# Technical Report to Accompany 'Functional Inference in Semiparametric Models Using the Piggyback Bootstrap'

John R. Dixon,
Department of Statistics, Florida State University,
Michael R. Kosorok,
Department of Statistics, University of Wisconsin-Madison,
and Bee Leng Lee,
Department of Statistics and Applied Probability,
Singapore National University

**Abstract:** This paper introduces the "Piggyback Bootstrap." Like the weighted bootstrap, this bootstrap procedure can be used to generate random draws that approximate the joint sampling distribution of the parametric and nonparametric maximum likelihood estimators in various semiparametric models, but the dimension of the maximization problem for each bootstrapped likelihood is smaller. This reduction results in significant computational savings in comparison to the weighted bootstrap. The procedure can be stated quite simply. First obtain a valid random draw for the parametric component of the model. Then take the draw for the nonparametric component to be the maximizer of the weighted bootstrap likelihood with the parametric component fixed at the parametric draw. We prove the procedure is valid for a class of semiparametric models that includes frailty regression models arising in survival analysis and biased sampling models that have application to vaccine efficacy trials. Bootstrap confidence sets from the piggyback and weighted bootstraps are compared for biased sampling data from simulated vaccine efficacy trials.

*KEYWORDS:* Biased sampling; Bootstrap; Censored data; Confidence sets; Empirical process; Monte Carlo Inference; Semiparametric efficiency; Survival analysis.

## 1. INTRODUCTION

We propose a computationally quick alternative to the weighted bootstrap in a general class of semiparametric models, which we call the "piggyback bootstrap." The semiparametric models for which the methodology is applicable include frailty regression models arising in survival analysis and biased sampling models that have application to vaccine efficacy trials. In the general set-up, the model consists of a parametric component $\theta$ and a nonparametric component $A$. More specifically, we consider likelihoods of the form $\prod_{i=1}^{n} \ell(\theta, A)(D_i)$. Here the contribution of the $i$th subject to the likelihood, $\ell(\theta, A)(D_i)$, depends upon the data vector $D_i$ corresponding to the $i$th subject, a vector $\theta \in \mathbb{R}^d$, and a nonnegative function of bounded variation $A(t)$ defined for $t$ in some finite interval $[0, \tau]$. In our survival analysis applications $A(t)$ is a cumulative hazard, and in our biased sampling application $A(t)$ is a cumulative distribution function (cdf). Our regularity conditions given in Section 5 will further constrain the likelihood.

Simulation methods such as the bootstrap provide a way to use information from a sample of size $n$ to generate random draws which accurately approximate the Gaussian limit process of the maximum likelihood estimates (MLEs), $\hat{\theta}_n$ and $\hat{A}_n$, of $\theta$ and $A$. These draws should be realizations of random variables $\theta_n$ and $A_n$ that satisfy the following asymptotic property: $\sqrt{n}(\theta_n - \hat{\theta}_n, A_n - \hat{A}_n)$ converges weakly, given the sample data, to the same distribution that $\sqrt{n}(\hat{\theta}_n - \theta_0, \hat{A}_n - A_0)$ does unconditional on the sample data, as $n \to \infty$. We make this statement precise in Section 2.

Computationally simple simulation methods have been developed for the Cox model for right censored data. Kim and Lee (2003) propose an empirical Bayes method. The regression coefficients $\theta_n$ are drawn from a posterior distribution for $\theta$ given the data. For each draw so obtained, one samples $A$ for from the posterior distribution of $A$ given $\theta_n$, which conveniently involves only generating gamma random variables. Unfortunately, this technique relies on the special structure of the Cox partial likelihood, a feature not shared by other semiparametric survival models. Shen (2002) gives conditions on appropriate priors for the parametric and nonparametric components in a more general semiparametric setting. However, no general method is given for finding these priors, if they even exist. And there is no guarantee that the posterior distributions will be computationally easy to sample from.

Lin, Fleming, and Wei (1994) propose a Monte Carlo method for the Cox Model. In their scheme the MLEs are computed, and then independent standard normal variables are plugged in to a simple expression involving the MLEs to obtain the random draws. Although this method is very simple computationally, it is unclear if an analogous approach can be devised for other semiparametric settings.

In contrast to these two simulation methods, the weighted bootstrap (Rubin 1981, Praestgaard and Wellner 1993) is broadly applicable to many semiparametric models. Each term of the likelihood is weighted by a positive random variable satisfying specified moment conditions. The resulting bootstrap likelihood is maximized over $\theta$ and $A$ to give a random draw. Unfortunately, this maximization is computationally intense in most semiparametric models, and must be repeated for each desired draw.

Tsodikov (2003) recommends using profile likelihood maximization in order to reduce the computational difficulty in obtaining the MLEs. Let $\ell_n(\theta, A) = n\mathbb{P}_n\ell(\theta, A)$ denote the log-likelihood function based on a sample of size $n$, and $\ell_n^\circ(A, \theta) = \mathbb{P}_n^\circ\ell(\theta, A) \equiv \mathbb{P}_n\eta\ell(\theta, A)/(\mathbb{P}_n\eta)$ denote a corresponding bootstrap log-likelihood. The profile bootstrap log-likelihood is defined by

$$p\ell_n^\circ(\theta) \equiv \sup_A \ell_n^\circ(\theta, A) = \ell_n^\circ(\theta, \hat{A}_\theta^\circ)$$

where $\hat{A}_\theta^\circ \equiv \text{argmax}_A \ell_n^\circ(\theta, A)$. Then the value of $\theta$ which maximizes $p\ell_n^\circ(\theta)$ coincides with the $\theta$ component of the joint maximizer of $\ell_n^\circ(\theta, A)$, and is thus the bootstrap MLE, which we will denote by $\hat{\theta}_n^\circ$. And for the full parameter $\psi \equiv (\theta, A)$, the bootstrap MLE is thus $\hat{\psi}_n^\circ \equiv (\hat{\theta}_n^\circ, \hat{A}_{\hat{\theta}_n^\circ}^\circ)$. Dropping the superscript "$\circ$", the above applies to the original likelihood.

In Tsodikov's scheme, one must use a search algorithm, such as Newton-Raphson or the Powell method (Press et al. 1994) employed by Tsodikov, to find the value of $\theta$ which maximizes $pl_n^\circ(\theta)$. And for each candidate search value $\theta^*$, one must compute $\hat{A}_{\theta^*}^\circ$, which we call a "profile computation". In the Cox Proportional Hazards model, $\hat{A}_\theta$ and $\hat{A}_\theta^\circ$ have explicit forms in terms of $\theta$. For instance, $\hat{A}_\theta$ is Breslow's estimator. In the other models we consider, the most explicit expression we have for $\hat{A}_\theta^\circ$ is a self-consistency equation:

$$\hat{A}_\theta^\circ(t) = f_n(t; \theta, \hat{A}_\theta^\circ),$$

where $f_n(t; \theta, A)$ is based on the $n$ observations. And so we must accomplish the profile computation by iterations of a fixed point algorithm, which we describe in detail in Section 2. Tsodikov (2003) presents ways to obtain the relevant self-consistency equation in a variety of semiparametric settings. He points out that this procedure can be viewed as an MM or EM algorithm. Since the fixed point algorithm often requires many iterations (see the simulation study in Section 4 for some specific average iterations), it ends up being the main source of computational cost in the weighted bootstrap.

The contribution of our paper is to take advantage of the profile structure elucidated by Tsodikov to perform computationally efficient inference for semiparametric models. In our proposed *piggyback bootstrap*, we assume that draws for the parametric component $\theta_n$ are readily available (we discuss methods of obtaining such draws in Section 2) and that $\sqrt{n}(\theta_n - \hat{\theta}_n)$, given the sample data, converges in distribution to the unconditional limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Then, for each parametric draw $\theta_n$, the piggyback bootstrap draw is $A_n = \arg\max_A \ell_n^\circ(\theta_n, A)$, resulting in the pair $(\theta_n, A_n)$.

Thus the piggyback bootstrap decreases the dimension of the maximization problem for each set of bootstrap weights. This method requires only one profile computation for each set of bootstrap weights. In contrast, the full weighted bootstrap requires a profile computation for each candidate value in the search for the parametric maximizer. Since the profile computations are accomplished using computationally costly iterations of a fixed point algorithm, we use the number of profile computations required by a procedure as a measure of its computational complexity. We show that the piggyback bootstrap is significantly less computationally complex than the weighted bootstrap.

Section 2 describes our method in more detail. Section 3 presents several example models for which the piggyback bootstrap works. In Section 4, we provide a numerical study comparing the weighted and piggyback bootstraps in a biased sampling analysis of simulated vaccine efficacy data. Sufficient regularity conditions are given in Section 5, and Section 6 contains a brief discussion. Details on the proofs of the results are given in Appendix A, and Appendix B discusses verification of the regularity conditions.

## 2. THE PIGGYBACK BOOTSTRAP

In this section we introduce our piggyback bootstrap approach to obtaining appropriate random draws for semiparametric inference.

The main idea is to first obtain valid random draws for the parametric component of the model. Usually, it is possible to do this in a manner that is computationally much less intense than repeatedly maximizing profile likelihoods. The second step is to piggyback the draws for the nonparametric component onto the parametric draws, by plugging the parametric draws into a bootstrap likelihood and maximizing over the nonparametric component holding the parametric part fixed. That is, for each $\theta_n^{(k)}$ drawn, $k = 1, \ldots, m$, we generate i.i.d.

random bootstrap weights $\eta_1, \ldots, \eta_n$, and compute $\hat{A}^\circ_{\theta_n^{(k)}} = \mathrm{argmax}_A \ell_n^\circ(\theta_n^{(k)}, A)$, where $\ell_n^\circ$ is the bootstrapped log-likelihood using the given bootstrap weights. We assume that these bootstrap weights are nonnegative, with mean and variance 1 and with $\int_0^\infty \sqrt{P[\eta_1 > x]}dx < \infty$. The bootstrap weights should be generated so that they are independent of the data. Further, the draws for $\theta_n$ should be i.i.d. and independent of the bootstrap weights.

Due to measurability issues which may arise in applying our theory to specific cases, we use empirical process results from van der Vaart and Wellner (1996) (hereafter abbreviated VW). The precise form of the bootstrap log-likelihood is $\ell_n^\circ(A, \theta) = \mathbb{P}_n^\circ \ell(\theta, A) \equiv \mathbb{P}_n \eta \ell(\theta, A)/(\mathbb{P}_n \eta)$. Hoffmann-Jørgensen (HJ) weak convergence is denoted by $\rightsquigarrow$.

Section 1.12 of VW gives a useful characterization of HJ weak convergence in a metric space $\mathbb{D}$ to a tight limit. For a metric space $\mathbb{D}$, define $BL_1$ to be the set of all $f : \mathbb{D} \mapsto \mathbb{R}$ with $||f||_{\mathbb{D}} \equiv \sup_{x \in \mathbb{D}} |f(x)| \leq 1$ and $|f(x) - f(y)| \leq d(x, y)$ for every $x, y$, where $d$ is the metric on $\mathbb{D}$. Then $X_n \rightsquigarrow X$, where $X$ is tight if and only if $\sup_{f \in BL_1} |E^* f(X_n) - Ef(X)| \to 0$.

Loosely speaking, we define $B_n \approx C_n$ to mean that $B_n$ has a limit law conditional on the data equal to the limit law of $C_n$. In specific, if $\mathbb{G}$ denotes the limit law of $C_n$, we require $\sup_{h \in BL_1} |E_\eta h(B_n) - Eh(\mathbb{G})| \to 0$ in outer probability and that $B_n$ is asymptotically measurable unconditionally. This has the form of the conclusion of Theorem 2.9.6 of VW, which we employ in Appendix A.

Now we state the main theorem. If we consider $\theta_n^{(1)}, \ldots, \theta_n^{(m)}$ to be realizations of a random vector $\theta_n$, then the following establishes the validity of the new approach:

THEOREM 1 *Under regularity conditions,* $\sqrt{n}(\theta_n - \hat{\theta}_n, \hat{A}^\circ_{\theta_n} - \hat{A}_{\hat{\theta}_n}) \approx \sqrt{n}(\hat{\theta}_n - \theta_0, \hat{A}_{\hat{\theta}_n} - A_0)$.

The regularity conditions are given and discussed in Section 5. Appendix B provides an outline for verifying the regularity conditions in applications, with specific detail given for the odds-rate and biased sampling models introduced in Section 3. Four of the six regularity conditions simply involve verifying the structure of derivatives of the log-likelihood, which should be simple in practice. The remaining two regularity conditions should be easy to verify when the MLEs are known to be efficient and the weighted bootstrap is known to be a valid simulation method for a model.

The proof of Theorem 1, given in Appendix A, is a consequence of expansion (6). Expansion (6) follows from a series of Lemmas, which can be established through technically detailed but rather straightforward proofs. In particular, Lemma 4 establishes a functional Taylor expansion using 2.4.8 of Abraham, Marsden, and Ratiu (1988). Then the regularity conditions, limit arguments, and a Multiplier Central Limit Theorem (Theorem 2.9.6 of van der Vaart and Wellner (1996)) are used to show this Taylor expansion implies (6). Since we can generate the bootstrap weights and the draws $\theta_n$ so that they are independent of each other, it is straightforward to show that the limiting covariance matrix of the right hand side of (6) is the inverse of the information operator, proving Theorem 1.

Before utilizing this result, it is necessary to obtain draws $\theta_n^{(k)}$, $k = 1, \ldots, m$, that have the right conditional distribution. The regularity condition that $\hat{\theta}_n$ is efficient implies that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically mean zero normal with variance $\tilde{I}_0^{-1}$, where $\tilde{I}_0$ is the efficient Fisher information for $\theta$. Thus one way to obtain the desired draws is to estimate $\tilde{I}_0^{-1}$ with a consistent estimator $\hat{V}_0$, and then let $\theta_n^{(k)} = \hat{\theta}_n + n^{-1/2}\hat{V}_0^{1/2}Z^{(k)}$, $k = 1, \ldots, m$, where the $Z^{(k)}$ are independent standard normal vectors of length $p$, where $p$ is the dimension of $\theta$.

In some settings, such as the Cox model for right-censored data, a consistent estimator of $\hat{V}_0$ is not difficult to construct. Corollary 3 of Murphy and van der Vaart (2000) can be used to consistently estimate this covariance matrix in a number of other semiparametric settings. For the biased sampling example, we verify in Appendix B that the hypotheses of Corollary 3 of Murphy and van der Vaart hold, justifying our estimate described in Section 4.

An alternative Monte Carlo approach is presented in Lee (2000). This method builds on the results of Murphy and van der Vaart (2000). The proposed MCMC scheme generates $\theta_n$ from the density proportional to $\exp\{p\ell_n(\theta)\}$. Provided the draws stay sufficiently close to $\hat{\theta}_n$, we have $\sqrt{n}(\theta_n - \hat{\theta}_n) \approx \sqrt{n}(\hat{\theta}_n - \theta_0)$. See Lee (2000) for numerical studies of this algorithm, and Dixon (2003) for an implementation of this method in applying the piggyback bootstrap to an odds-rate regression on a Non-Hodgkin's Lymphoma data set.

