [m]} \int_{z_{-j}} \prod_{l \in [m]} h(z_l) \left( \sum_{k \in [m]} h^{-1/2}(z_k) \beta(z_k) \right) dz_{-j} \\ &= mh^{-1/2}(z) \int_{z_{-1}} \prod_{l \in [m]} h(z_l) \left( \sum_{k \in [m]} h^{-1/2}(z_k) \beta(z_k) \right) dz_{-1}, \end{aligned}$$

where the third equality is by symmetry, and the last by recalling that  $z_1$  and  $z$  are identified in the integral over  $z_{-1}$ . Hence

$$\begin{aligned} B^*B\beta &= mh^{1/2}(z) \int_{z_{-1}} \prod_{l=2}^m h(z_l) \left( \sum_{k=1}^m h^{-1/2}(z_k) \beta(z_k) \right) dz_{-1} \\ &= mh^{1/2}(z) \int_{z_{-1}} \prod_{l=2}^m h(z_l) \left( h^{-1/2}(z) \beta(z) + \sum_{k=2}^m h^{-1/2}(z_k) \beta(z_k) \right) dz_{-1} \\ &= m\beta(z) \int_{z_{-1}} \prod_{l=2}^m h(z_l) dz_{-1} + mh^{1/2}(z) \int_{z_{-1}} \prod_{l=2}^m h(z_l) \left( \sum_{k=2}^m h^{-1/2}(z_k) \beta(z_k) \right) dz_{-1}. \end{aligned}$$

As  $h(z_l)$  is a density, the first term integrates to  $m\beta(z)$ . For the second term,

$$\begin{aligned} \int_{z_{-1}} \prod_{l=2}^m h(z_l) \sum_{k=2}^m h^{-1/2}(z_k) \beta(z_k) dz_{-1} &= \sum_{k=2}^m \int_{z_{-1}} \left( \prod_{l \notin \{1, k\}} h(z_l) \right) h^{1/2}(z_k) \beta(z_k) dz_{-1} \\ &= \sum_{k=2}^m \int_{z_{-1, k}} \left( \prod_{l \notin \{1, k\}} h(z_l) \right) dz_{-1, k} \int_{z_k} h^{1/2}(z_k) \beta(z_k) dz_k \\ &= \sum_{k=2}^m \int_{z_k} h^{1/2}(z_k) \beta(z_k) dz_k \\ &= 0 \end{aligned}$$

by (9), showing  $B^*B\beta = m\beta(z)$ , and the lemma.  $\square$

**Proof of Lemma 3.4** From (33) and (29), arguing as in the proof of Lemma 6.1 and applying Lemma 4.3 we obtain

$$\begin{aligned}
B^*\rho_0 &= \frac{1}{2} \sum_{\eta, i, r, j \in r} \int_{z_{-j}} \int_0^\infty f_0 h^{-1/2}(z_j) \left( z_i + \log \bar{G}(t) \sum_{k \in r} (z_k - EZ) + \eta \log \bar{G}(t) EZ \right) dt dz_{-j} \\
&= \frac{1}{2} \sum_{\eta} \varrho(\eta) \times \\
&\quad \left[ \sum_{j \in [m]} \int_{z_{-j}} \int_0^\infty \eta g(t) \bar{G}^{\eta-1}(t) h^{-1/2}(z_j) h(z_r) \left( z_1 + \log \bar{G}(t) \sum_{k \in [m]} (z_k - EZ) + \eta \log \bar{G}(t) EZ \right) dt dz_{-j} \right] \\
&= \frac{1}{2} \sum_{j \in [m]} \int_{z_{-j}} h^{-1/2}(z_j) h(z_r) E \left[ z_1 - \frac{1}{\eta} \sum_{k \in [m]} (z_k - EZ) - EZ \right] dz_{-j} \\
&= \frac{1}{2} \sum_{j \in [m]} \int_{z_{-j}} h^{1/2}(z_j) \prod_{k \neq j} h(z_k) E \left[ \left( \frac{\eta-1}{\eta} \right) (z_1 - EZ) - \frac{1}{\eta} \sum_{k=2}^m (z_k - EZ) \right] dz_{-j}.
\end{aligned}$$

For  $j = 1$  we obtain  $(1/2)h^{1/2}(z)E[(\eta-1)/\eta](z-EZ)$  from the first term in parenthesis, while each term in the second sum integrates to zero. For each of the  $m-1$  terms where  $j \neq 1$  the first term in parenthesis integrates to zero, but when  $k = j$  one term in the sum in the second term makes a nonzero contribution of  $-(1/2)h^{1/2}(z)(z-EZ)E[1/\eta]$ , for a total of

$$B^*\rho_0 = \frac{1}{2}h^{1/2}(z)(z-EZ)E \left[ \frac{\eta-1}{\eta} - \frac{m-1}{\eta} \right] = \frac{1}{2}h^{1/2}(z)(z-EZ)E \left[ \frac{\eta-m}{\eta} \right].$$

From Lemma 6.1 we clearly have,

$$(B^*B)^{-1}\beta = \frac{1}{m}\beta \quad \text{hence} \quad \hat{\beta} = (B^*B)^{-1}B^*\rho_0 = \frac{1}{2}h^{1/2}(z)(z-EZ)E \left[ \frac{\eta-m}{\eta m} \right],$$

proving (34), and applying  $B$  as in (31) to  $\hat{\beta}$  now yields (35).  $\square$

## 7 Operators and Parametric Score: Verification

The goal of the present section is to prove Lemma 3.1, verifying Assumption 2.2.

**Proof of Lemma 3.1.** The operators  $A$  and  $B$  are clearly linear. To see that their ranges are contained in  $L^2(\nu^+)$  and  $L^2(\nu)$ , respectively, we show that they are bounded. Write  $A = A_1 + A_2$  where  $A_1$  and  $A_2$  are given in (37). The operator  $A_1$  is bounded, as

$$\|A_1\alpha\|_\sigma^2 = \int_{\mathcal{X}} (g^{-1/2}f_0^{1/2}\alpha)^2 d\sigma = \int_{\mathcal{X}} K_{\eta, m} \bar{G}^{\eta-1}(t) \alpha^2(t) h(z_r) \varrho(\eta) d\sigma \leq E[\eta] \|\alpha\|_{\nu^+}^2.$$

Regarding  $A_2$ , first note that

$$\left| \int_t^\infty g^{1/2} \alpha du \right|^2 \leq \int_t^\infty g du \int_t^\infty \alpha^2 du \leq \|\alpha\|_{\nu^+}^2,$$

and hence

$$\|A_2\alpha\|_\sigma^2 \leq \int_{\mathcal{X}} K_{\eta,m}(\eta-1)^2 g(t) \overline{G}(t)^{\eta-3} \|\alpha\|_\nu^2 h(z_r) \varrho(\eta) d\sigma \leq E \left[ \frac{\eta(\eta-1)^2}{\eta-2} \right] \|\alpha\|_{\nu^+}^2,$$

showing  $A_2$ , and therefore  $A$ , is bounded. Next, for any  $j \in r$ ,

$$\int_{\mathcal{X}} \left( h^{-1/2}(z_j) \beta(z_j) f_0^{1/2} \right)^2 d\sigma = \int_{\mathcal{X}} K_{\eta,m} \beta^2(z_j) g(t) \overline{G}(t)^{\eta-1} h(z_{-j}) \varrho(\eta) d\sigma = \|\beta\|_\nu^2.$$

Hence  $B$  is a finite sum of bounded operators, and is therefore bounded.

To conclude the proof of the lemma we need to show that for any  $\alpha \in \mathcal{A}$  and  $\beta \in \mathcal{B}$  the term (12) tends to zero for an arbitrary

$$(\{\theta_n\}_{n \geq 0}, \{g_n\}_{n \geq 0}, \{h_n\}_{n \geq 0}) \in \Theta(\tau) \times \mathcal{C}_1(g, \alpha) \times \mathcal{C}_2(h, \beta).$$

By the triangle inequality (12) is upper bounded by the sum of the three terms

$$\|\sqrt{n} (f^{1/2}(\theta_n, g_n, h_n) - f^{1/2}(\theta_0, g_n, h_n)) - \tau \rho_0\|_\sigma \quad (46)$$

$$+ \|\sqrt{n} (f^{1/2}(\theta_0, g_n, h_n) - f^{1/2}(\theta_0, g, h_n)) - A\alpha\|_\sigma \quad (47)$$

$$+ \|\sqrt{n} (f^{1/2}(\theta_0, g, h_n) - f^{1/2}(\theta_0, g, h)) - B\beta\|_\sigma. \quad (48)$$

We show these three terms tend to zero in Lemmas 7.13, 7.6, and 7.5, respectively.  $\square$

## 7.1 Preliminaries

We first collect here some results, some of which are also used in later sections. We recall that

$$h(z_r) = \prod_{j \in r} h(z_j) \quad \text{for any } h \in \mathcal{H}.$$

When the measure space on which functions are defined is clear from context, we will use  $\|\cdot\|_q$  to denote their  $L^q$  norms on that space.