In order to piggyback the draws for $A$ onto the draws for $\theta$ we use the fixed point algorithm discussed in the Introduction. The same fixed point algorithm is used in implementing Tsodikov's (2003) method for computing the MLEs in the weighted bootstrap, as described in the Introduction. Before giving details on the algorithm, we mention that in the models we consider, we take the estimator for the nonparametric piece to be a step function with jumps only at observation points. In the biased sampling models $A$ is a cdf, and Vardi (1985) shows that the maximizer puts mass only at the observation points. In the survival analysis settings, the jumps are at observed failure times only. Murphy (1994) and Parner (1998) provide nice discussions of this issue. We will denote the jump at an observation $t$ by $\Delta\hat{A}_\theta(t)$.

Here and in what follows, for a right continuous function $f$ we define $\Delta f(t) \equiv f(t) - f(t-)$.

Then the fixed point algorithm for evaluating $\Delta \hat{A}_\theta(t)$ for a fixed $\theta$ at each observation $t$ is as follows:

- Step 1: Set $\Delta \hat{A}_\theta(t)^{(0)} \equiv g(t)$ for all observations $t$, where $g$ is some initial guess function.

- Step 2: Compute
$$\Delta \hat{A}_\theta(t)^{(r+1)} \equiv f_n(t; \theta, \hat{A}_\theta^{(r)})$$
at each observation $t$.

- Step 3: If $\sup_t |\Delta \hat{A}_\theta^{(r+1)}(t) - \Delta \hat{A}_\theta^{(r)}(t)| < \epsilon$, stop and set $\hat{A}_\theta \equiv \hat{A}_\theta^{(r+1)}$. Otherwise repeat step 2 with $r$ replaced by $r+1$. Where $\epsilon$ is some tolerance level. (We took $\epsilon = .0001$ in our simulation study in Section 4.)

In our simulation study the algorithm described above always converged. And for the biased sampling model we consider in that study, Vardi (1985) has shown the self-consistency equation has a unique solution. Thus we know the resulting step function is the nonparametric MLE.

The next section describes several models that satisfy the regularity conditions of Theorem 1.

## 3. EXAMPLES

In this section we describe several models for which the piggyback bootstrap is applicable. The first example is the Cox model for right censored data. While computationally efficient inference for this model can be done with empirical Bayes (Kim and Lee 2003) or the Monte Carlo approach of Lin, Fleming, and Wei (1994), the example provides a straightforward illustration of the methodology. The second example, odds-rate regression for right-censored data, is developed in detail and the regularity conditions are rigorously established in Appendix B. The next example is the correlated (and shared) gamma frailty model for right-censored data. A final example is the biased sampling model. The regularity conditions for this last example are also established in Appendix B. Technical restrictions which simplify or are necessary in the verification of the regularity conditions are given for the biased sampling and odds-rate models in Appendix B. See Dixon (2003) for technical restrictions on the other models, as well as a detailed verification of the regularity conditions for these models.

### 3.1 Cox Proportional Hazards Model

Efficient inference for $\theta$ in the Cox model for right censored data is based on the profile likelihood which is equivalent to the partial likelihood. The data consists of $n$ i.i.d. realizations of $(X, \delta, Z)$, where $X \equiv T \wedge C$ is the smaller of a failure time $T$ and right censoring time $C$, $Z(\cdot)$ is a $d$-dimensional caglad (its components are left-continuous with right-hand limits) covariate process, and $\delta \equiv \mathrm{I}\{T \leq C\}$ is the censoring indicator, where $\mathrm{I}\{\cdot\}$ is the indicator function. The survival function for this model is given by $S(t|Z) \equiv P(T > t|Z) = \exp\left\{-\int_0^t e^{\theta' Z(s)} dA(s)\right\}$. Here $\theta$ is a $d$-vector of regression coefficients, and $A(t)$, the cumulative baseline hazard, is a cadlag (right-continuous with left-hand limits) function of bounded variation on some finite interval $[0, \tau]$. Let $\hat{\theta}_n$ be the partial likelihood estimator for $\theta_0$ and let $\hat{V}_0$ be the corresponding estimator of the variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. The baseline hazard function can be estimated with Breslow's estimator $\hat{A}_{\hat{\theta}_n}$, where

$$\hat{A}_\theta(t) = \int_0^t \frac{\mathbb{P}_n dN(s)}{\mathbb{P}_n Y(s) e^{\theta' Z(s)}},$$

$N(t) = I\{X \leq t, \delta = 1\}$, $Y(t) = I\{X \geq t\}$, and $\mathbb{P}_n$ is the empirical measure based on the data. A piggyback bootstrap is accomplished by drawing $\theta_n = \hat{\theta}_n + n^{-1/2}\hat{V}_0^{-1/2} Z$, where $Z$ is a $p$-variate standard normal deviate. Then $A_n$ is obtained by computing Breslow's estimate with bootstrap weights $\mathbb{P}_n^\circ(\cdot) \equiv \mathbb{P}_n \eta(\cdot)/(\mathbb{P}_n \eta)$ replacing the empirical weights $\mathbb{P}_n(\cdot)$.

### 3.2 Odds-Rate Model

A flexible regression model for right-censored data, which includes the Cox proportional hazards model and the proportional odds model as special cases, is the odds-rate regression model considered in Dabrowska and Doksum (1988), Scharfstein, Tsiatis and Gilbert (1998), and Lee (2000). The survival function for this model is defined as $S(t|Z) \equiv P(T > t|Z) = E\left[\exp\left\{-W \int_0^t e^{\beta' Z(s)} dA(s)\right\}\Big| Z\right]$, where $\beta$ is a $d$-vector of regression coefficients, $W$ is an unobserved nonnegative gamma frailty with mean 1 and nonnegative variance $\gamma$, and $A(t)$, the cumulative baseline hazard, is a cadlag function of bounded variation on some finite interval $[0, \tau]$. After integrating over $W$, the survival function simplifies to $S(t|Z) = g_\gamma(\int_0^t e^{\beta' Z(s)} dA(s))$, where $g_\gamma(u) = (1 + \gamma u)^{-1/\gamma}$ for $\gamma \neq 0$ and $g_0(u) \equiv \lim_{\gamma \to 0} g_\gamma(u) = e^{-u}$. By multiplying the baseline hazard $e^{\beta' Z(t)} dA(t)$ by the random variable $W$, we are making an adjustment for misspecified or omitted covariates. Setting $\gamma = 0$ results in the Cox model, while setting $\gamma = 1$ yields the proportional odds model. We will focus on the case where $\gamma$ is unknown.

Letting $\theta \equiv (\gamma, \beta)$ and assuming that censoring is independent of $T$ given $Z$, the log-likelihood function for $\psi \equiv (\theta, A)$ in the odds-rate model is given by

$$\ell_n(\psi) = n\mathbb{P}_n \left\{ \int_0^\tau \left[ \gamma \log g_\gamma(H_\psi^s(Y)) + \beta' Z(s) + \log a(s) \right] dN(s) + \log g_\gamma(H_\psi^\tau(Y)) \right\}, \tag{1}$$

where $H_\psi^t(f) \equiv \int_0^t f(s) e^{\beta' Z(s)} dA(s)$ and $a \equiv dA/dt$.

As discussed in the Introduction, we consider $A$ constant except for jumps at the observed failure times, replacing $a(s)$ with $\Delta A(s) (\equiv A(s) - A(s-))$. Then the MLE may be computed in two steps via profile likelihood. For each $\theta$, define $\hat{\psi}_\theta = (\theta, \hat{A}_\theta)$. The maximizer $\hat{A}_\theta$ is a solution of the following self-consistency equation: $\hat{A}_\theta(t) = \int_0^t \left[ W_n(u; \hat{\psi}_\theta) \right]^{-1} dG_n(u)$, for all $t \in [0, \tau]$, where $G_n(u) = \mathbb{P}_n \delta \mathrm{I}\{X \leq u\}$ and

$$W_n(u; A, \theta) = \mathbb{P}_n \left[ (1 + \delta\gamma) e^{\beta' Z(u)} Y(u) / (1 + \gamma H_\psi^X(1)) \right]$$

(Scharfstein et al. 1998). The MLE for $\theta$ is $\hat{\theta}_n = \mathrm{argmax}_\theta p\ell_n(\theta)$, and the joint maximizer is thus $\hat{\psi}_n = (\hat{\theta}_n, \hat{A}_{\hat{\theta}_n})$.

The weighted bootstrap can be used to obtain draws for the parametric and nonparametric components of the model, as proven in Kosorok, Lee and Fine (2004) (the arguments are similar to those used in the proof of Lemma 8 in Appendix B). However, maximizing over both the parametric and nonparametric pieces, as required by the bootstrap, is computationally intense. The difficulty is that for each candidate value of the parametric maximizer, one must compute $\hat{A}_\theta^\circ$, which is the solution to the self consistency equation

$$\hat{A}_\theta^\circ(t) = \int_0^t (W_n^\circ(u; \theta, \hat{A}_\theta^\circ))^{-1} dG_n^\circ(u) \tag{2}$$

where $G_n^\circ(u) = \mathbb{P}_n \eta \delta \mathrm{I}\{X \leq u\}$ and $W_n^\circ(u; \theta, A) = \mathbb{P}_n \left[ \eta (1 + \delta\gamma) e^{\beta' Z(u)} Y(u) / (1 + \gamma H_\psi^X(1)) \right]$.

Our piggyback approach applies in this situation, simplifying computations. To obtain draws for the parametric component, $\theta$, we show in Appendix B that Corollary 3 of Murphy and van der Vaart (2000) holds. For each of the draws $\theta_n^{(k)}$, one generates bootstrap weights and plugs $\theta_n^{(k)}$ into the corresponding bootstrap likelihood. Then the bootstrap likelihood is maximized over the nonparametric component to obtain $A_n^{(k)}$. We verify in Appendix B that the resulting piggybacked pairs $(\theta_n^{(k)}, A_n^{(k)})$, $k = 1, \ldots, m$, satisfy the conditions of Theorem 1 and thus have the desired asymptotic property. In addition, these piggyback draws are relatively easy to compute.

Alternatively, one may use the MCMC scheme described in Lee (2000) to generate draws for $\theta$. See Dixon (2003) for an implementation of this method in applying the piggyback bootstrap to an odds-rate regression on a Non-Hodgkin's Lymphoma data set.

### 3.3 Correlated (and Shared) Gamma Frailty Model

Much of the theory for the odds-rate model is derived in a manner similar to the results of Parner (1998). Parner considers the correlated gamma frailty model for clustered data. The data is right censored with the observations for each of $n$ clusters (e.g., family, pair of twins, group) consisting of $m$ realizations of $(X_i, \delta_i, Z_i(u))$, $i = 1, \ldots, n$. We associate with the $ij$th individual, $i = 1, \ldots, n$, $j = 1, \ldots, m$, a frailty $W_i^{(j)} \equiv W_{i0} + W_{ij}$. The frailty components $W_{i0}, W_{i1}, \ldots, W_{im}$ are assumed to be independent, unobservable, gamma-distributed random variables with parameters $(\nu, \eta), (\nu^*, \eta), \ldots, (\nu^*, \eta)$, respectively. We adopt the parameterization used by Parner by considering the frailty parameters of the model to be $\theta \equiv \mathrm{var}(W_0) = \nu/\eta^2$ and $\theta^* \equiv \mathrm{var}(W_j) = \nu^*/\eta^2$. When $\theta^*$ is set equal to 0 we have the shared frailty model. The full parameter for the model is $\psi \equiv (\theta, \theta^*, \beta, A)$. The log-likelihood function is given by

$$\sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} [\delta_{ij}(\beta' Z_{ij}(X_{ij}) + \log a(X_{ij}))] + \log \sum_{\mathbf{k} \in K_i} \alpha_i(\mathbf{k}, \psi) \right\},$$

where $\mathbf{k} = (k_1, \ldots, k_m)$ and $K_i = \{\mathbf{k} | k_j \in \{0, \delta_{ij}\}, j = 1, \ldots, m\}$. And the $\alpha_i(\mathbf{k}, \psi)$ are given by

$$\frac{\theta_\cdot^{\delta_{i\cdot}} \Gamma(\theta\theta_\cdot^{-2} + \delta_{i\cdot} - k_\cdot)}{\Gamma(\theta\theta_\cdot^{-2})(1 + \theta_\cdot \Lambda_{i\cdot}(\psi))^{\theta\theta_\cdot^{-2} + \delta_{i\cdot} - k_\cdot}} \prod_{j=1}^{m} \left\{ \frac{\Gamma(\theta^*\theta_\cdot^{-2} + k_j)}{\Gamma(\theta^*\theta_\cdot^{-2})} \frac{1}{(1 + \theta_\cdot \Lambda_{ij}(\psi))^{\theta^*\theta_\cdot^{-2} + k_j}} \right\}$$

where $\Lambda_{ij}(\psi) \equiv \int_0^\tau Y_{ij}(s) \exp(\beta' Z_{ij}(s)) da(s)$, $Y_{ij}(t) \equiv \mathrm{I}\{X_{ij} \geq t\}$, $\theta_\cdot \equiv \theta + \theta^*$, and otherwise a subscript "$\cdot$" denotes summation over the corresponding index, e.g. $\delta_{i\cdot} = \sum_{j=1}^m \delta_{ij}$. Using the bootstrap which assigns weight to the $i$th cluster, $i = 1, \ldots, n$, the bootstrap self-consistency equation for $A$ is given by

$$\hat{A}_n^\circ(t) = \int_0^t \left( \frac{1}{n} \sum_{i=1}^{n} \eta_i \sum_{j=1}^{m} \hat{Z}_i^{(j)}(\theta, \theta^*, \beta, \hat{A}_n^\circ) Y_{ij}(s) \exp(\beta' X_{ij}(s)) \right)^{-1} d\bar{N}_{\cdot\cdot}^\circ(s).$$

Here $\hat{Z}_i^{(j)}(\psi) \equiv \hat{Z}_{i0}(\psi) + \hat{Z}_{ij}(\psi)$ with

$$\hat{Z}_{i0}(\psi) \equiv \frac{\sum_{\mathbf{k} \in K_i} \alpha_i(\mathbf{k}, \psi) b_{i0}(\mathbf{k}, \psi)}{\sum_{\mathbf{k} \in K_i} \alpha_i(\mathbf{k}, \psi)} \quad \text{and} \quad \hat{Z}_{ij}(\psi) \equiv \frac{\sum_{\mathbf{k} \in K_i} \alpha_i(\mathbf{k}, \psi) b_{ij}(\mathbf{k}, \psi)}{\sum_{\mathbf{k} \in K_i} \alpha_i(\mathbf{k}, \psi)},$$

where

$$b_{i0}(\mathbf{k}, \psi) = \frac{\theta\theta_\cdot^{-1} + \theta_\cdot(\delta_{i\cdot} - k_\cdot)}{1 + \theta_\cdot \Lambda_{i\cdot}(\psi)} \quad \text{and} \quad b_{ij}(\mathbf{k}, \psi) = \frac{\theta^*\theta_\cdot^{-1} + \theta_\cdot k_j}{1 + \theta_\cdot \Lambda_{ij}(\psi)},$$

$\bar{N}^{\circ}_{\cdot\cdot} = \sum_{i=1}^{n} \eta_i \sum_{j=1}^{m} N_{ij}$, and (analogously to the odds-rate model setting) $N_{ij}(s) \equiv \mathrm{I}\{X_{ij} \leq s, \delta_{ij} = 1\}$. Taking $\eta_i = 1$ with probability 1, we have the self consistency equation for $\hat{A}_n$ previously revealed by Parner (1998).

We omit verification of the conditions here, but with some additional technical restrictions, the piggyback bootstrap can be shown to be readily applicable to the correlated gamma frailty example. See Dixon (2003) for details.

**3.4 Biased Sampling Model**

As a final application, we consider a class of biased sampling models (Gilbert 2000). Here the data consists of $n$ i.i.d. realizations of $X = (I, Y)$. Here $I \in \{1, \ldots, s\}$ is a random stratum, taking on the value $i$ with selection probability $\lambda_i > 0$, $i = 1, \ldots, s$, with $\sum_{i=1}^{s} \lambda_i = 1$. Given $I = i$, $Y \in [0, \tau]$ has distribution $F_i$ defined on a sigma field of subsets $\mathcal{B}$ of $[0, \tau]$ by $F_i(B, \theta, A) \equiv W_i^{-1}(\theta, A) \int_B w_i(u, \theta) dA(u)$ for $B \in \mathcal{B}$. The $w_i$, $i = 1, \ldots, s$, are nonnegative (measurable) stratum weight functions assumed to be known up to the finite dimensional parameter $\theta \in \Theta$. $W_i(\theta, A) \equiv \int_0^{\tau} w_i(u, \theta) dA(u)$ is assumed to be finite for all $\theta \in \Theta$. The probability measure $A$ is the unknown infinite dimensional parameter of interest. The goal is to estimate it based on information from samples from the $F_i$ distributions, $i = 1, \ldots, s$. Thus the semiparametric log-likelihood is given by

$$n\mathbb{P}_n \left\{ \lambda_I \frac{w_I(Y, \theta)}{W_I(\theta, A)} dA\{Y\} \right\}.$$

Denote the number of observations belonging to the $i$th stratum by $n_i$, the total sample size by $n (= \sum_{i=1}^{s} n_i)$, and the $i$th sampling fraction by $\lambda_{ni} = n_i/n$. Furthermore, let $t_j$, $j = 1, \ldots, h$, denote the distinct observed realizations of $Y$, each with multiplicity $r_j$, and let $n_{ij}$ be the number of observations from the $i$th stratum with value $t_j$. The semiparametric likelihood is then proportional to

$$L_n(\psi|\text{data}) = \prod_{i=1}^{s} \prod_{j=1}^{h} \left[ \frac{w_i(t_j, \theta) dA\{t_j\}}{W_i(\theta, A)} \right]^{n_{ij}}.$$

As demonstrated in Vardi (1985), we may instead maximize the partial likelihood defined by

$$L_{n1}(\theta, \tilde{B}|\text{data}) = \prod_{i=1}^{s} \prod_{j=1}^{h} \left[ \frac{w_{ij}(\theta) B_i^{-1}}{\sum_{k=1}^{s} \lambda_{nk} w_{kj}(\theta) B_k^{-1}} \right]^{n_{ij}},$$

where $\tilde{B} = \{B_1, \ldots, B_s\}$, subject to $B_i > 0$, $i = 1, \ldots, s-1$ and $B_s = 1$, and $w_{ij}(\theta) \equiv w_i(t_j, \theta)$. Again the MLE is computed in two steps via the notion of profile (partial) likelihood. The profile partial log-likelihood is given by $p\ell_n(\theta) \equiv \sup_{\tilde{B}} \log L_{n1}(\theta, \tilde{B}) \equiv \log L_{n1}(\theta, \hat{B}_\theta)$ where $\hat{B}_\theta \equiv \text{argmax}_{\tilde{B}} \log L_{n1}(\theta, \tilde{B})$. The MLE for $\theta$ coincides with the profile-partial-MLE obtained by maximizing $p\ell_n(\theta)$ with respect to $\theta$. Then we take $\hat{\psi}_n \equiv (\hat{\theta}_n, \hat{A}_{\hat{\theta}_n})$.

As demonstrated in Vardi (1985), we may compute $\hat{A}_\theta$ by the following procedure. First find $\hat{B}$ defined as the solution to the self-consistency equation

$$B_i(\theta) = \sum_{j=1}^{h} \frac{r_j w_{ij}(\theta)}{\sum_{k=1}^{s} n_k w_{kj}(\theta) B_k^{-1}(\theta)}$$

$i = 1, \ldots, s$ subject to the constraints $B_i > 0$, $i = 1, \ldots, s-1$ and $B_s = 1$. Then we take

$$d\hat{A}_\theta(t_j) = \frac{r_j \left[\sum_{k=1}^{s} \lambda_{nk} w_{kj}(\hat{\theta}_n) \hat{B}_{nk}^{-1}(\hat{\theta}_n)\right]^{-1}}{\sum_{j=1}^{h} r_j \left[\sum_{k=1}^{s} \lambda_{nk} w_{kj}(\hat{\theta}_n) \hat{B}_{nk}^{-1}(\hat{\theta}_n)\right]^{-1}}. \quad (3)$$

In the weighted bootstrap case, we replace $n_{ij}$ with the sum $n_{ij}^\circ$ of $n_{ij}$ bootstrap weights in the above equations, and the other terms are modified accordingly, i.e., $n_i^\circ \equiv \sum_{j=1}^{h} n_{ij}^\circ$, $n^\circ \equiv \sum_{i=1}^{s} n_i^\circ = n$ (since all bootstrap weights add up to $n$ by construction), $r_j^\circ \equiv \sum_{i=1}^{s} n_{ij}^\circ$, and $\lambda_{ni}^\circ \equiv n_i^\circ / n$. Note that the use of the partial likelihood is for computational purposes only. It affects none of the theory because the nonparametric partial likelihood maximizer coincides with the nonparametric maximizer of the full likelihood.

The weighted bootstrap can be used to obtain draws for the parametric and nonparametric components of the model, as justified in Gilbert (1996). But, again, maximizing over both the parametric and nonparametric pieces in this approach is computationally intense. However, our piggyback approach applies in this situation, simplifying computations. To obtain draws for the parametric component, $\theta$, Murphy and van der Vaart's (2000) Corollary 3 can be used to consistently estimate $\tilde{I}_0$. Then the draws for the nonparametric piece are piggybacked on these draws as in the case of the odds-rate model discussed above. The next section provides a simulation study using this approach.

## 4. SIMULATION STUDY

We simulated vaccine efficacy trials in order to evaluate the coverage probabilities of confidence bands constructed using draws from the weighted and piggyback bootstraps. The simulated response represents percent divergence in the V3 loop amino acid sequence between the strain of HIV found in an infected subject versus the prototype strain used in the vaccine. These simulated data sets are modeled after ones found in Gilbert (1996).

We generated 200 data sets consisting of 400 independent observations. Half of the 400 observations in each data set come from a placebo group, and the other half from a vaccine group. The placebo group response was uniformly distributed on the interval $[0, 35]$, and the stratum bias functions are given by $w_1(y, \theta) = \exp(y\theta/35)$ and $w_2(y, \theta) = 1$. Here group 1 is the vaccine group and group 2 is the placebo group. Thus the cdf for the vaccine group is given by

$$F(y) = \frac{\int_0^y e^{s\theta_0/35} I_{[0,35]}(s)ds}{\int_0^\infty e^{s\theta_0/35} I_{[0,35]}(s)ds} = \frac{e^{y\theta_0/35} - 1}{e^{\theta_0} - 1},$$

and the cdf for the placebo group is the Uniform[0,35] cdf.

The choice of the Uniform[0,35] distribution is inspired by the use of this distribution in the simulation in Gilbert (1996). The value 35 corresponds to the suspected maximum possible percent divergence, and the true value of $\theta$ was taken to be $\theta_0 = 7.89$, arbitrarily.

The weight functions were chosen because they allow for straightforward estimation of relative risks, as shown in Gilbert, Self, and Ashby (1998). In particular, they show that if we assume that infection is possible from at most one strain during follow-up, that the relative prevalence of circulating strains is constant over $[0, \tau]$ $(\tau = 35)$, and that the probability of being exposed to a given strain is the same for vaccinated and unvaccinated trial participants, then $e^{y\theta/35} = RR(y)/RR(0)$, where

$$RR(y) \equiv \frac{P(\text{infected by y}|\text{one exposure to y, vaccine})}{P(\text{infected by y}|\text{one exposure to y, placebo})}$$

denotes the relative risk of a strain $y$.

For the parametric draws in our piggyback approach, we estimated the variance of the MLE $\hat{\theta}_n$ and then drew standard normal deviates with this variance and added them to $\hat{\theta}_n$ for each data set. To estimate the variance of $\hat{\theta}_n$, we applied Corollary 3 of Murphy and van der Vaart (2000). That is, we calculated

$$\lim_{h \to 0} -2 \frac{\log p\ell_n(\hat{\theta}_n + h) - \log p\ell_n(\hat{\theta}_n)}{400h^2} \equiv \hat{I}_0,$$

and took the variances of the normal deviates to be $(400\hat{I}_0)^{-1}$.

Numerical experimentation, with data sets generated like those used in the simulation, revealed that in a neighborhood of 0 the left hand side of the above expression is close to linear in $h$. Thus in the simulation we estimated the limit for each data set by averaging the values of the left hand side at two points equidistant from 0. These points were the same for each data set, and set in advance before the simulation based on the numerical experimentation. As a check on the validity of this method, we compare the variance estimates so obtained with the variances of the parametric draws obtained using two other methods. These methods are the MCMC method developed by Lee (2000) for parametric inference, and the full semiparametric weighted bootstrap. Neither of these methods estimates the variance before implementation as we do here. Instead, each generates parametric draws with the appropriate distribution for each of the 200 data sets, and afterwards we calculated the variance of the draws for each data set for the sake of this comparison. We found that the correlation of our variance estimates with the variances from Lee's method is .985. A scatterplot of the variance pairs does not show any outliers, and a least squares regression line estimates these variance estimates are $0.0210 + 0.9335 \times$ Lee's method variances with coefficient of determination $R^2 = 0.9699$. Since Lee's method is based on related results from Murphy and van der Vaart (2000), we should expect this close agreement. The correlation of the variances from this and Lee's method with the variances from the full weighted bootstrap are .815 and .798, respectively. Thus it appears the above procedure is a valid way of estimating the variance for the parametric draws.

In order to maximize over the parametric component of the model in the weighted bootstrap (and in finding $\hat{\theta}_n$ to apply the method described in the last two paragraphs for the piggyback bootstrap), we used a search algorithm that returns a value which is the parametric MLE plus or minus .01. This algorithm is a modified grid search algorithm and does not compute any derivatives.

We performed 2000 bootstrap repetitions for both the piggyback and weighted bootstraps for each of the 200 simulated data sets. For reasons discussed in the Introduction, we use the number of profile computations required by each bootstrap procedure as a measure of computational complexity. In the weighted bootstrap, we profiled (maximized) over the nonparametric component of the model on average 58189.755 times per data set. The mean number of iterations of the fixed point algorithm for computing these $A$ values was 35.758 with standard deviation 12.375. In order to obtain the variance estimates for the parametric component for use in the piggyback bootstrap, we profiled over the nonparametric component of the model on average 31.995 times per data set. (Note that this includes the profile computations needed to find the MLE for each data set.) The mean number of iterations for computing each $A$ was 36.909 with standard deviation 12.468. In piggybacking the draws for $A$ on the 2000 values of $\theta$ so drawn for each of the data sets, the mean number of iterations was 43.468 with standard deviation 9.867. Thus in the piggyback bootstrap we profiled over the nonparametric component of the model an average of 2031.995 times per data set compared to the average of 58189.755 times for the weighted bootstrap, a ratio of 1 to 29. It might be possible to streamline our maximization algorithm to decrease the number of times we must profile over the nonparametric component in the weighted bootstrap. But to beat the piggyback bootstrap in this sense, we would need to profile over the nonparametric component on average less than 2031.995/2000 times for each of the 2000 bootstrap likelihoods. This would require that the first candidate value we try in each bootstrap likelihood search happens to be the MLE (plus or minus some small error) almost all the time, which is clearly beyond even the best search algorithms.

In constructing confidence bands for the cdfs of the vaccine and placebo groups, we considered only the middle 75 percent of the observations. The reason for this is that the bands narrow towards the ends of the data set due to the nature of the estimate. Indeed, the estimated value of the cdf for the largest observation will always be 1, and so the confidence band at that point will never contain the true value. To construct a level $\alpha$ pointwise confidence interval at an observation $y$, we find the middle $\alpha$ percent of the 2000 cdf estimates induced by the draws at $y$. A 95 percent simultaneous confidence band is constructed by finding the smallest value of $\alpha$ so that 95 percent of the cdfs induced by the draws are contained within the level $\alpha$ pointwise confidence intervals for all observations $y$ in the middle 75 percent of the observations.

See Figures 1 and 2 for the resulting confidence intervals and bands for one of the simulated data sets. Since the piggyback and weighted bootstraps only specify the values of the intervals and bands at the observation points, we interpolated to create these figures. Let us denote the value of the upper bound of one of these confidence sets at an observation $y$ by $U(y)$ and the lower by $L(y)$. To extend the sets to the range of the data, the typical approach is to define $U(y) = U(y_1)$ and $L(y) = L(y_1)$ for all $y_1 \leq y < y_2$ for $y_1$ and $y_2$ adjacent observations. We call this the leftpoint interpolation scheme. In these simulations we find the coverage of the bands is better if we define $U(y) = U(y_2)$ and $L(y) = L(y_2)$ for all $y_2 - (y_2 - y_1)/2 \leq y < y_2 + (y_3 - y_2)/2$ for $y_1 < y_2 < y_3$ adjacent response values (where the left part of the inequalities does not apply for the smallest observation, and the right part does not apply for the largest). Thus we used this midpoint interpolation scheme in the figures. Note that asymptotically the leftpoint and midpoint methods will give the same result.

Table 1 presents the coverage of these bands so defined for the piggyback approach and the weighted bootstrap for three interpolation schemes: none, leftpoint, and midpoint. The values for "none" are obtained by only considering whether or not a confidence band contains the true cdf at each observation in the data set, and not between observations as is the case for interpolation. Note that we prove in this paper that the two methods are asymptotically equivalent. Even for this moderate sample size, we see that the two methods perform close to the same. Although the piggyback bootstrap does slightly better than the weighted bootstrap for the vaccine group, and slightly worse than the weighted bootstrap for the placebo group.