**Lemma 7.1.** *Let  $\{h_n\}_{n \geq 0} \in \mathcal{C}(h, \beta)$  for some  $\beta \in \mathcal{B}$ . Then the sequence of  $L^2(\nu)$  norms  $\|\beta_n\|_2, n = 1, 2, \dots$  of*

$$\beta_n = \sqrt{n}(h_n^{1/2} - h^{1/2}) \quad (49)$$

*is bounded. In addition, in  $L^2(\nu^m)$ ,*

$$\|\sqrt{n} (h_n^{1/2}(z_r) - h^{1/2}(z_r)) - h^{1/2}(z_r) \left( \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j) \right)\|_{\nu^m} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (50)$$

*and in particular  $\|h_n^{1/2}(z_r) - h^{1/2}(z_r)\|_{\nu^m} \rightarrow 0$  as  $n \rightarrow \infty$ .*



**Proof:** Denote the  $L^2(\nu)$  norm by  $\|\cdot\|_\nu$ . The condition that  $\{h_n\}_{n \geq 0} \in \mathcal{C}(h, \beta)$  implies  $\lim_{n \rightarrow \infty} \|\beta_n - \beta\|_\nu \rightarrow 0$ ; as  $\|\beta_n\|_\nu \leq \|\beta_n - \beta\|_\nu + \|\beta\|_\nu$ , the sequence  $\|\beta_n\|_\nu$  is bounded. Next, as (49) gives  $h_n^{1/2}(z) = h^{1/2}(z) + n^{-1/2}\beta_n(z)$ , we have

$$\begin{aligned}
& \sqrt{n} (h_n^{1/2}(z_r) - h^{1/2}(z_r)) \\
&= \sqrt{n} \left( \frac{h_n^{1/2}(z_r)}{h^{1/2}(z_r)} - 1 \right) h^{1/2}(z_r) \\
&= \sqrt{n} \left( \prod_{j \in r} \frac{h_n^{1/2}(z_j)}{h^{1/2}(z_j)} - 1 \right) h^{1/2}(z_r) \\
&= \sqrt{n} \left( \prod_{j \in r} (1 + n^{-1/2} h^{-1/2}(z_j) \beta_n(z_j)) - 1 \right) h^{1/2}(z_r) \\
&= \left( \sum_{s \subset r, |s| > 0} n^{-(|s|+1)/2} \prod_{j \in s} h^{-1/2}(z_j) \beta_n(z_j) \right) h^{1/2}(z_r) \\
&= h^{1/2}(z_r) \sum_{j=1}^m h^{-1/2}(z_j) \beta_n(z_j) + \left( \sum_{s \subset r, |s| > 1} n^{-(|s|+1)/2} \prod_{j \in s} h^{-1/2}(z_j) \beta_n(z_j) \right) h^{1/2}(z_r)
\end{aligned}$$

as the subset  $s = \emptyset$  gives the product 1, which cancels, and the subsets  $s$  of size 1 yield the first term above after the last equality. Now subtracting  $h^{1/2}(z_r) \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j)$  yields

$$h^{1/2}(z_r) \sum_{j=1}^m h^{-1/2}(z_j) (\beta_n(z_j) - \beta(z_j)) + \left( \sum_{s \subset r, |s| > 1} n^{-(|s|+1)/2} \prod_{j \in s} h^{-1/2}(z_j) \beta_n(z_j) \right) h^{1/2}(z_r).$$

Squaring the first term and integrating, we have

$$\begin{aligned}
\int_{z_r} h(z_r) \left( \sum_{j=1}^m h^{-1/2}(z_j) (\beta_n(z_j) - \beta(z_j)) \right)^2 dz_r &\leq m \sum_{j=1}^m \int_{z_r} h(z_r) h^{-1}(z_j) (\beta_n(z_j) - \beta(z_j))^2 dz_r \\
&= m^2 \|\beta_n - \beta\|_\nu^2 \rightarrow 0,
\end{aligned}$$

as  $n \rightarrow \infty$ . For the  $L^2(\nu^m)$  norm of the second term, again by the Cauchy-Schwarz inequality

$$\begin{aligned}
& \int_{z_r} \left( \sum_{s \subset r, |s| > 1} n^{-(|s|+1)/2} \prod_{j \in s} h^{-1/2}(z_j) \beta_n(z_j) \right)^2 h(z_r) dz_r \\
&\leq 2^m \int_{z_r} \sum_{s \subset r, |s| > 1} n^{-|s|+1} \prod_{j \in s} h^{-1}(z_j) \beta_n^2(z_j) h(z_r) dz_r \\
&= 2^m \sum_{k=2}^m n^{-k+1} \sum_{s \subset r, |s|=k} \int_{z_r} \prod_{j \in s} h^{-1}(z_j) \beta_n^2(z_j) h(z_r) dz_r \\
&= 2^m \sum_{k=2}^m n^{-k+1} \sum_{s \subset r, |s|=k} \|\beta_n\|_\nu^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty,
\end{aligned}$$

since  $|s| = k > 1$  and  $\|\beta_n\|_\nu, n = 1, 2, \dots$  is bounded, completing the proof of (50).

Finally, note that  $h^{1/2}(z_r) \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j)$  is bounded in  $L^2(\nu^m)$  as

$$\int_{z_r} \left( h^{1/2}(z_r) \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j) \right)^2 dz_r \leq m \|\beta\|_\nu^2,$$

and hence  $\|h_n^{1/2} - h^{1/2}\|_{\nu^m} \rightarrow 0$  as  $n \rightarrow \infty$  by (50). □

The following lemma will also be required in Sections 7.3 and 7.4.

**Lemma 7.2.** *Let  $f_n$  and  $f$  be density functions on a measure space. Then*

$$\|f_n^{1/2} - f^{1/2}\|_2 \rightarrow 0 \quad \text{implies} \quad \|f_n - f\|_1 \rightarrow 0.$$

Further, for any measurable set  $S$ , letting

$$A_n(S) = F_n(S) - F(S) \quad \text{where} \quad F_n(S) = \int_S f_n(x) \quad \text{and} \quad F(S) = \int_S f(x),$$

we have

$$A_n^2(S) \leq 2 \int_S (f_n^{1/2}(x) - f^{1/2}(x))^2 (F_n(S) + F(S)), \quad (51)$$

and

$$|A_n(S)| \leq \left( 2 \int_S (f_n^{1/2}(x) - f^{1/2}(x))^2 \right)^{1/2} (F_n^{1/2}(S) + F^{1/2}(S)). \quad (52)$$

**Proof:** Note that

$$f_n(x) - f(x) = (f_n^{1/2}(x) - f^{1/2}(x))(f_n^{1/2}(x) + f^{1/2}(x)). \quad (53)$$

Taking absolute value, integrating and then using the Cauchy Schwarz inequality followed by

$$|f_n^{1/2}(x) + f^{1/2}(x)|^2 \leq 2(f_n(x) + f(x)) \quad (54)$$

yields

$$\|f_n - f\|_1^2 \leq \|f_n^{1/2} - f^{1/2}\|_2^2 \left( \int_0^\infty |f_n^{1/2}(x) + f^{1/2}(x)|^2 \right) \leq 4 \|f_n^{1/2} - f^{1/2}\|_2^2,$$

yields the first claim.

Next, integrating (53) over  $S$  we have

$$A_n(S) = \int_S (f_n^{1/2}(x) - f^{1/2}(x))(f_n^{1/2}(x) + f^{1/2}(x)),$$

so squaring and using the Cauchy Schwarz inequality yields

$$A_n^2(S) \leq \int_S (f_n^{1/2}(x) - f^{1/2}(x))^2 \int_S (f_n^{1/2}(x) + f^{1/2}(x))^2.$$

Now applying (54) gives (51), and taking square root and using  $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$  for nonnegative  $x, y$  we obtain (52). □

For a sequence  $\{g_n\}_{n \geq 0} \subset \mathcal{G}$  and  $t \geq 0$  let

$$\alpha_n(t) = \sqrt{n}(g_n^{1/2}(t) - g^{1/2}(t)) \quad \text{and} \quad A_n(t) = \sqrt{n}(\overline{G}_n(t) - \overline{G}(t)). \quad (55)$$

**Lemma 7.3.** *Let  $\{g_n\}_{n \geq 0} \in \mathcal{C}_1(g, \alpha)$  for some  $\alpha \in \mathcal{A}$ . Then the sequence of  $L^2$  norms  $\|\alpha_n\|_2, n = 1, 2, \dots$  is bounded and  $\|g_n - g\|_1 \rightarrow 0$  as  $n \rightarrow \infty$ . In addition*

$$A_n^2(t) \leq 2 \left[ \int_t^\infty \alpha_n^2(u) du \right] (\bar{G}(t) + \bar{G}_n(t)), \quad (56)$$

$$|A_n(t)| \leq \left( 2 \int_t^\infty \alpha_n^2(u) du \right)^{1/2} (\bar{G}^{1/2}(t) + \bar{G}_n^{1/2}(t)), \quad (57)$$

$$|\sqrt{n}(\bar{G}_n^{1/2}(t) - \bar{G}^{1/2}(t))| \leq \left( 2 \int_t^\infty \alpha_n(u)^2 du \right)^{1/2} \leq \sqrt{2} \|\alpha_n\|_2, \quad (58)$$

and

$$\|\bar{G}_n - \bar{G}\|_\infty \rightarrow 0. \quad (59)$$

**Proof:** That  $\|\alpha_n\|_2, n = 1, 2, \dots$  is bounded follows as in Lemma 7.1. The next claim and inequalities (56) and (57) follow from Lemma 7.2. Now write

$$\sqrt{n}(\bar{G}_n^{1/2}(t) - \bar{G}^{1/2}(t)) = \frac{\sqrt{n}(\bar{G}_n(t) - \bar{G}(t))}{\bar{G}_n^{1/2}(t) + \bar{G}^{1/2}(t)} = \frac{A_n(t)}{\bar{G}_n^{1/2}(t) + \bar{G}^{1/2}(t)},$$

so (58) follows from (57), as does (59).  $\square$

For any measurable function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  and a sequence  $\{h_n\}_{n \geq 0} \subset \mathcal{H}$ , let

$$E_n \varphi(Z) = \int \varphi(z) h_n(z) dz \quad n = 0, 1, 2, \dots,$$

and define  $\text{Var}_n(\varphi(Z))$  similarly. The following result will be applied in Section 7.4.