We also constructed confidence intervals for the parametric component of the model. For each data set, the .975 and .025 percentiles of the draws for the parametric component were taken as the ends of the corresponding confidence interval. The coverage of these intervals was 94.5 percent using the weighted bootstrap, and 93 percent using the piggyback bootstrap, both of which are within 2 Monte Carlo standard deviations of 0.95 ($0.03 = 2\sqrt{0.95 \times .05/200}$).

## 5. REGULARITY CONDITIONS

In this section, we present sufficient technical conditions for the piggyback bootstrap to have the desired asymptotic property given in Theorem 1. This section may be skipped at first reading. Conditions (PB1) through (PB3) are conditions on the structure and smoothness of the information operator. Condition (PB4) involves asymptotic conditions on the Monte Carlo parameter estimates. Condition (PB5) requires sufficient smoothness of the score operator for the infinite dimensional parameter. Condition (PB6) requires the estimators to be efficient.

As noted in the Introduction, we consider likelihoods of the form $\prod_{i=1}^{n} \ell(\theta, A)(D_i)$. Here the contribution of the $i$th subject to the likelihood, $\ell(\theta, A)(D_i)$, depends upon the data vector $D_i$ corresponding to the $i$th subject, a vector $\theta \in \mathbb{R}^d$, and a nonnegative function of bounded variation $A(t)$ defined for $t$ in some finite interval $[0, \tau]$. Restrictions on the parameter spaces of $\theta$ and $A$ will be model specific. See Appendix B for restrictions on the odds-rate and biased sampling models, and Dixon (2003) for restrictions on the other example models.

Let $h = (h_1, h_2)$, where $h_1 \in \mathbb{R}^d$ and $h_2$ is a nonnegative function of bounded variation on $[0, \tau]$. Let $||\cdot||_i$ denote the Euclidean norm for $i = 1$, the total variation norm for $i = 2$, and $||\cdot||_1 + ||\cdot||_2$ for $i = 3$ (i.e. $||\psi||_3 = ||\theta||_1 + ||A||_2$). Let $\mathcal{S}$ be a set. We use $o_p^{\mathcal{S}}(R_n)(s)$ to denote a term $F_n(s) = R_n Q_n(s)$ such that $\sup_{s\in\mathcal{S}} ||Q_n(s)||$ converges in outer probability to 0 as $n \to \infty$ (here $||\cdot|| = ||\cdot||_i$ for $i = 1, 2$, or 3 as appropriate). We assume the score operator for the model,

$$\begin{aligned} U^n[\theta, A](h) &\equiv \frac{\partial}{\partial t}\ell_n\left(\theta + th_1, A + t\int_0^\tau h_2(u)dA(u)\right)\bigg|_{t=0} \\ &= \frac{\partial}{\partial t}\ell_n(\theta + th_1, A)\bigg|_{t=0} + \frac{\partial}{\partial t}\ell_n\left(\theta, A + t\int_0^\tau h_2(u)dA(u)\right)\bigg|_{t=0}, \end{aligned}$$

exists. Denote the first term in the last line $U_{n,1}[\theta, A](h_1) \equiv n\mathbb{P}_n U_1[\theta, A](h_1)$ and the second $U_{n,2}[\theta, A](h_2) \equiv n\mathbb{P}_n U_2[\theta, A](h_2)$. Let $\bar{h}_1 \in \mathbb{R}^d$, and $\bar{h}_2$ be a nonnegative function of bounded variation on $[0, \tau]$. We assume the following derivative exists

$$\begin{aligned} &\frac{\partial}{\partial s}U_i\left[\theta + s\bar{h}_1, A + s\bar{h}_2\right](h_i)\bigg|_{s=0} \\ &= \frac{\partial}{\partial s}U_i\left[\theta + s\bar{h}_1, A\right](h_i)\bigg|_{s=0} + \frac{\partial}{\partial s}U_i\left[\theta, A + s\bar{h}_2\right](h_i)\bigg|_{s=0}, i = 1, 2. \end{aligned} \tag{4}$$

The following are the regularity conditions:

(PB1) The negative of the first term of (4) is of the form $\bar{h}_1'\hat{\sigma}_{i,1}[\theta, A](h_i)$, and the negative of the second term of (4) is of the form $\int_0^\tau \hat{\sigma}_{i,2}[\theta, A](h_i)(u)d\bar{h}_2(u)$ for $i = 1, 2$, where $\hat{\sigma}_{i,1}[\theta, A](h_i)$ is a random $d$-vector depending upon an observation, and $\hat{\sigma}_{i,2}[\theta, A](h_i)(u)$ is a random function of bounded variation on $[0, \tau]$ depending upon an observation for $i = 1, 2$. Define $\hat{\sigma}_{n,i,j}[\theta, A](h_i) \equiv \mathbb{P}_n\hat{\sigma}_{i,j}[\theta, A](h_i), i, j = 1, 2$. Then $\hat{\sigma}_{n,i,j}[\theta, A](h_i)$ has total variation bounded by some $M$ over all $n$.

(PB2) Denote $\sigma_{i,j}[\theta, A](h_i) \equiv P_0\hat{\sigma}_{i,j}[\theta, A](h_i), i, j = 1, 2$, where $P_0$ is the expectation under the true model. Let $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$, where $\mathcal{H}_1 = \mathbb{R}^d$ and $\mathcal{H}_2$ is a set of functions which includes all $h(s) = \sigma_{2,2}[\theta_0, A_0]^{-1}(\mathrm{I}\{s \leq t\})$ with $t \in [0, \tau]$. And define $||\cdot||_j^{\mathcal{H}_2} \equiv \sup_{h\in\mathcal{H}_2} ||\cdot||_j$, $j = 1, 2$. Then $\lim_{\theta\to\theta_0, A\to A_0} ||\sigma_{2,j}[\theta, A](h) - \sigma_{2,j}[\theta_0, A_0](h)||_j^{\mathcal{H}_2} = 0, j = 1, 2$.

(PB3) For some $c > 0$, $\{\hat{\sigma}_{2,1}[\theta, A](h)_i,\ ||\theta - \theta_0||_1 \leq c, ||A - A_0||_2 \leq c, h \in \mathcal{H}_2\}$, $i = 1, \ldots, d$, and $\{\hat{\sigma}_{2,2}[\theta, A](h)(u),\ ||\theta - \theta_0||_1 \leq c, ||A - A_0||_2 \leq c, h \in \mathcal{H}_2, u \in [0, \tau]\}$ are Donsker and bounded.

(PB4) (i) $\sqrt{n}(\theta_n - \hat{\theta}_n) \approx \sqrt{n}(\hat{\theta}_n - \theta_0)$ and the common limiting distribution is tight.

(ii) $\sqrt{n}(\hat{A}^{\circ}_{\theta_0} - \hat{A}_{\theta_0}) \approx \sqrt{n}(\hat{A}_{\theta_0} - A_0)$ and the common limiting distribution is tight and cadlag.

(iii) $\sqrt{n}||\hat{A}^{\circ}_{\theta_n} - A_0||_2 \leq \tilde{O}_p(1) + o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\theta_n - \theta_0||_1)$

Here and in what follows, $\tilde{O}_p(1)$ denotes a random sequence which converges weakly to a tight, cadlag limit.

(PB5) $P_0 U_2(\theta_0, A_0)(\cdot) = 0$ and $\{U_2(\theta_0, A_0)(h), h \in \mathcal{H}_2\}$ is $P_0$-Donsker.

(PB6) $\sqrt{n}\left(\hat{\theta}_n - \theta_0, \hat{A}_{\hat{\theta}_n} - A_0\right)'$ is asymptotically linear, regular, and efficient. And the information operator

$$\begin{pmatrix} \sigma_{1,1} & \sigma_{2,1} \\ \sigma_{1,2} & \sigma_{2,2} \end{pmatrix}$$

is continuously invertible and onto. Here, and in other cases where the full notation is not necessary for clarity, we abbreviate $\sigma_{i,j}[\theta_0, A_0] \equiv \sigma_{i,j}$, $i, j = 1, 2$.

We assume the bootstrap weights $\eta_1, \ldots, \eta_n$ above are i.i.d., nonnegative, and with mean and variance 1. A technical condition on the bootstrap weights useful in establishing the validity of the weighted bootstrap and condition (PB4.ii) is that $\int_0^\infty \sqrt{P[\eta_1 > x]}dx < \infty$. Note that we assume the bootstrap weights are generated so that they are independent of the data, and that the draws $\theta_n$ generated for the parametric component are i.i.d. and independent of the bootstrap weights.

In Appendix B we give an outline of arguments one can use in verifying the regularity conditions in models such as our example models. We provide a brief overview here. In practice, verifying Conditions (PB1), (PB2), (PB3), and (PB5) mainly consists of taking derivatives of the log-likelihood and verifying that they have a specific structure. The Donsker conditions on the structure may be unfamiliar to some readers. Fortunately, certain "Permanence Properties" of Donsker classes of functions, as discussed in VW, usually makes verifying these conditions an easy task. Using these results from VW, to verify that a class of functions is Donsker, one simply needs to check if all member functions can be expressed in terms of Donsker preserving operations on functions from other known Donsker classes. As an example, see the proof of Lemma 7 in Appendix B. The condition of efficiency of the MLEs in (PB6) is often the justification for performing inference based on the MLEs. Thus in practice, this condition on the piggyback bootstrap will not add any theoretical challenge over other approaches to inference. In Section 2, we discussed methods of obtaining parametric draws satisfying condition (PB4.i). Condition (PB4.ii) follows from arguments similar to those which prove the validity of the weighted bootstrap (see Lemma 8 in Appendix B). And so this condition does not add any theoretical burden to the piggyback bootstrap compared to the weighted bootstrap. Verifying Condition (PB4.iii) requires relatively novel, but straightforward, arguments, which we give in the proof of Lemma 9.

We discuss verification of the regularity conditions in detail for the odds-rate and biased sampling examples in Appendix B. See Dixon (2003) for a detailed verification of the regularity conditions for the other example models. We now discuss the relevance of the above conditions in proving the main theorem.

Condition (PB1) specifies the structure of the likelihood. Note that the expression of the second term of (4) as an integral is crucial in the last step of the proof of (6) in Appendix A. In the survival analysis applications, the structure exists as a consequence of the likelihood depending upon the integral of a hazard function. And in the biased sampling application, it is a consequence of the dependence of the likelihood upon the integrals of stratum bias functions. These examples were discussed in detail in Section 3.

Condition (PB2) requires that the information operators are sufficiently smooth in the parametric and nonparametric components. This condition is useful in working with the generalized Taylor expansions (see 2.4.8 of Abraham, Marsden, and Ratiu, 1988) in the proof of the main theorem.

Condition (PB3) is included to guarantee the convergence in outer probability in the proof of Lemma 3. Note that we are assuming that the classes are Donsker, which is stronger than necessary to guarantee the convergence in outer probability with the empirical measure $\mathbb{P}_n$. However, we need the Lemma to hold when $\mathbb{P}_n$ is replaced by $\mathbb{P}_n^\circ$, and this follows easily if we make the stronger Donsker assumption. That is, if a class of functions $\mathcal{F}$ is Donsker, then so is $\eta\mathcal{F}$ for our choice of bootstrap weights, and the fact that a Donsker class is Glivenko-Cantelli then gives us the desired convergence in outer probability.

Condition (PB5) is crucial to the proofs in Appendix A, and is often proved in the process of proving (PB6). This is done, for instance, if one uses the same methods as Parner (1998) and Lee (2000) in verifying the conditions of the Z-Estimator Master Theorem (Theorem 3.3.1 of VW).

Condition (PB4.i) simply states that we have valid draws for the parametric component of the model. The tightness of the limit distribution allows us to apply Slutsky's Theorem (Example 1.4.7 of VW) in various parts of the proof. In Section 2 we discussed methods of obtaining such draws in practice. Condition (PB4.ii) requires that the bootstrap is valid when $\theta$ is fixed at $\theta_0$. This should hold whenever the bootstrap is valid for the full $\psi$. Note that we require the limit in (PB4.ii) to be tight and cadlag. This is useful in various parts of the proofs in light of Lemma 1 below.

LEMMA 1 *Suppose we have an expression of the form*

$$Q_n(h) = \sqrt{n} \int_0^\tau f_n(h)(u) dg_n(u) \tag{5}$$

*where* $f_n(h) = o_p^{\mathcal{H}_2}(1)(h)$ *has total variation bounded by some* $M$ *over all* $n$, *and* $\sqrt{n} g_n$ *converges weakly to a tight, cadlag limit. Then* $Q_n(h) = o_p^{\mathcal{H}_2}(1)(h)$.

See Appendix A for a proof.

For example we use Lemma 1 in the last step of the proof of Lemma 6, taking $g_n = \hat{A}_{\theta_0} - A_0$ and

$$f_n(h)(u) = \mathbb{P}_n \hat{\sigma}_{2,2}[\theta_0, A_0 + t_n(h)(\hat{A}_{\theta_0} - A_0)](h)(u) - \sigma_{2,2}[\theta_0, A_0](h)(u).$$

A similar lemma is useful in going from the second to last to the last line of the proof of Lemma 4. Note that we first observe $\hat{A}^\circ_{\theta_n} - \hat{A}_{\hat{\theta}_n} = (\hat{A}^\circ_{\theta_n} - A_0) - (\hat{A}_{\hat{\theta}_n} - A_0)$, and the integral on the second to last line is broken up accordingly. The integral with respect to $d[\sqrt{n}(\hat{A}_{\hat{\theta}_n} - A_0)]$ that results can be dealt with using Lemma 1 above. But we do not necessarily have that $\sqrt{n}(\hat{A}^\circ_{\theta_n} - A_0)$ converges weakly to a tight, cadlag limit. Condition (PB4.iii) is a weaker condition than this, which turns out to be easier to verify in our examples. In this case we use the following Lemma, taking $g_n = \hat{A}^\circ_{\theta_n} - A_0$ and

$$f_n(h)(u) = \mathbb{P}^\circ_n \left\{ \hat{\sigma}_{2,2}[\hat{\theta}_n + t_n(h)(\theta_n - \hat{\theta}_n), \hat{A}_{\hat{\theta}_n} + t_n(h)(\hat{A}^\circ_{\theta_n} - \hat{A}_{\hat{\theta}_n})](h)(u) \right\} - \sigma_{2,2}[\theta_0, A_0](h)(u).$$

LEMMA 2 *Suppose we have an expression of the form (5) where* $f_n(h) = o_p^{\mathcal{H}_2}(1)(h)$ *and* $||\sqrt{n} g_n||_2 \leq \tilde{O}_p(1) + o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\theta_n - \theta_0||_1) \equiv g_n^b$. *(The notation* $g_n^b$ *is referred to in the proof of this lemma.) Then* $Q_n(h) = o_p^{\mathcal{H}_2}(1)(h)$.

See Appendix A for a proof of this lemma.