**Lemma 7.4.** *Let  $\{h_n\}_{n \geq 0} \in \mathcal{C}_2(h, \beta)$  for some  $\beta \in \mathcal{B}$ . Then, with  $\theta_\xi$  as in (3), for all  $\bar{\theta} \in (0, \theta_\xi)$  and  $a \geq 0$ ,*

$$\sup_{n, |\theta| \leq \bar{\theta}} E_n |Z|^a e^{\theta Z} < \infty.$$

*In particular, for any  $r \in [\eta]$  and  $i \in r$  the functions  $z_i^a h_n(z_r)$  and  $z_i^a h(z_r)$  are in  $L^1(\nu^m)$ ,*

$$\|z_i^a h_n(z_r) - z_i^a h(z_r)\|_1 \rightarrow 0 \quad \text{and} \quad E_n Z^a \rightarrow EZ. \quad (60)$$

**Proof:** For all  $a \geq 0$  there exists a constant  $C$  such that  $|z|^a e^{\bar{\theta}|z|} \leq C(e^{\theta_\xi z} + e^{-\theta_\xi z})$  for all  $z \in \mathbb{R}$ . Hence for all  $\theta$  satisfying  $|\theta| \leq \bar{\theta}$ ,

$$E_n |Z|^a e^{\theta Z} \leq E_n |Z|^a e^{\bar{\theta}|Z|} \leq C E_n (e^{\theta_\xi Z} + e^{-\theta_\xi Z}) = C(M_{h_n}(\theta_\xi) + M_{h_n}(-\theta_\xi)) < C\xi, \quad n = 0, 1, \dots,$$

by (3), proving the first claim. Furthermore, for any  $K \in \mathbb{R}$ ,

$$\begin{aligned} E_n[|Z|^a \mathbf{1}(|Z| \geq K)] &= E_n[|Z|^a \mathbf{1}(e^{\bar{\theta}|Z|} \geq e^{\bar{\theta}K})] \leq e^{-\bar{\theta}K} E_n[|Z|^a e^{\bar{\theta}|Z|} \mathbf{1}(e^{\bar{\theta}|Z|} \geq e^{\bar{\theta}K})] \\ &\leq e^{-\bar{\theta}K} E_n[|Z|^a e^{\bar{\theta}|Z|}] \\ &\leq C\xi e^{-\bar{\theta}K}. \end{aligned} \quad (61)$$

Now given  $\epsilon > 0$ , choosing  $K$  such that (61) is less than  $\epsilon/2$ , we obtain

$$\begin{aligned}
\|z_i^a h_n(z_r) - z_i^a h(z_r)\|_1 &= \int |z_i|^a |h_n(z_r) - h(z_r)| dz_r \\
&= \int_{|z_i| < K} |z_i|^a |h_n(z_r) - h(z_r)| dz_r + \int_{|z_i| \geq K} |z_i|^a |h_n(z_r) - h(z_r)| dz_r \\
&\leq K^a \|h_n(z_r) - h(z_r)\|_1 + \int_{|z_i| \geq K} |z_i|^a h_n(z_r) dz_r + \int_{|z_i| \geq K} |z_i|^a h(z_r) dz_r \\
&= K^a \|h_n(z_r) - h(z_r)\|_1 + \int_{|z| \geq K} |z|^a h_n(z) dz + \int_{|z| \geq K} |z|^a h(z) dz \\
&\leq K^a \|h_n(z_r) - h(z_r)\|_1 + \epsilon.
\end{aligned}$$

By Lemma 7.1  $\|h_n^{1/2}(z_r) - h^{1/2}(z_r)\|_2 \rightarrow 0$ , and therefore by Lemma 7.2,  $\|h_n(z_r) - h(z_r)\|_1 \rightarrow 0$ . Thus, letting  $n$  tend to infinity and noting that  $\epsilon > 0$  is arbitrary completes the proof of the first claim in (60). The second claim now follows since

$$E_n[Z^a] = \int z^a h_n(z) dz = \int z_i^a h_n(z_r) dz_r \rightarrow \int z_i^a h(z_r) dz_r = \int z^a h(z) dz = E[Z^a].$$

□

In the following three subsections we show the three terms (48), (47) and (46) tend to zero, thus completing the proof of Lemma 3.1.

## 7.2 B Operator

The goal of this section is to prove Lemma 7.5, showing that the term (48) tends to zero.

**Lemma 7.5.** *For any  $\beta \in \mathcal{B}$  and  $\{h_n\}_n^\infty \in \mathcal{C}_2(h, \beta)$ , with  $B$  given by (31), we have*

$$\|\sqrt{n} \left( f(\theta_0, g, h_n)^{1/2} - f_0^{1/2} \right) - B\beta\|_\sigma \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

**Proof:** We have

$$\begin{aligned}
&\sqrt{n} \left( f(\theta_0, g, h_n)^{1/2} - f_0^{1/2} \right) - B\beta \\
&= \sqrt{n} \left( h_n^{1/2}(z_r) - h^{1/2}(z_r) \right) \left( K_{\eta, m} g(t) \overline{G}(t)^{\eta-1} \varrho(\eta) \right)^{1/2} - \left( \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j) \right) f_0^{1/2} \\
&= \left[ \sqrt{n} \left( h_n^{1/2}(z_r) - h^{1/2}(z_r) \right) - h^{1/2}(z_r) \left( \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j) \right) \right] \left( K_{\eta, m} g(t) \overline{G}(t)^{\eta-1} \varrho(\eta) \right)^{1/2}.
\end{aligned}$$

Squaring the above term, summing over  $i, r$  then integrating over  $t$ , summing over  $\eta$  and lastly integrating over  $z_r$ , we obtain

$$\left\| \sqrt{n} \left( h_n^{1/2}(z_r) - h^{1/2}(z_r) \right) - h^{1/2}(z_r) \left( \sum_{j=1}^m h^{-1/2}(z_j) \beta(z_j) \right) \right\|_{\nu^m},$$

which tends to zero by Lemma 7.1. □

### 7.3 A Operator

The goal of this section is to prove Lemma 7.6, showing that the term (47) tends to zero.

**Lemma 7.6.** *For any  $\alpha \in \mathcal{A}$  and  $\beta \in \mathcal{B}$  and  $\{g_n\}_{n \geq 0} \in \mathcal{C}_1(g, \alpha)$  and  $\{h_n\}_n^\infty \in \mathcal{C}_2(h, \beta)$ , with  $A$  as in (30),*

$$\|\sqrt{n} (f(\theta_0, g_n, h_n)^{1/2} - f(\theta_0, g, h_n)^{1/2}) - A\alpha\|_\sigma \rightarrow 0.$$

**Proof:** Adding and subtracting terms we may write

$$\begin{aligned} & \sqrt{n}(f(\theta_0, g_n, h_n)^{1/2} - f(\theta_0, g, h_n)^{1/2}) - A\alpha \\ = & (K_{\eta,m}h_n(z_r)\varrho(\eta))^{1/2}\sqrt{n}(g_n(t)^{1/2} - g^{1/2}(t)) (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2}) \\ + & (K_{\eta,m}\varrho(\eta))^{1/2} (h_n^{1/2}(z_r) - h^{1/2}(z_r)) \sqrt{n}(g_n(t)^{1/2} - g^{1/2}(t))\overline{G}(t)^{(\eta-1)/2} \\ + & (K_{\eta,m}h(z_r)\varrho(\eta))^{1/2}\sqrt{n}(g_n(t)^{1/2} - g^{1/2}(t))\overline{G}(t)^{(\eta-1)/2} - g^{-1/2}\alpha f_0^{1/2} \\ + & (K_{\eta,m}\varrho(\eta))^{1/2} \sqrt{n}(h_n^{1/2}(z_r) - h^{1/2}(z_r))g(t)^{1/2} (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2}) \\ + & (K_{\eta,m}h(z_r)\varrho(\eta))^{1/2}g(t)^{1/2}\sqrt{n} (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2}) - \frac{(\eta-1)\int_t^\infty g^{1/2}\alpha}{\overline{G}(t)}f_0^{1/2}. \end{aligned} \tag{62}$$

We show that each of the five expressions above tend to zero in  $L^2(\sigma)$ . Calculating the  $L^2(\sigma)$  norm of the first term, we obtain

$$\int_{\mathcal{X}} K_{\eta,m}h_n(z_r)\alpha_n^2(t) (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2})^2 d\sigma = E[\eta \int_0^\infty \alpha_n^2(t) (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2})^2 dt],$$

which tends to zero by the dominated convergence theorem and Lemma 7.3, in particular, using that  $E[\eta]$  is finite,  $\|\alpha_n\|_{\nu^+}$  is bounded and  $\|\overline{G}_n - \overline{G}\|_\infty \rightarrow 0$  as  $n \rightarrow \infty$ .

For the second term,

$$\int_{\mathcal{X}} K_{\eta,m} (h_n^{1/2}(z_r) - h^{1/2}(z_r))^2 \alpha_n(t)\overline{G}(t)^{\eta-1} d\sigma \leq E[\eta] \|h_n^{1/2}(z_r) - h^{1/2}(z_r)\|_{\nu_m}^2 \|\alpha_n\|_2^2,$$

which turns to zero by the uniform boundedness of  $\|\alpha_n\|_2$  and Lemma 7.1.

For the third term

$$\begin{aligned} \int_{\mathcal{X}} K_{\eta,m}\overline{G}^{\eta-1}(t)h(z_r)(\alpha_n(t) - \alpha(t))^2 d\sigma &= \sum_{\eta} \int_0^\infty \eta \overline{G}(t)^{\eta-1} (\alpha_n(t) - \alpha(t))^2 dt \varrho(\eta) \\ &\leq E[\eta] \|\alpha_n - \alpha\|_2^2 \rightarrow 0. \end{aligned}$$

For the fourth term, we note

$$\begin{aligned} & \int_{\mathcal{X}} K_{\eta,m}\varrho(\eta)g(t) (\sqrt{n}(h_n^{1/2}(z_r) - h^{1/2}(z_r)))^2 (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2})^2 d\sigma \\ = & \|\sqrt{n}(h_n^{1/2}(z_r) - h^{1/2}(z_r))\|_2^2 E[\eta \int_0^\infty g(t) (\overline{G}_n(t)^{(\eta-1)/2} - \overline{G}(t)^{(\eta-1)/2})^2 dt]. \end{aligned}$$

which tends to zero by virtue of (50), (59),  $E[\eta] < \infty$  and the dominated convergence theorem.

Moving now to the last term, let  $A_n(t)$  be as in (55), that is,

$$\begin{aligned} A_n(t) &= \sqrt{n}(\overline{G}_n(t) - \overline{G}(t)) = \sqrt{n} \int_t^\infty (g_n(u) - g(u))du \\ &= \int_t^\infty (g_n^{1/2}(u) + g^{1/2}(u))\alpha_n(u)du, \end{aligned} \quad (63)$$

and define

$$R_n(t) = r(A_n(t)) \quad \text{where} \quad r(x) = (1+x)^{(\eta-1)/2} - 1 - \frac{\eta-1}{2}x. \quad (64)$$

Now gathering terms and using  $\overline{G}_n(t)/\overline{G}(t) = 1 + A_n(t)/(\sqrt{n}\overline{G}(t))$ , we have that the last term in (62) equals

$$\begin{aligned} & f_0^{1/2} \left( \sqrt{n} \left( \left( \frac{\overline{G}_n(t)}{\overline{G}(t)} \right)^{(\eta-1)/2} - 1 \right) - \frac{(\eta-1) \int_t^\infty g^{1/2}\alpha}{\overline{G}(t)} \right) \\ &= (\eta-1) f_0^{1/2} \left( \frac{A_n(t)}{2\overline{G}(t)} - \frac{\int_t^\infty g^{1/2}\alpha}{\overline{G}(t)} \right) + \sqrt{n} f_0^{1/2} R_n(t) \\ &= (\eta-1) (K_{\eta,m} g(t) \overline{G}(t)^{\eta-3} h(z_r) \varrho(\eta))^{1/2} \left( \frac{A_n(t)}{2} - \int_t^\infty g^{1/2}\alpha \right) + \sqrt{n} f_0^{1/2} R_n(t). \end{aligned} \quad (65)$$

By (63) and adding and subtracting terms,

$$\begin{aligned} & \frac{A_n(t)}{2} - \int_t^\infty g^{1/2}\alpha = \int_t^\infty \left( \left( \frac{g_n^{1/2}(u) + g^{1/2}(u)}{2} \right) \alpha_n(u) - g^{1/2}\alpha \right) du \\ &= \int_t^\infty \left( \frac{g_n^{1/2} + g^{1/2}}{2} \right) (\alpha_n(u) - \alpha(u)) du + \int_t^\infty \left( \frac{g_n^{1/2} + g^{1/2}}{2} - g^{1/2} \right) \alpha(u) du. \end{aligned}$$

Applying the Cauchy Schwarz inequality, we see that the square of the first term is bounded by

$$\int_t^\infty \left( \frac{g_n^{1/2} + g^{1/2}}{2} \right)^2 \int_t^\infty (\alpha_n(u) - \alpha(u))^2 \leq 2 \int_0^\infty \left( \frac{g_n(u) + g(u)}{4} \right) du \|\alpha_n - \alpha\|_2^2 = \|\alpha_n - \alpha\|_2^2,$$

and similarly, for the square of the second term we have

$$\left( \int_t^\infty \left( \frac{g_n^{1/2} + g^{1/2}}{2} - g^{1/2}(u) \right) \alpha(u) du \right)^2 = n^{-1} \left( \frac{1}{2} \int_t^\infty \alpha_n(u) \alpha(u) du \right)^2 \leq \frac{1}{4n} \|\alpha_n\|_2^2 \|\alpha\|_n^2.$$

Hence, for all  $t \in \mathbb{R}^+$ ,

$$\left( \frac{A_n(t)}{2} - \int_t^\infty g^{1/2}\alpha \right)^2 \leq C_n \quad \text{where} \quad C_n = 2 \left( \|\alpha_n - \alpha\|_2^2 + \frac{1}{4n} \|\alpha_n\|_2^2 \|\alpha\|_n^2 \right)$$

and the first term in (65) has  $L^2(\sigma)$  norm bounded by

$$C_n \int_{\mathcal{X}} (\eta-1)^2 K_{\eta,m} g(t) \overline{G}(t)^{\eta-3} h(z_r) \varrho(\eta) d\sigma = C_n E \left[ \frac{\eta(\eta-1)^2}{\eta-2} \right] \leq C_n 2E[\eta^2],$$

which we see tends to zero, using  $E[\eta^2] < \infty$ ,

Lastly, we prove that the second term of (65) goes to zero in  $L^2(\sigma)$ . By Taylor's theorem applied to the function  $r(x)$  in (64), for every  $t \in \mathbb{R}^+$  there exists  $\xi_t$  between 0 and 1 such that

$$\begin{aligned} R_n(t) &= \frac{(\eta-1)(\eta-3)}{8} \left(1 + \xi_t \frac{A_n(t)}{\sqrt{n\bar{G}(t)}}\right)^{(\eta-5)/2} \frac{A_n^2(t)}{n\bar{G}^2(t)} \\ &= \frac{(\eta-1)(\eta-3)}{8} \bar{G}(t)^{(5-\eta)/2} \left(\bar{G}(t) + \xi_t \frac{A_n(t)}{\sqrt{n}}\right)^{(\eta-5)/2} \frac{A_n^2(t)}{n\bar{G}^2(t)} \\ &= \frac{(\eta-1)(\eta-3)}{8} \bar{G}(t)^{(5-\eta)/2} \left((1-\xi_t)\bar{G}(t) + \xi_t \bar{G}_n(t)\right)^{(\eta-5)/2} \frac{A_n^2(t)}{n\bar{G}^2(t)}. \end{aligned}$$

Since  $|(1-\xi_t)\bar{G}(t) + \xi_t \bar{G}_n(t)| \leq 1$ , we have

$$|R_n(t)|^2 \leq \frac{1}{64n^2} \eta^4 \bar{G}(t)^{1-\eta} A_n^4(t).$$

Now, multiplying by  $nf_0$ , integrating, and using (57) of Lemma 7.3 for the bound  $|A_n(t)| \leq 2\|\alpha_n\|_2$  we have

$$\left| \frac{1}{64n} \int_{\mathcal{X}} \eta^4 \bar{G}(t)^{1-\eta} A_n^4(t) f_0 d\sigma \right| = \left| \frac{1}{64n} \int_{\mathcal{X}} \eta^4 A_n^4(t) K_{\eta,m} g(t) h(z_r) \varrho(\eta) d\sigma \right| \leq \frac{1}{n} \|\alpha_n\|_2^4 E[\eta^5] \rightarrow 0.$$

## 7.4 Parametric Score

The goal of this subsection is to prove Lemma 7.13, showing that the term (46) tends to zero; we continue to occasionally write  $g_0$  and  $h_0$  for  $g$  and  $h$ , respectively, and let  $(\{g_n\}_{n \geq 0}, \{h_n\}_{n \geq 0}) \in \mathcal{C}_1(g, \alpha) \times \mathcal{C}_2(h, \beta)$  for some  $\alpha \in \mathcal{A}$  and  $\beta \in \mathcal{B}$ .

First we extend definitions (21) and (22) for  $\bar{G}_\theta(t; z)$ ,  $g_\theta(t; z)$  and  $\bar{G}_\theta(t)$ , the covariate  $z$  survival, density, and averaged survival, respectively, to a sequence  $g_n, h_n$ . Let  $\bar{G}_n(t) = \int_t^\infty g_n(u) du$ , the survival function corresponding to the density  $g_n$ , and define

$$\bar{G}_{n,\theta}(t; z) = \bar{G}_n^{e^{\theta z}}(t), \quad \bar{G}_{n,\theta}(t) = \int \bar{G}_{n,\theta}(t; z) h_n(z) dz, \quad (66)$$

and

$$g_{n,\theta}(t; z) = \begin{cases} e^{\theta z} g_n(t) \bar{G}_n^{e^{\theta z}-1}(t) & \text{when } \bar{G}_n(t) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

In a slight abuse of notation, unwritten  $n$  and  $\theta$  will correspond to the null, so we have

$$\bar{G}_n(t) = \bar{G}_{n,0}(t) \quad \text{while} \quad \bar{G}_\theta(t) = \bar{G}_{0,\theta}(t).$$

In the following lemma we collect some facts about the averaged survival function  $\bar{G}_{n,\theta}(t)$ ; again  $C$  will be used to denote a constant, not necessarily the same at each occurrence.