The fact that the information operator is continuously invertible and onto in condition (PB6) is particularly useful in the last step of the proof of (6) in Appendix A.

## 6. DISCUSSION

We have demonstrated that in semiparametric models where both the finite dimensional parameter $\theta$ and the infinite dimensional parameter $A$ are $\sqrt{n}$ consistent, it is possible to significantly decrease the dimensionality of the maximization required for Monte Carlo inference compared to the weighted bootstrap.

The proposed piggyback bootstrap algorithm is easy to implement. First, draws for the parametric component $\theta_n$ are generated with the same sampling distribution as the MLE $\hat{\theta}_n$. Under the regularity conditions, this can often be accomplished by estimating the covariance matrix of $\hat{\theta}_n$ and adding normal deviates with this covariance to $\hat{\theta}_n$. In some models, such as the Cox model, a theoretical estimate of the covariance is readily available. In other models, such as the biased sampling model in the simulation study, one can use Corollary 3 of Murphy and van der Vaart (2000) to accomplish the variance estimation. This involves computing a limit involving the profile likelihood evaluated at various values of $\theta$. We gave details on this method in Section 4. An alternative is Lee's (2000) MCMC approach, which is based on other results from Murphy and van der Vaart (2000). Next, for each parametric draw $\theta_n$, we piggyback the draw for $A$ onto the draw for $\theta$ by performing the profile computation $A_n = \arg\max_A \ell_n^\circ(\theta_n, A)$. This computation can be accomplished using the fixed point algorithm given in Section 2. Our main result is that the resulting pair $(\theta_n, A_n)$ has the correct limit distribution for joint semiparametric inference, as discussed in Section 2.

The regularity conditions necessary for the method, given in Section 5, are not difficult to check in practice. Conditions (PB1), (PB2), (PB3), and (PB5) can be verified by computing derivatives of the log-likelihood and verifying that they have a specified structure. Although the Donsker classes mentioned in these conditions may be unfamiliar to some practitioners, verifying that a class of functions is Donsker is often very simple using the "Donsker Permanence" properties discussed in van der Vaart and Wellner (1996). We expect that practitioners will be turning to the piggyback bootstrap for models for which they already know maximum likelihood estimation is efficient and that the weighted bootstrap is valid. Condition (PB6) is the result that the MLEs are efficient, and Conditions (PB4.ii) follows from arguments similar to those that can be used to verify the validity of the weighted bootstrap. Condition (PB4.iii) is verified by relatively novel, but straightforward, arguments given in the proof of Lemma 9. Condition (PB4.i) requires that appropriate draws for the parametric component of the model are available. In the previous paragraph we discussed obtaining these draws in practice.

For right censored survival data, the method applies to the Cox proportional hazards model, the proportional odds model, and the odds-rate model. In the case of clustered survival data, the procedure applies to the shared and correlated gamma frailty models. An application not arising in survival analysis is to biased sampling models, which arise in vaccine efficacy trials.

In the odds-rate model, the frailty $W$ is assumed to be gamma distributed with mean 1 and variance $\gamma$. We are confident the method can be proved valid for frailties following other distributions with unknown parameters. Asymptotic theory for some such models has been worked out by Kosorok, Lee, and Fine (2004), including the validity of the weighted bootstrap. In addition, the method should be applicable to the more general proportional hazards random effects regression models for clustered survival data considered in Vaida and Xu (2000). However, in the more complex cases, asymptotic theory has not yet been worked out. Perhaps the most complex case worked out thus far is the correlated gamma frailty model of Parner (1998).

While the precise choice of the distribution of $\{\eta_i\}$ in the weighted bootstrap has no effect asymptotically, the rate of convergence may be affected. Newton and Raftery (1994) discuss different choices in the context of parametric maximum likelihood. They demonstrate that unit exponential weights, which are Dirichlet after standardizing, perform well. Our own experience is that exponential weights also work well for semiparametric inference.

Our main theoretical result on the validity of the piggyback bootstrap for semiparametric inference was confirmed in the simulation study in Section 4. We saw that for simulated vaccine trial data from a biased sampling model, the piggyback bootstrap and the weighted bootstrap performed close to the same in semiparametric inference, with the piggyback bootstrap providing a dramatic improvement in computational efficiency.

## APPENDIX A: PROOF OF MAIN THEOREM

In this appendix we prove the main theorem and the intermediate lemmas.

**Proof of Theorem 1**. This theorem follows from the expansion

$$\begin{aligned}\sqrt{n}(\hat{A}^{\circ}_{\theta_n}(t) - \hat{A}_{\hat{\theta}_n}(t)) &= -\sqrt{n}(\theta_n - \hat{\theta}_n)'\sigma_{2,1}[\theta_0, A_0]\left[\sigma_{2,2}[\theta_0, A_0]^{-1}(\mathrm{I}\{u \le t\})\right] \\ &\quad +\sqrt{n}(\hat{A}^{\circ}_{\theta_0}(t) - \hat{A}_{\theta_0}(t)) + o_p^{[0,\tau]}(1)(t), \qquad (6)\end{aligned}$$

which we prove below. Let $U_n$ have the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and let $V_n$ have the distribution of $\sqrt{n}(\hat{A}_{\theta_0}(t) - A_0(t))$ and let $U_n$ and $V_n$ be independent. Since the bootstrap weights and the draws $\theta_n$ are independent given the data, (PB4) implies

$$\begin{aligned}
&-\sqrt{n}(\theta_n - \hat{\theta}_n)'\sigma_{2,1}\left[\sigma_{2,2}^{-1}(\mathrm{I}\{u \le t\})\right] + \sqrt{n}(\hat{A}^\circ_{\theta_0}(t) - \hat{A}_{\theta_0}(t)) \\
&\quad\approx \quad -U_n'\sigma_{2,1}\left[\sigma_{2,2}^{-1}(\mathrm{I}\{u \le t\})\right] + V_n \\
&\quad\rightsquigarrow \quad -Z_\theta'\sigma_{2,1}\left[\sigma_{2,2}^{-1}(\mathrm{I}\{u \le t\})\right] + Z_A(t),
\end{aligned}$$

where $Z_\theta$ is mean zero Gaussian with covariance $\left(\sigma_{1,1} - \sigma_{1,2}\sigma_{2,2}^{-1}\sigma_{2,1}\right)^{-1}$ and $Z_A$ is mean zero Gaussian with covariance $\sigma_{2,2}^{-1}$, and $Z_\theta$ is independent of $Z_A$. Thus

$$\sqrt{n}\left(\theta_n - \hat{\theta}_n, \hat{A}^\circ_{\theta_n} - \hat{A}_{\hat{\theta}_n}\right) \approx Z,$$

where $Z' \equiv (Z_\theta, Z_A - \sigma_{2,2}^{-1}\sigma_{1,2}Z_\theta)'$ is mean zero Gaussian with covariance

$$\begin{bmatrix}
\left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1} & -\left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1}\sigma_{2,1}\sigma_{2,2}^{-1} \\
-\sigma_{2,2}^{-1}\sigma_{1,2}\left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1} & \sigma_{2,2}^{-1} + \sigma_{2,2}^{-1}\sigma_{1,2}\left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1}\sigma_{2,1}\sigma_{2,2}^{-1}
\end{bmatrix}$$

which is the inverse of the information operator.□

In what sense the operators in the above proof are covariances deserves some explanation. Let $i, j = 1, \ldots, d$, and $t, s \in [0, \tau]$. Then $\mathrm{Cov}(Z_{\theta_i}, Z_{\theta_j}) = e_i'\left(\sigma_{1,1} - \sigma_{1,2}\sigma_{2,2}^{-1}\sigma_{2,1}\right)^{-1}e_j$, where $e_k$, $k = 1, \ldots, d$ denotes a vector whose $k$th component is one and all the others are zero; $\mathrm{Cov}(Z_A(t), Z_A(s)) = \int_0^\tau \mathrm{I}\{u \le t\}\sigma_{2,2}^{-1}(\mathrm{I}\{u \le s\})dA_0(u)$;

$$\mathrm{Cov}(Z_{\theta_i}, [Z_A - \sigma_{2,2}^{-1}\sigma_{1,2}Z_\theta](t)) = -e_i'\int_0^\tau \left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1}\sigma_{2,1}\sigma_{2,2}^{-1}(\mathrm{I}\{u \le t\})dA_0(u)$$

and

$$\mathrm{Cov}([Z_A - \sigma_{2,2}^{-1}\sigma_{1,2}Z_\theta](t), Z_{\theta_i}) = -\int_0^\tau \mathrm{I}\{u \le t\}\left[\sigma_{2,2}^{-1}\sigma_{1,2}\left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1}e_i\right](u)dA_0(u).$$

(That the last two covariances are equal is a consequence of the commutativity of repeated differentiation.) Finally,

$$\begin{aligned}
&\mathrm{Cov}([Z_A - \sigma_{2,2}^{-1}\sigma_{1,2}Z_\theta](t), [Z_A - \sigma_{2,2}^{-1}\sigma_{1,2}Z_\theta](s)) \\
&\quad= \int_0^\tau \mathrm{I}\{u \le t\}[\sigma_{2,2}^{-1} + \sigma_{2,2}^{-1}\sigma_{1,2}\left(\sigma_{1,1} - \sigma_{2,1}\sigma_{2,2}^{-1}\sigma_{1,2}\right)^{-1}\sigma_{2,1}\sigma_{2,2}^{-1}](\mathrm{I}\{u \le s\})dA_0(u).
\end{aligned}$$

Before verifying expansion (6), we need the following preliminary lemmas in addition to Lemmas 1 and 2 given in Section 5. (Note that the proofs of Lemmas 1 and 2 are given at the end of this section.)

LEMMA 3 *Under conditions (PB2) and (PB3), if $\tilde{\theta}_n \to_p \theta_0$ and $\tilde{A}_n \to_p A_0$, then*

$$\mathbb{P}_n \hat{\sigma}_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) - \sigma_{2,j}[\theta_0, A_0](h) = o_p^{\mathcal{H}_2}(1)(h), \; j = 1, 2.$$

*This result also holds with $\mathbb{P}_n^{\circ}$ replacing $\mathbb{P}_n$.*

**Proof**. Fix $\epsilon > 0$ and observe that for $j = 1, 2$,

$$\begin{aligned}
&\lim_{n\to\infty} P^* \left( \left|\left| \mathbb{P}_n \hat{\sigma}_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) - \sigma_{2,j}[\theta_0, A_0](h) \right|\right|_j^{\mathcal{H}_2} > \epsilon \right) \\
&\leq \lim_{m\to\infty} \lim_{n\to\infty} \left\{ P^* \left( ||\tilde{\theta}_n - \theta_0||_1 > \frac{1}{m} \right) + P^* \left( ||\tilde{A}_n - A_0||_2 > \frac{1}{m} \right) + \right. \\
&\quad \left. P^* \left( \left|\left| \mathbb{P}_n \hat{\sigma}_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) - \sigma_{2,j}[\theta_0, A_0](h) \right|\right|_j^{\mathcal{H}_2} > \epsilon, ||\tilde{\theta}_n - \theta_0||_1 \leq \frac{1}{m}, ||\tilde{A}_n - A_0||_2 \leq \frac{1}{m} \right) \right\} \\
&\leq \lim_{m\to\infty} \lim_{n\to\infty} P^* \Bigg( \left|\left| \mathbb{P}_n \hat{\sigma}_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) - \sigma_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) \right|\right|_j^{\mathcal{H}_2} \\
&\quad + \left|\left| \sigma_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) - \sigma_{2,j}[\theta_0, A_0](h) \right|\right|_j^{\mathcal{H}_2} > \epsilon, ||\tilde{\theta}_n - \theta_0||_1 \leq \frac{1}{m}, ||\tilde{A}_n - A_0||_2 \leq \frac{1}{m} \Bigg) \\
&\leq \lim_{m\to\infty} \lim_{n\to\infty} P^* \Bigg( \left|\left| \mathbb{P}_n \hat{\sigma}_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) - \sigma_{2,j}[\tilde{\theta}_n, \tilde{A}_n](h) \right|\right|_j^{\mathcal{H}_2} > \tilde{\epsilon}, \\
&\qquad ||\tilde{\theta}_n - \theta_0||_1 \leq \frac{1}{m}, ||\tilde{A}_n - A_0||_2 \leq \frac{1}{m} \Bigg),
\end{aligned}$$

for some $\tilde{\epsilon} > 0$ by condition (PB2). The last expression equals zero by condition (PB3) and the fact that a Donsker class is also Glivenko-Cantelli. For the case of $\mathbb{P}_n^{\circ}$ replacing $\mathbb{P}_n$, make the same replacement in the proof, and note that condition (PB3) implies $\{\eta \hat{\sigma}_{2,j}[\theta, A](h), \; ||\theta - \theta_0|| \leq c, ||A - A_0|| \leq c, h \in \mathcal{H}_2\}$ is Donsker and hence Glivenko-Cantelli for $j = 1, 2$.□

LEMMA 4

$$\begin{aligned}
\sqrt{n}(\mathbb{P}_n^{\circ} - \mathbb{P}_n) U_2(\hat{\theta}_n, \hat{A}_{\hat{\theta}_n})(h) &= \sqrt{n}(\theta_n - \hat{\theta}_n)' \sigma_{2,1}[\theta_0, A_0](h) + \\
&\qquad \sqrt{n} \int_0^{\tau} \sigma_{2,2}[\theta_0, A_0](h)(u) d(\hat{A}_{\theta_n}^{\circ} - \hat{A}_{\hat{\theta}_n})(u) + o_p^{\mathcal{H}_2}(1)(h)
\end{aligned}$$

**Proof**. By definition of $\hat{A}_{\theta}^{\circ}$ and $\hat{A}_{\theta_n}$ we have

$$0 = (\mathbb{P}_n^{\circ} - \mathbb{P}_n) U_2(\hat{\theta}_n, \hat{A}_{\hat{\theta}_n})(h) + \mathbb{P}_n^{\circ}(U_2(\theta_n, \hat{A}_{\theta_n}^{\circ})(h) - U_2(\hat{\theta}_n, \hat{A}_{\hat{\theta}_n})(h)).$$