**Lemma 7.7.** *The survival function  $\bar{G}_{n,\theta}(t)$  given in (66) has density*

$$g_{n,\theta}(t) = \int g_{n,\theta}(t; z) h_n(z) dz, \quad t \in \mathbb{R}^+,$$

and the partial derivative of  $\overline{G}_{n,\theta}(t)$  with respect to  $\theta$  exists for  $|\theta| < \theta_\kappa$  and is given by

$$\frac{\partial \overline{G}_{n,\theta}(t)}{\partial \theta} = E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta Z}}] \log \overline{G}_n(t). \quad (67)$$

**Proof:** Let  $\bar{\theta} \in (|\theta|, \theta_\kappa)$ . We have  $g_n(t) = 0$  a.e. on  $\overline{G}_n(t) = 0$ , so on this set the given  $g_{n,\theta}(t)$  is a version of the density of  $\overline{G}_{n,\theta}(t)$ . Otherwise, we have

$$\left| \frac{\partial}{\partial t} \overline{G}_{n,\theta}(t; z) \right| \leq (e^{\bar{\theta}z} + e^{-\bar{\theta}z}) \frac{g_n(t)}{\overline{G}_n(t)},$$

which is integrable with respect to  $h_n(z)$  since  $M_{h_n}(\bar{\theta}) + M_{h_n}(-\bar{\theta})$  is finite, and hence the derivative in  $t$  may be passed through the integral in definition (66) of  $\overline{G}_{n,\theta}(t)$  to obtain the density  $g_{n,\theta}(t; z)$ .

Next, it is easily verified that

$$\frac{\partial}{\partial \theta} \overline{G}_{n,\theta}(t; z) = z \overline{G}_{n,\theta}(t; z) \log \overline{G}_{n,\theta}(t; z) = z e^{\theta z} \overline{G}_n(t)^{e^{\theta z}} \log \overline{G}_n(t),$$

and hence there exists a neighborhood around  $\theta$  for which this partial is bounded by a constant times  $(e^{\bar{\theta}z} + e^{-\bar{\theta}z}) |\log \overline{G}_n(t)|$ , an integrable function in  $z$ . Hence, the partial in  $\theta$  may be passed through the integral in definition (66) of  $\overline{G}_{n,\theta}(t)$  yielding (67).  $\square$

In prove our results under the cohort size condition stated in Theorem 2.5, in this subsection only, we denote

$$f_{n,l}(\theta) = K_{\eta-l,m} e^{\theta z_i} g_n(t) \overline{G}_n(t)^{\sum_{j \in r} e^{\theta z_j} - 1} \overline{G}_{n,\theta}(t)^{(\eta-l)-m} h_n(z_r) \varrho(\eta) \quad \text{and} \quad f_n(\theta) = f_{n,0}(\theta). \quad (68)$$

In order for  $f_{n,l}(\theta)$  to be a density we require  $1 \leq m \leq \eta - l$ , as it then corresponds to the situation where an initial  $\eta$  is generated by  $\varrho(\eta)$ , but which then leads to a cohort of size  $\eta - l$ , from which  $m$  are sampled.

**Lemma 7.8.** *Let  $\theta_\xi$  be as in (3). If the positivity (boundedness) condition holds, then for all  $\bar{\theta} \in (0, \theta_\xi)$  the partial derivative of  $f_n^{1/2}(\theta)$  with respect to  $\theta$  exists for all  $0 \leq \theta \leq \bar{\theta}$  ( $|\theta| \leq \bar{\theta}$ ), and is given by*

$$\frac{\partial}{\partial \theta} f_n^{1/2}(\theta) = D_1(\theta) + D_2(\theta) + D_3(\theta)$$

where

$$\begin{aligned} D_1(\theta) &= \frac{1}{2} z_i f_n^{1/2}(\theta), \quad D_2(\theta) = \frac{1}{2} \log \overline{G}_n(t) \sum_{j \in r} z_j e^{\theta z_j} f_n^{1/2}(\theta) \\ \text{and } D_3(\theta) &= C_n \eta \log \overline{G}_n(t) f_n^{1/2}(\theta), \end{aligned} \quad (69)$$

where  $C_n$ , short for  $C_{n,\eta,m,\theta,t}$ , is uniformly bounded in  $n, \eta, m \in \{1, \dots, \eta\}$ ,  $0 \leq \theta \leq \bar{\theta}$ , ( $|\theta| \leq \bar{\theta}$ ) and  $t \geq 0$ .

In general, when  $1 \leq m \leq \eta - 2$  the same conclusion holds with  $f_{n,2}(\theta)$  replacing  $f_n(\theta)$  in (69), with the bound on the constant holding in the same range as for the bounded covariate case.

In any of the three scenarios, at the null  $\theta_0$ ,

$$\frac{\partial}{\partial \theta} f_n^{1/2}(\theta_0) = \frac{1}{2} \left[ z_i + \log \overline{G}_n(t) \sum_{j \in r} (z_j - E_n[Z]) + [\log \overline{G}_n(t)] \eta E_n[Z] \right] f_n^{1/2}(\theta_0). \quad (70)$$



**Proof:** The terms  $D_1(\theta)$ ,  $D_2(\theta)$  are obtained by differentiating  $f_n(\theta)$  in the terms  $e^{\theta z_i}$  and  $\overline{G}_n(t)^{\sum_{j \in r} e^{\theta z_j}}$ , respectively. Applying Lemma 7.7, the term which results from differentiating the term  $\overline{G}_n(t)^{\eta-m}$  is

$$\begin{aligned} D_3(\theta) &= \frac{1}{2}(\eta - m) \frac{\partial \overline{G}_{n,\theta}(t)}{\partial \theta} \overline{G}_{n,\theta}(t)^{-1} f_n^{1/2}(\theta) \\ &= \frac{1}{2}(\eta - m) E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}] \log \overline{G}_n(t) (\overline{G}_{n,\theta}(t)^{-1} f_n^{1/2}(\theta)). \end{aligned} \quad (71)$$

Hence, (69) holds with

$$C_n = \left( \frac{\eta - m}{2\eta} \right) \frac{E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}]}{\overline{G}_{n,\theta}(t)}.$$

Under the positivity condition  $z \geq 0$  and  $\theta \geq 0$ , the function  $z e^{\theta z}$  is increasing in  $z$  and  $\overline{G}_n(t)^{e^{\theta z}}$  is decreasing in  $z$ . Therefore these functions are negatively correlated and we have

$$E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}] \leq E_n[Z e^{\theta Z}] E_n[\overline{G}_n(t)^{e^{\theta z}}] = E_n[Z e^{\theta Z}] \overline{G}_{n,\theta}(t),$$

and in particular

$$|C_n| \leq \frac{1}{2} E_n[Z e^{\theta Z}],$$

which is a bounded sequence in  $n$  by Lemma 7.4.

Under the bounded covariate condition, with, say,  $|Z| \leq z_0$  almost surely, we have

$$|C_n| \leq \frac{1}{2} \frac{E_n[|Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}|]}{E_n[\overline{G}_n(t)^{e^{\theta z}}]} \leq \frac{1}{2} z_0 e^{|\theta| z_0} \leq \frac{1}{2} z_0 e^{\bar{\theta} z_0}.$$

Under the cohort size condition, write

$$D_3(\theta) = \frac{1}{2}(\eta - m) E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}] \left( \frac{K_{\eta,m}}{K_{\eta-2,m}} \right)^{1/2} \log \overline{G}_n(t) f_{n,2}^{1/2}(\theta),$$

which is (69) with  $f_{n,2}(\theta)$  replacing  $f_n(\theta)$  and

$$C_n = \left( \frac{\eta - m}{2\eta} \right) E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}] \left( \frac{K_{\eta,m}}{K_{\eta-2,m}} \right)^{1/2}.$$

Since  $|E_n[Z e^{\theta Z} \overline{G}_n(t)^{e^{\theta z}}]| \leq E_n[|Z| e^{\theta Z}]$  is bounded in  $n$  by Lemma 7.4, and  $K_{\eta,m}/K_{\eta-2,m} \leq 1$ , the constant is uniformly bounded.

At the null, expression (71) becomes

$$D_3(\theta_0) = \frac{1}{2}(\eta - m) E_n[Z] \log \overline{G}_n(t) f_n^{1/2}(\theta_0),$$

and adding this to  $D_1(\theta_0)$  and  $D_2(\theta_0)$  and rearranging gives (70).  $\square$

For  $a \in \mathbb{R}^m$  let  $z_r^a = \prod_{j \in r} z_j^{a_j}$  and  $a' z_r = \sum_{j \in r} a_j z_j$ . We will say an expression is a  $T_{n,l}$  term, and denote it by  $T_{n,l}$  for short, if there exists a bounded sequence  $M_n$ , a nonnegative integer  $\iota$ , a real number  $k$ , and  $a, b \in \mathbb{R}^m$  such that the expression can be written

$$T_{n,l}(M_n, \iota, k, a, b) = M_n \eta^\iota (\log \overline{G}_n(t))^k z_r^a e^{\theta b' z_r} f_{n,l}^{1/2}(\theta). \quad (72)$$

**Lemma 7.9.** *Let  $\theta_\xi$  be as in (3). Under the positivity (boundedness) condition, for all  $\bar{\theta} \in (0, \theta_\xi)$ , the partial derivative of  $T_{n,0}$  exists for all  $0 \leq \theta \leq \bar{\theta}$  ( $|\theta| \leq \bar{\theta}$ ), and can be written as a finite sum of  $T_{n,0}$  terms. In general, when  $l \geq 0$  and  $m \leq \eta - (l + 2)$  the partial derivative of  $T_{n,l}$  exists for all  $|\theta| \leq \bar{\theta}$ , and can be written as a finite sum of  $T_{n,l}$  and  $T_{n,l+2}$  terms.*

*In particular, under the positivity (boundedness) condition the first and second partial derivatives of  $f_n^{1/2}(\theta)$  with respect to  $\theta$  exist for all  $0 \leq \theta \leq \bar{\theta}$  ( $|\theta| \leq \bar{\theta}$ ), and can be written as a sum of  $T_{n,0}$  terms with  $0 \leq \iota \leq 2$ . In general, when  $m \leq \eta - 4$  almost surely, these first two partial derivatives exist for all  $|\theta| \leq \bar{\theta}$  and can be written as sums of  $T_{n,0}, T_{n,2}$  and  $T_{n,4}$  terms with  $0 \leq \iota \leq 2$ , so in particular in terms of functions  $f_{n,\iota}(\theta)$  which are densities.*

**Proof:** Letting  $e_j$  be the unit vector in direction  $j$  for  $j = 1, \dots, m$ , the partial derivative  $(\partial/\partial\theta)T_{n,0}(M_n, k, a, b)$  equals

$$\sum_{j \in r} T_{n,0}(b_j M_n, \iota, k, a + e_j, b) + M_n \eta^\iota (\log \bar{G}_n(t))^k z_r^a e^{\theta b' z_r} \frac{\partial}{\partial \theta} f_n^{1/2}(\theta). \quad (73)$$

By Lemma 7.8, under the positivity (boundedness) condition, and recalling the form (72), we may write

$$\frac{\partial}{\partial \theta} f_n^{1/2}(\theta) = T_{n,0} \left( \frac{1}{2}, 0, 0, e_i, 0 \right) + \sum_{j \in r} T_{n,0} \left( \frac{1}{2}, 0, 1, e_j, e_j \right) + T_{n,0}(C_n, 1, 1, 0, 0),$$

for  $0 \leq \theta \leq \bar{\theta}$  ( $|\theta| \leq |\bar{\theta}|$ ). Hence the last term in (73) equals

$$T_{n,0} \left( \frac{M_n}{2}, \iota, k, a + e_i, b \right) + \sum_{j \in r} T_{n,0} \left( \frac{M_n}{2}, \iota, k + 1, a + e_j, b + e_j \right) + T_{n,0}(M_n C_n, \iota + 1, k + 1, a, b),$$

demonstrating the first claim.

When taking the partial derivatives of  $T_{n,l}$  for  $m \leq \eta - (l + 2)$  the terms  $T_{n,0}$  in the sum in (73) are replaced by  $T_{n,l}$  and the second term by

$$T_{n,l} \left( \frac{M_n}{2}, \iota, k, a + e_i, b \right) + \sum_{j \in r} T_{n,l} \left( \frac{M_n}{2}, \iota, k + 1, a + e_j, b + e_j \right) + T_{n,l+2}(M_n C_n, \iota + 1, k + 1, a, b),$$

with equality for all  $|\theta| \leq \bar{\theta}$ .

The claims about the partial derivatives of  $f_n^{1/2}(\theta)$  now follow by noting that this function is a  $T_{n,0}$  term with  $\iota = 0$ , and that the value of  $\iota$  increases by at most one with every differentiation, and that, in general, the first partial of  $f_n^{1/2}$  is a sum of  $T_{n,0}$  and  $T_{n,2}$  terms, and the second partial of terms  $T_{n,0}, T_{n,2}$  and  $T_{n,4}$ .  $\square$

**Lemma 7.10.** *For  $1 \leq m \leq \eta$  and all  $\iota \in \{0, 1, 2\}, k \in \mathbb{R}$  and  $a, b \in \mathbb{R}^m$  there exists  $\bar{\theta} > 0$  such that*

$$\sup_{n, |\theta| \leq \bar{\theta}} \|\eta^\iota (\log \bar{G}_n(t))^k z_r^a e^{\theta b' z_r} f_n^{1/2}(\theta)\|_\sigma < \infty.$$

**Proof:**

$$\begin{aligned}
& \|\eta'(\log \bar{G}_n(t))^k z_r^a e^{\theta b' z_r} f_n^{1/2}(\theta)\|_\sigma^2 = \int_{\mathcal{X}} \eta'(\log \bar{G}(t))^{2k} z_r^{2a} e^{2\theta b' z_r} f_n(\theta) d\sigma \\
&= \sum_{\eta, i, r} K_{\eta, m} \eta' \int_{z_r} \int_t (\log \bar{G}_n(t))^{2k} z_r^{2a} e^{2\theta b' z_r} e^{\theta z_i} g_n(t) \bar{G}_n^{\sum_{j \in r} e^{\theta z_j} - 1}(t) \bar{G}_{n, \theta}^{\eta - m}(t) h_n(z_r) \varrho(\eta) dt dz_r \\
&= \sum_{\eta} \eta'^{+1} \int_{z_{[m]}} \int_t (\log \bar{G}_n(t))^{2k} z_{[m]}^{2a} e^{2\theta b' z_{[m]}} e^{\theta z_1} g_n(t) \bar{G}_n^{\sum_{j \in [m]} e^{\theta z_j} - 1}(t) \bar{G}_{n, \theta}^{\eta - m}(t) h_n(z_{[m]}) dt dz_{[m]} \varrho(\eta) \\
&\leq E[\eta'^{+1}] \int_{z_{[m]}} \int_t (\log \bar{G}_n(t))^{2k} z_{[m]}^{2a} e^{2\theta b' z_{[m]}} e^{\theta z_1} g_n(t) \bar{G}_n^{e^{\theta z_1} - 1}(t) h_n(z_{[m]}) dt dz_{[m]} \\
&= E[\eta'^{+1}] \prod_{j \neq 1} E_n[Z^{2a_j} e^{2\theta b_j Z}] \int_z z^{2a_1} e^{2\theta(b_1 - k)z} \int_t e^{2k\theta z} (\log \bar{G}_n(t))^{2k} e^{\theta z} g_n(t) \bar{G}_n(t)^{e^{\theta z} - 1} h_n(z) dt dz \\
&= E[\eta'^{+1}] \prod_{j \neq 1} E_n[Z^{2a_j} e^{2\theta b_j Z}] \int_z z^{2a_1} e^{2\theta(b_1 - k)z} \left( \int_t (\log \bar{G}_{n, \theta}(t; z))^{2k} g_{n, \theta}(t; z) dt \right) h_n(z) dz \\
&= E[\eta'^{+1}] (2k)! \prod_{j \neq 1} E_n[Z^{2a_j} e^{2\theta b_j Z}] \int_z z^{2a_1} e^{2\theta(b_1 - k)z} h_n(z) dz \\
&= E[\eta'^{+1}] (2k)! E_n[Z^{2a_1} e^{2\theta(b_1 - k)Z}] \prod_{j \neq 1} E_n[Z^{2a_j} e^{2\theta b_j Z}]
\end{aligned}$$

where Lemma 4.3 was applied to obtain the second to last equality. Now use Lemma 7.4 to verify there exists a sufficiently small interval around zero for  $\theta$  such that the supremum of this final quantity, over all  $n$  and such  $\theta$ , is finite.  $\square$

As in the proof of Lemma 1 of [2], we will make use of Lemma 7.11 below, there termed Vitali's theorem; see also, for instance, Chapter 2 of [12], Exercises 21 and 33.

**Lemma 7.11.** *If  $w_n, n = 0, 1, \dots$  are nonnegative functions in  $L^1(\mu)$  such that  $w_n \rightarrow w_0$  in  $\mu$  measure, then*

$$\int w_n d\mu \rightarrow \int w_0 d\mu \quad \text{implies} \quad w_n^{1/2} \rightarrow w_0^{1/2} \quad \text{in } L^2(\mu).$$

*In particular, the conclusion holds if  $w_n \rightarrow w_0$  in  $L^1(\mu)$ .*

Consistent with the notation in (68), we let  $f_0(\theta_0)$  denote the null density.

**Lemma 7.12.** *Let  $\eta \geq 2$ . Then for any  $k, a \in \mathbb{R}$  and any sequence of real numbers  $M_n$  converging to  $M$ ,*

$$M_n [\log \bar{G}_n(t)]^k z_i^a f_n^{1/2}(\theta_0) \rightarrow M [\log \bar{G}(t)]^k z_i^a f_0^{1/2}(\theta_0) \quad \text{in } L^2(\sigma) \text{ as } n \rightarrow \infty.$$

**Proof:** By Lemma 7.3 we have

$$\|g_n - g\|_1 \rightarrow 0 \quad \text{and} \quad \|\bar{G}_n - \bar{G}\|_\infty \rightarrow 0.$$

When  $\eta \geq 2$  the function  $\psi(x) = [\log x]^{2k} x^{\eta-1}$  is uniformly continuous on  $[0, 1]$ , and therefore

$$\lim_{n \rightarrow \infty} \|\psi(\bar{G}_n) - \psi(\bar{G})\|_\infty = 0 \quad \text{and} \quad \|\psi(\bar{G})\|_\infty < \infty, \quad (74)$$

implying

$$g_n(t)\psi(\overline{G}_n) \rightarrow g(t)\psi(\overline{G}) \quad \text{in } L^1(\nu^+).$$

Now Lemma 7.4 gives

$$z_i^{2a}h_n(z_r) \rightarrow z_i^{2a}h(z_r) \quad \text{in } L^1(\nu^m) \text{ for all } r \subset [\eta] \text{ of size } m \text{ and } i \in r,$$

and therefore, for any fixed  $\eta \geq 2$ ,

$$K_{\eta,m}[\log \overline{G}_n(t)]^{2k} z_i^{2a} g_n(t) \overline{G}_n(t)^{\eta-1} h_n(z_r) \varrho(\eta) \rightarrow K_{\eta,m}[\log \overline{G}(t)]^{2k} z_i^{2a} g(t) \overline{G}(t)^{\eta-1} h(z_r) \varrho(\eta)$$

in  $L^1(\nu^+ \times \nu^m)$  for each such  $i$  and  $r$ . As the choices of such  $i$  and  $r$  are finite, this convergence also holds in  $L^1(\text{counting} \times \text{counting} \times \nu^+ \times \nu^m)$ . Again invoking Lemma 7.4, and using (74), we see that the convergence holds in  $L^1(\sigma)$ , and clearly continues to hold when multiplying by a convergent sequence. Invoking Lemma 7.11 concludes the argument.  $\square$

**Lemma 7.13.** *Let  $\theta_n \in \Theta(\tau)$  and  $\rho_0$  be given by (29). Then*

$$\|\sqrt{n} (f_n^{1/2}(\theta_n) - f_n^{1/2}(\theta_0)) - \tau \rho_0\|_\sigma \rightarrow 0. \quad (75)$$

**Proof:** Let  $\tau_n = \sqrt{n}(\theta_n - \theta_0)$ . By Taylor expansion with remainder around  $\theta_0$  we may write the difference in (75) as

$$\tau \left( \frac{\partial}{\partial \theta} f_n^{1/2}(\theta_0) - \rho_0 \right) + (\tau_n - \tau) \frac{\partial}{\partial \theta} f_n^{1/2}(\theta_0) + \frac{1}{2} \tau_n (\theta_n - \theta_0) \frac{\partial^2}{\partial \theta^2} f_n^{1/2}(\xi_n) \quad (76)$$

where  $\xi_n$  lies between  $\theta_n$  and  $\theta_0$ ; we show each of these three terms tends to zero in  $L^2(\sigma)$ .