Using 2.4.8 of Abraham, Marsden, and Ratiu (1988) (hereafter abbreviated AMR) to expand the rightmost term of the preceding equation, under condition (PB1), we obtain

$$(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\hat\theta_n, \hat A_{\hat\theta_n})(h) = -\mathbb{P}_n^\circ \left\{ \frac{\partial}{\partial s} U_2 \left[ \hat\theta_n + s(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + s(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n}) \right](h) \Big|_{s=t_n(h)} \right\} =$$

$$-\mathbb{P}_n^\circ \left\{ \frac{\partial}{\partial s} U_2 \left[ \hat\theta_n + t_n(h)(\theta_n - \hat\theta_n) + s(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n}) + s(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n}) \right](h) \Big|_{s=0} \right\}$$

$$= \mathbb{P}_n^\circ \left\{ (\theta_n - \hat\theta_n)' \hat\sigma_{2,1}[\hat\theta_n + t_n(h)(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](h) \right\}$$

$$+\mathbb{P}_n^\circ \left\{ \int_0^\tau \hat\sigma_{2,2}[\hat\theta_n + t_n(h)(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](h)(u) d(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})(u) \right\},$$

for some $t_n(h) \in [0,1]$. Note that

$$\sigma_{2,j}[\theta_0, A_0](h) = \mathbb{P}_n^\circ(\hat\sigma_{2,j}[\hat\theta_n + t_n(h)(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](h)) + o_p^{\mathcal{H}_2}(1)(h)$$

by Lemma 3 and conditions (PB2), (PB3) and (PB4). This, condition (PB4), and Lemmas 1 and 2 imply,

$$\sqrt{n}(\theta_n - \hat\theta_n)'\sigma_{2,1}(h) =$$
$$\sqrt{n}\mathbb{P}_n^\circ \left\{ (\theta_n - \hat\theta_n)' \hat\sigma_{2,1}[\hat\theta_n + t_n(h)(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](h) \right\} + o_p^{\mathcal{H}_2}(1)(h),$$

and

$$\sqrt{n}\mathbb{P}_n^\circ \left\{ \int_0^\tau \hat\sigma_{2,2}[\hat\theta_n + t_n(h)(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](h)(u) d(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})(u) \right\}$$
$$= \int_0^\tau \mathbb{P}_n^\circ \left\{ \hat\sigma_{2,2}[\hat\theta_n + t_n(h)(\theta_n - \hat\theta_n), \hat A_{\hat\theta_n} + t_n(h)(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](h)(u) \right\} d[\sqrt{n}(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})](u)$$
$$= \sqrt{n} \int_0^\tau \sigma_{2,2}[\theta_0, A_0](h)(u) d(\hat A^\circ_{\theta_n} - \hat A_{\hat\theta_n})(u) + o_p^{\mathcal{H}_2}(1)(h). \square$$

LEMMA 5

$$\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\hat\theta_n, \hat A_{\hat\theta_n})(h) \approx \sqrt{n}(\mathbb{P}_n - P_0)U_2(\theta_0, A_0)(h).$$

**Proof**. If we show that

$$\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\hat\theta_n, \hat A_{\hat\theta_n})(h) = \sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\theta_0, A_0)(h) + o_p^{\mathcal{H}_2}(1)(h),$$

then the conclusion follows as a consequence of Theorem 2.9.6 of VW. To see this, note that by Slutsky's Theorem (Example 1.4.7 of VW) and condition (PB5),

$$\begin{aligned}
\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\theta_0, A_0)(h) &= \frac{\sqrt{n}\mathbb{P}_n \eta U_2(\theta_0, A_0)(h)}{\mathbb{P}_n \eta} - \sqrt{n}\mathbb{P}_n U_2(\theta_0, A_0)(h) \\
&= \sqrt{n}\mathbb{P}_n \eta U_2(\theta_0, A_0)(h) - \sqrt{n}\mathbb{P}_n U_2(\theta_0, A_0)(h) + o_p^{\mathcal{H}_2}(1)(h) \\
&= \sqrt{n}\mathbb{P}_n(\eta - 1)(U_2(\theta_0, A_0)(h) - P_0 U_2(\theta_0, A_0)(h)) + o_p^{\mathcal{H}_2}(1)(h) \\
&\approx \sqrt{n}(\mathbb{P}_n - P_0)U_2(\theta_0, A_0)(h).
\end{aligned}$$

The last line follows by the aforementioned theorem from VW. Thus all that remains is to prove the first display of the proof.

To establish the first display of the proof, note that 2.4.8 of AMR reveals

$$\begin{aligned}
&\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\hat{\theta}_n, \hat{A}_{\hat{\theta}_n})(h) - \sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)U_2(\theta_0, A_0)(h) = \\
&-\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)\left\{(\hat{\theta}_n - \theta_0)'\hat{\sigma}_{2,1}[\theta_0 + t_n(h)(\hat{\theta}_n - \theta_0), A_0 + t_n(h)(\hat{A}_{\hat{\theta}_n} - A_0)](h)\right\} + \\
&-\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)\left\{\int_0^\tau \hat{\sigma}_{2,2}[\theta_0 + t_n(h)(\hat{\theta}_n - \theta_0), A_0 + t_n(h)(\hat{A}_{\hat{\theta}_n} - A_0)](h)(u)d(\hat{A}_{\hat{\theta}_n} - A_0)(u)\right\},
\end{aligned}$$

for some $t_n(h) \in [0,1]$. Note that $(\mathbb{P}_n^\circ - \mathbb{P}_n)\hat{\sigma}_{2,j}[\theta_0 + t_n(h)(\hat{\theta}_n - \theta_0), A_0 + t_n(h)(\hat{A}_{\hat{\theta}_n} - A_0)](h) = o_p^{\mathcal{H}_2}(1)(h)$ for $j = 1, 2$ by by conditions (PB2), (PB3) and (PB4), and Lemma 3. This, Lemma 1, and condition (PB4) completes the proof of the first display, and hence, the lemma.□

LEMMA 6

$$\sqrt{n}(\mathbb{P}_n - P_0)U_2(\theta_0, A_0)(h) = \sqrt{n}\int_0^\tau \sigma_{2,2}[\theta_0, A_0](h)(u)d(\hat{A}_{\theta_0} - A_0)(u) + o_p^{\mathcal{H}_2}(1)(h)$$

**Proof**. By condition (PB5), the definition of $\hat{A}_{\theta_0}$, and 2.4.8 of AMR,

$$\begin{aligned}
&\sqrt{n}(\mathbb{P}_n - P_0)U_2(\theta_0, A_0)(h) = \sqrt{n}\mathbb{P}_n(U_2(\theta_0, A_0)(h) - U_2(\theta_0, \hat{A}_{\theta_0})(h)) \\
&= \sqrt{n}\mathbb{P}_n\left\{\int_0^\tau \hat{\sigma}_{2,2}[\theta_0, A_0 + t_n(h)(\hat{A}_{\theta_0} - A_0)](h)(u)d(\hat{A}_{\theta_0} - A_0)(u)\right\} + o_p^{\mathcal{H}_2}(1)(h),
\end{aligned}$$

for some $t_n(h) \in [0,1]$. Note that

$$\mathbb{P}_n\hat{\sigma}_{2,2}[\theta_0, A_0 + t_n(h)(\hat{A}_{\theta_0} - A_0)](h) = \sigma_{2,2}[\theta_0, A_0](h) + o_p^{\mathcal{H}_2}(1)(h)$$

by conditions (PB2), (PB3) and (PB4), and Lemma 3. This, Lemma 1, and condition (PB4) completes the proof. □

**Proof of (6)**. Note that by condition (PB4),

$$\sqrt{n}\int_0^\tau \sigma_{2,2}[\theta_0, A_0](h)(u)d(\hat{A}^\circ_{\theta_0} - \hat{A}_{\theta_0})(u) \approx \sqrt{n}\int_0^\tau \sigma_{2,2}[\theta_0, A_0](h)(u)d(\hat{A}_{\theta_0} - A_0)(u).$$

This and the conclusions of Lemmas 4, 5, and 6 give

$$\sqrt{n}\int_0^\tau \sigma_{2,2}[\theta_0, A_0](h)(u)d(\hat{A}^\circ_{\theta_0} - \hat{A}_{\theta_0})(u) =$$
$$\sqrt{n}(\theta_n - \hat{\theta}_n)'\sigma_{2,1}[\theta_0, A_0](h) + \sqrt{n}\int_0^\tau \sigma_{2,2}[\theta_0, A_0](h)(u)d(\hat{A}^\circ_{\theta_n} - \hat{A}_{\hat{\theta}_n})(u) + o_p^{\mathcal{H}_2}(1)(h).$$

Since $\mathcal{H}_2$ includes all functions of the form $h(u) = \sigma_{2,2}[\theta_0, A_0]^{-1}(\mathrm{I}\{u \le t\})$, for all $t \in [0, \tau]$, we have established (6), and the proof is complete.□

**Proof of Lemma 1**. We employ arguments from the proof of Lemma A.3 of Bilias et. al (1997). Define $g$ to be the weak limit of $\sqrt{n}g_n$ and $G_n = \sqrt{n}g_n - g$. Note that since the $f_n(h)(u)$ has total variation bounded by $M < \infty$, we can write $f_n(h)(u) = f_n^+(h)(u) + f_n^-(h)(u)$, where $f_n^+(h)(u)$ is an increasing function of $u$, and $f_n^-(h)(u)$ is a decreasing function of $u$. Further, each of these functions can be written as the sum of a cadlag and a caglad function: $f_n^+(h)(u) = f_n^{+,d}(h)(u) + f_n^{+,g}(h)(u)$ and $f_n^-(h)(u) = f_n^{-,d}(h)(u) + f_n^{-,g}(h)(u)$. And since $f_n(h) = o_p^{\mathcal{H}_2}(1)(h)$, the latter four functions are $o_p^{\mathcal{H}_2}(1)(h)$ as well. Let $f_n^a(h)(u)$ denote one of these four functions. Observe that

$$\sqrt{n}\int_0^\tau f_n^a(h)(u)dg_n(u) = f_n^a(h)(\tau)\sqrt{n}g_n(\tau) - f_n^a(h)(0)\sqrt{n}g_n(0)$$
$$+\sqrt{n}\sum \Delta g_n(u)\Delta f_n^{a,r}(h)(u) - \sqrt{n}\int_0^\tau g_n(u)df_n^{a,r}(h)(u) \qquad (7)$$

where $f_n^{a,r}(h)(u)$ denotes the right continuous version of $f_n^a(h)(u)$. The first three terms on the right hand side of (7) are $o_p^{\mathcal{H}_2}(1)(h)$ by Slutsky's Theorem (Example 1.4.7 of VW). The last term is $o_p^{\mathcal{H}_2}(1)(h)$ by the following argument.

Abbreviate $f_n^b(h)(u) \equiv f_n^{a,r}(h)(u)$. Observe that $G_n(u) = o_p^{[0,\tau]}(1)(u)$ and so, since $f_n^b(h)(u)$ has variation bounded by $M < \infty$,

$$\int_0^\tau G_n(u)df_n^b(h)(u) = o_p^{\mathcal{H}_2}(1)(h). \qquad (8)$$

Let $\tilde{f}_n^b$, $\tilde{g}_n$ and $\tilde{g}$ be almost sure representations of $f_n^b$, $g_n$ and $g$, respectively such that $\tilde{g}(u)$ is cadlag and tight, and $\tilde{f}_n^b(h)(u)$ has variation bounded by $M$ over all $n$ and has the same monotonicity (increasing or decreasing) as $f_n^b(h)(u)$ (see Theorem 1.10.4 and Addendum 1.10.5 of VW). Let $(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{P})$ be the probability space corresponding to $\tilde{f}_n^b$, $\tilde{g}_n$ and $\tilde{g}$. Use $o_{a.s.*}^{\mathcal{H}_2}(1)(h)$

to denote a term $F_n(h)$ such that $\sup_{h\in\mathcal{H}_2}||F_n(h)||$ converges outer almost surely to 0 as $n\to\infty$. Since $\tilde{f}_n^b(h)=o_{a.s.*}^{\mathcal{H}_2}(1)(h)$ there are measurable $\Delta_n$ such that $\sup_{h\in\mathcal{H}_2}||\tilde{f}_n^b(h)||\le\Delta_n$ and $\Delta_n=o_{a.s.*}^{\mathcal{H}_2}(1)(h)$. Since $\tilde{g}(u)$ is cadlag, for arbitrary $\epsilon>0$ there is a set $B_\epsilon\in\tilde{\mathcal{B}}$ that has $\tilde{P}$-probability 1 with the following property. For each fixed $\omega\in B_\epsilon$ we can find a partition $0=u_0<u_1<\cdots<u_k$ and constants $\tilde{g}_i$ such that the simple function

$$\tilde{g}_\epsilon(u)\equiv\sum_{i=1}^k\tilde{g}_iI_{u\in(u_{i-1},u_i]}$$

satisfies $\sup_{u\in[0,\tau]}|\tilde{g}_\epsilon(u)-\tilde{g}(u)|<\epsilon$. Thus for the fixed $\omega\in B_\epsilon$,

$$\begin{aligned}\left|\int_0^\tau\tilde{g}(u)d\tilde{f}_n^b(h)(u)\right|&\le\left|\int_0^\tau[\tilde{g}(u)-\tilde{g}_\epsilon(u)]d\tilde{f}_n^b(h)(u)\right|+\left|\int_0^\tau\tilde{g}_\epsilon(u)d\tilde{f}_n^b(h)(u)\right|\\&\le\epsilon M+2\sum_{i=1}^k|\tilde{g}_i|\sup_{u\in[0,\tau]}|\tilde{f}_n^b(h)(u)|\le\epsilon M+2\sum_{i=1}^k|\tilde{g}_i|||\tilde{f}_n^b(h)(u)||\\&\le\epsilon M+2\sum_{i=1}^k|\tilde{g}_i|\Delta_n\equiv\epsilon M+Q_n(\epsilon,\omega).\end{aligned}$$

Since $g$ and $\Delta_n$ are measurable as functions of $\omega$, so is $Q_n(\epsilon)$. Thus, since for each fixed $\omega\in B_\epsilon$, $\sup_{h\in\mathcal{H}_2}Q_n(\epsilon,\omega)\to0$ as $n\to\infty$, it follows that $Q_n(\epsilon)=o_p^{\mathcal{H}_2}(1)$. And so for $\delta>\epsilon M$,

$$\lim_{n\to\infty}P^*\left(\sup_{h\in\mathcal{H}_2}\left|\int_0^\tau\tilde{g}(u)d\tilde{f}_n^b(h)(u)\right|>\delta\right)\le\lim_{n\to\infty}P^*\left(\sup_{h\in\mathcal{H}_2}|Q(\epsilon)|>\delta-\epsilon M\right)=0.$$

Since $\epsilon>0$ was arbitrary, this gives $\int_0^\tau\tilde{g}(u)d\tilde{f}_n^b(h)(u)=o_p^{\mathcal{H}_2}(1)(h)$. And hence

$$\left|\int_0^\tau\tilde{g}(u)d\tilde{f}_n^b(h)(u)\right|\wedge C\rightsquigarrow0$$

for some $0<C<\infty$. By property (ii) of Theorem 1.10.4 of VW this gives that the last term of (7) is $o_p^{\mathcal{H}_2}(1)(h)$ □.