For the first term, applying (70) and (29) we have

$$\begin{aligned} & 2 \left( \frac{\partial}{\partial \theta} f_n^{1/2}(\theta_0) - \rho_0 \right) \\ &= z_i \left( f_n^{1/2}(\theta_0) - f_0^{1/2}(\theta_0) \right) \\ &+ \sum_{j \in r} \left( \log \overline{G}_n(t)(z_j - E_n[Z]) f_n^{1/2}(\theta_0) - \log \overline{G}(t)(z_j - E[Z]) f_0^{1/2}(\theta_0) \right) \\ &+ \eta \left( \log \overline{G}_n(t) E_n[Z] f_n^{1/2}(\theta_0) - \log \overline{G}(t) E[Z] f_0^{1/2}(\theta_0) \right) \end{aligned}$$

all of which tend to zero in  $L^2(\sigma)$  by Lemma 7.12. For the second term on the right hand side of (76), since  $\tau_n \rightarrow \tau$ , it suffices that

$$\left\| \frac{\partial}{\partial \theta} f_n^{1/2}(\theta_0) \right\|_\sigma < \infty,$$

but this is a consequence of Lemmas 7.9 and 7.10. For the third term of (76), since  $\tau_n \rightarrow \tau$ ,  $\theta_n \rightarrow \theta_0$ , and  $\xi_n \rightarrow \theta_0$ , it suffices to show that for some  $\bar{\theta} > 0$ ,

$$\sup_{|\theta| \leq \bar{\theta}} \left\| \frac{\partial^2}{\partial \theta^2} f_n^{1/2}(\theta) \right\|_\sigma < \infty.$$

Again, this bound follows from Lemmas 7.9 and 7.10.  $\square$

## 8 Verification of Assumptions 2.1 and 2.3

This following lemma shows that Assumption 2.1 is satisfied.

**Lemma 8.1.** *Let  $g \in \mathcal{G}$  and  $h \in \mathcal{H}$  where  $\mathcal{G}$  and  $\mathcal{H}$  are as given in (2) and (3), respectively. Then the set of perturbations  $\mathcal{A}$  and  $\mathcal{B}$  defined in (6) and (10) are given by*

$$\mathcal{A} = \{\alpha \in L^2(\nu^+) : \langle \alpha, g^{1/2} \rangle_{\nu^+} = 0\} \quad \text{and} \quad \mathcal{B} = \bigcap_{|\theta| < \theta_\kappa} \{\beta \in L^2(\nu_\theta) : \langle \beta, h^{1/2} \rangle_\nu = 0\}, \quad (77)$$

which are subspaces of  $L^2(\nu^+)$  and  $L^2(\nu)$ , respectively.

**Proof:** Let  $\mathcal{A}'$  and  $\mathcal{B}'$  denote the sets given in (77), respectively. We first show  $\mathcal{B} = \mathcal{B}'$  by demonstrating inclusion in both directions. Let  $\beta \in \mathcal{B}$  and  $\{h_n\}_{n \geq 0} \subset \mathcal{H}$  a sequence such that

$$\|\beta_n - \beta\|_{\nu_\theta} \rightarrow 0 \quad \text{for all } |\theta| < \theta_\kappa, \quad \text{where } \beta_n = \sqrt{n}(h_n^{1/2} - h^{1/2}). \quad (78)$$

Since the squares of  $h^{1/2}$  and  $h_n^{1/2} = h^{1/2} + n^{-1/2}\beta_n$  are densities with respect to  $\nu$ ,

$$\int h^{1/2}\beta_n d\nu + \frac{1}{2}n^{-1/2} \int \beta_n^2 d\nu = 0 \quad \text{for all } n = 1, 2, \dots$$

and hence

$$\int h^{1/2}\beta d\nu = \int h^{1/2}\beta_n d\nu - \int h^{1/2}(\beta_n - \beta) d\nu = - \left( \frac{1}{2}n^{-1/2} \int \beta_n^2 d\nu + \int h^{1/2}(\beta_n - \beta) d\nu \right).$$

The right hand side tends to zero as  $n \rightarrow \infty$  as

$$n^{-1/2} \int \beta_n^2 d\nu \rightarrow 0 \quad \text{and} \quad \left( \int h^{1/2}(\beta_n - \beta) d\nu \right)^2 \leq \|\beta_n - \beta\|_\nu^2 \rightarrow 0,$$

using Lemma 7.1 for the first term and (78) with  $\theta = 0$  for the second. Therefore,  $\int h^{1/2}\beta d\nu = 0$ , that is,  $\langle \beta, h^{1/2} \rangle_\nu = 0$ . Next, let  $|\theta| < \theta_\kappa$ . Writing  $\beta = \beta_n + (\beta - \beta_n)$  we have  $\beta \in L^2(\nu_\theta)$  by virtue of

$$\beta_n^2 \leq 2n(h_n + h) \in L^1(\nu_\theta),$$

and  $\beta - \beta_n \in L^2(\nu_\theta)$  by (78). Hence  $\mathcal{B} \subset \mathcal{B}'$ .

Now let  $\beta \in \mathcal{B}'$  and define

$$\bar{h}_n^{1/2} = h^{1/2} + n^{-1/2}\beta \quad \text{and} \quad h_n^{1/2} = \frac{\bar{h}_n^{1/2}}{\|\bar{h}_n^{1/2}\|_\nu}.$$

Clearly  $h_n$  is a density with respect to  $\nu$ . In addition, by  $\langle h^{1/2}, \beta \rangle_\nu = 0$  we have

$$\|\bar{h}_n^{1/2}\|_\nu^2 = 1 + n^{-1} \int \beta^2 d\nu \quad \text{and therefore} \quad \sqrt{n} \left( 1 - \frac{1}{\|\bar{h}_n^{1/2}\|_\nu} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (79)$$

Let  $|\theta| < \theta_\kappa$ . By the triangle inequality in  $L^2(\nu_\theta)$ ,

$$\|\bar{h}_n^{1/2}\|_{\nu_\theta} \leq \|h^{1/2}\|_{\nu_\theta} + \|\beta\|_{\nu_\theta} \quad \text{for all } n = 1, 2, \dots,$$

and therefore

$$\|\sqrt{n}(\bar{h}_n^{1/2} - h_n^{1/2})\|_{\nu_\theta} = \|\bar{h}_n^{1/2}\|_{\nu_\theta} \sqrt{n} \left(1 - \frac{1}{\|\bar{h}_n^{1/2}\|_{\nu}}\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Rewriting, we obtain

$$\|\sqrt{n}(h_n^{1/2} - h^{1/2}) - \beta\|_{\nu_\theta} = \|\sqrt{n}(h_n^{1/2} - \bar{h}_n^{1/2})\|_{\nu_\theta} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

showing that there exists a sequence of densities having  $\beta$  as their limiting perturbation in  $L^2(\nu_\theta)$  for all  $|\theta| < \theta_\kappa$ . Lastly, at  $\theta_\xi$

$$\begin{aligned} M_{\bar{h}_n}(\theta_\xi) &= \int h(z) d\nu_{\theta_\xi} + 2n^{-1/2} \int \beta(z) h(z)^{1/2} d\nu_{\theta_\xi} + n^{-1} \int \beta^2(z) d\nu_{\theta_\xi} \\ &\leq M_h(\theta_\xi) + 2n^{-1/2} \|\beta\|_{\nu_{\theta_\xi}} M_h^{1/2}(\theta_\xi) + n^{-1} \|\beta\|_{\nu_{\theta_\xi}}^2. \end{aligned}$$

Since  $\beta \in L^2(\nu_{\theta_\xi})$  and  $M_h(\theta_\xi) < \infty$  the last two terms tends to zero as  $n \rightarrow \infty$ . Since  $\|\bar{h}_n^{1/2}\|_{\nu}^2 \rightarrow 1$  by (79), we have

$$\limsup_{n \rightarrow \infty} M_{h_n}(\theta_\xi) = \limsup_{n \rightarrow \infty} \frac{M_{\bar{h}_n}(\theta_\xi)}{\|\bar{h}_n^{1/2}\|_{\nu}^2} = \limsup_{n \rightarrow \infty} M_{\bar{h}_n}(\theta_\xi) \leq M_h(\theta_\xi).$$

As this same argument obtains at  $-\theta_\xi$ , we conclude

$$\limsup_{n \rightarrow \infty} (M_{h_n}(\theta_\xi) + M_{h_n}(-\theta_\xi)) \leq M_h(\theta_\xi) + M_h(-\theta_\xi) < \xi,$$

thus proving that  $\beta$  can be obtained by a sequence of density functions in  $\mathcal{H}$ . Hence,  $\mathcal{B}' \subset \mathcal{B}$ . The argument that  $\mathcal{A} = \mathcal{A}'$  is similar, but simpler.  $\square$

We remark that the support condition found in [1], that the support of the densities in  $\mathcal{G}$  must be contained in the support of  $g$ , is not needed here; the term in the proof of Lemma 1 in [2] on which it is used does not appear in the null case considered here.