**Proof of Lemma 2**.

A similar argument to the proof of Lemma 1 above works with $f_n^b(h)(u)$ taken to be $f_n^{a,r}(h)(u)$ if $f_n^{a,r}(h)(u)$ is increasing or $-f_n^{a,r}(h)(u)$ if $f_n^{a,r}(h)(u)$ is decreasing, and $\sqrt{n}g_n$ replaced by its $\tilde{O}_p(1)+o_p^{\mathcal{H}_2}(1+\sqrt{n}||\theta_n-\theta_0||_1)(h)$ bound $g_n^b$. And so, since in these cases

$$\left|\int_0^\tau\sqrt{n}g_n(u)df_n^b(h)(u)\right|\le\int_0^\tau||\sqrt{n}g_n(u)||_2df_n^b(h)(u)\le\int_0^\tau g_n^b(u)df_n^b(h)(u)$$

it follows that (7) is $o_p^{\mathcal{H}_2}(1)(h)$. □

# APPENDIX B: VERIFYING THE REGULARITY CONDITIONS FOR EXAMPLES

In this appendix we verify that the regularity conditions hold for the odds-rate and biased sampling models, adding some technical restrictions for each model as necessary. See Dixon (2003) for a verification of the regularity conditions and the additional technical restrictions for the other example models. We start with a general outline that the arguments follow in all of the examples, and which should serve as a guide to verifying the regularity conditions in other applications.

As our approach is an alternative to the weighted bootstrap, we note that the validity of the weighted bootstrap for the Cox proportional hazards model, the proportional odds model, and the odds-rate model follows from Kosorok, Lee and Fine (2004). For the shared and correlated gamma frailty models, the validity of the weighted bootstrap which assigns weight to clusters (as described in Section 3) follows by similar arguments. And Gilbert (1996) establishes the validity of the weighted bootstrap for the biased sampling example.

**General Outline**

Verification of the regularity conditions follows the same outline in each of our examples. The same outline should be applicable to other semiparametric models.

Conditions (PB1) and (PB2) can be readily verified by inspection of the score and information operators, which we give for the odds-rate and biased sampling models below. Condition (PB6) can be verified in our examples using the Z-Estimator Master Theorem (Theorem 3.3.1 of VW) and Convolution Theorems (Theorems 5.1-5.3 of Bickel et al. 1993). The fact that the information operator is continuously invertible and onto is established using a standard result from functional analysis (see Theorem 3.4 and Corollary 3.8 of Kress 1989).

Verification of Conditions (PB3) and (PB5) is routine in our examples using Donsker theorems from VW. As an example, consider the piece $H_\psi^X(f) \equiv \int_0^t f(s)e^{\theta' Z(s)}dA(s)$ which appears in many of our survival analysis examples. We have the following Lemma.

Lemma 7 *For each fixed* $p < \infty$ *and each fixed* $\psi \in \Theta_q \times BV_r$ *with* $q, r < \infty$*, the class* $\{H_\psi^\tau(f) : f \in BV_p\}$ *is* $P_0$*-Donsker. Here* $\Theta_q \equiv \{\theta : ||\theta|| \leq q\}$.

**Proof**

Classes of cells and uniformly bounded classes of functions of bounded variation are standard examples of Donsker classes. Thus $Y(\cdot)(\equiv \mathrm{I}\{X \geq \cdot\})$, $\beta' Z(\cdot)$, and $f(\cdot)$ are uniformly bounded Donsker classes on $[0,\tau]$. The exponential function is Lipschitz on compact sets, therefore the Lipschitz Transformation Donsker Permanence Theorem (Theorem 2.10.6 of VW) gives that $\exp(\beta' Z(\cdot))$ is Donsker on $[0,\tau]$. Since the product of two Lipschitz functions is Lipschitz on compact sets, another application of this theorem gives that $f(\cdot)e^{\beta' Z(\cdot)}Y(\cdot)$ is Donsker. Thus the claim follows from the continuous mapping theorem. □

Condition (PB4.ii) can be verified in our examples by the following argument, which is similar to the proof of the validity of the weighted bootstrap for proportional hazards frailty regression models given in Kosorok, Lee, and Fine (2004).

LEMMA 8 *Condition (PB4.ii) holds in our examples.*

**Sketch of Proof**. Define $\hat{\psi}^\circ_\theta \equiv (\theta, \hat{A}^\circ_\theta)$ and $\hat{\psi}_\theta \equiv (\theta, \hat{A}_\theta)$. Applying the Z-Estimator Master Theorem of VW gives that $\sqrt{n}(\hat{\psi}^\circ_{\theta_0} - \psi_0) = \sqrt{n}\mathbb{P}^\circ_n U^\tau(\psi_0)(\sigma^{-1}(\cdot)) + o_p(1)$ unconditionally, and $\sqrt{n}(\hat{\psi}_{\theta_0} - \psi_0) = \sqrt{n}\mathbb{P}_n U^\tau(\psi_0)(\sigma^{-1}(\cdot)) + o_p(1)$, where $o_p(1)$ denotes a quantity $\to 0$ uniformly in outer probability. Thus $\sqrt{n}(\hat{\psi}^\circ_{\theta_0} - \hat{\psi}_{\theta_0}) = \sqrt{n}(\mathbb{P}^\circ_n - \mathbb{P}_n)U^\tau(\psi_0)(\sigma^{-1}(\cdot)) + o_p(1)$ unconditionally. Finally, note that

$$\begin{aligned}
\sqrt{n}(\mathbb{P}^\circ_n - \mathbb{P}_n)U^\tau(\psi_0)(h) &= \frac{\sqrt{n}\mathbb{P}_n \eta U^\tau(\psi_0)(h)}{\mathbb{P}_n \eta} - \sqrt{n}\mathbb{P}_n U^\tau(\psi_0)(h) \\
&= \sqrt{n}\mathbb{P}_n \eta U^\tau(\psi_0)(h) - \sqrt{n}\mathbb{P}_n U^\tau(\psi_0)(h) + o_p^{\mathcal{H}_2}(1)(h) \\
&= \sqrt{n}\mathbb{P}_n(\eta - 1)(U^\tau(\psi_0) - P_0 U^\tau(\psi_0))(h) + o_p^{\mathcal{H}_2}(1)(h) \\
&\approx \sqrt{n}(\mathbb{P}_n - P_0)U^\tau(\psi_0)(h),
\end{aligned}$$

where the last relation follows by the Multiplier Central Limit Theorem (Theorem 2.9.6 of VW).□

Condition (PB4.iii) can be verified in our examples by a similar argument to one used in the proof of Theorem 3.4 of Lee (2000), which we give in the proof of the following lemma:

LEMMA 9 *Condition (PB4.iii) holds in our examples.*

**Sketch of Proof**. Define

$$\tilde{D}_n(A)(h) \equiv \mathbb{P}^\circ_n U^\tau \begin{pmatrix} \theta_n \\ A \end{pmatrix} \begin{pmatrix} 0 \\ h \end{pmatrix}, \quad D_n(A)(h) \equiv \mathbb{P}^\circ_n U^\tau \begin{pmatrix} \theta_0 \\ A \end{pmatrix} \begin{pmatrix} 0 \\ h \end{pmatrix},$$

and

$$D_0(A)(h) \equiv P_0 U^\tau \begin{pmatrix} \theta_0 \\ A \end{pmatrix} \begin{pmatrix} 0 \\ h \end{pmatrix}.$$

Note that $\tilde{D}_n(\hat{A}^\circ_{\theta_n}) = 0$ and $D_0(A_0) = 0$. We can use Lemma 3.3.5 of VW to obtain

$$\sqrt{n}(\tilde{D}_n - D_0)(\hat{A}^\circ_{\theta_n}) - \sqrt{n}(\tilde{D}_n - D_0)(A_0) = o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\hat{\psi}^\circ_n - \psi_0||_3)(h),$$

and

$$\begin{aligned}\sqrt{n}(\tilde{D}_n - D_n)(A_0) &= \sqrt{n}(\tilde{D}_n - D_0)(A_0) - \sqrt{n}(D_n - D_0)(A_0)\\ &= o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\theta_n - \theta_0||_1)(h),\end{aligned}$$

from which we have

$$\begin{aligned}\sqrt{n}(D_0(\hat{A}^\circ_{\theta_n}) - D_0(A_0)) &= \sqrt{n}(D_0(\hat{A}^\circ_{\theta_n}) - \tilde{D}_n(\hat{A}^\circ_{\theta_n}))\\ &= -\sqrt{n}(\tilde{D}_n - D_0)(A_0) + o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\hat{\psi}^\circ_n - \psi_0||_3)(h)\\ &= -\sqrt{n}(D_n - D_0)(A_0) + o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\hat{\psi}^\circ_n - \psi_0||_3)(h)\\ &= \tilde{O}_p(1) + o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\hat{\psi}^\circ_n - \psi_0||_3)(h).\end{aligned}$$

The last equality in the preceding display follows from the Z-Estimator master theorem of VW. Since $\sigma$ is continuously invertible, we have for some $c > 0$ that

$$\|D_0(A) - D_0(A_0)\| \geq c||A - A_0||_2 + o(||A - A_0||_2).$$

Thus, $\sqrt{n}||\hat{A}^\circ_{\theta_n} - A_0||_2(c + o_p^{\mathcal{H}_2}(1)(h)) \leq \tilde{O}_p(1) + o_p^{\mathcal{H}_2}(1 + \sqrt{n}||\theta_n - \theta_0||_1)(h)$. Multiplying each side by $(c + o_p^{\mathcal{H}_2}(1)(h))^{-1}$ and applying Slutsky's Theorem (Example 1.4.7 of VW), we have the desired result. □

Draws $\theta_n$ satisfying Condition (PB4.i) can be generated in our examples by applying results from Murphy and van der Vaart (2000).

LEMMA 10 *The hypotheses of Corollary 3 of Murphy and van der Vaart (2000) hold, and so we can obtain draws for the parametric component of the model satisfying (PB4.i).*

**Sketch of Proof**. Based on the discussion in Chapter 25 of van der Vaart (2000) an approximately least favorable submodel for estimating $\theta$ in the presence of $A$ is

$$dA_t(\theta, A) = (1 + (\theta - t)^{\mathrm{T}}\nu_0)\, dA, \tag{9}$$

where $\nu_0 : \mathbb{R} \mapsto \mathbb{R}^d$ is the least favorable direction at $(\theta_0, A_0)$ defined by

$$h^{\mathrm{T}}\nu_0 = \sigma_{2,2}^{-1}\sigma_{1,2}h, \quad h \in \mathbb{R}^d.$$

Let $\tilde{\theta}_n \to \theta_0$ almost surely under $P_0$. It follows from the definition of $\hat{A}_\theta$ that $\mathrm{L}_n(\tilde{\theta}_n, \hat{A}_{\tilde{\theta}_n}) \geq \mathrm{L}_n(\tilde{\theta}_n, \hat{A}_{\hat{\theta}_n})$. Therefore, noting that $(\tilde{\theta}_n, \hat{A}_{\hat{\theta}_n})$ is consistent for $(\theta_0, A_0)$, the method for proving consistency of the MLE in Theorem 3.2 of Lee (2000) (see also Theorem 1 of Parner, 1998 or Theorem 3 of Kosorok, Lee and Fine, 2004) can be adapted in a straightforward manner to show that $\hat{A}_{\tilde{\theta}_n} \to A_0$ almost surely under $P_0$. Upon substituting $\theta = t$ and $A = A_t(\theta, A)$ into the log-likelihood function, it is seen that the path is smooth in $t$ and is continuously differentiable. Arguments from the proof of Lemma 7 give that the Glivenko-Cantelli and Donsker conditions in Murphy and van der Vaart (2000) are satisfied. All that remains to be established is that their equation (11) holds. Murphy and van der Vaart note that this equation holds whenever

$$\|\hat{A}_{\tilde{\theta}_n} - A_0\| = O_P\left(\|\tilde{\theta}_n - \theta_0\|\right) + o_P(n^{-1/2}).$$

This is certainly implied by

$$\sqrt{n}\|\hat{A}_{\tilde{\theta}_n} - A_0\| \leq O_P(1) + o_P(1 + \sqrt{n}\|\tilde{\theta}_n - \theta_0\|),$$

which can be proven in a similar manner to Lemma 9. □

### B.1 Odds-Rate Regression

The assumptions we make are somewhat standard and are closely related to the assumptions made in Parner (1998). We assume for the true parameter $\psi_0 = (\gamma_0, \beta_0, A_0)$ that $\gamma_0 \in [0, M_0)$, where $M_0 < \infty$, that $\beta_0 \neq 0$ is in a compact subset of $\mathbb{R}^p$, and that $A_0$ is continuous on $[0, \tau]$ with derivative $a_0$ satisfying $0 < a_0(t) < \tau$ for all $t \in [0, \tau]$. We also assume that censoring is independent of $T$ given $Z$ and uninformative of $\psi = (\gamma, \beta, A)$; that the analysis is restricted to the interval $[0, \tau]$ with $\tau < \infty$ such that $\mathrm{P}[C \geq \tau | Z] = \mathrm{P}[C = \tau | Z] > 0$ almost surely; that the total variation of $Z(\cdot)$ on $[0, \tau]$ is $\leq M_1 < \infty$; and that $\mathrm{var}[Z(0+)]$ is positive definite.