The next lemma shows that Assumption 2.3 is satisfied.

**Lemma 8.2.** *The elements*

$$\hat{\alpha} = \frac{EZ}{2} [1 + \log \bar{G}(t)] g^{1/2}(t) \quad \text{and} \quad \hat{\beta} = \frac{1}{2} h^{1/2}(z) \frac{\eta - m}{m\eta} (z - EZ)$$

are members of  $\mathcal{A}$  and  $\mathcal{B}$ , respectively.

**Proof:** By Lemma 8.1, to show  $\hat{\alpha} \in \mathcal{A}$  it suffices to verify that  $\hat{\alpha} \in L^2(\nu^+)$  and  $\langle \hat{\alpha}, g^{1/2} \rangle_{\nu^+} = 0$ . The first claim follows from Lemma 4.3, and, by applying that same lemma with  $\eta = k$  and  $j = 1$ , the second claim from

$$\int_0^\infty \log \bar{G}(t) g(t) dt = -1.$$

By Lemma 8.1, for  $\hat{\beta} \in \mathcal{B}$  it suffices that  $\hat{\beta} \in L^2(\nu_\theta)$  for all  $|\theta| < \theta_\kappa$  and  $\langle \hat{\beta}, h^{1/2} \rangle_{\nu} = 0$ . For the first condition, fix a  $\theta$  satisfying  $|\theta| < \theta_\kappa$ . Letting  $\bar{\theta} \in (|\theta|, \theta_\kappa)$ , there exists a constant  $C$  such that

$$(z - EZ)^2 e^{\theta z} \leq C \left( e^{\bar{\theta} z} + e^{-\bar{\theta} z} \right) \quad \text{for all } z \in \mathbb{R}.$$

As  $M_h(\theta) < \infty$  whenever  $|\theta| < \theta_\kappa$ ,

$$\int h(z) (z - EZ)^2 e^{\theta z} d\nu \leq C (M_h(\bar{\theta}) + M_h(-\bar{\theta})) < \infty,$$

showing that  $\hat{\beta} \in L^2(\nu_\theta)$ . The orthogonality condition is simply that  $E(Z - EZ) = 0$ .  $\square$

## 9 Appendix

**Proof of Theorem 2.1** The set  $\mathbb{H}$  given in (14) is a subspace of  $L^2(\sigma)$ , being the image of the subspace  $\mathbb{R} \times \mathcal{A} \times \mathcal{B}$  under the linear transformation  $(\tau, \alpha, \beta) \rightarrow \tau\rho_\theta + A\alpha + B\beta$ . Hence, with  $\hat{\alpha}$  and  $\hat{\beta}$  as in Assumption 2.2, by that assumption,

$$\tau\hat{\zeta} \in \mathbb{H} \quad \text{for all } \tau \in \mathbb{R}, \text{ where } \hat{\zeta} = \rho_\theta - A\hat{\alpha} - B\hat{\beta}, \quad (80)$$

and  $4\|\tau\hat{\zeta}\|_\sigma^2 = \tau^2 I_*$ . Now let  $\{f_n\}_{n \geq 0} \in \mathcal{F}(f, \tau\hat{\zeta})$  and continue as in the proof of Theorem 3.1 in [1]. In particular, with  $L_n$  the log likelihood ratio for  $f_n$  vs  $f_0$ ,  $S$  the limiting distribution of  $n^{1/2}(\hat{\theta}_n - \theta_0)$ , guaranteed to exist by the regularity of  $\hat{\theta}_n$ , and  $Z \sim \mathcal{N}(0, I_*)$ , the random vector  $(n^{1/2}(\hat{\theta}_n - \theta), L_n)$  converges weakly under  $f$  to  $(S, \tau Z - 1/2\tau^2 I_*)$ , and the characteristic function of  $S$  factors into the product of the characteristic functions of  $S - Z/I_*$  times that of  $Z/I_*$ .  $\square$

**Proof of Theorem 2.2** Note that with  $\hat{\zeta}$  as in (80), as  $\hat{\zeta} \in \mathbb{H}$  we have that  $\mathcal{F}(f, \hat{\zeta}) \subset \mathcal{F}(f)$ , and therefore, for all  $c \geq 0$

$$B_n^*(c) = \{f_n \in \mathcal{F}(f, \hat{\zeta}) : n^{1/2}\|f_n^{1/2} - f^{1/2}\|_\sigma \leq c\} \subset \{f_n \in \mathcal{F}(f) : n^{1/2}\|f_n^{1/2} - f^{1/2}\|_\sigma \leq c\} = B_n(c).$$

Hence the argument for the proof of Theorem 3.2 of [1] obtains.  $\square$

**Proofs of Theorems 2.3 and 2.4** The application of the results of [14] and [3], as in the proofs of Theorem 4.1 and 4.2 in [1] apply with minimal changes. In particular, for any element of  $\mathbb{H}$  given by

$$\zeta = \tau\rho_\theta + A\alpha + B\beta,$$

let  $\mathcal{T} : H \rightarrow B_0 = \{x \in C[0, 1] : x(0) = x(1) = 0\}$  be defined by

$$\mathcal{T}\zeta = \int_0^t 2\alpha g^{1/2} d\nu^+ = \langle \alpha, 2g^{1/2} I_{[0,t]} \rangle_{\nu^+}.$$

Writing  $\zeta$  as

$$\zeta = \tau(\rho_\theta - A\hat{\alpha} - B\hat{\beta}) + A(\tau\hat{\alpha} + \alpha) + B(\tau\hat{\beta} + \beta),$$

the orthogonality provided by Assumption 2.3 and Lemma 4.1 yield  $A^*\zeta = A^*A(\tau\hat{\alpha} + \alpha)$  and therefore

$$\alpha = C^*\zeta - \langle \zeta, 4(\rho_\theta - A\hat{\alpha} - B\hat{\beta})/I_* \rangle \hat{\alpha} \quad \text{where } C = A(A^*A)^{-1}.$$

Continuing, one may verify that the adjoint  $\mathcal{T}^*$  of  $\mathcal{T}$  is given by a formula analagous to that in Lemma 5.2 of [1], and that

$$\frac{1}{4}\|\mathcal{T}^*v\|_\sigma^2 = E \left( \int_0^t \mathbb{Z}_* d\nu \right)^2.$$

$\square$

We remark that though the subspace  $\mathbb{H}$  is not assumed to be closed in  $L^2(\sigma)$ , and hence the projection theorem cannot be applied, as long  $\mathbb{H}$  contains the approach to  $f$  along the ‘worst case’ direction  $\hat{\zeta}$  the proof of [1] carries through. Moreover, this holds true independently of the number of factors in the model, one more here than in [1]. The other difference between the situation here and that of [1], that  $\mathcal{B}$  consists of the perturbations which approximate  $n^{1/2}(h_n^{1/2} - h^{1/2})$  in  $L^2(\nu_\theta)$  for all  $|\theta| < \theta_\kappa$  rather than in the weaker  $L^2(\nu)$  sense, is handled by Assumption 2.2, which gives, in particular, that the critical  $\hat{\beta}$  lies in  $\mathcal{B}$  even when insisting on the stronger form of convergence.

## Bibliography

1. Begun, J., Hall, J., Huang, W.M., and Wellner, J. (1983) Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.*, **11**, pp. 432-452.
2. Begun, J. and Wellner, J. (1983) Asymptotic efficiency of relative risk estimates. In *Contributions to statistics: Essays in Honor of Norman L. Johnson* ed by P.K. Sen, 47-62, North-Holland, Amsterdam, 1983.
3. Beran, R. (1977) Estimating a distribution function. *Ann. Statist.*, **5**, pp. 400-404.
4. Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1998) Efficient and adaptive estimation for semiparametric models. Springer-Verlag, New York.
5. Borgan, Ø., Goldstein, L., and Langholz, B. (1995) Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann. Statist.*, **23**, pp. 1749-1778.
6. Cox, D. R. (1972) Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, **34**, 187-220.
7. Chen, K. (2004) Statistical estimation in the proportional hazards model with risk set sampling. *Ann. Statist.*, **32**, pp. 1513-1532.
8. Goldstein, L. and Langholz, B. (1992) Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann. Statist.*, **20**, pp. 1903-1928.
9. Goldstein, L. and Langholz, B. (2008) The highly stratified Cox model. *Preprint*
10. Greenwood, P. and Wefelmeyer, W. (1990) Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.*, **36**, pp. 353-370.
11. Floderus, B., Persson, T., Stenlund, C., Wennberg, A., and Knave, B. (1993) Occupational Exposure to Electromagnetic Fields in Relation to Leukemia and Brain Tumors: A Case-Control Study in Sweden. *Cancer Causes and Control*, **4**, pp. 465-476.
12. Folland, G. (1999) Real Analysis: Modern Techniques and Their Applications. Wiley & Sons, New York.
13. Luenberger, D. (1969) *Optimization by vector space methods*, John Wiley & Sons, Inc., New York.
14. Millar, P. (1979) Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete*, **48**, pp. 233–252.
15. Robins, J., Rotnitzky, A., and Zhao, L. P. (1994) Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, **89**, pp. 846-866.
16. Thomas, D.C. (1977). Addendum to a paper by Liddel, F.D.K., McDonald, J.C., and Thomas, D.C. *J. R. Statist. Soc. A* **140**, pp. 483-485.
17. Wellner, J.A. (1985) Asymptotic optimality of the product limit estimator. *Ann. Statist.*, **10**, pp 595-602.