Using empirical process techniques, Lee (2000) and Kosorok, Lee and Fine (2004) establish that $\hat{\psi}_n$ is uniformly consistent for $\psi_0$ outer almost surely, that $\sqrt{n}(\hat{\psi}_n - \psi_0)$ converges weakly to a tight Gaussian process (in the uniform metric), and that $\hat{\psi}_n$ is regular and fully efficient. Thus condition (PB6) is established. Condition (PB4.ii) is established using the arguments from the proof of Lemma 8, and similar arguments establish the validity of the weighted bootstrap. Arguments from the proof of Lemma 9 establish that (PB4.iii) holds. For $h = (h_1, h_2)$, with $h_1 = (h_{11}, h_{12})^T \in \mathbb{R} \times \mathbb{R}^p$ and $h_2 \in BV([0, \tau])$, the score and information operators have the following form:

$$
\begin{aligned}
U_1[\theta,A](h_1) &= \left\{\log(1+\gamma H_\psi^X(1)) - \frac{(1+\delta\gamma)\gamma H_\psi^X(1)}{1+\gamma H_\psi^X(1)}\right\}\frac{h_{11}}{\gamma^2} \\
&\quad + \int_0^\tau h_{12}' Z(u)dN(u) - \left\{\frac{(1+\delta\gamma)H_\psi^X(Z'h_{12})}{1+\gamma H_\psi^X(1)}\right\}, \\
U_2[\theta,A](h_2) &= \int_0^\tau h_2(u)dN(u) - \left\{\frac{(1+\delta\gamma)H_\psi^X(h_2)}{1+\gamma H_\psi^X(1)}\right\}, \\
\hat\sigma_{1,1}[\theta,A](h_1) &= \begin{bmatrix} \hat\sigma_{11,11}[\theta,A](h_{11}) + \hat\sigma_{11,12}[\theta,A](h_{12}) \\ \hat\sigma_{12,11}[\theta,A](h_{11}) + \hat\sigma_{12,12}[\theta,A](h_{12}) \end{bmatrix} \\
\hat\sigma_{2,1}[\theta,A](h_2) &= \begin{bmatrix} \hat\sigma_{2,11}[\theta,A](h_2) \\ \hat\sigma_{2,12}[\theta,A](h_2) \end{bmatrix},
\end{aligned}
$$

where

$$
\hat\sigma_{11,11}[\theta,A](h_{11}) = \left\{2\log(1+\gamma H_\psi^X(1)) - \frac{(3+\delta\gamma)\gamma H_\psi^X(1)}{1+\gamma H_\psi^X(1)} + \frac{(1+\delta\gamma)\gamma H_\psi^X(1)}{\left[1+\gamma H_\psi^X(1)\right]^2}\right\}\frac{h_{11}}{\gamma^3},
$$

$$
\hat\sigma_{11,12}[\theta,A](h_{12}) = \frac{(\delta - H_\psi^X(1))H_\psi^X(Z'h_{12})}{\left[1+\gamma H_\psi^X(1)\right]^2},
$$

$$
\hat\sigma_{12,11}[\theta,A](h_{11}) = \frac{(\delta - H_\psi^X(1))H_\psi^X(Z)h_{11}}{\left[1+\gamma H_\psi^X(1)\right]^2},
$$

$$
\hat\sigma_{12,12}[\theta,A](h_{12}) = \frac{(1+\delta\gamma)H_\psi^X(ZZ'h_{12})}{1+\gamma H_\psi^X(1)} - \frac{(1+\delta\gamma)\gamma H_\psi^X(Z)H_\psi^X(Z'h_{12})}{\left[1+\gamma H_\psi^X(1)\right]^2},
$$

$$
\hat\sigma_{2,11}[\theta,A](h_2) = \frac{(\delta - H_\psi^X(1))H_\psi^X(h_2)}{\left[1+\gamma H_\psi^X(1)\right]^2},
$$

$$
\hat\sigma_{2,12}[\theta,A](h_2) = \frac{(1+\delta\gamma)H_\psi^X(Zh_2)}{1+\gamma H_\psi^X(1)} - \frac{(1+\delta\gamma)\gamma H_\psi^X(Z)H_\psi^X(h_2)}{\left[1+\gamma H_\psi^X(1)\right]^2},
$$

and also

$$
\begin{aligned}
\hat\sigma_{1,2}[\theta,A](h_1)(u) &= Y(u)e^{\beta' Z(u)}\left\{\frac{(\delta - H_\psi^X(1))h_{11}}{\left[1+\gamma H_\psi^X(1)\right]^2}\right. \\
&\quad \left. + \frac{(1+\delta\gamma)Z'(u)h_{12}}{1+\gamma H_\psi^X(1)} - \frac{(1+\delta\gamma)H_\psi^X(Z'h_{12})}{\left[1+\gamma H_\psi^X(1)\right]^2}\right\}, \quad \text{and} \\
\hat\sigma_{2,2}[\theta,A](h_2)(u) &= \frac{Y(u)e^{\beta' Z(u)}h_2(u)}{\left[1+\gamma H_\psi^X(1)\right]^2}.
\end{aligned}
$$

Conditions (PB1) and (PB2) are readily verified by inspecting the information operators given above. Arguments from the proof of Lemma 7 verify that conditions (PB3) and (PB5) hold (the mean zero condition is proved in Kosorok, Lee and Fine's Proposition 3).

Arguments from the proof of Lemma 10 show that Corollary 3 of Murphy and van der Vaart (2000) can be used to generate random draws for the parametric component of the model satisfying condition (PB4.i). Alternatively, an MCMC scheme developed by Lee (2000) which builds on the results of Murphy and van der Vaart (2000) can be used to generate the required parametric draws. See Dixon (2003) for an implementation of the piggyback bootstrap using Lee's MCMC scheme for an odds-rate regression on a Non-Hodgkin's Lymphoma data set.

**B.2 A Biased Sampling Model**

Observe that for the biased sampling models

$$\begin{aligned}
U_1[\theta, A](h_1) &= h_1\left(\frac{\dot{w}_i(y,\theta)}{w_i(y,\theta)} - \frac{\int \dot{w}_i(z,\theta)dA(z)}{\int w_i(z,\theta)dA(z)}\right), \\
U_2[\theta, A](h_2) &= h_2(y) - \frac{\int h_2(z)w_i(z,\theta)dA(z)}{\int w_i(z,\theta)dA(z)}, \\
\hat{\sigma}_{1,1}[\theta, A](h_1) &= -h_1\left(\frac{\ddot{w}_i(y,\theta)w_i(y,\theta) - (\dot{w}_i(y,\theta))^2}{(w_i(y,\theta))^2}\right. \\
&\quad \left. - \frac{\int \ddot{w}_i(z,\theta)dA(z)\int w_i(z,\theta)dA(z) - \left(\int \dot{w}_i(z,\theta)dA(z)\right)^2}{\left(\int w_i(z,\theta)dA(z)\right)^2}\right),
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}_{1,2}[\theta, A](h_1)(v) &= h_1\left(\frac{\dot{w}_i(v,\theta)\int w_i(z,\theta)dA(z) - w_i(v,\theta)\int \dot{w}_i(z,\theta)dA(z)}{\left(\int w_i(z,\theta)dA(z)\right)^2}\right), \\
\hat{\sigma}_{2,1}[\theta, A](h_2) &= \frac{\int h_2(z)\dot{w}_i(z,\theta)dA(z)\int w_i(z,\theta)dA(z)}{\left(\int w_i(z,\theta)dA(z)\right)^2} \\
&\quad - \frac{\int h_2(z)w_i(z,\theta)dA(z)\int \dot{w}_i(z,\theta)dA(z)}{\left(\int w_i(z,\theta)dA(z)\right)^2},
\end{aligned}$$

and

$$\hat{\sigma}_{2,2}[\theta, A](h_2)(v) = \frac{h_2(v)w_i(v,\theta)\int w_i(z,\theta)dA(z) - w_i(v,\theta)\int h_2(z)w_i(z,\theta)dA(z)}{\left(\int w_i(z,\theta)dA(z)\right)^2}.$$

For the biased sampling model examined in the numerical studies, we used $w_1(z,\theta) = e^{z\theta}$ and $w_2(z,\theta) = 1$, and we assume $\theta$ lies in a known compact subset of $\mathbb{R}$. Then Conditions (PB1) and (PB2) are readily verified. Conditions (PB3) and (PB5) hold by arguments similar to the proof of Lemma 7 and the proof of Theorem 5.8 of Gilbert (1996). The mean zero assertion of Condition (PB5) follows upon taking expectations in the above expression for $U_2$. Conditions (PB4.i), (PB4.ii) and (PB4.iii) hold by the arguments of Lemmas 10, 8 and 9. Here the necessary identifiability for using arguments from the proof of Theorem 3.2 of Lee (2000), as described in the sketch of the proof of Lemma 10, follows from Theorem 4.4 of Gilbert (1996). Theorem 5.11 of Gilbert (1996) proves the validity of the bootstrap for bootstrap weights that satisfy our aforementioned conditions. Finally, Condition (PB6) follows from Theorem 5.8 and the discussion on pages 106–107 of Gilbert (1996). Note that Gilbert takes $\mathcal{H}_2$ to be the set of functions of bounded supremum norm. Since this contains the set of functions of bounded total variation norm, the proofs in Gilbert immediately imply that our regularity conditions hold, except one must establish that the information operator is continuously invertible and onto for this smaller space of functions. This holds by the same arguments Gilbert uses for the larger class of functions.

## REFERENCES

Abraham, R., Marsden, J. E., and Ratiu, T. (1988), *Manifolds, Tensor Analysis, and Applications.* New York: Springer-Verlag.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins University Press.

Bilias, Y., Gu, M., and Ying, Z. (1997), "Towards a General Asymptotic Theory for Cox Model with Staggered Entry," *Annals of Statistics*, 25, 662-682.

Dabrowska, D. M., and Doksum, K. A. (1988), "Estimation and Testing in a Two-Sample Generalized Odds-Rate Model," *Journal of the American Statistical Association*, 83, 744–749.

Dixon, J. R. (2003), *The Piggyback Bootstrap for Functional Inference in Semiparametric Models.* PhD Dissertation, Department of Statistics, University of Wisconsin-Madison.

Gilbert, P. B. (1996), *Sieve Analysis: Statistical Methods for Assessing Differential Vaccine Protection Against HIV Types.* PhD Dissertation, Department of Biostatistics, University of Washington.

——— (2000), "Large Sample Theory of Maximum Likelihood Estimates in Semiparametric Biased Sampling Models," *Annals of Statistics*, 28, 151–194.

Gilbert, P. B., Self, S. G., and Ashby, M. A. (1998), "Statistical Methods for Assessing Differential Vaccine Protection Against Human Immunodeficiency Virus Types," *Biometrics*, 54, 799–814.

Kim, Y., and Lee, J. (2003), "Bayesian Bootstrap for Proportional Hazards Models," *Annals of Statistics*, 31, 1905-1922.

Kosorok, M. R., Lee, B. L. and Fine, J. P. (2004). "Robust Inference for Proportional Hazards Univariate Frailty Regression Models," *Annals of Statistics*, 32, 1448–1491.

Kress, R. (1989), *Linear Integral Equations*, Springer-Verlag, Berlin.

Lee, B. L. (2000), *Efficient Semiparametric Estimation Using Markov Chain Monte Carlo.* PhD Dissertation, Department of Statistics, University of Wisconsin-Madison.

Lin, D. Y., Fleming, T. R., and Wei, L. J. (1994), "Confidence Bands for Survival Curves Under the Proportional Hazards Model," *Biometrika*, 81, 73–81.

Murphy, S. A. (1994), "Consistency in a Proportional Hazards Model Incorporating a Random Effect," *Annals of Statistics*, 22, 712–731.

Murphy, S. A., and van der Vaart, A. W. (2000), "On Profile Likelihood," *Journal of the American Statistical Association*, 95, 449–465.

Newton, M. A., and Raftery, A. E. (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series B*, 56, 3–26.

Parner, E. (1998), "Asymptotic Theory for the Correlated Gamma-Frailty Model," *Annals of Statistics*, 26, 183–214.

Praestgaard, J., and Wellner, J. A. (1993), "Exchangeably Weighted Bootstraps of the General Empirical Process," *Annals of Probability*, 21, 2053–2086.

Press, W., Flannery, B., Teukolsky, S. and Vetterling, W. (1994), *Numerical Recipes in Pascal: the Art of Scientific Computing.* Cambridge University Press, New York.

Rubin, D. B. (1981), "The Bayesian Bootstrap," *Annals of Statistics.* 9. 130-134.

Scharfstein, D. O., Tsiatis, A. A., and Gilbert, P. B. (1998), "Semiparametric Efficient Estimation in the Generalized Odds-Rate Class of Regression Models for Right-Censored Time-to-Event Data," *Lifetime Data Analysis*, 4, 355–391.

Shen, X. (2002), "Asymptotic Normality of Semiparametric and Nonparametric Posterior Distributions," *Journal of the American Statistical Association*, 97, 222 – 235.

Tsodikov, A. (2003), "Semiparametric Models: A Generalized Self-Consistency Approach," *Journal of the Royal Statistical Society, Series B*, 65, 759-774.

van der Vaart, A. W. (2000), *Asymptotic Statistics.* Cambridge University Press, New York.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer.

Vaida, F., and Xu, R. (2000), "Proportional Hazards Model with Random Effects," *Statistics in Medicine*, 19, 3309–3324.

Vardi, Y. (1985), "Empirical Distributions in Selection Bias Models," *Annals of Statistics*, 13, 178–203.

Table 1: Coverage of 95% confidence bands generated with the piggyback bootstrap (pb) and weighted bootstrap (wb) approaches for three different interpolation schemes.

| Interpolation | pb placebo | wb placebo | pb vaccine | wb vaccine |
|---|---|---|---|---|
| none | 92.5% | 93.5% | 93% | 92.5% |
| leftpoint | 91% | 92% | 91.5% | 90.5% |
| midpoint | 92% | 93% | 93% | 92.5% |

Figure 1: MLE cdf for placebo group with 95% confidence intervals and confidence bands, for a simulated biased sampling data set.

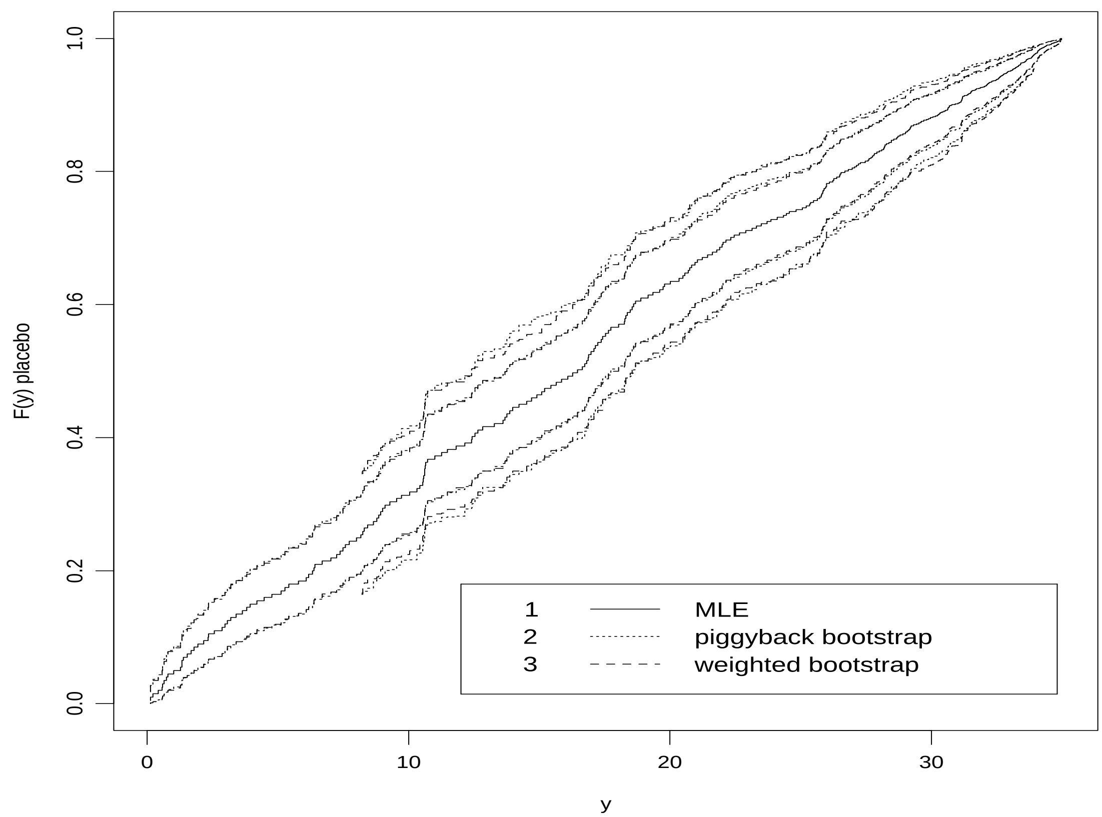

Figure 2: MLE cdf for vaccine group with 95% confidence intervals and confidence bands, for the same simulated biased sampling data set used in Figure 1.